Applications of Language & Knowledge Engineering

Research in Computing Science

Research in Computing Science es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. **Volumen 148, No. 3,** marzo de 2019. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No.: 04-2005-121611550100-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Queda prohibida la reproducción total o parcial, por cualquier medio, sin el permiso expreso del editor, excepto para uso personal o de estudio haciendo cita explícita en la primera página de cada documento. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

Editor responsable: *Grigori Sidorov, RFC SIGR651028L69*

Research in Computing Science is published by the Center for Computing Research of IPN. **Volume 148**, **No. 3**, March 2019.

Applications of Language & Knowledge Engineering

Beatriz Beltrán
Yolanda Moyao
Andrés Vázquez
David Pinto (eds.)

Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2019

ISSN: 1870-4069

Instituto Politécnico Nacional (IPN)
Centro de Investigación en Computación (CIC)
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal
Unidad Profesional "Adolfo López Mateos", Zacatenco
07738, México D.F., México

http://www.rcs.cic.ipn.mx
http://www.ipn.mx
http://www.cic.ipn.mx

Indexed in LATINDEX, DBLP and Periodica

Electronic edition

ISSN 1870-4069

Editorial

This volume of the “Research in Computing Science” journal contains selected papers related to the topic of Language and Knowledge Engineering and their applications, with topics covering natural language processing, computational linguistics, knowledge engineering, pattern recognition, artificial intelligence and other paper covering computer science in general.

Language engineering is an area of artificial intelligence and applications aiming to bridge the gap between traditional computational linguistics research and the implementation of potentially real-world applications. Knowledge engineering, on the other hand, refers to all technical, scientific and social aspects involved in designing, building, maintaining and using knowledge-based systems.

The papers published in this special issue were carefully chosen by the editorial board on the basis of the at least two double blind reviews by the members of the reviewing committee or additional reviewers. The reviewers took into account the originality, scientific contribution to the field, soundness and technical quality of the papers. It is worth noting that various papers for this special issue were rejected (rejected rate was 60%).

We would like to thank Mexican Society for Artificial Intelligence (Sociedad Mexicana de Inteligencia Artificial) and the Thematic Academic Network named “Language Technologies” (Red Temática en Tecnologías del Lenguaje) for their invaluable support in the construction of this volume.

The entire submission, reviewing, and selection process, as well as preparation of the proceedings, were supported for free by the EasyChair system (www.easychair.org).

Beatriz Beltrán
Yolanda Moyao
Andrés Vázquez
David Pinto
Guest Editors
Benemérita Universidad Autónoma de Puebla,
LKE-FCC-BUAP, Mexico

March 2019

ISSN 1870-4069

Table of Contents

Page

ISSN 1870-4069

A Formal Technique for Text Summarization from Web Pages by using Latent Semantic Analysis

J. Guadalupe Ramos[1], Isela Navarro-Alatorre[1], Georgina Flores Becerra[2], Omar Flores-Sánchez[2]

[1] Tecnológico Nacional de México, Instituto Tecnológico de La Piedad, Michoacan, Mexico

[2] Tecnológico Nacional de México, Instituto Tecnológico de Puebla, Puebla, Mexico

jgramos@pricemining.com, isela.na@piedad.tecnm.mx, kremhilda@gmail.com, omar.flores@itpuebla.edu.mx

Abstract. Web is the more attractive media for information consulting of, practically, whatever theme; humanity considers the Web, in the facts, the standard source of information. However as content grows, effort for discriminating and filtering increases too. Orthogonally, users employ each time smaller devices with reduced screens for web reviewing. Both considerations suggest the neediness of software tools for information acquiring and reduction, i.e., text summarization. There are several methods for text summarization, however, majority of them are based on techniques who considere plain documents in contrast with tree like structures of web pages, other are settled on the existence of keywords ignoring relations among words. In this work we present a formal method for the preparation of text summaries based on latent semantic analysis (LSA), which exploits the implicit relationships between the words that appear in a common context. In this way, text summaries are enriched with a certain semantic flavor incorporated by LSA. Furthermore we prepare the text summary induced by the query of an user and retrieving text excerpts more semantically similar to user's interest. Additionally we define a formula called semantic similarity which encapsulates the properties of LSA and determines the best text web page node for producing summaries.

Keywords. Latent semantic analysis, LSA, summarization, web pages.

1 Introduction

Automatic summarization from web pages has many clear motivations, by one side, available information in the Web is in constant growing, and in the other side, the most popular device for web consulting is the smartphone, which disposes of very small screens as the main interface for user interaction, this is obviously a non comfortable interface. In this sense, summarization, filtering and reduction of information techniques are required in order to choose only the meaningful information to be presented to an user. For text reduction from web pages there are not many approaches, we could consider that mature techniques for automatic summarization (such as those surveyed in [3]) could be employed. However those approaches lack of some aspects of our interest. For instance [3] considers procedures which do not use the tree like structure implicit in a web page.

We consider than tree like structure is important because can be employed for transforming a web page in a fragment of itself (by modifying tree structure), what could be read easily in a small device, this is called *filtering* [8]. Scheme of [8], employs techniques which are more based on keywords appearing in documents, i.e., the statistical information of co-occurrences between terms in a collection of documents, which can be gathered by formal methods like for instance *Latent Semantic Analysis (LSA)*, is not taken into account.

Certainly there are many works in the literature like [11, 5] and cited there, who produce automatic text summaries, inclusive by using LSA, those methods treat the text source as plain documents. However, web text is formatted using (X)HTML, i.e., the information is organized in formatting nodes which frequently provide certain implicit unity (all the information in a node is related) imposed by the web designer. In this way, main inspiration of this work is going to be established in the context of approaches that exploit the text fragments within a web page, for instance those devoted to web filtering such as [2, 8]. Regardless this focus, formulas here developed for calculus of semantic similarity among text fragments can be applied in a seamless way to any pair of text fragments, and thereby we could produce summaries from sentences instead of web page text fragments.

We consider that methods preserving the unity of (X)HTML nodes are more acceptable. Some works following this idea are, for instance [6, 15, 13, 2]. Indeed, in [2] a method for information extraction from web pages considering the distance between (X)HTML nodes is introduced. However standard tests of similarity as a basis for producing web summarization are not employed in [2], we believe that it is necessary to test classical techniques in order to establish an adequate comparison framework.

In this work we present a formal method for automatic creation of text summaries from a set of URLs, considering a web query, based on techniques of similarity employed in natural language processing and inspired on the notion of latent semantic analysis. Our formal method requires only once the calculations of LSA and then it can prepare a summary based on any web user query.

The rest of the paper is organized as follows. In Section 2, we overview standard concepts of vector representation of text in natural language and latent semantic analysis. In Section 3, we introduce a formal technique for text extraction based on semantic similarity. Then, in Section 4, we describe a set of experiments. Finally, we present our conclusions in Section 5.

2 Vector Space and Latent Semantic Analysis

In this subsection we introduce a formal standard representation for text documents written in natural language. The vector space model for automatic indexing was originally introduced by Salton et al., in [14] and it is considered a standard representation technique in information retrieval setting, where stored entities (documents) are compared with each other.

Given a text document d, a dictionary of *terms* is a set whose elements are the different *words* in the document d. Formally:

Definition 1 (dictionary). *Given a text document* d *let the set* $dict(d) = \{t_0, \ldots, t_{n-1}\}$ *be the dictionary of* d*, where* $\{t_0, \ldots, t_{n-1}\}$ *are the* n *terms in* d*.*

$\overrightarrow{V}(d)$ denotes the vector associated to document d, whose components are the weights for each element in the dictionary. It is assumed that element weights are computed using the *tf* weighting scheme, i.e., the value of a particular component is given according to the number of times the corresponding word occurs in document d. The set of documents in a collection then may be viewed as a set of vectors in a vector space, in which there is one axis for each term. This representation loses the relative ordering of the terms in each document [9]. In this view of a document, known in the literature as the *bag of words model*, the ordering of the terms in a document is ignored but the number of occurrences of each term is considered. Nevertheless, it seems intuitive that two documents with similar bag of words representations are similar in content [9].

Example 1. Given $text$ = "Web sites, Web services or Web-based applications", the dictionary (without lower-upper case distinction) is composed by $\{$web, sites, services, based, applications$\}$, then $\overrightarrow{V}(text)$ is $\langle 3, 1, 1, 1, 1\rangle$. Here the term Web appears 3 times, sites 1, and so on.

In Example 1 there are typographic symbols such as "," or "-", regularly this kind of text elements are ignored when a vectorial representation is prepared. A convenient set of particular terms is treated in the same way, in the example the term "or" was ignored, since words in a document are not equally important, i.e., some extremely common words provide little value in helping to distinguish the meaning of a text. These words are called *stop words*. The steps in natural language processing for analyzing a text document are the following.

- Filtering: the subject text must be filtered, dropping typographic symbols, for instance: ",.:;?".
- Removing stop words: those terms which do not provide meaning must be removed.
- Preparing the bag of words: the set of different terms is gathered, i.e., the dictionary properly.
- Vector representation: following the terms in the dictionary a vector is computed for each document.

Additionally, for web documents a step for removing format labels is mandatory.

2.1 Latent Semantic Analysis

Latent Semantic Analysis (LSA) is a theory and technical method for extracting and representing the contextual-usage meaning of words by means of statistical computations applied to a large corpus of text [4]. Hence, the underlying idea is that the aggregate of all the word contexts in which a given word does and does not appear, provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other.

LSA transports the statistical findings from heap of documents and gathers numerical information in a convenient model to be exploited in methods of natural language processing. Basically numerical information is useful to reason about the likeness of terms in the same textual space.

First step of LSA consists of the construction of a matrix representation of text, i.e., the matrix M, in which columns are employed for modeling documents and rows for terms (words). Each row i represents a specific term as well as each column j represents a document.

Thus, each cell $M_{i,j}$ stands for the frequency in which every term i appears in the document j. Term frequency tf can be substituted by some other scheme for measuring the significance of each term in a document, which is called term weighting scheme. A common weighting scheme for terms is for instance $tf.idf$ [9].

Next, LSA peforms the *Singular Value Decomposition* process (SVD) on the matrix M. SVD is the core of LSA, is a standard technique which is applied in linear algebra over matrixes, it is a specific form of factorial analysis.

In the original matrix M terms and documents are mutually dependent between them. In SVD, a rectangular matrix M is decomposed in the product of other three matrixes, i.e., $M = USV^T$. New matrixes will be formed by *singular vectors* or *singular values*. Resultant matrix U will contain a vector representation of the terms, which will have linear independency w.r.t. the relationship with the documents, while V will contain the vector representation of the documents whose components will be linearly independent w.r.t. the relationships with terms in M. Finally S is a diagonal matrix in which singular values are found in descendent order, and they represent the relationships between the other matrixes. The highest values in S represent the relations with major variance among terms and documents.

After SVD decomposition, the original matrix M can be rebuilt as of the matrix product of the resultant three matrixes. When a reconstruction over matrixes is performed it is possible to choose only the first k elements of the matrixes, i.e., $M' = U_k S_k {V_k}^T$, with this, a new matrix M' is obtained, in which the noise introduced by irrelevant relations is eliminated. Thus, the new values $M'_{i,j}$ unveil latent relationships among terms and documents, the so called human cognitive relations in [4]. The reason is that the reconstructed matrix employs the singular values representing the major variance, in this way, we gain a matrix that models the best relationships in data. In this work SVD is considered a core tool for LSA, the goal of this paper is to exploit the advantages of these formal techniques for producing technology. Hence the paper's main focus is presentation of summarization method. For a deep explanation of SVD the reader can consult, for instance [7, 4].

Example 2. Let us consider the following four sentences.

1. $d0 =$ My computer has branded software
2. $d1 =$ A PC is useful only with branded software
3. $d2 =$ PC (as computer) hardware can be generic

4. $d3 =$ Branded software and generic hardware go well with my computer

Hence, the dictionary of the document collection is {computer, software, branded, PC, hardware, generic}. According to above speech, the first row in M is for the representation of the term $computer$ (second one for software, and so on w.r.t. the dictionary) and the column 0 will be for the first document, then $M_{0,3}$ stands for the number of times that computer appears in document 3, and so on. By applying the technique SVD $M = USV^T$ is obtained:

$$M = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 0.49 & 0.21 & 0.35 & 0.77 \\ 0.47 & -0.50 & 0.02 & -0.17 \\ 0.47 & -0.50 & 0.02 & -0.17 \\ 0.26 & 0.14 & -0.93 & 0.22 \\ 0.35 & 0.47 & 0.08 & -0.39 \\ 0.35 & 0.47 & 0.08 & -0.39 \end{bmatrix}$$

$$S = \begin{bmatrix} 3.19 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.74 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.19 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.6 \end{bmatrix} \quad V^T = \begin{bmatrix} 0.45 & 0.38 & 0.46 & 0.67 \\ -0.46 & -0.50 & 0.73 & 0.08 \\ 0.32 & -0.75 & -0.36 & 0.45 \\ 0.70 & -0.22 & 0.35 & -0.58 \end{bmatrix}$$

Reconstructing M and considering only the first $k = 2$ elements we obtain:

$$M' = \begin{bmatrix} 0.542 & 0.413 & 0.986 & 1.085 \\ 1.067 & 0.993 & 0.044 & 0.929 \\ 1.067 & 0.993 & 0.044 & 0.929 \\ 0.265 & 0.195 & 0.557 & 0.577 \\ 0.133 & 0.019 & 1.115 & 0.821 \\ 0.133 & 0.019 & 1.115 & 0.821 \end{bmatrix}$$

If the similarity between vectors representing the rows 0 and 3 of M is calculated, i.e., the comparison of similarity among terms "PC" and "computer", result is 0.4, while, calculation of the same operation of both vectors in M' there is a result of 0.99. This new value is interesting due to the collection uses in a seamless way "PC" and "computer". Although the coincidence of both terms is given only in one document, correlations in the rest of documents allow to unveil the major *latent* similarity. This insights of relationships will be exploited later by the method exposed in this document.

3 A Formal Technique for Text Summarization

Now, we describe the main contribution of this work, i.e., a formal extraction technique useful for producing text summaries from a set of URLs.

The goal of the technique is to find out and then to extract web page fragments (from several URLs) that present the best semantic similarity w.r.t. a given web user query. Hence a *web user query* is a fragment of text in natural language, for instance *techniques applied to Web sites, Web services or Web-based applications*, which is regularly sent towards a web search engine.

3.1 Cosine Similarity

Similarity is a typical measurement between fragments of text in natural language. The standard way of quantifying the similarity between two texts $t1$ and $t2$ is to compute the cosine similarity of their vector representations $\overrightarrow{V}(t1)$ and $\overrightarrow{V}(t2)$ [9].

Definition 2 (similarity [9]). *The similarity between fragments of text* $t1, t2$ *is defined by*
$sim(t1, t2) = \frac{\overrightarrow{V}(t1)\cdot\overrightarrow{V}(t2)}{|\overrightarrow{V}(t1)||\overrightarrow{V}(t2)|}$, *where the numerator represents the dot product of the vectors* $\overrightarrow{V}(t1)$ *and* $\overrightarrow{V}(t2)$*, and the denominator is the product of their Euclidean lengths. The dot product of the two vectors* v, w *is* $\sum_{i=1}^{n} v_i w_i$ *while the Euclidean length of* $t1$ *is* $\sqrt{\sum_{i=1}^{n} \overrightarrow{V}_i^2(t1)}$. n *is the maximum number of different words between* $t1$ *and* $t2$. *A total similarity is 1.*

Example 3. If $t1 =$"World Wide Web" and $t2 =$"Web sites, Web services or Web-based applications", then the dictionary for $t1$ and $t2$ as a whole is "world, wide, web, sites, services, based, applications". The vector representation is: $\overrightarrow{V}(t1) = \langle 1, 1, 1, 0, 0, 0, 0\rangle$, $\overrightarrow{V}(t2) = \langle 0, 0, 3, 1, 1, 1, 1\rangle$. Therefore $sim(t1, t2)$ is $\frac{3}{\sqrt{3}\sqrt{13}}$, i.e., 0.48.

Roughly speaking our method will take a web user query and then will compute many similarity tests between the query and the text into web pages, searching by similar information in order to produce an automatic summary.

3.2 Semantic Similarity Induced by LSA

Certainly LSA provides valuable information w.r.t. similarity between whether documents or terms. For instance, in order to attend a query, [10] applies a similarity calculus by considering a vector from the query and a vector from the document. There, a vector document is each one of the columns of M'. However a summary of web pages requires being constructed from many text pieces from different web pages. In this setting a procedure for calculation of similarity among text excerpts instead of documents is needed. Granularity is a property commonly employed for referencing the size of a piece of data. Here, granularity is going to be used for talking about the number of terms in a text fragment. Methods for automatic summarization composed by text fragments of different granularity are required. A first approach for taking into account LSA through M' and multi-granularity in text fragments is the following:

Definition 3 (relative similarity). *The relative similarity between web user query text* Q *and* f_d *, a fragment of text from a document* d*, is defined by :*
$relsim(Q, f_d) = \frac{\overrightarrow{V}(Q)\cdot\overrightarrow{V}(f_d)}{|\overrightarrow{V}(Q)||\overrightarrow{V}(f_d)|}$,
with $\overrightarrow{V}(f_d) = \langle v_0, \ldots, v_i, \ldots, v_{n-1}\rangle$*, for* n *terms in the collection, and* v_i *is obtained from*

$$v_i = \begin{cases} M'[i][d] & t_i \in dict(collection) \text{ and} \\ & t_i \in dict(f_d) \\ 0 & \text{otherwise,} \end{cases}$$

where t_i *is a term in the collection of documents.*

In simple words, if a text fragment f_d from document d is being analyzed, for each term i in collection we put in vector $\overrightarrow{V}(f_d)$ value 0 if that word does not appear in f_d, and we put value $M'[i][d]$ if the word appears in f_d. Practically, the corresponding values in vector $\overrightarrow{V}(f_d)$ are mapped from column d in M'.

This mapping must be pointed out, because values from M' have been adjusted by LSA and then, are more appropriate. Indeed standard cosine similarity does not produce directly proportional values than relative similarity, for instance, similarity of fragment 21 in Table 1 is the lowest w.r.t. the rest of similarity measurements, nevertheless its relative similarity is the second one ranked in that column. Let us remember that we are comparing similarities among text fragments of several granularity, this is the reason why we apply a mapping of values from document in M' to $\overrightarrow{V}(f_d)$, i.e., we do not take the whole document terms.

So far, we could use relative similarity as a main tool for comparing text fragments, however, when there are not common terms in vectors, then result produced is 0, look for instance column $relsim$ in Table1 of tests. Hence, the strategy of only relative similarity is not completely well.

We must consider that LSA gathers information w.r.t. correlations among terms in a collection. A certain kind of measurement of term behavior should be conveniently incorporated into a new scheme of comparisons among text fragments. This is going to be treated in the rest of section.

First, we require a data structure which collects measurements of similarities among terms depicted in M'. This we call *Mutual similarity matrix* $\mathcal{M}$.

Intuitively speaking $\mathcal{M}$ is a data structure which contains the similarity values between all the terms in the reconstructed matrix M'. Hence $\mathcal{M}_{i,j}$ contains the similarity value among the term i and the term j of M', and the vector of term i is composed by the values in the row i in M' and so on.

Example 4. Let us consider the Example 2, the mutual similarity matrix of M' is as follows:

$$\mathcal{M} = \begin{bmatrix} 1.000 & 0.730 & 0.730 & 0.999 & 0.920 & 0.920 \\ 0.730 & 1.000 & 1.000 & 0.692 & 0.405 & 0.405 \\ 0.730 & 1.000 & 1.000 & 0.692 & 0.405 & 0.405 \\ 0.999 & 0.692 & 0.692 & 1.000 & 0.940 & 0.940 \\ 0.920 & 0.405 & 0.405 & 0.940 & 1.000 & 1.000 \\ 0.920 & 0.405 & 0.405 & 0.940 & 1.000 & 1.000 \end{bmatrix}$$

Here $\mathcal{M}[0][3] = 0.999$ which is the result of similarity among term 0 and 3 in M', i.e., the similarity between *computer*, row 0, $\langle 0.542, 0.413, 0.986, 1.085\rangle$ and *PC*, row 3, $\langle 0.265, 0.195, 0.557, 0.577\rangle$ of M'. These numbers, referenced particularly in the collection of Example 2. Obviously the diagonal matrix is composed by only ones.

Now, in the mutual similarity matrix the information that shows explicitly the numerical similarities among terms in the collection of documents is located. Quantities in $\mathcal{M}$ are the numerical measures of relations of terms w.r.t. a whole collection of documents, however, in order to prepare a text summary, composed from diverse text pieces we require the processing of smaller fragments of text instead of complete documents. Hence our method will perform comparisons among pieces of text (queries and web page text blocks) rather than documents and including the information gathered from LSA.

To compute text comparisons we should not apply only the $relsim$ calculation because we would not take advantage of the likeness between terms. Indeed in [12] a report is prepared by considering only the presence of literal words in text fragments and ignoring relations between them, we consider this approach loses a certain level of information for summary production.

In this way, it is necessary to exploit the measurements of relations contained in $\mathcal{M}$ to compute the calculus of similarity of small fragments of text and guaranteeing better summaries. For this, we apply a simple idea: if a term $t0$ is related with term $t4$ according to $\mathcal{M}$ and, if we have a text fragment f which contains $t0$ and does not contain $t4$ then, in order to execute the calculation of similarity of f we will aggregate to the vector $\overrightarrow{V}(f)$ the values for $t0$ and also for $t4$, i.e., $\overrightarrow{V}(f) = \langle 1, 0, 0, 0, threshold\rangle$ where $threshold$ is an arbitrary value in order to identify the limit for considering a *meaningful relation* between terms.

Basically, threshold define the least numerical limit of similarity measure for considering a relation between terms as important. Each row in M' represents a term in the collection of documents, and the numerical relation between the terms t_i, t_j is gathered in $\mathcal{M}_{i,j}$. In order to incorporate meaningful relations in the vectors of text fragments we define the concept of inflated vector.

Definition 4 (inflated vector). *Let $\overrightarrow{V\triangle}(f)$ be the inflated vector of a text fragment f in the setting of a collection of documents, such that $\overrightarrow{V\triangle}(f) = \langle v_0, \ldots, v_{m-1}\rangle$ where v_i is the corresponding weight of $t_i \in dict(collection)$ in the fragment f, which is obtained from:*

$$v_i = \begin{cases} weight(t_i, f) & t_i \in dict(f) \\ threshold & t_i \notin dict(f) \text{ but there is a} \\ & \text{meaningful relation with} \\ & \text{some } t_j \in dict(f) \\ 0 & \text{otherwise}, \end{cases} \quad \text{where } i, j \in \{0, \ldots, m-1\}, m \text{ is}$$

the number of rows in M' and $weight(t_i, f)$ is a function that returns the corresponding weight of t_i in f.

When a vector is inflated, following the dimensions of the collection, not only appears the weight of those terms present in a fragment of text, threshold is incorporated in the position corresponding for terms, which are not present in the fragment, but they maintain a meaningful relation (according to the collection) with other terms present in the fragment. In other words, in an inflated vector there appear the weight values of their terms and threshold value for their meaningful relationships. Hence, let us call *inflated similarity* $inflsim(t_1, t_2)$ to the similarity between two inflated vectors.

Here, inflated similarity returns the similarity calculus of two inflated vectors, where vectors were augmented from mutual closeness among terms. However the numerical importance of cosine similarity or relative similarity computations cannot be discarded, the reason is that cosine similarity and relative similarity are affected by text dispersion, i.e., big text fragments with few coincidences of terms have as a consequence low values of similarity, while inflated similarity is affected by the number of relationships between terms in the mutual similarity matrix. It appears more coincidences of terms. In this way, we introduce semantic similarity in order to take into account properties of previous calculus.

Particularly, relative similarity presents best properties because it performs the calculus by considering LSA.

Definition 5 (semantic similarity). *Given* $t1$ *and* $t2$*, let* $semsim(t_1, t_2) = relsim(t_1, t_2) \times inflsim(t_1, t_2)$ *be the semantic similarity between text fragments* t_1 *and* t_2*.*

Fundamentally we apply the product of relative similarity as a factor of correction for inflated similarity, in two ways, in one hand, it takes advantage of LSA and in the other hand, it computes greater values in the case of coinciding terms. When relative similarity value is 0, semantic similarity result is 0 too, in this case we can substitute product by a sum of logarithms, i.e., $semsim(t_1, t_2) = log(relsim(t_1, t_2)) + log(inflsim(t_1, t_2))$, and instead of a 0 argument we can put 0.001. In this way, semantic similarity combines properties of both measurements, relative similarity is more affected by coincidence of terms in the query and analyzed fragment, takes advantage of LSA and is sensible to fragment size.

3.3 Treatment of Text Fragments from Web Pages

A convenient web page representation is needed. DOM model is an adequate scheme, it stands for Document Object Model, which is a W3C standard platform—and language—neutral interface that allows programs and scripts to dynamically access and update the content, structure and style of documents [1].

Example 5. Let us consider the following toy web page:

```
<html>
 <head>   <title> Information Retrieval </title>  </head>
 <body>
   IR stands for information retrieval
  <div>
    There are two approaches for information retrieving:
    <span> a) When metadata is present.   </span>
    <span> b) By using soft-computing techniques.
              One of them is <strong>natural language processing.</strong></span>
  </div>
 </body>
</html>
```

Its corresponding DOM representation is visualized in a graphical mode in Figure 1. We can see a web page as a tree-like data structure where each node is an (X)HTML element, i.e., a (X)HTML tag with its contained text and its attributes, furthermore, its children are the embedded (X)HTML labels. If a node contains nested (X)HTML labels, there is a relation of *embedding* between the node and its children. For instance, `div` node embeds two `span` node children.

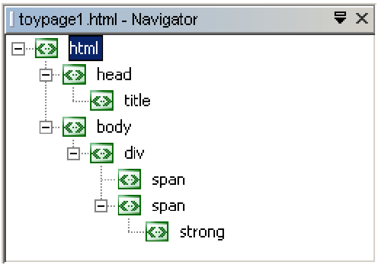


Fig. 1. The DOM tree of Example 5, visualized in a graphical way.

Table 1. Measurements for `http://lsa.colorado.edu/whatis.html` text fragments.

f	sim	$relsim$	$inflsim$	$comp$	***semsim***	terms	% terms	sem term	% sem term	text of f
1	0.14	0.137	0.336	0.046	-1.332	508	30.71	1654	100	What is LSA? What is ...
2	0	0	0.091	0	-4.039	1	0.06	123	7.43	What is LSA?
3	0	0	0.862	0	-3.064	17	1.02	1605	97.03	Note: If you linked ...
4	0	0	0.892	0	-3.050	9	0.54	1515	91.59	click here to open ...
5	0	0	0.997	0	-3.001	3	0.18	1221	73.82	The information on this page is based
6	0	0	0.867	0	-3.062	9	0.54	1605	97.03	Landauer, T. K., Foltz, ...
7	0	0	0.888	0	-3.051	4	0.24	1532	92.62	which is available for ...
8	0	0	0.780	0	-3.108	68	4.11	1654	100	Latent Semantic Analysis (LSA) is a ...
9	0	0	0.995	0	-3.002	3	0.18	1226	74.12	Latent Semantic Analysis
10	0	0	0.091	0	-4.039	1	0.06	123	7.43	(LSA)
11	0.21	0.154	0.804	0.124	-0.905	70	4.23	1654	100	Research reported in, and ...
12	0	0	0.997	0	-3.001	2	0.12	1219	73.70	semantic space
13	0	0	0.833	0	-3.079	33	1.99	1654	100	LSA can be construed ...
14	0	0	0.753	0	-3.123	72	4.35	1654	100	As a practical method ...
15	0	0	0.735	0	-3.134	76	4.59	1605	97.03	Of course, LSA, as ...
16	0	0	0.853	0	-3.069	24	1.45	1605	97.03	However, LSA as currently...
17	0	0	0.750	0	-3.125	98	5.92	1654	100	LSA differs from other...
18	0	0	0.729	0	-3.137	105	6.34	1605	97.03	However, as stated above...
19	0	0	0.874	0	-3.058	6	0.36	1579	95.46	Preliminary Details about ...
20	0	0	0.842	0	-3.075	47	2.84	1605	97.03	Latent Semantic Analysis is ...
21	0.05	0.163	0.822	0.134	-0.869	29	1.75	1605	97.03	The first step is to ...
22	0.45	0.234	0.754	0.176	**-0.752**	62	3.74	1654	100	**Next, LSA applies singular ...**
23	0	0	0.859	0	-3.066	17	1.02	1605	97.03	Landauer, T. K., ...
24	0	0	0.867	0	-3.062	4	0.24	1605	97.03	Basic and applied memory...
25	0	0	0.863	0	-3.064	15	0.90	1605	97.03	Landauer, T. K., & Dumais...
26	0	0	0.182	0	-3.740	2	0.12	400	24.18	Psychological Review, ...

4 Experiments

In this section, we describe an experiment performed upon a prototype, and correspondingly upon the formal technique. The collection was composed by text documents from a set of web pages whose links are the following.

1. `http://lsa.colorado.edu/whatis.html`
2. `https://en.wikipedia.org/wiki/Latent_semantic_analysis`
3. `http://recommender-systems.org/latent-semantic-indexing/`

Each web page was downloaded, and their DOM nodes were extracted by means of a DOM parser, then a text file with the set of nodes from each URL was prepared. The whole process of the formal technique introduced in Section 3 was computed in order to determine semantic similarity of every fragment for producing text summarization.

The launched query was: "singular value decomposition (SVD) to the matrix", for this we developed a tool that receives a query and returns each text fragment and its corresponding measurements. Summarization is constructed by taking fragments with higher semantic similarity. In the Table 1 a set of measurements are presented, there, a series of 26 text fragments f from `http://lsa.colorado.edu/whatis.html` are put through testing.

The first calculus shown is `sim` which represents the standard cosine similarity between the web user query an the analyzed fragment. The second one is the result of the relative similarity `relsim`, the third one represents the inflated similarity `inflsem`. Next column presents results of composed `comp` similarity, i.e., the product of `inflsem` and `relative`, then `semsim` is calculated by using the logarithmic approach. `semsim` unveils the result of the query, i.e., fragment 22 from the `http://lsa.colorado.edu/whatis.html` URL. For each fragment of web page the maximum semantic similarity w.r.t. the query is calculated and then the summary is produced.

A previous determination of a threshold of 0.95 was done, i.e., those relationships in the mutual similarity matrix equal or greater than 0.95 were considered important for the method. Value of threshold is thoroughly related with the experiment. Ideally, LSA should be fed with a huge stack of documents in order to harvest the more representative relationships between documents.

In the rest of columns, number of terms in the fragment is presented and their corresponding percent representation w.r.t. total terms in the collection. Next, the number of semantic terms is shown and its corresponding percent in the collection. Let us observe the percent increasing of semantic terms w.r.t. the real number of terms, this illustrates the great quantity of relationships that words implies in a text. Hence, methods which take into account the occurrence of relationships among terms are welcomed. Finally the last column shows the beginning text of the fragment.

In Table 1 text fragments with higher semantic similarity are chosen for composing the text summary.

5 Conclusions

We developed an extractive, multi web page, query-based, unsupervised technique for automatic summarization of web documents. We focused in approaches based on DOM tree structure. In this way previous works of [2, 12] were improved, for instance [12] computations would have returned 0 in the absence of common terms, here, semantic relationships provide numerical information for answer producing.

We have produced a formal technique which presents several advantages: always returns a value, is independent of the size of text fragment, privileges (numerically) the existence of common words in text fragment and query, outperforms results of cosine similarity, only once calculation of LSA is required to produce any number of summaries from a query. The formula discovered of semantic similarity can be applied to sentences of language which is useful for other kind of source documents.

The potential applications of the technique are the following: Summarization of web pages, summarization of documents, filtering of web pages (since we rank DOM nodes with semantic similarity), transformation of web pages, determining of hot sections in a web page (hot sections), production of industrial tools, and other more.

Acknowledgements. This work was partially supported by Tecnológico Nacional de México.

References

1. Arquitecture Domain, W.: Document Object Model (DOM) (2018), "Available at http://www.w3.org/DOM/"
2. Castillo, C., Valero, H., Ramos, J., Silva, J.: Information Extraction from Webpages Based on DOM Distances. In: CICLing (2). pp. 181–193 (2012)
3. Das, D., Martins, A.: A Survey on Automatic Text Summarization. Engineering and Technology 4, 192–195 (2007)
4. Furnas, G.W., Deerwester, S.C., Dumais, S.T., Landauer, T.K., Harshman, R.A., Streeter, L.A., Lochbaum, K.E.: Information retrieval using a singular value decomposition model of latent semantic structure. SIGIR Forum 51(2), 90–105 (2017)
5. Gambhir, M., Gupta, V.: Recent automatic text summarization techniques: A survey. Artif. Intell. Rev. 47(1), 1–66 (Jan 2017)
6. Gupta, S., Kaiser, G., Neistadt, D., Grimm, P.: DOM-based Content Extraction of HTML Documents. In: Proceedings of the 12th International Conference on World Wide Web. pp. 207–214. WWW '03, ACM (2003)
7. Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W. (eds.): Handbook of Latent Semantic Analysis. Lawrence Erlbaum Associates (2007)
8. López, S., Silva, J.: A new information filtering method for webpages. In: Database and Expert Systems Applications, DEXA, International Workshops, 2010. pp. 32–36 (2010)
9. Manning, C., Raghavan, P., Schütze, H.: An Introduction to Information Retrieval. Cambridge University Press (2008)
10. N. Underhill, T.: An introduction to information retrieval using singular value decomposition and principal component analysis. `http://buzzard.ups.edu/courses/2007spring/projects/underhill-paper.pdf` (01 2007)
11. Ozsoy, M.G., Alpaslan, F.N., Cicekli, I.: Text summarization using latent semantic analysis. J. Inf. Sci. 37(4), 405–417 (2011)
12. Ramos, J.G., Campoy, L., Ruiz, S., Jasso, N., Solorio, J.: Preparing text reports from web pages employing similarity tests. In: Mexican International Conference on Computer Science, ENC 2013. pp. 13–19 (2013)
13. Ramos, J.G., Silva, J., Arroyo, G., Solorio, J.: A Technique for Information Retrieval from Microformatted Websites. Lecture Notes in Computer Science 5947/2010, 344–351 (2010)
14. Salton, G., Wong, A., Yang, C.S.: A Vector Space Model for Automatic Indexing. Commun. ACM 18(11), 613–620 (1975)
15. Silva, J.: A Program Slicing Based Method to Filter XML/DTD Documents. In: SOFSEM (1). pp. 771–782 (2007)

ISSN 1870-4069

Classification of Barking Context of Domestic Dog using High-Level Descriptors

Benjamín Gutiérrez-Serafín[1], Humberto Pérez-Espinosa[1,2], Juan Martínez-Miranda[1,2], Ismael Espinosa-Curiel[1]

[1] CICESE-UT3, Nayarit, Mexico

[2] CONACyT Research Fellow, Mexico

{hperez,jmiranda,ecuriel}@cicese.mx, begutierrezse@outlook.com

Abstract. Barking has been a controversial topic that has been studied from different points of view. While some authors argue that dog barking is a non-communicative vocalization, others believe that barking plays a significant role in the human-dog interaction. Among the studies that take the last perspective, one of the most recent methods is to implement machine learning algorithms to classify single barks in different behavioral context by evaluating low-level descriptors. However, these research works do not incorporate the analysis of temporal structure or other dog vocalizations. In the present study, we proposed a broader approach by taking into account these relevant features that are currently not considered in the analysis of single barks for the classification of the context. By implementing an automatic process that segments long recordings of dog vocalizations and extracts both low-level and high-level descriptors, promising results were obtained for the barks' context classification from long recordings, where the highest value of F-measure was 0.71.

Keywords. Machine learning, barking classification, acoustic analysis, high-level descriptors.

1 Introduction

Within the repertoire of dog vocalizations, barking is typified by being the most characteristic sound of this species. Because of its constant presence in every context, barking has been considered as a non-functional way of communication [2].

Nevertheless, the study conducted by [5] contrasts this idea and suggests that the process of dog domestication affected to improve the barking of dogs to become the best communicative component. Similarly, the findings of [18] support the functionality of barking since they managed to prove with spectrographic analysis that bark structure varies depending on the context.

Research regarding this matter [11, 12], proved that people can identify the emotional content in which a dog barks, regardless of the age and level of experience of the participants with dogs. These results indicate that dog barking works as a communication system within the dog-human interaction. An effective identification of what the dogs are trying to communicate can improve the one-of-a-kind relationship between humans and dogs. In the same way, this could also benefit the welfare of the dogs, as well as take advantage of their vocalizations, for instance, to use them as security systems in houses. For these reasons, further research on this topic is required.

In recent years, modern approaches have been employed, such as applying artificial intelligence techniques, to enhance the dog barking classification. For example, in [8] the authors developed an algorithm to detect and select the best feature sets for the identification of context and individual, which were later tested in a classification task. Larranaga et al. [7] went even further by adding the prediction of sex and age of domestic dogs in its research. These problems, coupled with the individual recognition, have shown better results in contrast with the ones obtained in the context classification. As determined by [9] the context classification of barks relies highly on the barking individual. It was reported that when a dog-independent classification model is used, a decrease in accuracy occurs, which explains the reason for low performance in the problem of context recognition.

Additionally, these studies have centered just on the analysis of single barks and have overlooked other dog vocalizations and temporal patterns. According to the findings of [1] when a dog vocalizes, it emits not only barks but also other types of sounds such as growls and whines among others. This could be either by consecutive emissions of two or more sort of sounds or superimposing them, as well as a mix of these two manners. Looking at research work of other dog vocalizations, T. Faragó et al. [4] confirmed that growls also convey affective and contextual content to humans. In addition, they remarked that the natural temporal structure of this vocalization influences the understanding of the emotional state of the dogs, similarly to the inter-bark time intervals in the studies carried out by Pongrácz et al. [12, 13].

Due to these last-mentioned facts, it is reasonable to consider not only the acoustic features of single barks as the single factor of assessment in the bark's context classification problem through machine learning, but also the temporal structure and other dog vocalizations present in the barking sequences. In this paper, we report our research on the analysis of long barking recordings, in which high-level descriptors providing temporal structure and number of dog vocalizations are evaluated by using an automated method. The research questions that we are aiming to solve in this analysis are:

Does the inclusion of high-level descriptors add relevant information to the context classification?

What effect does an individual-independent validation have in the different stages of our proposed method?

2 Experimental Data

For this study, it was required to utilize the Mudi barking database for the classification of the contexts of barks and the UT3 dog vocalizations database for the identification of the type of dog sound.

2.1 Mudi Barking Database

The Mudi barking database was collected by Pongrácz et al. [13] from 12 Mudi dogs and consists of 244 recordings divided into 7 different behavioral contexts, recorded both inside the owners' dwelling and outdoor. These recordings last from 1 second to 6 minutes with a mean time of 44.78 seconds and a standard deviation of 48.19 seconds. They mainly include barking sequences with inter-bark intervals, as well as some other dog vocalizations and environmental sounds. Additionally, this repository also contains a total of 6614 single barks that were manually segmented from these long recordings and range from 0.1 to 0.8 seconds.

According to [13], the barks were recorded while performing a set of actions to encourage the dogs to bark. The compilation of every recording was made with a tape recorder and a microphone, where the experimenter was stood in front of the dog while holding the microphone within 1 to 4 meters from the dog. As reported in [13], the bark recordings were collected in the following 7 type of situations:

1. Alone: The dog was tied with its leash to a tree by its owner, then he/she walked away from the sight of the dog.
2. Ball: The owner held a ball or the dog's favorite toy and showed it to the dog 1.5 meters in front of it.
3. Fight: The trainer enraged the dog to bark aggressively and to bite his special glove while the owner kept the dog on a leash.
4. Food: The owner held the dog's food bowl 1.5 meters in front of the dog.
5. Play: The owner played a typical game with the dog.
6. Stranger: The experimenter presented in the owner's garden or front door in the absence of the owner.
7. Walk: The owner pretended like he or she was preparing to go for a walk with the dog.

2.2 UT3 Dog Vocalizations Database

The UT3 dog vocalization database was also employed to classify the type of sounds that the dogs emit. Audio and video of eighty dogs, mainly Chihuahuas, Schnauzers and French Poodles, were recorded to obtain the collection of audio constituted of a total of 30907 segments, of which 911 were whines, 666 growls, 5645 barks and the rest were background sounds (TV, human voices, bird sounds, etc.). The data collection protocol was designed to obtain aggressive and not aggressive vocalizations as following:

- Aggression to unfamiliar people inside the house.
    1. Normal alert due to the presence of a stranger: The experimenter (stranger) repeatedly knocked on the door to make the dog start barking.
    2. Aggression to the dog: The experimenter entered the home and provoked aggressive barks by making menacing movements in front of the dog.
    3. Aggression to the owner: An attack on the owner was simulated by making noises such as clapping and beating on the floor.
- Reaction towards enjoyment, happiness, and satisfaction.

4. Words of affection: The owner spoke in a high-pitched and affectionate tone of voice to the dog, with the intention of generating vocalizations.
5. Scratching: Both the owner and experimenter (if the dog allowed it) caressed or scratched the back, abdomen, chest, and sides of the dog until it produced vocalizations.
6. Play: The owner stimulated the dog using the objects or toys with which it usually plays, hoping to cause vocalizations.
7. Arrival at home: The owner was requested to leave the house with the experimenter and, after a while, returned home and talked to the dog affectionately from the outside without opening the door.
8. Simulation of going for a walk: The owner performed the routine that precedes taking the dog for a walk.

– Outdoor behavior.

9. Sadness/Anxiety by separation: The owner tied the dog with the leash to a tree and moved away from his sight.
10. Aggression to the dog: While the dog was still tied to the tree, the experimenter or some other stranger threateningly approached the dog, such as in the second stimulus.
11. Aggression to the owner: The experimenter pretended to attack the owner, such as in the third stimulus while walking his/her dog through the park.

After the audio compilation, individual sounds were automatically segmented using a Python script created by us that takes as input the energy and spectral centroid thresholds to detect audio activity. In the end, the automated method generated segments of a distinct span that range from very short individual barks (0.4 seconds) to longer segments (4 seconds) that consist of a group of a different number of individual barks emitted very quickly with the absence of pauses.

3 Methodology

In order to classify long barking recordings, we devised an automated method that segments the complete recordings and extracts low-level descriptors from the individual segments to classify both the type of vocalization and the context of each bark. Furthermore, high-level descriptors of each long barking audio file are also extracted to evaluate them and recognize the context in which the long recordings were induced. It should be pointed out that the number of recordings and barks used in our analysis differ from those in the original database due to the phases of automatic segmentation and dog vocalizations classification implemented in our approach. A set of the segments obtained were misclassified. Besides, some of the recordings could not be segmented properly using the optimal parameters, which precluded from being an automated process. In the end, we decided to discard this set of sub-optimal recordings and maintain those recordings useful to perform our analysis. Tables 1 and 2 show an overview of the number of recordings and single barks in the original database and those used in our experiments according to the contexts and the dogs, respectively. In the following subsections, we explain the details of the process that was carried out.

Table 1. Number of recordings and single barks for each context.

Original Database		Samples in our Analysis		Contexts
Long Recordings	Single Barks	Long Recordings	Single Barks	
22	758	19	706	Alone
53	1004	48	1000	Ball
30	1056	29	995	Fight
41	833	36	746	Food
23	752	23	977	Play
46	1425	44	1680	Stranger
29	786	28	1093	Walk
244	6614	227	7197	Total

Table 2. Number of recordings and single barks for each dog.

Original Database		Samples in our Analysis		Dogs
Long Recordings	Single Barks	Long Recordings	Single Barks	
9	275	9	317	d05
25	693	25	753	d09
5	108	5	140	d10
37	1007	31	908	d12
24	465	20	492	d14
18	336	17	336	d16
19	219	15	351	d18
29	686	29	736	d20
32	968	30	1041	d23
40	1650	40	1976	d24
2	83	2	59	d26
4	124	4	88	d27
244	6614	227	7197	Total

3.1 Overview of the Proposed Method

The general objective of this research is to take a broader approach by considering other variables that are not contemplated in the analysis of single barks. For this reason, we proposed to evaluate the long barking recordings of the Mudi database under the systematic procedure shown in Figure 1, that consists of 5 main stages, of which, three independent models had to be trained using machine learning algorithms. The functions of each of the main phases of this study are the following:

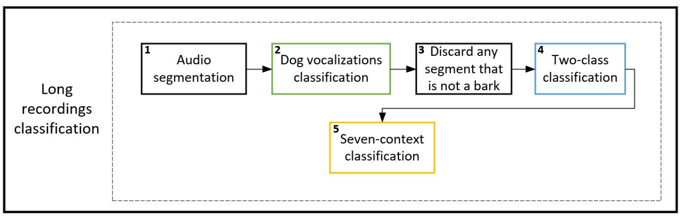


Fig. 1. Schematic figure for the proposed experimental procedure of long barking recordings classification.

1. Audio segmentation: The long barking classification model takes into account every dog vocalization within the complete recordings. Therefore, it is essential to have a segmentation stage to be able to inspect individually each of the sounds emitted by the dog. In this first step, the long barking recordings are inputted for segmentation. In addition, the span of inter-bark intervals, the duration of the vocalizations and the number of groups are registered.
2. Dog vocalizations classification: Having obtained each independent sound, the process shown in Figure 2(A) takes place to identify barks, whines, and growls with the assistance of the labeled data of the UT3 dog vocalization database. Background sounds are eliminated because they do not provide relevant information.
3. Discarding of segments that were not identified as barks: The barks are segregated from other canine vocalizations to continue with a second classification. The number of whines and growls are registered as part of other features needed in the final examination.
4. Two-class classification: To simplify the long recordings classification, a supplementary classification as illustrated in Figure 2(B), is implemented to improve the efficiency by labeling the single barks in one of the following two classes used to group the 7 different contexts according to the pleasant/unpleasant and activated/deactivated dimensions based on the circumplex model of affect [14] (see Fig. 3). The results of this assortment are added to the global features. A deeper explanation about the use of the two classes is provided in the Section 3.3.3.
5. Seven-class classification: Finally, the classification stage presented in Figure 2(C) is performed where the high-level descriptors previously identified are evaluated to complete the classification of the long recordings.

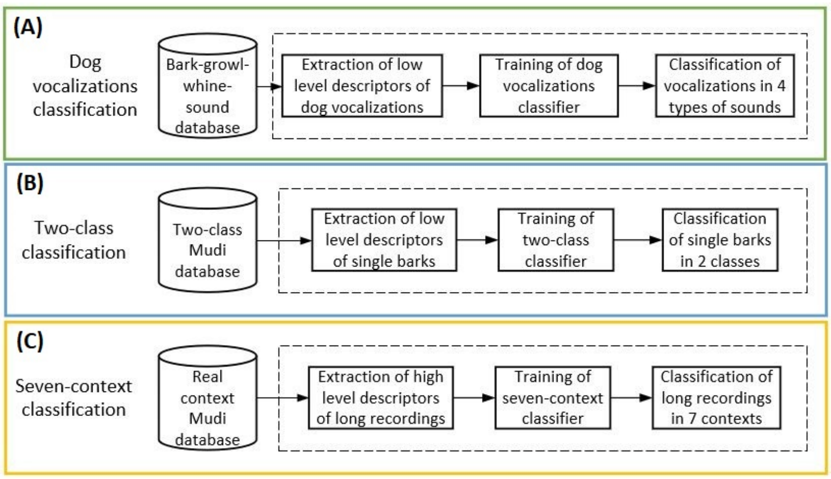


Fig. 2. Subsystems for accomplishing the classification of dog vocalizations.

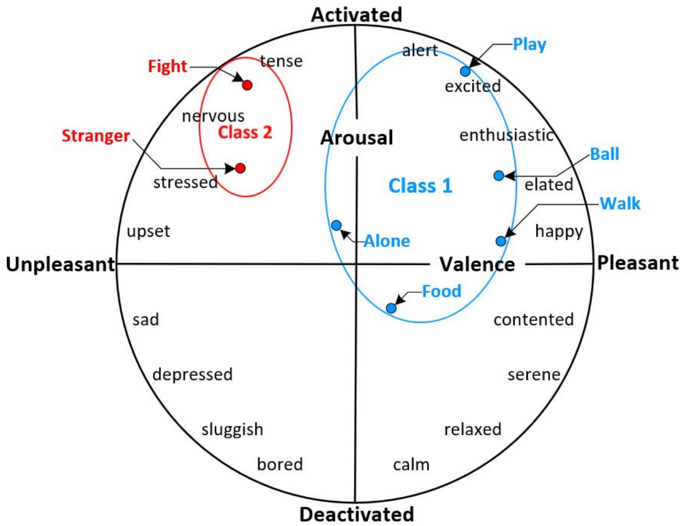


Fig. 3. Distribution of the seven barks' contexts within the circumplex model of affect.

3.2 Automatic Segmentation of Recordings

The long Mudi barking recordings were automatically segmented in the first stage by each canine vocalization emitted in their respective audio file. The duration of each recording in the Mudi database varies from a couple of seconds up to 6 minutes. An automated procedure was implemented for the segmentation process where was established an energy threshold of the signal to differentiate sound from silence and background noise. The threshold was detected by calculating the short time energy and spectral centroid. Precisely, the segmentation algorithm creates an array with the beginning and end time of the periods with sound. Then, it extracts from the long recordings the periods with sound for a further detailed examination to differentiate dog vocalizations. Figure 4 shows a fragment of a segmented long recording, in which both the periods' durations where there are no sound and the individual barks are registered to generate some of the High-Level Descriptor (HLDs). During the audio segmentation interval, the same algorithm extracts other HLDs from the long recordings in the background. More information about these features is explained in the next subsection.

3.3 Audio Characterization

Extraction of Low-Level Descriptors After segmentation stage, it was necessary to apply some signal processing techniques to acquire the acoustic properties of the segmented audio files in order to classify them, by type of vocalization (bark, growl or whine) and the context of the bark (alone, ball, fight, food, play, stranger or walk), according to the correlation of their features. To this end, the tool openSMILE [3] was used to extract Low-Level Descriptors (LLDs) for both databases. It was decided to use

Fig. 4. A recording containing five single bark sounds. Each single bark is stored in a separated audio file.

the next feature sets that have been proposed by different authors to conduct various experiments.

- IS-09 [15]: This feature set is extracted using a frame size of 25 ms and a frame step of 10 ms applying a Hamming window. Then, data contours are smoothed by a moving average filter to process the signal. Subsequently, 39 statistical functions are calculated over the values of the LLDs and their deltas and double deltas coefficient in each frame of the audio samples. In the end, a total of 5148 attributes are obtained.
- IS-10 [16]: The LLDs values are extracted using a frame size of 60 ms, and a frame of 25 ms applying a Gauss window for some features and a Hamming window for others. Then, the signal is processed by a moving average filter for smoothing data contours. In the next stage, it computes 21 statistical functions in some features and 19 in others over the values of the LLDs and their deltas and double deltas coefficient in each frame of the audio samples. As a result, an amount of 1582 attributes is obtained.
- IS-11 [17]: LDDs are computed using a frame size of 60 ms, and a frame of 25 ms applying a Gauss window for some features and a Hamming window for others. Then, the signal is processed by a moving average filter for smoothing data contours. After that, it calculates 37 statistical functions in some features and 36 in others over the values of the LLDs and their deltas and double deltas coefficient in each frame of the audio samples. Eventually, 4368 acoustic features are obtained.

Low-Level Descriptors Selection To improve the performance of the classifiers, it was determined to apply the Subset Evaluation method to every feature set with the assistance of the software WEKA [6]. This process evaluated all the original attributes from the IS-09, IS-10, and IS-11 sets to reduce the dimensionality of their feature vectors to 192, 140 and 217, respectively.

Extraction of High-Level Descriptors The classification model of long barking recordings was created based on the high-level characteristics registered during some processes that form our automated method. The feature vector of each long recording is constituted of 9 values, of which, 7 of them were extracted in the stages of segmentation and dog vocalizations classification. Moreover, these features were normalized due to the recordings do not have the same time duration. The other two remaining HLDs were obtained in the two-class classification phase and are represented in percentage, with respect to the number of barks classified in their respective class. The arrangement of the two classes was decided after evaluating the segments of our analysis in different combinations of the seven barks' contexts according to their distribution within the circumplex model of affect (see Fig. 3). We grouped the barks induced by the stimuli of stranger and fight in one class, and the barks caused by the rest of the stimuli in a second class. This grouping, as shown in Table 3, obtained the best classification results with 80/20 validation and using the machine learning algorithms mentioned in the next subsection. The HLDs of the classification of dog barking recordings are listed below:

1. Number of barks: The collection of segments that were identified as barks during the dog vocalizations classification.
2. Number of whines: The number of segments that were recognized as whines during the second stage of the automated process.
3. Number of growls: The quantity of segments that were labeled as growls in the phase two.
4. Number of groups: The number of segments grouped that were detected in the segmentation process according to their temporal proximity.
5. Number of pauses: The number of gaps that exists between the segments that were perceived in the segmentation process.
6. Total duration of barks: The sum time in seconds of each bark that was calculated during the segmentation process.
7. Total duration of pauses: The total amount of time in seconds of each space between the segments that was calculated during the first part of the process.
8. Class 1: The percentage of barks classified in the two-class classification as a context within the class 1 (alone, ball, food, play and walk).
9. Class 2: The proportion of barks classified in the two-class classification as a context within the class 2 (fight and stranger).

Table 3. Average F-measure of every experiment carried out for the Two-class classification using the barks obtained in our analysis.

			80/20						
IS-09			IS-10			IS-11			Arrangements
RF	NB	SVM	RF	NB	SVM	RF	NB	SVM	
0.33	0.35	0.35	0.38	0.40	0.35	0.35	0.36	0.39	Alone, Ball, Play ǀ Food, Walk ǀ Fight, Stranger
0.45	0.42	0.46	0.42	0.40	0.43	0.52	0.41	0.47	Alone, Food ǀ Play, Ball, Walk ǀ Fight, Stranger
0.37	0.38	0.38	0.41	0.39	0.35	0.43	0.43	0.42	Alone, Food, Walk ǀ Play, Ball ǀ Fight, Stranger
0.52	0.61	0.64	0.62	0.58	0.63	0.66	0.47	0.69	Ball, Food, Play, Walk ǀ Alone, Fight, Stranger
0.60	0.52	0.59	0.66	0.51	0.56	0.70	0.70	0.67	Alone, Ball, Food, Play, Walk ǀ Fight, Stranger

3.4 Classification and Validation Methods

Due to the positive results in previous similar classification analyzes [8–10], we used Support Vector Machine (SVM), Random Forest (RF) and Naive Bayes (NB) to conduct every experimentation in this study and store their results within Table 4 to Table 7. In addition, we also validated the stability of these machine learning algorithms with two validation methods:

– 10-Fold Cross-Validation (10FCV): In this technique, a 90% of the samples in the dataset is trained to create a model that is then tested with the remaining 10%. This process is repeated ten times using different training and testing sets on each occasion.
– 80/20 Validation: In this validation scheme, a classification model is trained using 80% of the samples in the dataset. The 20% left is used to test the model. In this validation process, we created a training model in which were not included the dogs belonging to the test model, generating as a consequence, a dog-independent training model.

Table 4. F-measure of single barks from the Mudi database for each 7 contexts in 10FCV settings.

10FCV									
IS-09			IS-10			IS-11			Context
RF	NB	SVM	RF	NB	SVM	RF	NB	SVM	
0.80	0.44	0.75	0.82	0.46	0.74	0.81	0.46	0.76	Alone
0.69	0.34	0.61	0.69	0.33	0.60	0.67	0.26	0.61	Ball
0.81	0.72	0.82	0.82	0.67	0.82	0.81	0.67	0.84	Fight
0.72	0.51	0.63	0.70	0.51	0.61	0.67	0.47	0.60	Food
0.69	0.57	0.68	0.69	0.59	0.68	0.65	0.54	0.66	Play
0.77	0.59	0.74	0.76	0.62	0.73	0.75	0.53	0.75	Stranger
0.63	0.44	0.62	0.65	0.46	0.60	0.60	0.45	0.61	Walk
0.74	0.52	0.70	0.74	0.53	0.69	0.72	0.49	0.70	Weighted Average

4 Results

It has been reported in previous studies related to the analysis of barking classification with a machine learning approach that, when samples of the same dog are used in the training and test sets, favorable results are obtained. However, the opposite occurs when both sets have samples from different dogs. To prove this, we conducted two experiments. Using the single context audio recordings of the Mudi database, a classification model was created, in which the barks of the 12 dogs were considered to train it. Looking at the results of the first experiment in Table 4, a good classification performance can be observed as it more than half of the outcomes shown 0.70 or higher in terms of F-measure. On the other hand, when using the same data but this time evaluating a barking test set of the dogs d18, d23 and d27 with a training set comprised from barks of the remaining dogs, poor results were obtained. As it is presented in Table 5, the experiment of the dog-independent classification model provided us with F-measure values less than 0.30 in all the cases.

When the number of contexts that had to be classified was reduced, an increase in F-measure values was presented. This occurred even if the training and test sets

Table 5. F-measure of single barks from the Mudi database for each 7 contexts in 80/20 validation settings.

80/20									
IS-09			IS-10			IS-11			Context
RF	NB	SVM	RF	NB	SVM	RF	NB	SVM	
0.08	0.13	0.06	0.02	0.10	0.02	0.05	0.29	0.06	Alone
0.06	0.10	0.18	0.08	0.04	0.09	0.11	0.06	0.13	Ball
0.61	0.63	0.49	0.50	0.46	0.34	0.56	0.63	0.55	Fight
0.29	0.18	0.14	0.27	0.24	0.14	0.07	0.15	0.09	Food
0.30	0.35	0.27	0.39	0.19	0.47	0.33	0.23	0.16	Play
0.35	0.30	0.28	0.39	0.38	0.32	0.31	0.37	0.31	Stranger
0.22	0.18	0.27	0.28	0.19	0.28	0.28	0.19	0.22	Walk
0.27	0.26	0.24	0.27	0.22	0.22	0.24	0.27	0.22	Weighted Average

differ from barking individuals. We conducted experimentation regarding this matter in a dog-independent classification model using the same data sets of the previous independent analysis. The improvement of the results can be noticed in Table 6, wherein almost all of the cases are above 0.75. In this experiment, the seven contexts were regrouped into two classes with the purpose of facilitating the classification process. Due to its efficiency, this procedure was included as the two-class classification stage in our method to allocate the segmented barks of long recordings in Class 1 and Class 2 features for the subsequent classification of long barking recordings in 7 contexts.

Table 6. F-measure of single barks from the Mudi database for each 2 contexts in 80/20 validation settings.

80/20										
IS-09			IS-10			IS-11			Context	
RF	NB	SVM	RF	NB	SVM	RF	NB	SVM		
0.84	0.75	0.81	0.85	0.76	0.83	0.83	0.83	0.84	Class 1	Alone Ball Food Play
0.56	0.61	0.60	0.55	0.50	0.63	0.59	0.64	0.68	Class 2	Walk Fight Stranger
0.76	0.71	0.75	0.76	0.69	0.77	0.76	0.78	0.79	Weighted Average	

The last part of this study was to implement the automatic method explained in Section 3. We used the long bark audio recordings of the Mudi database as the input files of the process. As in the past two experiments, the same arrangement of barking individuals was implemented to create the dog-independent classification model. In this case, the training set was used to train the two-class classifier to designate the single barks in the long recordings to their respective category. Once the process concluded, the features extracted from the complete recordings of the dogs d18, d23 and d27 were evaluated. It was complicated to create again effective training and test sets for an individual-independent validation due to the small number of instances and the variation of samples of each dog in this step. Nevertheless, we managed to get some promising results by carrying out a 10FCV to the final data frame. It can be contemplated in Table 7 that we obtained a positive outcome when we trained all the classification models with the Random Forest and Naive Bayes algorithms. However, poor performance was presented in the experiments where Support Vector Machine algorithm was used for training and caused that some of the values could not be captured.

Table 7. F-measure of long barking audio files from the Mudi database for each 7 contexts in 10FCV settings.

10FCV									
IS-09			IS-10			IS-11			Context
RF	NB	SVM	RF	NB	SVM	RF	NB	SVM	
0.67	0.67	-	0.73	0.75	-	0.73	0.80	-	Alone
0.55	0.56	0.44	0.50	0.58	0.36	0.67	0.58	0.46	Ball
0.33	0.40	0.00	0.33	0.46	-	0.50	0.31	-	Food
0.92	1.00	0.83	0.86	0.80	0.67	0.83	0.77	0.73	Play
0.77	0.77	-	0.80	0.86	-	0.86	0.86	-	Walk
1.00	0.80	0.50	0.89	0.75	0.00	1.00	0.80	0.22	Fight
0.44	0.43	0.00	0.44	0.38	0.00	0.44	0.43	0.00	Stranger
0.65	0.64	-	0.62	0.64	-	0.71	0.64	-	Weighted Average

5 Conclusion and Future Work

In our study, we explored the idea of evaluating high-level descriptors of long recordings of domestic dogs for context classification. Our findings demonstrate that the inclusion of these descriptors provide useful information for this classification problem. Additionally, our results support the idea that the presence of other dog vocalizations and temporal structure contain valuable patterns to assist machine learning algorithms in determining the real context of barking sequences. Concerning the model independence affair, a satisfactory performance with a dog-independent model was presented during most part of the proposed method. Nonetheless, the last part of the evaluation of this process was difficult to perform, since the data sample did not have a required diversity of contexts to demonstrate the effectiveness of the method. Thus, the results indicated poor accuracy in an individual-independent validation. Even though the final results were obtained only with a dependent classification model because of the lack of data, further study with this approach and more recordings may provide better results in the dog-independent model classification. Therefore, future research will concentrate on addressing this problem with a more extensive database, as well as the possibility of including more features to add more useful high-level information to the data frame to obtain better results.

Acknowledgment. This research work has been carried out in the context of the “Catedras CONACyT" programme funded by the Mexican National Research Council (CONACyT).

References

1. Cohen, J., Fox, M.: Vocalizations in wild canids and possible effects of domestication. Behavioural Processes 1(1), 77–92 (1976)
2. Coppinger, R., Feinstein, M.: Hark! hark! the dogs do bark...’and bark and bark. Smithsonian 21(10), 119–127 (1991)
3. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: the munich versatile and fast open-source audio feature extractor. In: Proceedings of the 18th ACM international conference on Multimedia. pp. 1459–1462. ACM (2010)
4. Faragó, T., Takács, N., Miklósi, Á., Pongrácz, P.: Dog growls express various contextual and affective content for human listeners. Royal Society open science 4(5), 170134 (2017)
5. Feddersen-Petersen, D.: Vocalization of european wolves (canis lupus lupus l.) and various dog breeds (canis lupus f. fam.). Archives Animal Breeding 43(4), 387–398 (2000)
6. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD explorations newsletter 11(1), 10–18 (2009)
7. Larranaga, A., Bielza, C., Pongrácz, P., Faragó, T., Bálint, A., Larranaga, P.: Comparing supervised learning methods for classifying sex, age, context and individual mudi dogs from barking. Animal cognition 18(2), 405–421 (2015)
8. Molnár, C., Kaplan, F., Roy, P., Pachet, F., Pongrácz, P., Dóka, A., Miklósi, Á.: Classification of dog barks: a machine learning approach. Animal Cognition 11(3), 389–400 (2008)
9. Pérez-Espinosa, H., Pérez-Martınez, J.M., Durán-Reynoso, J.Á., Reyes-Meza, V.: Automatic classification of context in induced barking. Research in Computing Science 100(6), 63–74 (1976)
10. Pérez-Espinosa, H., Reyes-Meza, V., Aguilar-Benitez, E., Sanzón-Rosas, Y.M.: Automatic individual dog recognition based on the acoustic properties of its barks. Journal of Intelligent & Fuzzy Systems 34(5), 3273–3280 (2018)
11. Pongrácz, P., Molnár, C., Dóka, A., Miklósi, Á.: Do children understand man’s best friend? classification of dog barks by pre-adolescents and adults. Applied animal behaviour science 135(1), 95–102 (2011)
12. Pongrácz, P., Molnár, C., Miklósi, Á.: Acoustic parameters of dog barks carry emotional information for humans. Applied Animal Behaviour Science 100(3), 228–240 (2006)
13. Pongrácz, P., Molnár, C., Miklósi, A., Csányi, V.: Human listeners are able to classify dog (canis familiaris) barks recorded in different situations. Journal of Comparative Psychology 119(2), 136 (2005)
14. Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology 39(6), 1161–1178 (1980)
15. Schuller, B., Steidl, S., Batliner, A.: The interspeech 2009 emotion challenge. In: Tenth Annual Conference of the International Speech Communication Association (2009)
16. Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., Narayanan, S.: The interspeech 2010 paralinguistic challenge. In: Proc. INTERSPEECH 2010, Makuhari, Japan. pp. 2794–2797 (2010)
17. Schuller, B., Steidl, S., Batliner, A., Schiel, F., Krajewski, J.: The interspeech 2011 speaker state challenge. In: Twelfth Annual Conference of the International Speech Communication Association (2011)
18. Yin, S.: A new perspective on barking in dogs (canis familaris.). Journal of Comparative Psychology 116(2), 189 (2002)

ISSN 1870-4069

To Beep or not to Beep: On the Influence of Some Interaction Design Variables on Its Performance

Ulrich J. Bröckl[1], Belém Priego Sánchez[2]

[1] Hochschule Karlsruhe,
Fakultät für Informatik und Wirtschaftsinformatik, Germany

[2] Universidad Autónoma Metropolitana unidad Azcapotzalco,
Departamento de Sistemas, unidad Azcapotzalco, Mexico

Ulrich.Broeckl@hs-karlsruhe.de,abps@azc.uam.mx

Abstract. During the construction and development of graphical interfaces, it is important to decide among a huge number of possible feedbacks (colors, sounds, etc.) with graphical variants (circular or rectangular shapes) and graphical behaviors (Close- and X-Button in title-bar e.g. to close a window). In this paper, the results of a long-term experiment comparing mouse- and touch-based interaction are presented. The aim of this work is to compare design variables of interaction in order to explore the influence of these variables on the interactions performance. The obtained results, for mouse interaction, reveal that only one studied design variable (clicking or not when reaching a target) is independent from target size. Color feedback showed to have negative influence for large objects and positive influence for small objects. For small targets circle-shaped targets and sound feedback proved to be negative, on the contrary to touch interaction. Here sound feedback saves for small objects 14.5% of interaction time. Finally the best results obtained are condensed to simple to apply design rules.

Keywords. Graphical interfaces, variables of interaction, mouse interaction.

1 Introduction

When building a Graphical User Interface, designers need to decide for a huge variety of possible feedbacks and decide among many graphical variants. Besides of graphical feedbacks like using color as a feedback for successful interaction sound feedback may be used there. Other graphical decisions are whether to take rectangular, or more fashionable, circular targets. Or whether to offer several targets for the same functionality (Close- and X-Button in title-bar e.g. to close a window) or restrict the design to a single target for each functionality.

These design variables are of interest in several areas. Firstly, GUI- and Web-Designers need that knowledge when designing new controls, for instance new button or menu control elements. Secondly and more generally the correct decision of these design variables also may support better design of any 2D pointing instrument based interaction since large control elements like panels and boxes were inestigated too. Thirdly the results presented in that article may serve GUI- and Web-Designers to calculate the operating times in advance and thus to predict or compare the performance of designs with freely chosen variables.

In this work the results of a long-term experiment comparing mouse- and touch-based interaction are presented. Especially, the influence of above mentioned design variables is studied. Fitts' Law was used to differentiate among small and large targets. The results reveal for mouse interaction that only one studied design variable (clicking or not when reaching a target) is independent from target size. Color feedback showed to have negative influence for large objects and positive influence for small objects. For small targets circle-shaped targets and sound feedback proved to be negative. On the contrary to touch interaction. Here sound feedback saves for small objects: 14.5% of interaction time. Finally the won results are condensed to simple to apply design rules.

2 Related Work

A computer is made up of different devices that allow a user to interact and control it. In this paper, it is carried out a study of the "mouse", a device that allows a computer user to control a screen pointer or cursor and to do click in screen positions that determine a given flow of information. During the 60s, one of the pioneers of the human-computer interaction area, Douglas C. Engelbart invented the mouse device with the help of Bill K. English.

The first mention of the term "mouse" in the literature, as an input device, was made when the input device at Stanford Research Institute, Menlo Park, CA developed by [4] was created. A comparison of this input device with other ones has been done throughout the years. For example, in 1967 a comparison among mouse, joystick and a fight pen was performed by [4]. The aim of that comparison was to select characters and words in the computer screen; the results obtained shown that the use of the mouse device is faster than the other ones. Other comparisons against the mouse device has been done, for example, with rare-controlled isometric joystick, step keys and text keys [1], or with other input devices such as touch screens, panel and keywords [8], with graphic tablet and trackball [10], with absolute touchpad, relative touchpad, trackball, displacement joystick, and force joystick [5], with high precision touch screens [18], with a pen device [2], with the finger-controlled isometric joystick [12], with touchpad and multitouch input technologies [19], just to mention some of the diverse literature dedicated to performance comparison of the mouse device with respect to other ones. Here we can see the numerous research works in which the mouse device is compared with different input devices, however, it is also important to mention the manner in which the mouse device has influenced the use and construction of graphical interfaces. Perry and Voelcker [15], for instance, present a perspective of the development of the mouse device and user-friendly interfaces. A button size and spacing on touch screen buttons experiment was made by [9]. They compared performance and input accuracy between older adults and younger adults. Their results have shown that younger adults required significantly less amount of time to complete a given input task. Input accuracy did not show significant different between older and younger adults. Authors found not any button size or spacing configuration where younger or older adults were stronger or weaker with. However, although spacing difference did not affect time performance, it significantly affects input accuracy.

3 Methods

3.1 Experimental Design

The experiments were conducted in fairs, exhibitions or laboratory sessions using a Windows program as shown in the Figure 1. After a short poster-based introduction into the experiments and privacy issues, the participants performed the test undisturbed.

When the program is started the user is first asked to convey some socio-demographic and basic data (age, gender, device type) in a pop-up window. Furthermore, binary test variables were set in that window: sound feedback on/off, color feedback on/off, circular or rectangular targets, and display of one or several targets at once (only one of these needed to be hit). These are considered as test variables for this article.

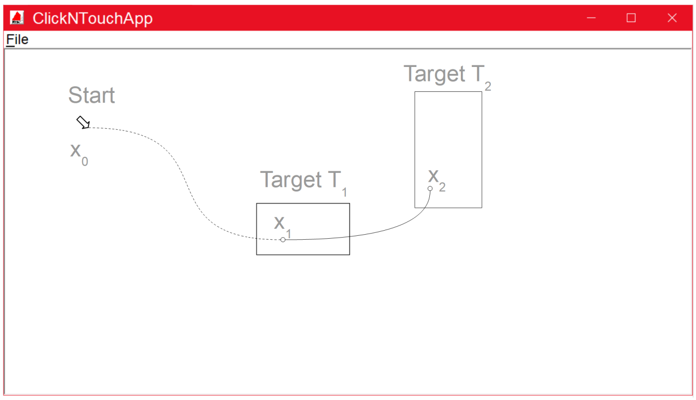


Fig. 1. Basic experimental design.

Right after that basic data acquisition, the user is shown targets by the program (rectangles or circles, see Figure 1 and the user clicks or touches on these. As soon as the target is hit, and only if it is hit, the next target is shown. So in Figure 1 the user starts at position x_0, begins to move to position x_1 and clicks in target $T1$. Directly after that Target $T1$ vanishes and Target $T2$ is shown. Thereafter the user moves from x_1 to x_2 and clicks in Target $T2$.

The picture's inscriptions for explanation in the picture $(Start, x_0,)$ and the mouse-traces were not shown in the real experiment. The application was opened in full-screen mode, so the title bar and File-menu were not visible. Actually only the targets and the mouse pointer became visible. Depending on the experiments context (sessions in fairs and exhibitions or laboratory sessions) the users had to hit between 50 up to 300 different targets (average 99.6 targets). In total 15,689 hits were stored, 35 were peaked out before (0.22%) since the users used more than 3000 ms to hit the target.

The targets areas were sized randomly between 440 and 84213 pixels (0.012% 4.4% of the screen's area). The x- and y-positions for the next targets were as well chosen randomly; the cases that the next target does not move or does move only little in comparison to the last target was not excluded.

3.2 Subjects

The experiments took place from 2013 to 2017; the subjects used either standard PCs with HDTV Monitors (27%) or smaller Laptops (73%, Asus Slate and Microsoft Surface) where the touch capacities of these were used for touch/mouse comparisons. In total 163 experiments were conducted (38 female, 125 male). Since the experiments took place several times on fairs and exhibitions a wider range of ages is covered (average age 31 years, standard deviation 18 years).

3.3 Analysis

The data of the experiments was stored as CSV-files on disk and afterwards an import tool was used to integrate all data into an SQL database. SQL queries were used to peak out and to calculate the Fitts Law regression parameters as defined in [10].

In order to check the significance of the differences furthermore the average interaction time and its standard deviation for the lower and upper half of the ID-interval was calculated. In the variable-comparing plots (e.g. in Figure 3) these IDs are depicted by diamond symbols.

The R program for Statistical Computing was used as an SQL interface and it was used to calculate the significance of differences using Rs Welch Modified Two-Sample t-Test. Furthermore R was used to plot the results.

Fitts Law. The analysis of the given data is based on Fitts Law [6]. There are several variants of that law available, see [3] for details on the ongoing discussion. We used the definition of the Index of Difficulty (ID) as being proposed in [10] since it is standardized in [13].

That definition of ID and the measured interaction times (in ms) were used to calculate Fitts Law linear regression as shown in the figure 2.

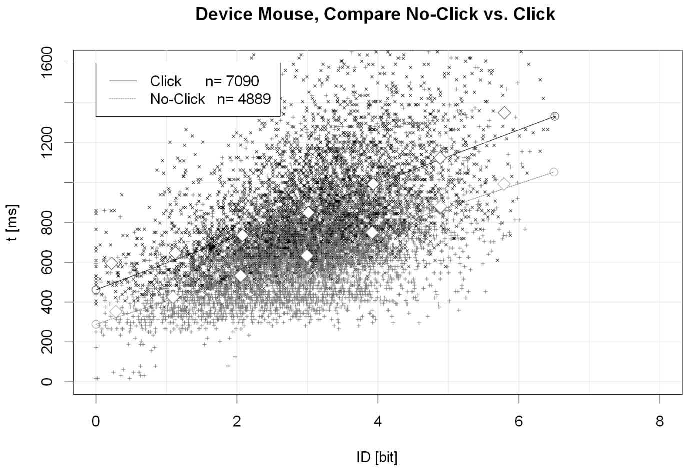


Fig. 2. Scatter plot to compare click and no-click interaction all scatters.

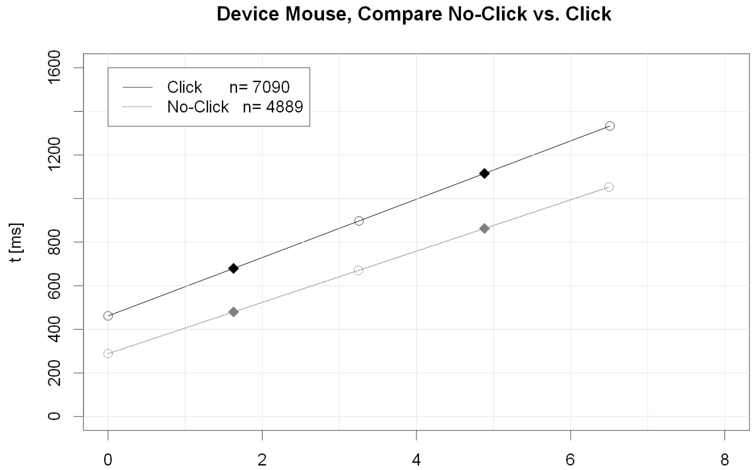


Fig. 3. Regression plot to compare click and no-click interaction.

The small x- and +-symbols depict a single interaction. The regression lines are ended by circle-symbols. In order to show the linearity of the averaged IDs these are shown in that figure too (white diamond symbols). For the sake of clearness these and the x- and +-symbols for a single interaction are omitted in the subsequent figures.

Significance Tests. Figure 3 shows that reduced variant of the scatter plot in Figure 2. The diamond symbols depict the points on the x–axis (thus ID–values) for which a significance tests of the differences between the different outcomes of the studied variable(s) were conducted. This was done by using R's Welch Modified Two–Sample t–Test for both pairs of diamond symbols.

For the left pair ($ID = 1.63bit$) we measured an average width and height of the targets of 113x75 Pixels (0.46% of screen size). For the right pair ($ID = 4.88bit$) we have 79x61 pixels (0.19%). The left pair is denoted from now on as "Large Targets" and the right pair as "Small Targets".

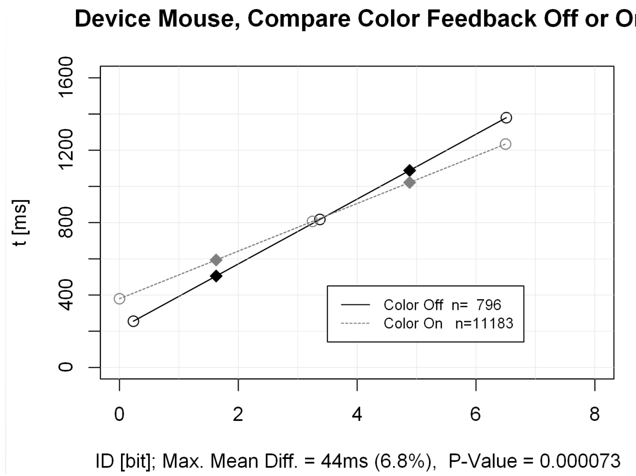


Fig. 4. Using color feedback.

4 Results

For all investigated devices (touch and mouse) and variables (circular or rectangular targets, sound feedback on/off, click needed or not, single or multiple targets on screen, color feedback on/off) the Fitts Law regressions and the average interaction times for small and large targets were calculated. Furthermore the p-value of the t-test was computed to check if the differences of the averages are significant. In the following paragraphs only those variables are presented, for which the average interaction times for small and large targets differ more than 6% and have a $p - value < 0.02$.

4.1 Mouse Interaction

Variables of Influence for Small and Large Targets. The only variable that exhibited advantages independently of the target size is shown in Figure 3. The difference here is that the users do not have to click in order to get the next target. That difference is for small targets 27.40%, for large targets as much as 33.20%.

This advantage is already used in several occasions in current GUIs: on the one hand for onmouseover-events in web pages [16] or for tooltips, e.g. in Microsofts ribbons [11]. Here the mouse interaction is used to trigger a functionality that does not cause an irreversible action. On the other hand, the design study in [7] shows that a click-free interaction is possible also for irreversible action – though it exhibits lacks in conformity with user expectations and suitability for learning.

For mouse interaction the usage of color feedback influences the performance depending on the target's size: small targets are clicked 6.80% faster. If we omit color feedback, large targets gain 5.60% performance (Fig. 4).

For mouse interaction the usage of several targets simultaneously showed near to no influence on the interactions performance, especially if large targets are considered. Here we noted only a difference of 3.9% Fig. 5).

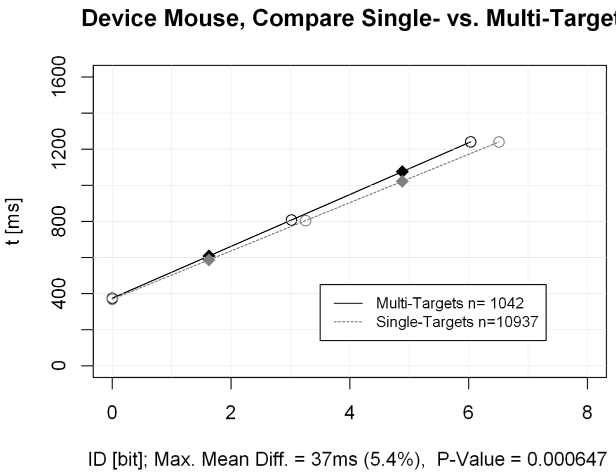


Fig. 5. Using multiple clickable targets.

Variables of Influence for Small Targets. With mouse interaction we detected a difference for small objects of 7.5% of interaction time when circles are not used instead of rectangles (Fig. 6)

For small targets using sound feedback proved to be negative: 6.8% of performance are lost. For large targets still 2.7% are lost (Fig. 7).

Combination of Variables of High Influence. In order to study the influence of the combinations of several variables we tested all possible combinations. The only exception here was the variable whether the user needs to click to get the next target. This variable was excluded since the large influence of that variable heavily shadows the influence of the other variables.

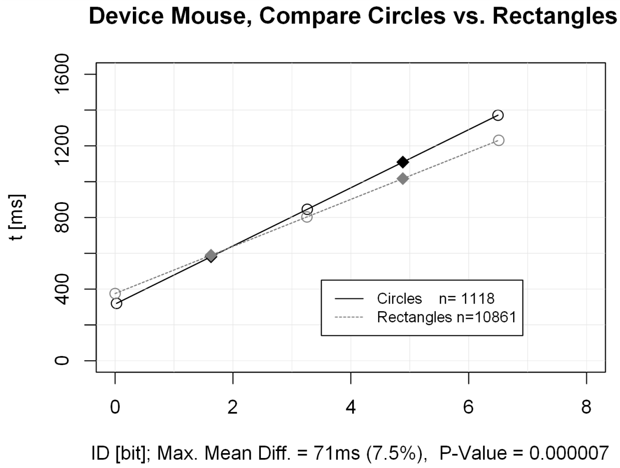


Fig. 6. Circles for interaction.

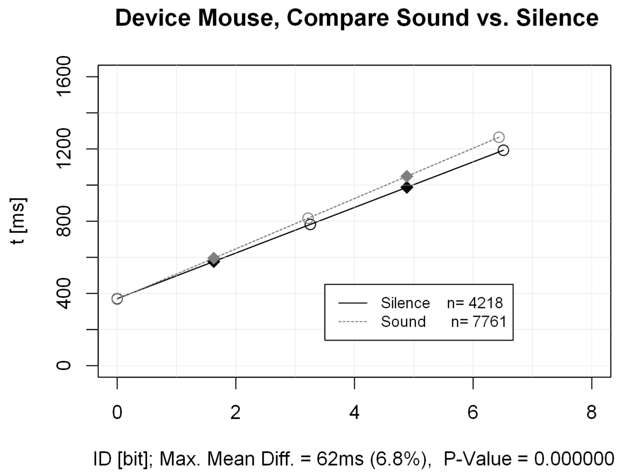


Fig. 7. Sound feedback.

The best combination was using color and multiple targets, and restrict from using sound. That delivered for small targets an advantage of 23.2% and for large targets 19.8% (Fig. 8).

4.2 Touch Interaction

Variables of High Influence for Small Targets. For touch interaction we detected a difference for small objects: 14.5% of interaction time can be won if we use sound feedback for small objects. This is a clear difference to mouse

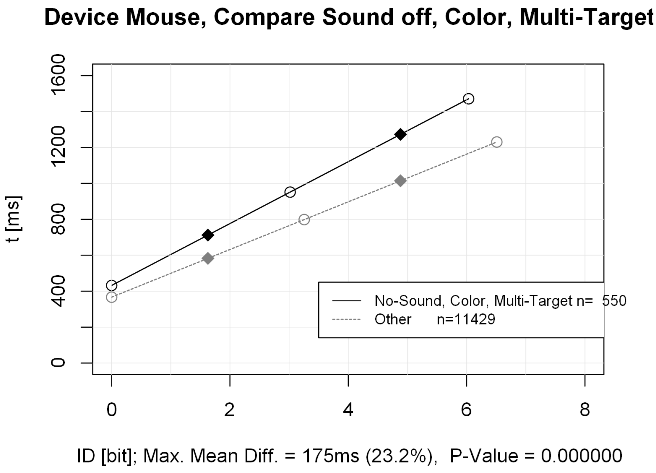


Fig. 8. Sound feedback for touch interaction.

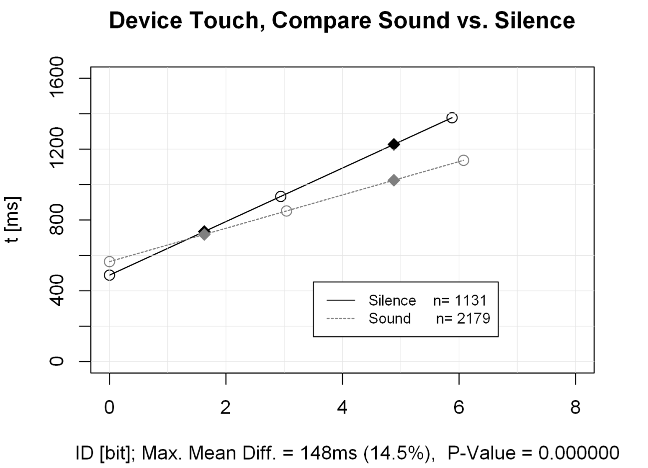


Fig. 9. Sound feedback for touch interaction.

interaction: in paragraph 4.1.2 we saw a performance loss of 6.8% if sound is used (Fig. 9).

For other variables (circular or rectangular targets, single or multiple targets on screen, color feedback) we could find no clearly significant influences, probably due to insufficient number of experiments.

5 Discussion and Conclusion

One meta result of this study is the fact that when studying influences of different feedbacks the size of the targets matters. Only one variable (whether to click or not for the next target) was not influenced by the size. Others variables like using rectangles instead of circles as targets only have positive influence if the targets are small. Other variables like color feedback heavily depend on the targets' sizes: color feedback only serves for better performance for small targets, large targets are influenced negatively.

The variables also proved to be interdependent: Unfortunately, when studying the different variable combinations a simple combination of the best outcomes of each single variable in order to obtain an optimal combination did not work.

Finally, we found one variable (sound feedback) that delivered contradictive results for mouse- and touch-based interaction. Here design strategies like Mobile First as being proposed for example in [17] should be reconsidered.

To sum it up: Following main rules can be derived from the study:

- Use sound feedback for touch interaction, avoid it for mouse interaction.
- Better do not use circular targets for small targets.
- For mouse interaction a combination of color feedback, silence and using multiple targets worked best.
- If possible think about avoiding clicks to trigger an intended functionality. One hot candidate seen by the authors to do so: the annoying but necessary cookie messages in websites could be switch off by a simple onmouseover-event. See for instance [14] for details on cookies law.

For some variables of touch interaction there are needed more experiments in future to identify further significant differences.

References

1. Card, S.K., English, W.K., Burr, B.J.: Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT. Ergonomics 21(8), 601–613 (1978), `http://dx.doi.org/10.1080/00140137808931762`
2. Charness, N., Bosman, E.A., Elliott, R.G.: Senior-Friendly Input Devices: Is the Pen Mightier than the Mouse? In: 103rd Annual Convention of the American Psychological Association Meeting. New York, New York (Aug 1995)
3. Drewes, H.: Only one fitts' law formula please! In: CHI '10 Extended Abstracts on Human Factors in Computing Systems. pp. 2813–2822. CHI EA '10, ACM, New York, NY, USA (2010), `http://doi.acm.org/10.1145/1753846.1753867`
4. Engelbart, D.C., English, W.K.: A research center for augmenting human intellect. In: Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I. pp. 395–410. AFIPS '68 (Fall, part I), ACM, New York, NY, USA (1968), `http://doi.acm.org/10.1145/1476589.1476645`
5. Epps, B.W.: Comparison of six cursor control devices based on fitts' law models. Proceedings of the Human Factors Society Annual Meeting 30(4), 327–331 (1986)
6. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental PSychology 74, 381–391 (1954)

7. Frank, A.: Dontclick. urlhttp://www.dontclick.it/ (2017), accessed 12-09-2017
8. Karat, J., M.J.E.A.M.: A comparison of menu selection techniques: touch, panel, mouse, and keyboard. In: International Journal of Man-Machine Studies. pp. 73–88
9. KIM, J.M., ISHIBASHI, K., IWANAGA, K.: The effect of invisible touch sensitive area on target acquisition. Proceedings of the Annual Meeting of Japan Ergonomics Society 48spl, 188–189 (2012)
10. MacKenzie, I.S., Buxton, W.: Extending fitts' law to two-dimensional tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 219–226. CHI '92, ACM, New York, NY, USA (1992), http://doi.acm.org/10.1145/142750.142794
11. Microsoft: "ribbons" windows dev center. urlhttps://msdn.microsoft.com/en-us/library/windows/desktop/dn742393(v=vs.85).aspx (2017), accessed 12-09-2017
12. Mithal, A.K., Douglas, S.A.: Differences in movement microstructure of the mouse and the finger-controlled isometric joystick. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 300–307. CHI '96, ACM, New York, NY, USA (1996), http://doi.acm.org/10.1145/238386.238533
13. Natapov, D., Castellucci, S.J., MacKenzie, I.S.: Iso 9241-9 evaluation of video game controllers. In: Proceedings of Graphics Interface 2009. pp. 223–230. GI '09, Canadian Information Processing Society, Toronto, Ont., Canada, Canada (2009), http://dl.acm.org/citation.cfm?id=1555880.1555930
14. Optanon: The cookie law explained. urlhttps://www.cookielaw.org/the-cookie-law/ (2017), accessed 24-10-2017
15. Perry, T.S., Voelcker, J.: Of mice and menus: Designing the user-friendly interface. IEEE Spectr. 26(9), 46–51 (Sep 1989), https://doi.org/10.1109/6.90184
16. Refsnes, D.: Onmouseover event. urlhttps://www.w3schools.com/jsref (2017), accessed 12-09-2017
17. Reich, M.: The importance of designing for mobile first. urlhttps://www.commonplaces.com/blog/the-importance-of-designing-for-mobile-first (2017)
18. Sears, A., Shneiderman, B.: High precision touchscreens: Design strategies and comparisons with a mouse. Int. J. Man-Mach. Stud. 34(4), 593–613 (Apr 1991), http://dx.doi.org/10.1016/0020-7373(91)90037-8
19. Shanis, J.M., Hedge, A.: Comparison of mouse, touchpad and multitouch input technologies. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 47(4), 746–750 (2003), https://doi.org/10.1177/154193120304700418

ISSN 1870-4069

Evaluating Predictive Techniques in Educational Data Mining: An Unbalanced Data Set Case of Study

Lourdes Sánchez-Guerrero, Josué Figueroa-González,
Beatriz González-Beltrán, Silvia González-Brambila

Universidad Autónoma Metropolitana, System Department,
CDMX, Mexico

{lsg,jfgo,bgonzalez,sgb}@azc.uam.mx

Abstract. This work presents an evaluation of the predictive techniques decision trees using CART algorithm, Naïve Bayes Classifier, Gradient Boosting Machine and Support Vector Machine for predicting whether a student will successfully complete a programming course or not. Factors considered for prediction were university-entrance and personal criteria like entrance age, gender, scholarship, high school GPA, mark in admission exam and other related with student's performance in a pre-requisite introductory programming course. The predicted variable takes two values, 'Approved' or 'Not Approved', and the data record contains an unbalanced portion of the class 'Approved'. For the analysis were considered two data sets, unbalanced and balanced. Evaluation of algorithms was performed considering the concepts of accuracy and ROC area. Results show that accuracy is bigger for the unbalanced data set, but its ROC area was very poor. Using the balanced data set, results were more reliable because accuracy and ROC area are closer. Best results were obtained with Naïve Bayes and Support Vector Machine algorithms. The most important factor in the prediction was whether a student had a scholarship or not.

Keywords. Educational data mining, predictive techniques evaluation, unbalanced data set, students' performance prediction, ROC curve evaluation.

1 Introduction

The increment in the use of technology has allowed gathering and storing a lot of data in many fields; as time passed, was observed that these data could be analyzed using several techniques for obtaining information. This process of analyzing large sets of data for finding patterns that lead to knowledge is known as Data Mining (DM) [8], and nowadays is applied in a lot of fields. In an educational environment, also is gathered a lot of information from many aspects of students, like their academic performance, personal characteristics or even social and affective ones. All these data can be exploited using DM techniques, for finding valuable information about them and the aspects that could affect their academic performance. Applying DM techniques for analyzing educative information received the name of Educational Data Mining (EDM) [15].

The main goal of educational institutions is ensuring that their students obtain a good preparation and finish their studies. Use of EDM has grown in the last decades, and right now it is applied for analyzing several aspects related with education [11], being one of the most studied the prediction of students' achievement at different levels, from a single work or exam, to the complete path of their higher education. Students' performance has been analyzed mainly with DM regression and classification techniques [2].

One common problem in EDM is that data sets are not big enough compared with other fields. Small data sets are common with educational data, and it is possible that these data sets are unbalanced, having a big difference between the amount of data of one class and another. This generates a problem when creating predictive models because obtained results could not be totally reliable, this is known as the *unbalanced data set problem*.

For dealing with unbalanced data sets, there exists some methods known as 'Sampling Methods' that balance the distribution of classes modifying the size of the original data. This work presents an analysis and comparison of the accuracy of predictive techniques: decision trees using CART algorithm [4], Naïve Bayes classifier (NB) [12], Support Vector Machines (SVM) [9] and Gradient Boosting Machine (GBM) [13] applied to an small educational data set with an unbalanced distribution of classes.

In particular, the goal of this paper is comparing the accuracy of different predictive algorithms for determining the most suitable for predicting whether a student will successfully complete a programming course or not, taking into account personal characteristics and academic performance in a previous programming course, and considering an unbalanced data set and a balanced one.

The structure of this paper is: Section 2 presents the related work concerning predicting student performance. Section 3 shows the experiments and the obtained results. Section 4 presents the analyzed results, and finally, Section 5 contains conclusions and future work.

2 Related Work

Predicting students' performance is the most studied topic in EDM because there are a lot of different approaches and factors to consider.

In [5], the authors compared the effectiveness to identify early fail of students in an introductory programming course using Neural Networks, Decision Tree, Support Vector Machine and Naïve Bayes classifiers, using data from one distance course (considering 22 variables) and the other from on-campus (considering 16 variables).

They conclude that the SVM technique outperformed the other ones by predicting with 92% and 83% of effectiveness, the failures of students that have performed at least 50% of the courses by distance education and on-campus, respectively. They conclude that the information is useful for teachers to take decisions, but is necessary make more experiments that permit the generalization of their results.

In [14], CART, J48, C4.5 and C5.0 decision trees algorithms and artificial neural network are applied for predicting study branch enrollment decision, future grades and satisfaction level of students in an Indian University. Processed data varied depending the classification. For enrollment decision (i.e. school and engineering), were considered national and state ranks of student and personal and social factors. For future grades, were considered academic criteria like: midterm grades, participation, and understanding of practice and theory. Finally, for satisfaction level were considered obtained grades, selected branch, social and family factors and student expectations about its studies. Results shows that enrollment decision prediction was the best, with a level of confidence over 98% for every technique. Other decisions obtain 60% and 67% of accuracy. The algorithm with the highest accuracy was C5.0.

In [6] was predicted students' performance using linear regression and matrix factorization approaches. They predicted a) students' next-term course grades, and b) within-class assessment performance. In particular, they investigated four methods: the course-specific regression (CSpR), the personalized linear multi-regression (PLMR) methods, the standard matrix factorization (MF) and the MF method based on factorization machines (FM) to predict the grade that a student will achieve in a specific course. They shown that PLMR and MF can predict next-term grades with lower error rates than traditional methods. PLMR were also useful for predicting grades on assessments within a traditional class or online course.

In [16], was developed a method for predicting student performance based in the cumulative Grade Point Average in a certain area. Main contribution of the paper is focusing on three main problems: differences of students at the moment of taking courses, importance of different courses at the moment of the prediction and incorporation of student progress in the prediction. Proposed method, which consists in a structure of base predictors that later were assembled in a cascade of predictors which incorporated the progress of student performance in the prediction, obtained better performance than traditional methods which gives different courses the same importance and do not include the evolution of students performance in later predictions.

In [7], is presented a predictive analysis for determining students performance in a public school in Brazil considering two aspects: first, personal and before entrance school criteria, and then, other set considering academic factors in a certain moment of the scholar period. Data sets were taken separately from years 2015 and 2016 and then combined. Classification technique was Gradient Boosting Machine and the process was performed with CRISP methodology. Results showed that academic factors like grades and absences are the most important, but also some personal factors, specially the neighborhood and the age of students have an important role over their performance at the end of the scholar year.

3 Experiments and Results

This section presents the analysis performed to the data set and the obtained results. We considered the steps of CRISP-DM methodology [1]: business understanding, data understanding, data preparation and modeling for obtaining the accuracy of considered algorithms. The evaluation stage is presented in Section 4 and the deployment step is out the scope of the present work.

3.1 Business Understanding

The first step involved the understanding and defining the goal of the analysis. The proposed goal for the study case was predicting whether a student would successfully complete a programming course or not, considering personal, entrance and academic factors. Also, it is presented the importance of different criteria at the moment of predicting grades and measured the effect of processing balanced and unbalanced data set.

It was used a study case related with predicting whether a student will approve the programming course "Object Oriented Programming" (OOP) or not, based in its personal characteristics, university entrance factors and performance in the prerequisite introductory course "Structured Programming" (SP) in the Mexican Universidad Autónoma Metropolitana Azcapotzalco (UAM-A). Prerequisite or seriation relationship means that for taking OOP, a student must have approved SP. This prerequisite relationship in the study programs is based on the idea of linking a set of courses so, one or more give the appropriate knowledge to students for having a good performance in the next ones.

The main goal of this work was measuring the accuracy of different algorithms applied to balanced and unbalanced data sets, for determining the algorithms more suitable, according to the characteristics of data.

3.2 Data Understanding

For accomplishing the goals, the student data set is about academic, personal and entrance data. Data for the analysis were obtained from two sources, the General File of Students (GFS) that contains personal and entrance information of the student like: age of entrance, score in the entrance exam, high school GPA, gender and scholar period of entrance.

Another source was the historical record of marks file (called *kardex* at UAM-A), which contains: student identifier, course identifier, obtained mark, scholar period where the grade was obtained and the way a course was approved, at UAM-A exist two: taking the course (called Global evaluation - GLO.) or through an extraordinary exam (called Recuperation Exam - REC.); these data were considered as part of the performance in SP.

It was only considered the grade obtained the first time a student took OOP (not mattering whether it was approved or not).

Grades at UAM-A are assigned with letters: MB (Very Good), B (Good), S (Sufficient) and NA (Not Approved). Some criteria are expressed in scholar periods, which at UAM-A represents three months.

Were only considered Computing Engineering students that are currently studying and for which OOP is mandatory. The total of students processed for analyzing the relationship between SP and OOP was 258.

3.3 Data Preparation

Considered criteria for predicting whether a student will approve OOP or not were:

- *Entrance age.* Age of the student at the moment of being accepted at UAM-A.
- *Gender of the student.* Male or Female.
- *High school GPA.* Average of the student in the former scholar level.
- *Score in entrance exam.* Score in the admission exam, maximum score is 1000 points.
- *Scholarship.* Whether the student has a scholarship or not.
- *Time before SP.* Time elapsed before the student took SP, in the first opportunity. Notice that SP it is supposed to be taken 2 or 3 scholar periods after a student enters the university.
- *Grade in SP.* Grade obtained at the moment of approving SP.
- *Tries for approving SP.* Number of attemps that took the student approving SP.
- *Time invested in approving SP.* Time invested by the student in approving SP, notice that this time and the number of tries it is not necessarily the same.
- *Time elapsed before taking OOP.* Number of scholar periods elapsed since the student approved SP and took OOP, in its first opportunity, approving it or not.

The predicted variable was the *performance of student in OOP in its first opportunity*; that is, whether the student approved OOP or not OOP, in its first opportunity.

Possible values and acronym for each criteria used in predicting if a student approved or not OOP are presented in Table 1.

Kardex did not contain neither the time invested in approving SP nor the number of tries. First value was obtained considering the entrance scholar period and the one in which the student approved SP. Number of tries needed for approving SP was obtained considering the amount of not approved marks obtained taking the SP before approving it. Time elapsed before taking OOP after approving SP was calculated considering the scholar period in which SP was approved and the one in which the OOP was taken. According the study program of Computing Engineering, OOP should be taken the next scholar period after approving SP.

Table 1. Criteria, acronyms and possible values used in predicting grade "Object Oriented Programming".

Criteria	Acronym	Values
Entrance age	AGE	17 to 37
Gender	GEN	Male or Female
High school GPA	MLA	7.7 to 10.0
Score in entrance exam	EXM	548 to 909
Scholarship	SCH	Yes or No
Progress level	PRO	0 to 15
Mark in SP	MSP	S, B or MB
Number of tries needed for approving SP	TRSP	1 to 5
Time invested in approving SP	TMSP	1 to 10
Time elapsed before taking OOP	TIMEP	0 to 14

Initially, grades to be predicted had four possible values: MB, B, S and NA; however after some tests, accuracy of different algorithms was very low, about 30%. For improving this accuracy, the approved marks (MB, B and S) were grouped in a single criteria. This also allows obtaining a binary classification where the predicted variable could have one or two possible values: AP, when the student successfully complete the course, and NAP when the student did not approve the course.

Distribution of grades in OOP was: 181 students approved (AP) in their first try and 77 did not approve (NAP). This represents a 45.5% of NAP. Difference it is not very large, but considering the reduced amount of data could be treated as an unbalanced data set.

For balancing the data, it was applied the Random Over Sampling Examples (ROSE) [10] package of R software.

3.4 Modeling and Algorithm Accuracy

The modeling stage was focused on measuring accuracy for each algorithm.

As already mentioned, the data set were analyzed with the predictive techniques Decision Tree using CART algorithm, NB, GBM and SVM.

Considering that the amount of data was small, we tested different percentages for training and testing models. Best results for predicting OOP grade were obtained with 90% for training and 10% for testing. A cross-validation process with 10-fold was applied for all predictive algorithms.

Accuracy was measured using the concept of Receiver Operating Characteristics (ROC) [3] which represents the rate of Positive values classified as positive, True Positve (TP), against the Negative values classified as positive, False Positive (FP), rate. Accuracy of the model is represented by the area under a curve, as more area, more will be the efficiency of the predictive model.

Accuracy of unbalanced data set. First, it was measured the accuracy of algorithms considering the original data set. The distribution of AP and NAP classes for training and testing sets is shown in Table 2.

The distribution of classes were: 68.9% of AP and 31.1% NAP for the training set and 80.7%for AP and 19.3% for the testing set.

Table 2. Class distribution in unbalanced data set.

	Approved	Not Approved
Training set	160	72
Testing set	21	5

The best accuracy and the ROC area for each algorithm is presented in Table 3.

Table 3. Accuracy and ROC area for unbalanced data set.

Algorithm	Accuracy	ROC area
CART	65.38%	0.557
SVM	80.76%	0.5
NB	76.9%	0.629
GBM	80.76%	0.5

ROC curves for each algorithm are presented in Figure 1.

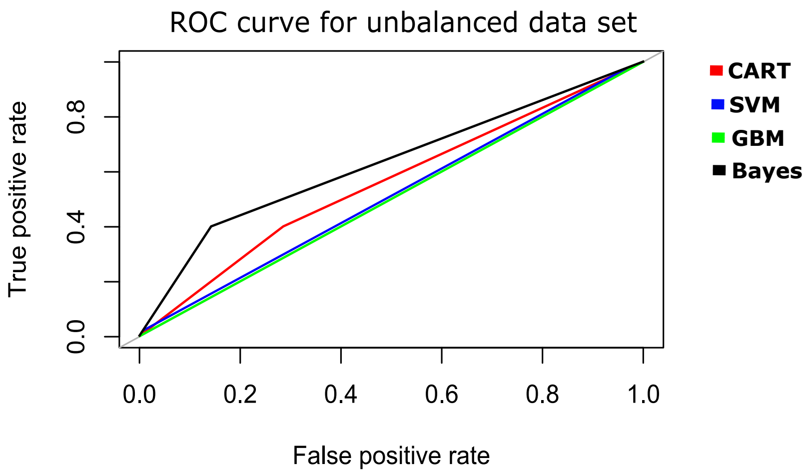


Fig. 1. ROC curves for algorithms using unbalanced data set.

Accuracy of balanced data set. After balancing the data set, the difference between the amount of AP and NAP was reduced for training and testing sets, as shown in Table 4. Distribution of classes were: 53.8% of AP and 46.2% NAP for both training and testing set.

Table 4. Class distribution in balanced data set.

	Approved	Not Approved
Training set	125	107
Testing set	14	12

Best accuracy for each algorithm is presented in Table 5.

ROC curves for each algorithm processing the balanced data set are presented in Figure 2.

Table 5. Accuracy and ROC area for balanced data set.

Algorithm	Accuracy	ROC area
CART	38.46%	0.643
SVM	76.92%	0.774
NB	80.76%	0.815
GBM	57.69%	0.601

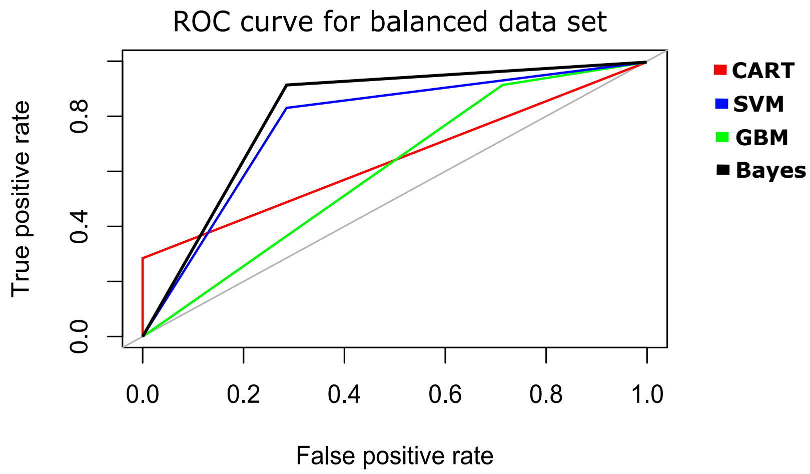


Fig. 2. ROC curves for algorithms using balanced data set.

For unbalanced and balanced data sets, it was created a confusion matrix which is presented in Table 6.

Table 6. Confusion matrix of prediction for unbalanced and balanced data.

		CART		SVM		Bayes		GBM	
Data set		AP	NAP	AP	NAP	AP	NAP	AP	NAP
Unbalanced	AP	15	6	21	0	18	3	21	0
	NAP	3	2	5	0	3	2	5	0
Balanced	AP	10	4	10	2	10	4	4	10
	NAP	12	0	2	10	1	11	1	11

Importance of criteria over approving OOP or not. The importance of each criteria in the prediction was obtained from the model produced by each algorithm. All of them showed the same results for balanced and unbalanced data sets. Figure 3 shows the obtained results.

4 Results Analysis

We evaluated the accuracy and ROC value for each algorithm for both data sets, balanced and unbalanced.

4.1 Unbalanced Data Set

As shown in Table 3, accuracy of SVM and GBM has a good value, more than 80%; however, the ROC area of these algorithms is very low, 50% that is the value of random prediction, this is also presented in Figure 1 where is clear that CART and NB have a bigger area even if their accuracy is smaller.

Analyzing Table 6, it is clear that accuracy of SVM and GBM is because of the classification of AP cases. Both classified correctly total of AP cases, but also classified incorrectly total of NAP cases. This is because there are few cases of NAP class and do not contribute with enough information for prediction. CART and NB classified more NAP cases correctly but less AP ones. The four algorithms are efficiently classifying AP cases correctly, mainly for the mentioned reason about class distribution.

4.2 Balanced Data Set

From Table 5, accuracy obtained analyzing the balanced data set is smaller than the unbalanced one; however, ROC area is bigger. Consider SVM with a 76.9% of accuracy and 0.774 of ROC area and NB with 80.76% of accuracy and 0.815

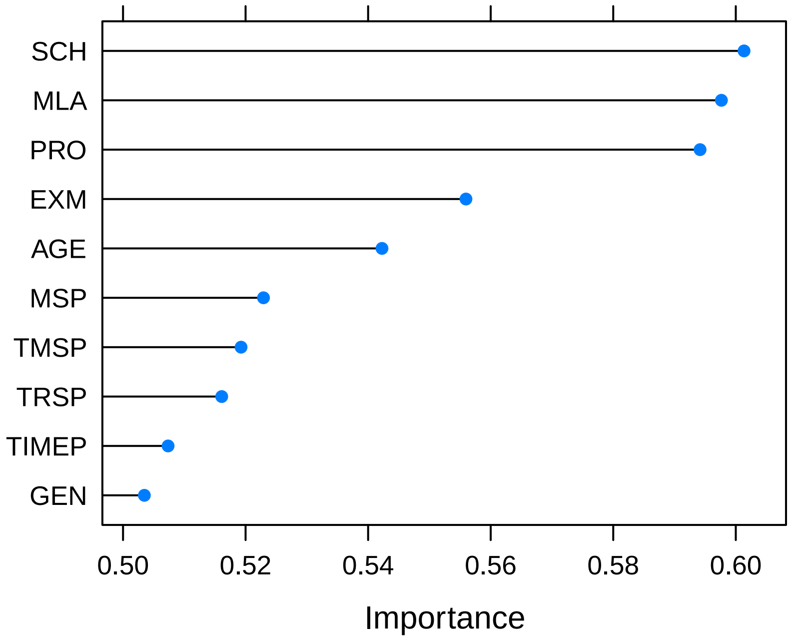


Fig. 3. Importance of criteria for predicting grade in "Object Oriented Programming".

of ROC area. For both algorithms, ROC area and accuracy are closer than using unbalanced data. This results also are presented in Figure 2 where is clear that SVM and NB have a bigger ROC curve area than CART and GBM.

Also, as is shown in the rows corresponding to the balanced results in Table 6, the distribution of correct predictions is bigger for NAP cases which was the main problem in the unbalanced data set. Here, for SVM and NB the distribution of correct predicted classes is more balanced, meanwhile for CART is good for AP cases, but very low for NAP and for GBM is good for NAP and very poor for AP.

4.3 Importance of Criteria in Prediction

Finally, Figure 3 shows that the most important factor in predicting whether a student will approve a programming course or not is related with having a scholarship (SCH), followed by high school GPA (MLA), level of progress (PRO), entrance exam (EXM) and entrance age (AGE).

This represents that most important factors are not related with the performance in a previous course, but with the general performance of the student.

Most important factor related with prerequisite course is the obtained mark (MSP).

5 Conclusions

The goal of this work was evaluating predictive techniques for predicting whether a student will approve a programming course or not, considering both personal and performance criteria over a prerequisite course. Also, analyzing the impact of working with an unbalanced data set has over the predictions and determining if balancing the data set helps for obtaining better results.

Results show that using unbalanced data could generate good results; however these can be misleading because the minority class could not contribute to the prediction and cases of that class could be wrong classified, or if there are not enough cases, could lead to a false high accuracy depending only in the majority class.

Using a balanced data set reduced a little the accuracy of the algorithms; however, results are more reliable because now both classes contribute to the prediction and eliminate the problem produced by the minority class.

About measuring the efficiency of algorithms, using ROC area and its graphic representation (ROC curve) is a better way of evaluating predictions when are analyzed unbalanced data sets. Using only accuracy, comparing right predictions against total of cases is affected by the lack of minority class cases, meanwhile ROC area consider false positive and true positive classifications, so its results are more reliable.

Study case presented the opportunity of working with an unbalanced data set; however the amount of information it is very reduced. This impacts in the efficiency of the algorithms obtaining as best results 80.76% of accuracy and 0.815 for ROC area using Naïve Bayes Classifier and 76.92% of accuracy and 0.774 for ROC area for Support Vector Machine.

From the results, was observed that it is more important when the student has a scholarship and its general performance, than its mark in the prerequisite course.

As future works, a similar analysis for a course that offers more data and that also could have a unbalanced relationship in its classes will help to verify the importance of balancing the data set and considering other evaluating criteria more than accuracy.

References

1. Azevedo, A.I., Santos, M.F.: Kdd, semma and crisp-dm: a parallel overview. IADS-DM (2008)
2. Bakhshinategh, B., Zaiane, O.R., ElAtia, S., Ipperciel, D.: Educational data mining applications and tasks: A survey of the last 10 years. Education and Information Technologies 23(1), 537–553 (2018)

3. Bradley, A.P.: The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern recognition 30(7), 1145–1159 (1997)
4. Breiman, L.: Classification and regression trees. Routledge (2017)
5. Costa, E.B., Fonseca, B., Santana, M., de Araújo, F., Rego, J.: Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior 73, 247–256 (2017)
6. Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., Rangwala, H.: Predicting student performance using personalized analytics. Computer 49(4), 61–69 (Apr 2016)
7. Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., Van Erven, G.: Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. Journal of Business Research (2018)
8. Hand, D.J.: Principles of data mining. Drug safety 30(7), 621–622 (2007)
9. Hsu, C., Chang, C., Lin, C.: A practical guide to support vector classification (2003)
10. Lunardon, N., Menardi, G., Torelli, N.: Rose: A package for binary imbalanced learning. R Journal 6(1) (2014)
11. Manjarres, A.V., Sandoval, L.G., Suárez, M.: Data mining techniques applied in educational environments: Literature review. Digital Education Review (33), 235–266 (2018)
12. Murphy, K.P.e.a.: Naive bayes classifiers. University of British Columbia 18 (2006)
13. Natekin, A., Knoll, A.: Gradient boosting machines, a tutorial. Frontiers in neurorobotics 7, 21 (2013)
14. Rajni, J., Malaya, D.B.: Predictive analytics in a higher education context. IT Professional (4), 24–33 (2015)
15. Romero, C., Ventura, S.: Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 40(6), 601–618 (2010)
16. Xu, J., Moon, K.H., Van Der Schaar, M.: A machine learning approach for tracking and predicting student performance in degree programs. IEEE Journal of Selected Topics in Signal Processing 11(5), 742–753 (2017)

ISSN 1870-4069

Captura de atributos discriminativos

Alberto Tapia-Palacios, Darnes Vilariño, Beatriz Beltrán, María Josefa Somodevilla, José De Jesús Lavalle

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias de la Computación, Puebla, México

alberto.tapia.p@gmail.com, {darnes, bbeltran, mariasg, jlavalle}@cs.buap.mx

Resumen. En el siguiente artículo se describe como se resolvió la tarea 10 de SemEval 2018 llamada captura de atributos discriminativos, la cual consiste en encontrar la diferencia semántica entre la relación de dos conceptos como una característica. El modelo desarrollado se basa sobre el significado de la semántica y léxica de cada palabra para ampliar su conjunto de datos.

Palabras clave: Similitud semántica, semántica léxica, atributos discriminativos.

Capture of Discriminative Attributes

Abstract. This paper describes how the task 10 of SemEval 2018 called capture of discriminative attributes was solved, this consists to find the semantic difference between the relation of two concepts as a feature. The developed model is based on the meaning of the semantics and lexical of each word to expand the data set.

Keywords. Semantic similarity, lexical semantic, discriminative attributes.

1. Introducción

El avance de las tecnologías de hoy en día ha impulsado el desarrollo de mejorar la comunicación humano-computadora; además de que gran parte de la información se encuentra de forma digital en diferentes tipos de colecciones de datos, desde foros, *blogs*, *wikis* hasta redes sociales. Estas colecciones son inmensas, además de que se encuentran creciendo exponencialmente día a día gracias al impulso de internet, donde la información en su mayoría se encuentra sin ningún tipo de clasificación, generando que las investigaciones de la compresión y el uso de forma automática de los lenguajes naturales se vea incrementado en los últimos años.

Sin duda una tarea que es de las más vitales y con un alto grado de complejidad dentro del procesamiento del lenguaje natural (PLN) es la detección de similitud semántica textual entre pares de sentencias. La semántica es la disciplina que estudia el significado de las expresiones lingüísticas, por lo que es común separar el estudio de significado de las palabras y de expresiones más complejas, lo que se distingue como semántica léxica y semántica composicional respectivamente [10]. La semántica léxica es el estudio de todo lo relativo al significado de las palabras lo que quiere decir que existe una colección de datos como un diccionario léxico que incluye el significado de las palabras de la lengua [1].

Faruqui, Tsvetkov, Rastogi y Dyer [3] describen algunos de los problemas más comunes con las tareas de tipo de evaluación de similitud para los modelos de vectores de palabras, sugiriendo que el uso de estos modelos sin supervisión puede conducir a resultados incorrectos, de manera que se requiere especial atención a la hora de evaluación y desarrollo de la tarea [2].

Una posible solución a este problema propuesta por Krebs y Papermo [5] consiste en extraer diferencias semánticas entre las palabras, al referirse a sus atributos que componen a cada palabra, de manera que una diferencia se puede expresar como la presencia o ausencia de un atributo en específico, así que ocupar atributos discriminativos puede brindar mejores resultados.

Tratando los atributos de un objeto o entidad como variable proporciona la ventaja de utilizar etiquetas de clasificación para diferenciar la información y poder predecir resultados más fiables [11,4,6].

El articulo está conformado por una sección de la descripción de la tarea, la siguiente sección trata acerca de la metodología, continuando con el análisis de resultados y finalmente tener la conclusión de la investigacón desarrollada.

2. Descripción de la tarea y conjunto de datos

La tarea de captura de atributos discriminativos consiste en una tarea de clasificación, que dado un par de palabras y un atributo discriminativo permite clasificar si tiene relación o no con la primera palabra. Entonces para cada par de palabras se verifica que la primera (pivote) tenga relación con el atributo proporcionado, pero la segunda (comparación) no tenga relación alguna con el atributo, para agregarlo a lista de ejemplos candidatos positivos.

- *palabra1* (pivote),
- *palabra2* (comparación),
- *atributo*,
- etiqueta (1 si el atributo es característica de *palabra1* pero no de *palabra2*, 0 en otro caso).

Un ejemplo de lo anterior seria *palabra1*: airplane (avión), *palabra2*: helicopter (helicóptero) y *atributo*: wings (alas), donde la *palabra1* tiene relación con el *atributo* alas, pero no la *palabra2*, ya que un helicóptero no tiene alas; para el caso de *palabra1*: helicóptero, *palabra2*: avión, *atributo*: alas se tomará en otra lista, ya que la *palabra1* no tiene relación con el *atributo*, se pueden ver algunos ejemplos en la Tabla 1. Para la lista de ejemplos negativos se tomarán en cuenta los casos donde ambas palabras tengan relación con el atributo y donde las dos palabras no tienen relación alguna con el atributo, para este último caso la cantidad de ejemplos es muy grande.

Tabla 1. Algunos ejemplos del conjunto de datos.

Palabra1	Palabra2	Atributo
airplane	helicopter	wings
bagpipe	accordion	pipes
canoe	sailboat	fibreglass
dolphin	seal	fins
gorilla	crocodile	bananas
oak	pine	leaves
octopus	lobster	tentacles
pajamas	necklace	silk
skirt	jacket	pleats
subway	train	dirty

Dentro del conjunto de datos proporcionado se van a utilizar una parte para el entrenamiento (training), otra para la validación (validation) y finalmente para las pruebas (tests) para su desarrollo, esta información se encuentra detallada en la Tabla 2.

Tabla 2. Información acerca del conjunto de datos.

Conjunto de datos	Ejemplos	Negativos	Positivos	Atributos	Negativos	Positivos
Entrenamiento	17,782	11191	6591	1,292	1290	333
Validación	2,722	1358	1364	576	383	410
Pruebas	2,340			577		

Los conjuntos de datos proporcionados fueron de entrenamiento y validación, las pruebas para entrenar el modelo fueron con el conjunto de datos de entrenamiento y se ajustaron algunos parámetros con el conjunto de datos de validación, el conjunto de datos de pruebas no contiene etiquetas, este se tiene que enviar a la página de *CodaLab* para su evaluación. A contiuación, se explica la metodología utilizada.

3. Metodologia

Los resultados presentados para la tarea 10 de evaluación se obtuvieron a partir de un modelo propuesto de clasificación, el cual se basa sobre los principios de semántica léxica descritos por Oana [1] y Escandell [10], asi como los problemas que mencionan Faruqui et al. [3].

Algo fundamental que utiliza el modelo propuesto son diccionarios de datos, que a partir del conjunto de entrenamiento se encarga de clasificar los datos para su correcto almacenamiento en los diccionarios asignados, como se mencionó en la sección de descripción de la tarea y conjunto de datos, estos tienen una etiqueta que los diferencia (0,1), donde 1 corresponde a que *palabra1* tiene relación con el *atributo* pero no la *palabra2*, sin embargo para cualquier otro caso la etiqueta es 0, lo que tiene como consecuencia inmediata las opciones donde *palabra1* y *palabra2* no tienen ninguna relación con el *atributo*, también donde la *palabra1* y el atributo no tienen relación alguna, asi como la situación donde la *palabra1* y *palabra2* están relacionadas con el *atributo*, dentro de este conjunto también se encuentran combinaciones clave, que si son ignoradas puede disminuir el resultado en la precisión del modelo.

Cabe hacer mención que durante la extracción de datos para el entrenamiento se encontró que había mucho ruido en los datos proporcionados por lo que no eran consistentes, es decir que para el conjunto de entrenamiento (training) se encuentran entradas que no pueden ser clasificadas correctamente, una de las razones es que su orden de magnitud es mayor, dado que la mayoría de los ejemplos son negativos, además de ser generado automáticamente por lo que algunas entradas pueden ser incorrectas, esto con el fin de tener un entrenamiento del sistema rico en parámetros [5].

Una vez que los diccionarios de datos contienen la información del conjunto de entrenamiento estos se utilizan de referencia para futuras predicciones del conjunto de pruebas, los diccionarios principalmente utilizados son los de las listas de palabras-atributos positivos (*dicPos*) y negativos (*dicNeg*), los cuales tienen una llave primaria (*Key*) conformada por una palabra y un atributo, esto con el fin de agilizar las búsquedas; también hay un control sobre las palabras a buscar con el diccionario *dicWeb*, este consiste en que al hacer una búsqueda de una palabra-atributo esta quede registrada y no se repita el proceso de clasificación, ver Tabla 3.

Tabla 3. Elementos utilizados.

Tipo	**Lista**	**Key**
Diccionario	positivo	Palabra, Atributo
Diccionario	negativo	Palabra, Atributo
Diccionario	Web	Palabra

Cada entrada del conjunto de pruebas se busca en los diccionarios creados a partir del conjunto de entrenamiento, para poder determinar si existe y si es un atributo discriminativo palabra-atributo en alguno de los diccionarios, en caso contrario de no encontrar ninguna coincidencia en los diccionarios se procederá a realizar una búsqueda con los siguientes criterios:

- Significado de la palabra, la definición de la palabra con palabras de la misma lengua.
- Primer Sinónimo de la palabra.
- Hiperónimo de la palabra, es decir es la relación existente de una palabra cuyo significado engloba otras palabras, ejemplo Árbol.
- Hipónimos son aquellas palabras que señalan, de una manera. específica y precisa, a todos los seres que pertenecen al mismo conjunto, género o clase, ejemplo Álamo, roble, pino (son tipos de árboles).
- Miembros holónimos Son aquellas palabras que señalan el todo de una estructura, ejemplo bicicleta.
- Miembros merónimos son aquellas partes que representan algunas partes, pero no todas tienen el mismo tipo de cohesión con respecto al conjunto, ejemplo ruedas, manubrio, pedal (son partes de una bicicleta).

La búsqueda de los criterios anteriores devuelve un diccionario con el contenido encontrado, para el caso del significado de la palabra se eliminan las palabras vacías, este diccionario contiene la palabra de la búsqueda y sus atributos encontrados por el modelo, para posteriormente añadir las combinaciones a la lista positiva y ampliar el conjunto de datos lo más posible para futuras predicciones. En caso de no encontrar la combinación con los criterios descritos anteriormente, seguirá realizar una búsqueda en los sitios *dictionary* y *Wikipedia* con la ayuda de las herramientas *request* y *Wikipedia API*, los cuales amplían considerablemente la lista de elementos positivos, de igual manera los resultados solo envían palabra-atributo, eliminando las palabras vacías encontradas en la búsqueda; se agrega la palabra al diccionario *dicWeb* para no repetir nuevamente el proceso.

compareWorAtr Extrae todas las características (Atributos) de una palabra de acuerdo con el significado del diccionario del corpus de *WordNet*, elimina las palabras vacías y agrega sinónimos (actualiza el diccionario principal y muestra si tienen relación las palabras).

comparewordfeature Busca el significado de la palabra en sitios web de dictionary.com y wikipedia.org, elimina palabras vacías y devuelve un diccionario con la palabra y atributos encontrados.

A continuación, se presenta el algoritmo del modelo propuesto:

Algoritmo 1. Modelo propuesto.

```
Inicio:
 dicPos={} //tipo diccionario
 dicNeg={}
 dicWeb={}
 archivo=abrirArchivo(validation.txt)
 para linea en archivo leerlinea:
```

```
   //linea es de la forma palabra1,palabra2,atributo
   linea=separar(',') //ahora es palabra1, palabra2, atributo
   w1aPos=(existe (palabra1,atributo) en dicPos) //verdad ó falso
   w2aPos=(existe (palabra2,atributo) en dicPos)
   w1aNeg=(existe (palabra1,atributo) en dicNeg)
   w2aNeg=(existe (palabra2,atributo) en dicNeg)
   Si not(w1aPos): //w1a = falso
     w1a=compareWorAtr(palabra1,atributo)
     //buscar la palabra con requests y wikipedia api
     Si not(w1a) & not(existe (palabra1) en dicWeb):
       w1a,features=compare_word_feature(palabra1,atributo)
       dicPos = actualizaDiccionario(features)
       dicWeb[palabra1]=1 //fue buscado en web
       Si not(w1a) & not(existe (atributo) en dicWeb):
         w1a,features=compare_word_feature(atributo,palabra1)
         dicPos = actualizaDiccionario (features)
         dicWeb[atributo]=1 //fue buscado en web
   sino: //w1a = verdadero
     w1a=Verdadero
 Si not(w2aPos): //w2a = falso
   w2a=compareWorAtr(palabra2,atributo)
   Si not(w2a) & not(existe (palabra2) en dicWeb):
     w2a,features=compare_word_feature(palabra2,atributo)
     dicPos = actualizaDiccionario(features)
     dicWeb[palabra2]=1 //fue buscado en web
     Si not(w2a) & not(existe atributo en dicWeb):
       w2a,features=compare_word_feature(palabra2,atributo)
       dicPos = actualizaDiccionario (features)
       dicWeb[palabra2]=1 //fue buscado en web
 sino:
   w2a=Verdadero
 Si ((w1aPos & w2aNeg) || (w1aPos & not(w2a)) ||
     (w1a & w2aNeg) || (w1a & not(w2a)) ):
   imprimir("palabra1,palabra2,atributo,1") //Es atributo discriminativo
 sino: //w1aNeg || (w1aPos and w2aPos)
   imprimir("palabra1,palabra2,atributo,0") //No es atributo discriminativo
 archivo cerrarArchivo()
Fin
```

Para la implementación de lo mencionado anteriormente se utilizaron las herramientas y bibliotecas basadas en Python.

WordNet es una gran base de datos léxica de inglés. Los sustantivos, verbos, adjetivos y adverbios se agrupan en conjuntos de sinónimos cognitivos (synsets), cada uno expresando un concepto distinto. Los sintonizadores están interrelacionados por medio de relaciones semántico-conceptuales y léxicas. Utilizándose como un recurso principal para ampliar el conjunto de datos con los elementos mencionados anteriormente [8].

NLTK[1] es una plataforma que trabaja con datos de lenguaje humano que proporciona un corpus, junto con un conjunto de bibliotecas de procesamiento de texto para clasificación, tokenización, derivación, etiquetado, análisis y razonamiento semántico, envoltorios para bibliotecas de PLN, entre otros. Utilizándose como herramienta indispensable para implementar las funciones y corpus de *WordNet*, para la obtención de la lista de palabras vacías, sinónimos, hiperónimos, hiponimos por mencionar algunos[7].

Spacy[2] tiene modelos de etiquetado, análisis sintáctico y reconocimiento de entidades, se utilizó junto con el modelo "en vectors web lg" de lenguaje inglés que contiene 300 vectores dimensionales de palabras entrenadas en rastreo común con Glove[3] [9].

Textblob[4] proporciona métodos simples de implementar la integración de *WordNet*, así como etiquetado parcial, extracción de frase nominal, análisis de sentimiento, clasificación, traducción, por mencionar algunos.

Requests[5] es una biblioteca HTTP que se utilizó para obtener la información de una página web de manera simple.

BeautifulSoup[6] es una biblioteca para extraer datos de archivos html y xml. Se utilizó en conjunto con la biblioteca *Requests* para obtener solo la información necesaria de una página web, es decir solo su contenido relevante.

Wikipedia Api[7] obtiene el contenido de una consulta desde la página de Wikipedia, lo que ayuda a obtener las definiciones de las palabras o atributos que no encuentre el modelo, ampliando de manera considerable el conjunto de datos.

4. Análisis de resultados

Como mencionan Faruqui y Dyer [2] la evaluación de similitud semántica entre palabras supone que existe una sola noción de similitud, por lo que utilizar algunos de los diferentes modelos existentes mostrara resultados diferentes independientemente del problema a resolver, de igual manera no existe un estándar para la representación de palabras en vectores, lo que tiene como consecuencia que en algunos conjuntos de datos resulte difícil encontrar diferencias significativas.

El entrenamiento del modelo utilizo el conjunto de datos *training* por otra parte el conjunto de pruebas utiliza el conjunto de datos *validation*, en la Tabla 4

[1] http://www.nltk.org
[2] https://spacy.io/
[3] Algoritmo de aprendizaje no supervisado para obtener representaciones de vectores para palabras.
[4] https://textblob.readthedocs.io/en/dev/
[5] http://docs.python-requests.org/en/master/user/quickstart/
[6] https://www.crummy.com/software/BeautifulSoup/
[7] http://wikipedia-api.readthedocs.io/

se muestran los resultados obtenidos. Como se puede observar se hicieron pruebas con los modelos de similitud de:

WordNet La relación principal entre las palabras en WordNet es la sinonimia, los sinónimos (palabras que denotan el mismo concepto y son intercambiables en muchos contextos) se agrupan en conjuntos no ordenados (synsets). Cada uno de los 117 000 synsets de WordNet está vinculado a otros synsets por medio de un pequeño número de relaciones conceptuales.
Spacy 300 vectores de palabras tridimensionales entrenados en el rastreo común con Glove, con 1.1m de llaves y 1.1m de vectores únicos (300 dimensiones).

Los modelos anteriormente mencionados utilizan los mismos conjuntos de datos, con lo que muestran los resultados de 50.44 y 53.41 de precisión para NLTK y Spacy respectivamente, mientras que el modelo propuesto tiene un 62.41 de precisión, con esto se puede observar la mejora en los resultados obtenidos al utilizar atributos discriminativos en la tarea de similitud.

Tabla 4. Resultados *train - validation*.

Total del archivo						
	WordNet		**Spacy**		**Modelo**	
Validación	**Líneas**	**Porcentaje**	**Líneas**	**Porcentaje**	**Líneas**	**Porcentaje**
Total de líneas	2722		2722		2722	
Líneas positivas	1364	50.11	1364	50.11	1364	50.11
Líneas negativas	1358	49.89	1358	49.89	1358	49.89
Líneas erróneas						
	WordNet		**Spacy**		**Modelo**	
Validación	**Líneas**	**Porcentaje**	**Líneas**	**Porcentaje**	**Líneas**	**Porcentaje**
Líneas positivas (1,0)	1308	95.90	1189	87.18	901	66.06
Líneas negativas (0,1)	41	3.02	79	5.82	122	8.99
Total de líneas erróneas	1349	49.56	1268	46.59	1023	37.59
Líneas acertadas						
	WordNet		**Spacy**		**Modelo**	
Validación	**Líneas**	**Porcentaje**	**Líneas**	**Porcentaje**	**Líneas**	**Porcentaje**
Líneas positivas	56	4.10	175	12.82	463	33.94
Líneas negativas	1317	96.98	1279	94.18	1236	91.01
Total de líneas acertadas	1373	50.44	1454	53.41	**1699**	**62.41**

Para los resultados obtenidos con el modelo propuesto se utiliza el conjunto de entrenamiento *training* y *validation*, para el conjunto de pruebas se utiliza test, sin embargo, este no contiene las etiquetas de respuesta, por lo que se tiene que enviar a *CodaLab* para su evaluación en línea. Los resultados de la evaluación del modelo en la página de CodaLab[8] obtuvieron una calificación correcta de 0.63 de precisión.

5. Conclusión

En este artículo presentamos nuestro modelo como solución a la tarea de captura de atributos discriminativos entre pares de palabras, basado principal-

[8] `https://competitions.codalab.org/competitions/17326#results`

mente en la extracción de características según el significado o definiciones de una palabra, basado en semántica léxica. Los resultados obtenidos muestran que el tamaño del diccionario o conjunto de datos que utiliza cada modelo depende mucho de sus resultados, por ese motivo la precisión que resulta de cada uno varía según su tamaño y forma en que clasifica la similitud entre pares de palabras. De igual manera el conjunto de entrenamiento *training* es inconsistente, al ser generado automáticamente, además de tener una gran cantidad de ejemplos negativos, por lo cual el diccionario negativo es de extensas dimensiones. Se planea extender el modelo propuesto al añadir entrenamiento con vectores de palabras, con lo cual se espera aumentar el grado de precisión.

Referencias

1. Duţă, O.: Semántica léxica y oposiciones de sentido: un enfoque teórico. Annals of the University of Craiova. Series Philology. Linguistics. Recuperado de www. ceeol. com (2009)
2. Faruqui, M., Dyer, C.: Community evaluation and exchange of word vectors at wordvectors.org. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. pp. 19–24. Association for Computational Linguistics (2014), `http://www.aclweb.org/anthology/P14-5004`
3. Faruqui, M., Tsvetkov, Y., Rastogi, P., Dyer, C.: Problems with evaluation of word embeddings using word similarity tasks. CoRR abs/1605.02276 (2016), `http://arxiv.org/abs/1605.02276`
4. Guo, Y., Ding, G., Jin, X., Wang, J.: Learning predictable and discriminative attributes for visual recognition. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. pp. 3783–3789. AAAI'15, AAAI Press (2015), `http://dl.acm.org/citation.cfm?id=2888116.2888241`
5. Krebs, A., Paperno, D.: Capturing discriminative attributes in a distributional space: Task proposal. In: RepEval@ACL (2016)
6. Lazaridou, A., Pham, N.T., Baroni, M.: The red one!: On learning to refer to things based on their discriminative properties. CoRR abs/1603.02618 (2016), `http://arxiv.org/abs/1603.02618`
7. Loper, E., Bird, S.: Nltk: The natural language toolkit. In: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1. pp. 63–70. ETMTNLP '02, Association for Computational Linguistics, Stroudsburg, PA, USA (2002), `https://doi.org/10.3115/1118108.1118117`
8. Miller, G.A.: Wordnet: A lexical database for english. COMMUNICATIONS OF THE ACM 38, 39–41 (1995)
9. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Empirical Methods in Natural Language Processing (EMNLP). pp. 1532–1543 (2014), `http://www.aclweb.org/anthology/D14-1162`
10. Vidal, M.: Apuntes de semántica léxica. Cuadernos de la UNED, Universidad Nacional de Educación a Distancia (2007), `https://books.google.com.mx/books?id=RWOGPgAACAAJ`
11. Wang, Y., Mori, G.: A discriminative latent model of object classes and attributes. In: Proceedings of the 11th European Conference on Computer Vision: Part V. pp. 155–168. ECCV'10, Springer-Verlag, Berlin, Heidelberg (2010), `http://dl.acm.org/citation.cfm?id=1888150.1888163`

ISSN 1870-4069

Evolution of Modern Deep Learning Methods of Object Recognition

Pritom Hazarika, Ms Madhu Kumari

NIT Hamirpur, Himachal Pradesh, India

mimikaan@gmail.com, madhu.jaglan@gmail.com

Abstract. Vision is a very complex work of our brain. 90% of the information entered to our brain is related to vision. In computer vision engineers and scientists are trying to give the computers the ability of vision. Object recognition is one of the most exciting fields of computer vision and AI. There is no rarity of problems and challenges in object recognition, from image classification to key-point detection. But like many other problems in the world, there is still no obvious and "Best" way to resolve these problems. But the recent advancement in GPU technology propelled the success and accuracy of deep learning algorithms. In this paper, we will study about the evolution of deep learning methods in object recognition. We will also go through some of the classical machine learning methods for object recognition and talk about multimodal approaches to solving these issues.

Keywords. Computer vision, AI, object recognition, GPU, deep learning, machine learning.

1 Introduction

Object recognition is one of the most exciting fields of computer vision and AI. There are countless practical application of object recognition to solve real-world problems. Face detection, people counting, anomaly detection, web image classification, self-driving cars, video surveillance are some practical example where object recognition is used. In recent years image classification techniques surpassed the human ability.

There are multiple subfields of object recognition. Image classification, where an image is classified into many different categories. Object Localization, which is similar to classification. Localization predicts the location of a dominant object inside the image. Object detection, it is the combination of localization and classification. It is the process of finding and classifying multiple numbers of objects on an image. Instance Segmentation, here we not only find objects inside an image but label each pixel of an image by object class and object in the region. Key-point detection, detect locations of a set of predefined key points of an object in an image, such as the human body or a human face.

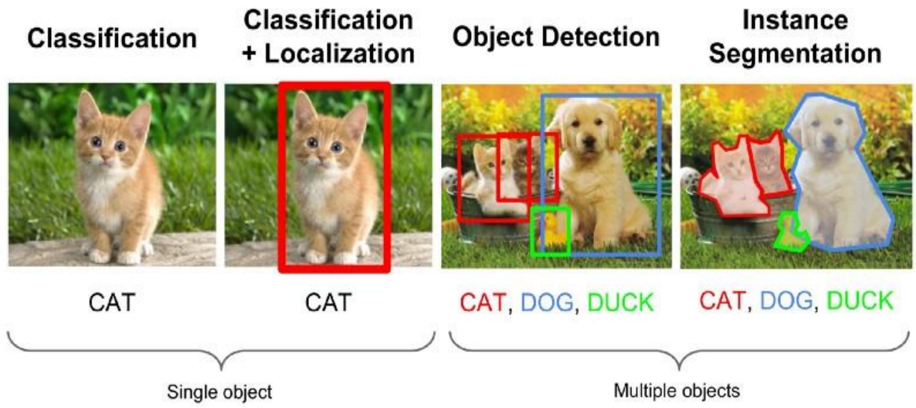


Fig. 1. Object recognition problems.

In 2001 Paul Viola and Michael Jones [10] invented a simultaneous face detection algorithm allowing for a human figure to be identified through their facial traits. Navneet Dalal and Bill Triggs [11] published a Histograms of Oriented Gradients (HOG) in 2005 which theories a feature detector for the recognition of pedestrians in security system circuits. The modern era of object recognition starts with the development of the convolutional neural network (CNN).

In 2012 Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton introduce a new algorithm called AlexNet [1] ensuring an 85% level of accuracy. One of the first deep learning method used for object detection was Overfeat [7] in 2013 where P Sermanet proposed a multi-scale sliding window algorithm using CNN. Quickly after that Ross Girshick, et al. proposed a Region based convolutional neural network which is a combination of heuristic region proposal method and CNN feature extractor [2].

In 2015 Ross Girshick proposed Fast R-CNN [3]. It applied the CNN on the complete image and then used both Region of Interest (RoI) Pooling on the feature map with a final feedforward network for classification and regression to generate object proposal, unlike R-CNN which use selective search independently and use Support Vector machine as a classifier.

YOLO is published in 2015 by Joseph Redmon which use CNN [5]. YOLO's level of accuracy in facial recognition exceeded 95% and it was also very fast, which allows for the very first time to use facial recognition in real time. Subsequently, Shaoqing Ren co-authored by Girshick proposed the third iteration of the R-CNN series the Faster R-CNN [4]. In Faster R-CNN they make the model trainable end to end by adding region proposal network (RPN).

2 Challenges

There are a lot of issues and challenges related to object recognition.

2.1 Variable Number of Objects

Usually while training a machine learning modal we need to represent data into fixed-sized vectors. If we don't know the number of object in an image, we can't tell the correct number of output. It will create a problem determining the vector size. Post-processing is required to solve this problem but post-processing increase the complexity of the model.

2.2 Sizing

All the objects in an image are not of the same size. The size difference of the objects is a big challenge in object recognition model. When doing classification we generally want to recognize the object covering the large portion of the image. But sometimes some object covers a comparatively small portion of an image. This problem could be solved using variable size sliding windows but this solution is very inefficient.

2.3 Modelling

Solving two problem at the same time is another challenge. Combining classification and localization into a single model is a challenge.

2.4 Illumination

Depending on the lighting condition the same object may look different on different images. The system must be able to recognize the object irrespective of the lighting condition.

2.5 Occlusion

Sometimes objects on the image are not completely visible. Some objects are partially covered by some other object. Model must be able to handle these situation.

Noise, blurry picture, deformation, interclass variation, background clutter etc. are also some problems faced in object recognition.

3 Object Recognition Methods

3.1 Classical Methods

Over the years there are many different types of methods proposed to solve object recognition problems. But two methods stands out of all. First on is in 2001 by Paul Viola. He published a paper "Robust Real-Time Object Detection". This approach is fast and relatively simple. It is being applied in point and shoot cameras to detect faces in real time. This method creates different binary classifier by using Haar feature then these binary classifiers are assessed with a multi-scale sliding window in cascade and dropped early in case of a negative classification.

The second method is the Histogram of oriented Gradient or HOG feature proposed by Navneet Dalal and Bill Triggs. They use Support Vector Machine (SVM) for classification. It involves a multiscale sliding window similar to Paul's method. In terms of accuracy, it is superior to the first method but it is much slower than Paul's method.

3.2 Deep Learning Methods

Deep learning revolutionize the field of machine learning, especially computer vision. Deep learning models have outperformed the other classical methods of object recognition. Modern history or object recognition stared in 2012 with the development of the convolutional neural network. It all started when AlexNet won the ILSVRC 2012 by a large margin. AlexNet was based on the decades-old LeNet, combined with data augmentation, rectified linear unit (ReLU), dropout, and GPU implementation. It proved the effectiveness of a convolutional neural network, it opened a new era for computer vision.

OverFeat: In 2013 P Sermanet from NYU proposed a multiscale sliding window algorithm using AlexNet to extract feature from an input image.

R-CNN: Region-based convolutional neural network (R-CNN) is a natural combination of heuristic region proposal method and CNN feature extractor. From an image, possible objects are extracted using a region proposal method like selective search. Those regions are then cropped and warped to a fixed size. CNN is used to extract features from each region. Then a support vector machine model is trained to classify each region. Training an R-CNN is a difficult process although it can achieve great results.

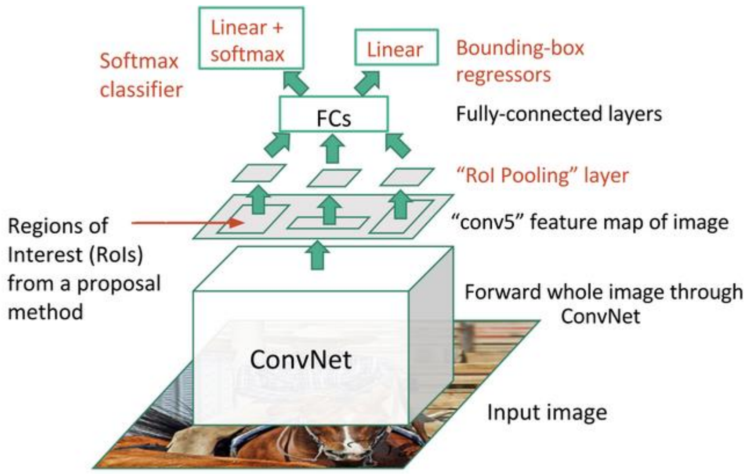


Fig. 2. Fast R-CNN.

Fast R-CNN: Fast R-CNN is similar to R-CNN. Like R-CNN, it also uses selective search to extract possible object. But difference comes in feature extraction step. Instead of applying SVM in individual region it applies CNN in the entire image to extract feature. And applies the Region of Interest (ROI) pooling on the feature map with a final feed forward network for classification and regression. The biggest disadvantage of this system is, it is still relied on selective search for region proposal.

YOLO: In 2015 Joseph Redmon published a paper You Only Look Once: Unified, Real-Time Object Detection (YOLO). YOLO is a development of multibox which is a CNN based region proposal solution. It Convert multibox from a region proposal system to object recognition system by adding softmax layer parallel to the box regressor and box classifier layer, to directly predicts the object class. It gives great result as well as high speed.

Faster R-CNN: Faster R-CNN is a Fast R-CNN where selective search is replaced by Region Proposal Network (RPN) for region proposal. RPN is also inspired by multibox. This makes the modal completely trainable from end to end.

SSD: Single shot detector uses the RPN of Faster R-CNN [6]. In Faster RCNN, RPN is used to give object confidence score but here it directly uses the RPN to classify object inside the prior box.

Mask R-CNN: It is a modified Faster R-CNN for segmentation. Here a Branch is added for predicting class specific object mask [8]. Mask RCNN replace RoIPooling with RoIAlign since the previous technique was not designed for pixel to pixel alignment.

3.3 Multimodal Methods

Deep learning methods of object recognition have impressive results. But to train a deep learning model we need a lot of data. For classification problems we need labelled data. But labelling image is time consuming task. Images from internet are sometimes not very well labelled or inaccurately. Multimodal machine learning is the solution to this problem. Modality refers to the type of information or data representation format in which information is stored. The way we perceived the world multimodal. We see things, we hear sounds, smell odours, feel texture. This is how we infer knowledge from the world. Multimodal machine learning is based on the same concept. Information from different modality compliments each other. For example, image classification and image captioning models relies on labelled input data. But labels maybe incorrect or unavailable. In such cases descriptions, tags available along with the image could be used to train the model. There are some core technical challenges in multimodal learning [9]:

- Representation,
- Alignment,
- Translation,
- Fusion,
- Co-learning.

Due to the heterogeneity of multimodal data, it is difficult to construct a representation to represent the data from different modal which can exploits the complementarity and redundancy of multimodality. There are two major types of multimodal representation –joint representation and coordinated representation. Joint representation projects uni-modal representations together into a multimodal space.

Mathematically, joint representation is expressed as

$$X_m=f(x_1,x_2,\ldots\ldots,x_n). \tag{1}$$

Instead of projecting modality into joint space, coordinated representation learn separate representation for each modality but coordinate them through a constraint.
Mathematically, coordinated representation is expressed as

$$f(x_1)\sim g(x_2). \tag{2}$$

Multimodal deep Boltzmann Machine is an example of joint representation. It is a graphical model based representation which stack restricted Boltzmann machines as building blocks. Canonical correlation analysis (CCA) is a correlated representation model. It computes linear projection which maximize the correlation between two random modalities and enforces orthogonality of the new space.

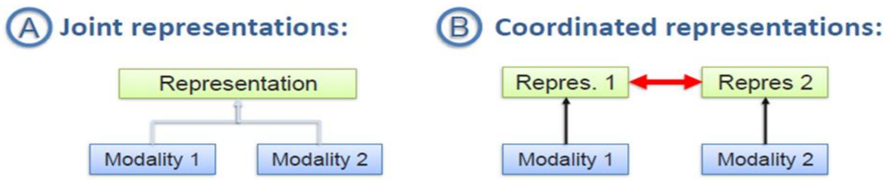


Fig. 3. Multimodal representation.

Task of identifying direct relationship between elements from different modality is known as alignment. Multimodal alignment are categorize into two parts – implicit and explicit. In explicit alignment the goal is to directly find correspondences between elements of different modalities, whereas in implicit alignment the alignment is used as an intermediate step for another task. There are two types of algorithm that handles explicit alignment- supervised and unsupervised. There are two types of implicit alignment – graphical model and neural network.

Fusion is the task of joining information from two or more modality to perform prediction. Multimodal fusion could be classified in to two broad category. Model-agnostic and model based. Model –agnostic fusion can be divided as early fusion, late fusion and hybrid fusion. Model based fusion can be categories as multiple kernel learning (MKL), graphical models and neural networks. MKL approach is a popular method for fusing visual descriptor for object detection.

Translation is the task of translating data from one modality to another. Data from different modality are heterogeneous and relationship between modalities is often open ended and subjective. There could be multiple correct answers for the same problem. Translation could be categorized as example based and generative. Example based model used a dictionary while translating between modals. Generative models constructs models that is able to produce translation.

The last challenge of multimodal machine learning is Co-Learning it is concern with transfer of knowledge between multiple modality. Co-learning is categorized into parallel, non-parallel and hybrid data. For object recognition task primarily non-parallel and hybrid data is used. Frome et al. used transfer learning method and used text to improve visual representations for image classification by synchronizing CNN visual feature with word2vec textual feature. Zero shot learning is another popular algorithm for image classification. Here to recognize any concept the model need not see any explicit example of that concept.

4 Important Dataset

Table 1. Dataset table.

Name	Images	Classes
ImageNet	450K	200
COCO	120K	80
Pascal VOC	12K	20
Oxford-IIIT Pet	7K	37
KITTI Vision	7K	3
MNIST	70K	10
Open Images Dataset	900K	5K
SVHN	630K	10

5 Conclusion

Object recognition gives the computer and robots the ability to see. Computer scientists have been working in computer vision technology since 1966 when students of MIT are asked to solve the human vision problem as a summer project. Since then there has been a lot of progress in this field. At the core of all computer vision problem is the task of classification. Since the images are represented as a 3D array of numbers, with integers from 0 to 255, there is a semantic gap. Besides, there are challenges such as Illumination, Deformation, Occlusion, Background Clutter and Interclass variation.

In recent years classical machine learning methods for object recognition are being outperformed by deep learning methods which have achieved accuracies that are far beyond that of classical ML methods. Deep learning methods scale effectively with data. The feature engineering is not required in deep learning as it was required in classical ML methods. Classical ML methods also have some advantages over deep learning methods. Classical methods work better on small data. It is also cheap financially and computationally compared to deep learning methods. Also, classical ML methods are easier to understand.

Multimodal methods took advantage of redundant information from different modalities. The use of new sensing modalities, in particular depth and thermal cameras, has seen some development in the recent years [e.g., Fehr and Burkhardt (2008) and Correa et al. (2012)]. Classical ML and deep learning methods are applied to multiple modalities of data to solve new problems such as caption generation, depth sensing, segmentation using thermal and depth camera.

Still, there are some problems which we believe have not been addressed, or addressed partially, and may be interesting relevant research directions. Open world learning and active vision is one such area. Another area which is not being addressed completely is Pixel-Level Detection (Segmentation) and Background Objects. Another basic dilemma during the detection process is, should we detect the object first or the parts first? And there are no clear solution exists yet.

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12), Vol. 1, Curran Associates Inc., USA, 1097–1105 (2012)
2. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14), IEEE Computer Society, Washington, DC, USA, 580–587 (2014) DOI: https://doi.org/10.1109/CVPR.2014.81
3. Girshick, R.: Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) (ICCV '15), IEEE Computer Society, Washington, DC, USA, 1440–1448 (2015) DOI=http://dx.doi.org/10.1109/ICCV. 2015.169
4. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'15), Vol. 1, MIT Press, Cambridge, MA, USA, 91–99 (2015)
5. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 779–788 (2016)
6. Liu., W. Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C., Berg, A.C.: SSD: Single Shot MultiBox Detector. ECCV (2016)
7. Sermanet, P., Eigen, D., Zhang, X., et al.: OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks CoRR. Vol. abs/1312.6229 (2013)
8. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In Proceedings IEEE International Conference on Computer Vision (ICCV), Venice, pp. 2980–2988 (2017)
9. Baltrušaitis, T., Ahuja, C., Morency, L.: Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence
10. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. Int. J. Comput. Vision 57, 2, 137–154 (2004) DOI: https://doi.org/10.1023/B:VISI.0000013087.49260.fb

11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In Proceeding IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, vol. 1, pp. 886–893 (2005) DOI: 10.1109/CVPR.2005.177

ISSN 1870-4069

Hill Algorithm Decryption using Parallel Calculations by Brute Force

Bárbara Emma Sánchez Rinza, Juan Carlos García Lezama

Benemérita Universidad Autónoma de Puebla,
Facultad de ciencias de la computación, Puebla, Mexico

brinza@hotmail.com

Abstract. Hill coding, based on linear algebra, by the American mathematician Lester S. Hill in 1929 in this method we use a square matrix A of integers as a key, which determines the linear transformation Y = A * X where Y, X they are the column vectors. Using this encryption method, a text was encrypted to later decrypt it with the use of brute force, that is, to test each of the possible combinations of keys to find the original text in this article. A 2x2 key was used to encrypt the text with a limit from 1 to 256 for each element in the matrix 256 x 256 x 256 x 256 permutations were found that is 4,294,967,296 possible keys for this decipher this text as it can be clearly seen there are too many operations to perform that can consume a considerable time for the CPU since he must decipher the text for each of these combinations and find the correct one, that is why to do this arduous task, parallel programming was used to generate each of the keys and work with each one of them.

Keywords. Cryptography, encrypt, keys, parallel programming.

1 Introduction

Cryptography comes from an etymological word Kriptos means “hidden”, Graphos means “writing”, which would mean “hidden writings”, or in its broadest sense it would be to apply some technique to make a message unintelligible [1].

The main objective is to encrypt and / or protect the information with an algorithm using keys, without them would be really difficult to obtain the original text. In these years the protection of the information is an indispensable need once saved on a computer, due to its use in great part of daily life.

Even worse, the Internet makes available a large number of people, devices that contain confidential information for each one of us, such as addresses, telephone numbers and financial information, among others.

2014 was a great scenario of major attacks on companies by hackers, which mostly represented great losses for companies, and a big a risk for millions of clients, including eBay, HOME DEPOT, SONY, CASINOS SANDS [2]

For a company, its information is the most important asset, without it, might bring a bankruptcy for example, a company can lose all the products due to a natural disaster and might be able to recover itself by looking for investors, loans, mortgaging its properties, etc., but if it loses all the customer lists, suppliers, debtors, etc. it can be catastrophic because of the great value of this information that would not be recovered.

2 Development

Encrypting data means altering them, through the use of a key or pair of keys, so that they are not readable for those who should not have access to information, that is to say, intruders. Through the decryption process for those who have the key, they can use it to obtain the original information.

This technique protects sensitive information such as personal data or bank accounts of an organization, guaranteeing its authenticity, integrity, and confidentiality, if the encrypted data is intercepted, it cannot be read or modified by intruders.

The cryptographic systems where the encryption and decryption key match, are called “symmetric encryption” that can be seen as the lock of a door where it is possible to open and close with the same key.

There are 3 basic encryption techniques from which all the classic systems of secret key are generated: **Transposition** (units order alteration of the original text according to a given key), **Substitution** (replacement of the original text units by others according to a key), and **Product** (Composition of several ciphers, substitution and / or transposition, each of which will depend on a key).

In this article we will focus on the Polyalphabetic Substitution Method since the algorithm to be treated is this type.

3 Substitution Ciphers

The substitution cipher consists in units of plain text (text without encryption) are replaced by units of encrypted text, there are different cipher substitution types. If the cipher operates on simple letters it is term simple substitution cipher; if the cipher operates on larger groups of letters it is called, polygraphic. A cipher is monoalphabetic when a character of the plaintext is replaced by one and only one of the ciphertext or polyalphabetic. An element in the plaintext can be represented by more than one character. [3]

3.1 Polyalphabetic

A polyalphabetic substitution system is when each character is not always replaced by the same character that is in the system there are several characters that could replace it and according to the circumstances would apply one or the other.

3.2 Hill Cipher

It is based on linear algebra developed by the mathematician Lester S. Hill in 1929 in his article Cryptography in an Algebraic Alphabet, published in The American Mathematical Monthly [4] is a cryptographic system of polyalphabetic substitution.

It consists in associate each alphabet letter with a number, for this article were used 27 characters of the alphabet. In this case, we did it for 256 characters.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
0	1	2	3	4	5	6	7	8	9	10	11	12	13
Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z	
14	15	16	17	18	19	20	21	22	23	24	25	26	

Fig. 1. Table of characters.

In Hill's cipher, a square matrix of integers A is used as a key, which determines the linear transformation Y = A X, where Y, X are column vectors.

Let's see an example. Consider the 3x3 square matrix (square matrices of any size can be taken) and the corresponding linear transformation Y = A X:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 1 & 0 & 6 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$\left.\begin{aligned} y_1 &= 1 \cdot x_1 + 2 \cdot x_2 + 3 \cdot x_3 \\ y_2 &= 0 \cdot x_1 + 4 \cdot x_2 + 5 \cdot x_3 \\ y_3 &= 1 \cdot x_1 + 0 \cdot x_2 + 6 \cdot x_3 \end{aligned}\right\}$$

Fig. 2. Multiplication of a matrix by a column vector.

Taking the plain text "HOLA MUNDO"

Whose numeric transcription according to the PREVIOUS table would be: 7,15,11,0,12,21,13,3,15

Since the linear transformation is in sequences of 3, we are going to group the numbers in three, then we will apply the linear transformation (72,111,108), (97, 32, 74), (117, 97,110).

Next, we are going to transform in a trigraph the previous numbers, through the linear transformation given by the key, into new trigraph that will be the cipher numeric message.

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 1 & 0 & 6 \end{pmatrix} \cdot \begin{pmatrix} 7 \\ 15 \\ 10 \end{pmatrix} = \begin{pmatrix} 67 \\ 110 \\ 67 \end{pmatrix} \equiv \begin{pmatrix} 13 \\ 2 \\ 13 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 1 & 0 & 6 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 12 \\ 21 \end{pmatrix} = \begin{pmatrix} 87 \\ 153 \\ 126 \end{pmatrix} \equiv \begin{pmatrix} 6 \\ 18 \\ 18 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 1 & 0 & 6 \end{pmatrix} \cdot \begin{pmatrix} 13 \\ 3 \\ 15 \end{pmatrix} = \begin{pmatrix} 412 \\ 393 \\ 361 \end{pmatrix} \equiv \begin{pmatrix} 21 \\ 6 \\ 22 \end{pmatrix}$$

Fig. 3. Multiplicative column vector and operation module 256.

Although the linear transformation of the trigraph (7, 15, 10) is initially (67, 110, 67), since we are working with integers module 27, this trigraph becomes (13, 2, 13), since 67 = 2 x 27 + 13 and 110 = 4 x 27 + 2. Results the same for the rest.

Therefore, the cipher numeric message is "13, 2, 13, 6, 18, 18, 21, 6, 22" by transforming the numbers again into their corresponding letters, it becomes into the cipher message: NCNGRRUGV

In order to decode encrypted messages using the Hill method, the matrix of the linear transformation used, the key, must be an invertible matrix. Our matrix example is, since its determinant is non-zero, | A | = 22. In addition, the inverse matrix of A, is the one needed to decode an encrypted message, is:

$$A^{-1} = \begin{pmatrix} \frac{24}{22} & \frac{-12}{22} & \frac{-2}{22} \\ \frac{5}{22} & \frac{3}{22} & \frac{-5}{22} \\ \frac{-4}{22} & \frac{2}{22} & \frac{4}{22} \end{pmatrix}$$

Fig. 4. Inverse Matrix key.

We are working with the integers module 27 and we are going to transform the previous inverse matrix into a matrix with integers modulo 27. To begin, we need the inverse of the number 22. We look for a number that multiplied by 22 the module is equal to 1 in the following way: 22 x 16 = 352 is equal to 1, module 27, then 1/22 = 16. And the inverse matrix is transformed, module 27 would be equal to:

$$A^{-1} = \begin{pmatrix} \frac{24}{22} & \frac{-12}{22} & \frac{-2}{22} \\ \frac{5}{22} & \frac{3}{22} & \frac{-5}{22} \\ \frac{-4}{22} & \frac{2}{22} & \frac{4}{22} \end{pmatrix} = \begin{pmatrix} 24\times16 & -12\times16 & -2\times16 \\ 5\times16 & 3\times16 & -5\times16 \\ -4\times16 & 2\times16 & 4\times16 \end{pmatrix}$$

$$= \begin{pmatrix} 384 & -192 & -32 \\ 80 & 48 & -80 \\ -64 & 32 & 64 \end{pmatrix} = \begin{pmatrix} 6 & 24 & 22 \\ 26 & 21 & 1 \\ 17 & 5 & 10 \end{pmatrix}$$

Fig. 5. Inverse matrix module numbers 27.

In order to decode the message it is necessary to use the same previous method, Hill's cipher, but using as inverse key matrix A-1 (module 27) of the coding matrix A.

In the same way, the encrypted message is written in terms of the associated number in Figure 1 (13, 2, 13), (6, 18, 18), (21, 6, 22) they are transformed by the linear transformation with matrix A ^ (-1), that is, Y = A ^ (- 1) · X.

$$\begin{pmatrix} 6 & 24 & 22 \\ 26 & 21 & 1 \\ 17 & 5 & 10 \end{pmatrix} \cdot \begin{pmatrix} 13 \\ 2 \\ 13 \end{pmatrix} = \begin{pmatrix} 412 \\ 393 \\ 361 \end{pmatrix} \equiv \begin{pmatrix} 7 \\ 15 \\ 11 \end{pmatrix}$$

$$\begin{pmatrix} 6 & 24 & 22 \\ 26 & 21 & 1 \\ 17 & 5 & 10 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ 18 \\ 18 \end{pmatrix} = \begin{pmatrix} 508 \\ 552 \\ 372 \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 12 \\ 21 \end{pmatrix}$$

$$\begin{pmatrix} 6 & 24 & 22 \\ 26 & 21 & 1 \\ 17 & 5 & 10 \end{pmatrix} \cdot \begin{pmatrix} 21 \\ 6 \\ 22 \end{pmatrix} = \begin{pmatrix} 736 \\ 694 \\ 739 \end{pmatrix} \equiv \begin{pmatrix} 13 \\ 3 \\ 15 \end{pmatrix}$$

Fig. 6. Inverse matrix transformation and encrypted vector.

The original sequence of the numerical trigraphs associated with the previous coded message is (7, 15, 10), (22, 12, 21), (7, 19, 10). And by translating the numbers to their corresponding letters of the alphabet you get that the original message sent is: HOLA MUNDO

As you can see it is a simple encryption and decryption algorithm knowing the correct key but what would happen if the key is not known? It could take too much time to find the correct key that is why this algorithm was implemented in a program written in the Java language with the use of thread to generate the possible permutations keys and at the same time to test each one of them.

4 Decryption in Java Language

In this application developed in Java language, the decryption of the Hill algorithm by brute force was implemented, that consists in testing in the worst case each of the possible keys of a set in a parallel way with 3 thread, the main shows the window featured in figure 7.

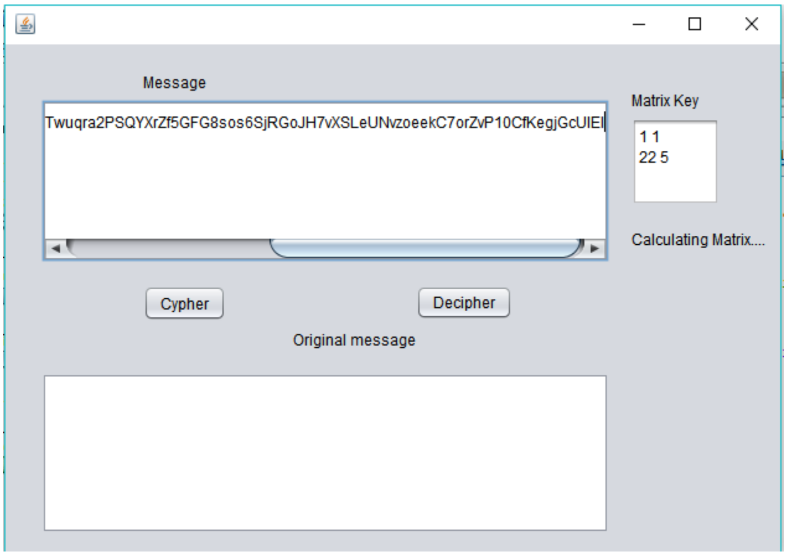


Fig. 7. Main window application.

```java
public static double determinante(double[][] matriz)
{
    double det;
    if(matriz.length==2)
    {
        det=(matriz[0][0]*matriz[1][1])-(matriz[1][0]*matriz[0][1]);
        return det;
    }
    double suma=0;
    for(int i=0; i<matriz.length; i++){
    double[][] nm=new double[matriz.length-1][matriz.length-1];
        for(int j=0; j<matriz.length; j++){
            if(j!=i){
                for(int k=1; k<matriz.length; k++){
                int indice=-1;
                if(j<i)
                indice=j;
                else if(j>i)
                indice=j-1;
                nm[indice][k-1]=matriz[j][k];
                }
            }
        }
        if(i%2==0)
        suma+=matriz[i][0] * determinante(nm);
        else
        suma-=matriz[i][0] * determinante(nm);
    }
    return suma;
}
```

Fig. 8. Algorithm where the determinant of a square matrix of n * n is calculated.

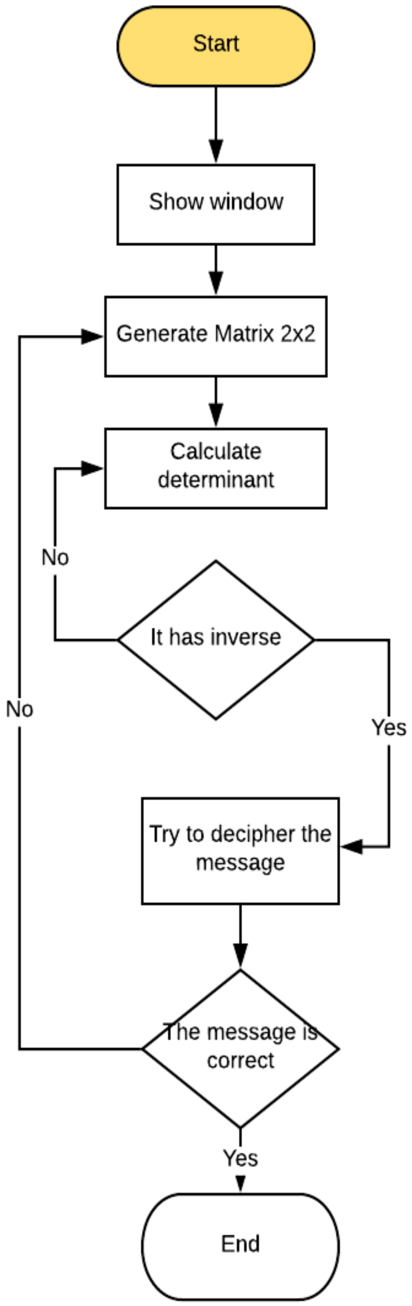


Fig. 9. Decryption process Flowchart.

In thread 2 the possible permutations were generated. 4,294,967,296 possible keys for a square 2x2 matrix from 1 to 256 in each of its elements and verify that this matrix has an inverse, which means that its determinant is different from 0.

In thread 3, the decryption process is done but only if the generated matrix has an inverse if it does not, goes back to generate a new matrix within the established range.

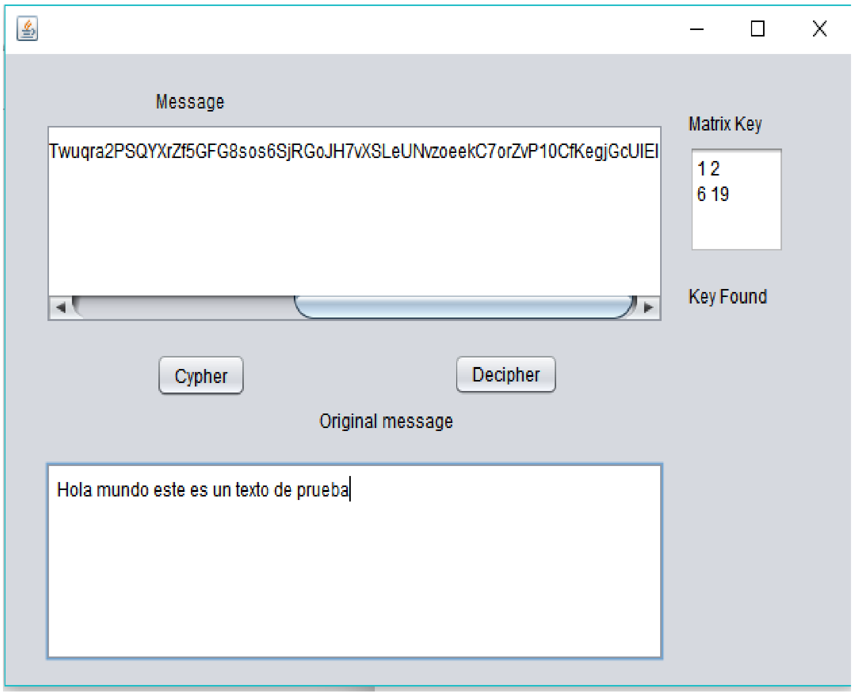


Fig. 10. Window with a test executed.

Enter the encrypted text in the central text box, the possible keys are generated in the upper left corner, after that in the lower text input box, the decrypted text will be displayed after having the correct key.

Test executed in 346800 milliseconds corresponding to 5.78 minutes which is the total time it takes to find the key and decrypt the text with a 2x2 matrix.

5 Conclusions

This article shows how to take advantage of the parallelism to find a correct key in the deciphering process through brute force into an algorithm. This test was developed with a 2x2 matrix considering integers numbers from 1 to 256 however this case is one of the best since the algorithm can be applied to a square matrix of n * n the decryption process can grow exponentially in time if the size of the matrix is not known or even the set of values used in each element.

References

1. Granados, P.G.: Introducción a la criptografía. Digital University Journal, 7 (2006)
2. López, J.: 5 hackeos que sacudieron a empresas en 2014. http://www.elfinanciero.com.mx/tech/hackeos-que-sacudieron-a-empresas-en-2014.html
3. Castañeda, C.B. Caballero, G.P.: Sistemas criptográficos de sustitución. Didactic Mathematics magazine, 30, 15–30 (1997)
4. Lester S.H.: Cryptography in an Algebraic Alphabet. The American Mathematical Monthly, 36(6) 306–312 (1929)
5. Sánchez-Rinza, B.E., Garcia-Ramirez, J., Rossainz, M.: Decodificación de texto en español utilizando frecuencias de palabras mediante cómputo paralelo. IEEE Computer society, 16, 1–6 (2018)
6. Rossainz, M., Capel, M.I., Sánchez-Rinza, B.E.: Uso de objetos paralelos para la decodificación de texto mediante frecuencias de palabras. Avances en arquitectura y tecnología de computadores, 19–28 (2018)

ISSN 1870-4069

Sistema hipermedia adaptativo para el apoyo del aprendizaje autónomo del idioma inglés

Eugenia Erica Vera[1], Yadira Navarro[2], Josefina Guerrero[1]

[1] Facultad de Ciencias de la Computación,
Benemérita Universidad Autónoma de Puebla, Puebla, México
[2] Facultad de Ciencias de la Electrónica,
Benemérita Universidad Autónoma de Puebla, Puebla, México

{eevclibra, ynavarro44, joseguga01}@gmail.com

Resumen. Hoy en día es más evidente la diversidad de los alumnos en el aprendizaje de inglés, encontramos estudiantes con facilidad de comprensión tanto en lectura como en gramática compartiendo el aula con estudiantes que requieren más tiempo para alcanzar un desenvolvimiento adecuado en el manejo del idioma. El potencial de las TIC puede favorecer en diversas formas los procesos de autoaprendizaje y posibilitar distintas modalidades educativas para apoyar la enseñanza aprendizaje del idioma. El propósito de este artículo es presentar el diseño y desarrollo de un Sistema Hipermedia Adaptativo para el apoyo del aprendizaje autónomo del idioma inglés y en particular promover la comprensión auditiva y oral del mismo. El sistema considera las características del estudiante (historial), para adaptar y presentar los contenidos educativos. También se presentan los resultados obtenidos del sistema al realizar una prueba piloto con una muestra de estudiantes.

Palabras clave: Sistema Hipermedia Adaptativo, multimedia, aprendizaje de inglés.

Adaptive Hypermedia System for the Support of Autonomous Learning of the English Language

Abstract. Presently the diversity of the students in the learning of English is becoming more evident, we found students with easy comprehension, both in reading and in grammar, sharing the classroom with students that require more time to reach an adequate development in the use of the language. The potential of ICTs can favor self-learning processes in different ways and enable different educational modalities to support the teaching and learning of the language. The purpose of this article is to present the design and development of an Adaptive Hypermedia System for the support of the autonomous learning of the English language, particularly to promote the auditory and oral comprehension of it. The system considers the characteristics of the student (record), to adapt and present the educational contents. The results obtained from the system when conducting a pilot test with a sample of students are also presented.

Keywords. Adaptive Hypermedia System, multimedia, learning English.

1. Introducción

A la mayoría de las personas se les dificulta entablar y mantener una conversación en el idioma inglés, a pesar de que es una de las materias que han cursado en su preparación básica. Entre las principales causas se tienen, el tiempo que los alumnos dedican al estudio y practica del idioma, la estructura de los planes curriculares que se basan en enseñar la gramática en forma tradicional, lo que ocasiona que los alumnos abandonen el curso una y otra vez. Por lo anterior es importante desarrollar nuevas herramientas que faciliten el aprendizaje de este idioma.

Este proyecto se centra en desarrollar una herramienta que estimula la comprensión auditiva y la comunicación del idioma inglés, tomando en cuenta que actualmente la generación de estudiantes tiene acceso a numerosa información que se encuentra en internet.

El objetivo es estimular y desarrollar el oído fonemático, para mejorar la comprensión auditiva a través de actividades audiovisuales, para lo cual se desarrollaron actividades multimedia con las que el alumno trabaja de manera autónoma y éstas se apegan a sus necesidades y disponibilidad, logrando facilitar el aprendizaje.

Los estudios revelan que el estilo docente no se relaciona de manera significativa con el rendimiento académico de los alumnos, es por ello que trabajamos en apoyar al estudiante en el contexto del aprendizaje autónomo como proceso individual que depende del trabajo, de los hábitos de estudio y de las estrategias del alumno.

El documento está conformado y organizado de la siguiente manera: en la sección 2, se presenta la fundamentación teórica del Marco Común Europeo, acerca de la competencia lingüística, la explicación del concepto de un Sistema Hipermedia Adaptativo (SHA), la conceptualización de aprendizaje autónomo, así como los temas y; subtemas a considerar y competencias lingüísticas a desarrollar; en la sección 3, se describe la metodología de la investigación; en la sección 4, se presenta el desarrollo y pruebas del SHA y, finalmente en la sección 5, se exponen las conclusiones y trabajos a futuro.

2. Fundamentación teórica

2.1. Marco común europeo de referencia para las lenguas

El Marco Común Europeo de Referencia para las Lenguas es, principalmente, el estándar internacional que define la competencia lingüística. Se utiliza para la elaboración de programas de lenguas, orientaciones curriculares, exámenes, manuales, etc., en todo el mundo. En él se describe lo que tienen que aprender a hacer los estudiantes de lenguas con el fin de utilizar una lengua para comunicarse, así como los conocimientos y destrezas que tienen que desarrollar para poder actuar de manera eficaz.

El enfoque que ha adoptado el Marco Común se centra en la acción y en la medida en que considera a los usuarios y alumnos que aprenden una lengua principalmente como agentes sociales, es decir, como miembros de una sociedad que tienen tareas (no sólo relacionadas con la lengua) que llevar a cabo en una serie determinada de circunstancias, en un entorno específico y dentro de un campo de acción concreto.

El marco común señala seis niveles que cubren adecuadamente el proceso de aprendizaje de una lengua. Los niveles son: 1) Acceso (Breakthrough), 2) Plataforma (Waystage), 3) Umbral (Threshold), 4) Avanzado (Vantage), 5) Dominio operativo eficaz (Effective Operational Proficiency) y 6) Maestría (Mastery).

Por lo que los tres grados comunes: Básico, Intermedio y Avanzado ó A, B, y C sitúan a los seis niveles como sigue:

Básico (A): A1: Acceso (Breakthrough), A2: Plataforma (Waystage).

Intermedio (B): B1:Umbral (Threshold), B2: Avanzado (Vantage).

Avanzado (C): C1: Dominio operativo eficaz y C2: Maestría (Mastery).

2.2. Sistema Hipermedia Adaptativo (SHA)

Brusilovsky define un SHA como un sistema basado en hipertexto e hipermedia que refleja algunas características del usuario y aplica un modelo para adecuar varios aspectos del sistema al usuario [4]. Aquellos sistemas de hipermedia capaces de ajustar su presentación y navegación a los diferentes usuarios reducen los problemas de desorientación y falta de comprensión de estos mismos usuarios (propios de los sistemas hipermedia no adaptativos) [5]. Un sistema se considera adaptativo cuando se ajusta de forma automática y personalizada a las necesidades del usuario [4]. Por lo que un SHA permite personalizar la información almacenada y la presenta a los usuarios según sus preferencias, conocimientos e intereses. El proceso de personalización muestra la información que es apropiada para algún tipo de conocimiento y de aprendizaje de cada usuario. El modelo posee un conjunto de reglas que permiten adaptar los contenidos al perfil del usuario, considerando sus características y el tipo de contenido que debe aprender para cumplir con los objetivos, logrando manifestar relaciones interesantes a partir de la información existente.

Los SHA Educativos son sistemas que, basados en el grado de conocimiento del usuario construyen rutas de aprendizaje adaptativas, personalizando la adquisición de conocimiento y adaptando el material que es recomendado.

Según P. Brusilovsky, un SHA debe cumplir los siguientes criterios [4]:

1. Ser un sistema hipertexto o hipermedia,
2. Tener un modelo de usuario
3. Ser capaz de adaptar el hipermedia usando este modelo.

Uno de los objetivos más relevantes para el cual un SHA es desarrollado, es que el sistema se adapte al usuario y no sea el usuario quien deba adaptarse al sistema, como sucede regularmente en los hipermedia “clásicos”, los cuales muestran el mismo contenido y los mismos enlaces a todos los usuarios. Las soluciones con la hipermedia adaptativa son: mejor comprensión de la información (rutas de navegación seleccionadas), y orientación en el hiperespacio. Es importante considerar el nivel de conocimiento y ritmo de aprendizaje del usuario.

En el SHA que se presenta en este trabajo considera el perfil académico con el que cuenta el estudiante, considerando el nivel de conocimientos y habilidades en el idioma inglés y canal preferente de aprendizaje.

2.3. Método eclecticismo de enseñanza

Desde una perspectiva ecléctica (The New Principled Eclecticism Method), que da la oportunidad a los docentes de utilizar y tomar metodologías y enfoques, así como seleccionar técnicas de cada método que consideren efectivas para aplicarlas de acuerdo al contexto y los objetivos aprendizaje [15], en el SHA se consideran dos diferentes métodos o estrategias de enseñanza; el auditivo lingual y el método basado en tareas.

El método Audio lingual sugiere que a los estudiantes se les debe de enseñar el idioma de manera directa, es decir sin utilizar la lengua materna del estudiante para explicar nuevas palabras o gramática del idioma que están aprendiendo [16].

Con base en las necesidades del SHA, este método es muy útil porque las habilidades se enseñan en el siguiente orden: escuchar, hablar, leer y, escribir. Considerando que, la habilidad auditiva es importante para desarrollar la competencia del habla y de esta manera, recibe un énfasis en particular.

De acuerdo con Brown las características principales de este método son [8]:

- La enseñanza se presenta por medio del diálogo.
- Hay poca o nula explicación gramatical.
- La gramática enseña de manera inductiva.
- El vocabulario se aprende en contexto.
- Se utilizan audios y material visual.
- Se le da prioridad a la pronunciación.

El método basado en tareas (Task-based method), es el método que ve el proceso de aprendizaje como un set de tareas de comunicación que están directamente conectadas con una meta curricular y propósitos que se extienden más allá de la práctica del idioma; esto significa que los alumnos usan el lenguaje ya aprendido para completar tareas o actividades [17].

La mayoría de los métodos que se conocen en el área de la enseñanza del inglés funcionan con las cuatro habilidades principales y cada una de ellas depende de la otra para conseguir sus objetivos, por lo que no es común que se enseñe una sola habilidad, es decir aislada de las demás; por ejemplo, hablar y escuchar se necesitan para poder comprender con exactitud una conversación, y en algunos contextos, leer o escuchar y tomar notas son parte una conversación. Las actividades de integración de habilidades en el SHA incluyen, por ejemplo, escuchar audios, aprender con juegos, o trabajar con paquetes de información y ejercicios para resolver problemas; este tipo de práctica requieren que los estudiantes se involucren en el idioma y que además se involucren en la integración.

2.4. Aprendizaje autónomo

Una característica importante del SHA es el aprendizaje autónomo que se considera como un aprendizaje estratégico en el que la persona toma decisiones claves sobre su propio aprendizaje: autodirigiéndolo en función de unas necesidades, metas o propósitos, auto regulándolo (seleccionando alternativas, acciones, tiempos) y autoevaluándolo, de acuerdo con los recursos y escenarios de que dispone y de las exigencias y condiciones del contexto. Con el aprendizaje autónomo la persona aprende a aprender gracias al entrenamiento y desarrollo de competencias o habilidades cognitivas, afectivas e interactivas, pero también, y de manera esencial, gracias al desarrollo de habilidades metacognitivas [10].

Tabla 1. Temas y subtemas para el SHA.

Tema	Subtemas
1. Simple Present	1.1. Introduce yourself
	1.2. Greetings
	1.3. Habits and routines
2. Present Continuous	2.1. What are you doing?
3. Past Simple	3.1. What did you do last week?
	3.2. Regular and irregular verbs
	3.3. Time Expressions
4. Past Continuous	4.1. How I met my best friend
	4.2. My childhood
5. Future: Will / Shall	5.1. What will you do when you graduate?
6. Future: Going to	6.1. What are you going to do on your vacations?
7. Modal Verbs	7.1. Can: abilities
	7.2. Could: possibilities
	7.3. Should: advices
8. Perfect present	8.1. Extreme sports and food
	8.2. Have you ever…?
9. Wh Questions	9.1. News
10. Nouns	10.1 Clothes

2.5. Temas y subtemas para el SHA

El contenido temático que se consideró en el SHA considera las funciones que el alumno debe realizar con el idioma a un nivel A2, conforme al Marco Común Europeo (ver tabla 1). Como objetivo general de aprendizaje se establece que el alumno será capaz de comunicar necesidades simples y básicas, así como desarrollar funciones sociales tales como: saludar, presentarse, preguntar como están y conseguir información sencilla, manejando los tres tiempos: presente, pasado y futuro.

2.6. Competencias

Las competencias a desarrollar mediante el SHA para la comunicación en el idioma inglés se presentan en la tabla 2.

Tabla 2. Competencias a desarrollar para la comunicación en el idioma inglés.

Competencias	Atributos
2. Se expresa y comunica	El alumno será capaz de presentarse, así como presentar a alguien más, sabrá reconocer distintas maneras de saludar y hablará sobre sus hábitos y rutinas. El alumno será capaz de describir las actividades que otras personas están realizando. El alumno será capaz de narrar las actividades que realizó la semana pasada, y reconocerá los verbos regulares e irregulares. El alumno será capaz de narrar un momento específico de su vida. El alumno expresará sus planes a futuro. El alumno será capaz de describir lo que hará en sus vacaciones.
3. Piensa crítica y reflexivamente	El alumno identifica los diferentes tempos de las oraciones. El alumno identificara las diferencias entre los verbos modales, así como las situaciones en las que debe emplearlos. El alumno será capaz de preguntar y responder sobre actividades inusuales. El alumno será capaz de describir alguna noticia relevante que haya escuchado.
4.Aprende de forma autónoma	El alumno desarrollara habilidades auditivas y orales. El alumno será capaz de seleccionar la técnica para comprender y producir información en idioma inglés.
5.Trabaja en forma colaborativa	El alumno será capaz de comunicar necesidades simples y básicas, así como desarrollar funciones sociales en forma colaborativa.

3. Metodología

El SHA se diseñó y desarrolló dentro de la metodología de innovación aplicada, propuesta por el grupo CIDER (creatividad e innovación para el desarrollo regional) de la Universidad Politécnica de Valencia (UPV) [2], por las ventajas que representaba el permitir generar referentes para buscar nuevas estrategias para la mejora. Esta metodología interpreta y combina las tendencias del mercado, la tecnología disponible y los movimientos sociales para resolver las necesidades, en nuestro caso la enseñanza del inglés como lengua extranjera y lograr superar las expectativas de los usuarios [3].

En este estudio se consideró una muestra de 308 adultos de 15 a 55 años, 123 niños y 308 adolescentes de 12 a 14 años; del instituto DMSV de capacitación para el trabajo, en el primer trimestre de otoño 2016.

En el paso de la Abstracción, el estudio consistió en analizar el servicio y se concluye que el instituto ofrece una gran variedad de cursos en inglés, y los costos están al alcance de la mayoría de las personas, el 78.64% de alumnos adultos que ingresan al instituto, se ubican en el nivel Básico; el 18.93% en el nivel intermedio y sólo el 2.43% en un nivel avanzado. En el grupo de niños el 72% se ubican en el nivel Básico, el 28% en el nivel intermedio y ninguno en avanzado. En el grupo de adolescentes el 87% que ingresa al instituto se ubican en el nivel Básico, el 13% en el nivel intermedio y ninguno en avanzado. Estos niveles se ajustan de acuerdo al Marco Común Europeo. Este paso permitió definir el contenido temático para el cual desarrollamos el SHA.

En el paso de Conocimiento que es el que pretende monitorizar a los actores, nuestro estudio se dirige al grupo de los adultos, para el cual es el producto a desarrollar. Mediante un cuestionario pretendemos detectar las necesidades de estos actores. El cuestionario está formado de nueve preguntas, con respuesta de opción múltiple documentado en [3]. De los resultados obtenidos, se consideraron en el SHA, los siguientes: el 64% de la muestra considera que la prioridad es lograr mantener una conversación, un 30% considera que es el poder redactar textos en inglés, y un 4% requieren de entender las lecturas en este idioma, el menor porcentaje 2% y como última prioridad requieren de certificaciones como del MARCO COMÚN EUROPEO o el TOFEL. El 80% de los estudiantes desea dominar la lengua en menos de un año. Los estudiantes revelan que no tienen tiempo de estudiar el idioma y la dedicación al estudio de este es de 2 horas cada tercer día o ningún momento, son pocos los libros de textos que leen y si los leen es porque se los dejaron de tarea y la mayoría de las veces son artículos relaciones con sus respectivas carreras, ven videos y películas de vez en cuando. Lo que cabe resaltar es que la mayoría prefiere material didáctico en línea por que pueden acceder.

En el paso de análisis de tendencias intersectorial: Existen institutos que ofrecen cursos que cubren sin problema las necesidades marcadas por los alumnos encuestados, omitimos nombres por cuestiones de derechos de autor, sin embargo estos cursos, comúnmente se pagan con anticipación a un precio no muy accesible para el público en general.

Los curso online están ganado terreno, sin embargo se requieren competencias tecnológicas previas de los estudiantes, comúnmente las tareas y objetivos de aprendizaje, de estos cursos, no están bien descritos y es fácil que los estudiantes dejen los estudios a medias o no aprovechen al máximo la formación.

En el paso de definición de escenarios se tiene que una de las prioridades de la población es mantener una conversación, ya que piensan que en la conversación pone en práctica la gramática y pueden comprobar si realmente entiendes la conversación de la otra persona que está hablando, es por esto que se debe seleccionar, de los planes de estudio, las partes que realmente aportan al aprendizaje y desarrollo de la habilidad de habla y de escucha del idioma. Requieren lograr estos objetivos en menos de un año, es por ello que se necesita priorizar y jerarquizar los aprendizajes, programar las actividades e incluir en estas varias competencias para lograr en menos tiempo los objetivos.

Solo le dedican una hora a estudiar, esto nos lleva a crear actividades que motiven al estudiante, enseñarle a utilizar todo lo que le rodea para aprender. La mayoría prefiere material didáctico en línea por que pueden acceder a la hora que tengan disponible, por lo que debemos programar los materiales y actividades en línea. Los alumnos cuentan con diferentes conocimientos por lo que es importante reforzar los nuevos conocimientos y saltarse los ya aprendidos, en esta parte el SHA adapta las actividades según el perfil del estudiante.

4. Desarrollo y pruebas del Sistema Hipermedia Adaptativo para el apoyo del aprendizaje autónomo del idioma inglés

El sistema se desarrolló bajo el modelo cliente-servidor donde todo comienza con una petición o requerimiento HTTP iniciado en un navegador por un cliente que quiere acceder a un recurso de nuestro sitio web por medio de una dirección URL. La dirección URL apunta a la localización física de una página de extensión .ASP.

4.1. Actividades de aprendizaje autónomo

Las actividades de aprendizaje autónomo son acciones específicas que facilitan la ejecución de la estrategia de aprendizaje autónomo y el desarrollo de competencias. Estas actividades hacen énfasis en el proceso intelectual que se encuentra especificado para cada idea principal y además son la base del diagnóstico académico así que las preguntas se presentan de la manera más clara posible, para que el alumno no caiga en confusiones. En cada actividad se le presenta al alumno la información del tema y subtema que está cursando, involucren los niveles cognoscitivos, para lo cual se tomaron como referencia los cinco niveles cognoscitivos que describe Bloom (tabla 3).

Es importante disponer de información suficiente para resaltar los aprendizajes relevantes de cada unidad, es decir diferenciar un aprendizaje importante de uno secundario.

Para definir cuáles son los aprendizajes relevantes de las unidades, temas o áreas que se seleccionaron para ser diagnosticadas, se tiene el siguiente método que consta de tres pasos [12]:

a) Seleccionar las ideas principales o esenciales de cada una de las unidades, temas o áreas elegidas.
b) Determinar los procesos cognoscitivos deseables para cada idea esencial. Una vez que se seleccionaron las ideas principales del contenido de la enseñanza, se realiza un análisis de cada una de ellas para determinar los procesos intelectuales que el alumno pondrá en juego al haberlas aprendido.

   Los procesos intelectuales característicos del aprendizaje escolar utilizados son [12]: Conocimiento (repetir y reconocer información), comprensión (ejemplificar, comparar e interpretar la idea esencial), aplicación (integrar varias ideas, aplicar la idea esencial en situaciones nuevas o en la solución del problema), análisis (analizar la idea principal), síntesis y evaluación cognitivo.

   Al final obtendremos una puntuación de las ideas principales (conceptos, hechos, procedimientos), junto con el señalamiento de los procesos cognitivos que se esperan para cada uno de ellos.
c) Señalar el tipo de aprendizaje referido en cada idea esencial, en función al programa de estudios. El aprendizaje se clasifica en diferentes formas. Quesada propone hacerlo en tres categorías: indispensable, esencial y antecedente [12].

   Los aprendizajes relevantes conforman parte del perfil académico deseado del alumno. Una vez que se tiene definido por completo el perfil académico deseado (aprendizajes relevantes) del alumno, así como sus antecedentes, se comienza la elaboración del instrumento diagnóstico, mediante el cual se le evaluará.

Se diseñaron cinco tipos de actividades: Booleana, Opción Múltiple, Completar Texto, Relacionar y Arrastrar. En cada actividad es posible que el alumno se encuentre con un poco de indecisión al momento de responder a su actividad, en este caso todas las actividades cuentan con la opción de regresar a la clase que observaron antes de dicha actividad, como ayuda para resolver dudas, sin otorgar directamente la respuesta.

Tabla 3. Atributos de las actividades de aprendizaje autónomo.

Actividad			**Competencia**			**Tipo de aprendizaje (Quesada,1996)**			**Proceso cognitivo (Bloom, 2004) (perfil deseable)**						**Perfil actual del estudiante**	
Tema	Subtema	Actividad	Genérica	Específica	Atributo	Antecedente	Indispensable	Esencial	Conocimiento	Comprensión	Aplicación	Análisis	Síntesis	Evaluación	Canal preferente de aprendizaje	Nivel
1.	1.1	1	4	7	28	X					X				A	B
2.	2.1	2	4	7	28	X					X				A	M
2.	1.2	3	4	7	28		X		X						V	B
2.	1.3	4	4	7	28						X				V	A

Nota: A = Auditivo, V= Visual, K= Kinestésico, B= Bajo, M=medio, A= Alto.

4.2. Desarrollo del Sistema educativo

En la Fig. 1 se observa que el usuario tiene acceso al Sistema educativo a través de links que le permitirán irse desplazando por todo el Sistema.

Fig. 1. Entrada al Sistema educativo para el aprendizaje del idioma inglés.

Una vez dentro de la escuela, se muestran vínculos, a los que se podrán acceder a: Aula, Cafetería, Biblioteca, Secretaria, Tablón de anuncios, Correo, Tareas Programadas, Ayuda y desplazarse por el sitio. En la sección Aula encontraremos foros, contenidos de la materia, ideas clave, herramientas multimedia, actividades de aprendizaje, un archivador, y tablón de anuncios (ver Fig. 2).

Fig. 2. Dentro del aula en el Sistema educativo.

Una vez que seleccionamos la materia y el tema, el sistema pide que se identifique el estudiante, mediante una clave de acceso.

El Sistema educativo tiene programadas las clases de todo el curso, permitiendo al alumno seleccionar la clase de aprendizaje que desea aprender o repasar. En la figura 3 se muestra una clase mediante un video.

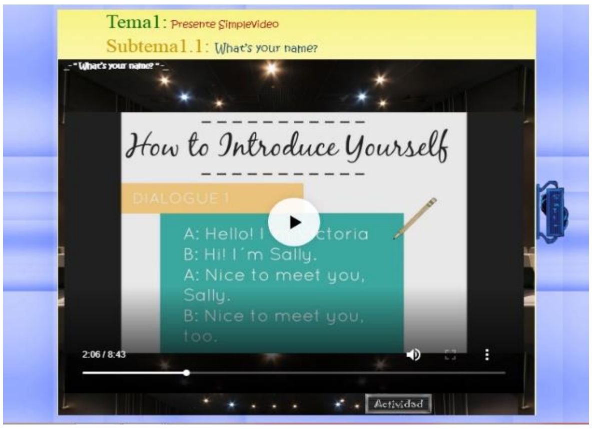


Fig. 3. Pantalla de una clase de aprendizaje de inglés básico en el Sistema educativo.

El Sistema educativo tiene programadas las actividades en ambientes virtuales para el aprendizaje, en donde los estudiantes reciban una instrucción y al mismo tiempo son partícipes de ella jugando un rol activo para fomentar la colaboración de los mismos y enriqueciendo el acervo que el sistema ofrezca para dicho aprendizaje. En la figura 4 se muestra un ejemplo de actividad de aprendizaje, de tipo opción múltiple.

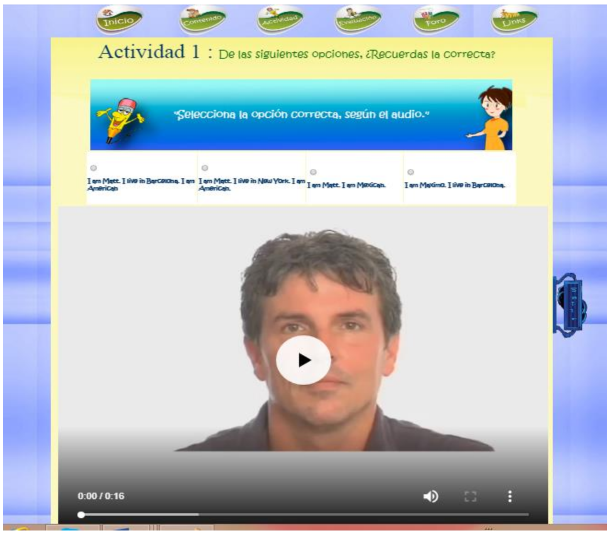


Fig. 4. Pantalla de una actividad de aprendizaje, de tipo opción múltiple, de inglés básico en el Sistema educativo.

En la Fig. 5 se observa una actividad de relacionar el concepto con la imagen. En caso de que el alumno falle en su respuesta, se determinó que no se le puede decir literalmente que no tiene validez, así que en los dos tipos de respuesta, correcta o no, se guiará con un tipo de apoyo para aclararle cual camino era el indicado, esto sin darles la respuesta, simplemente como una idea que despejará sus dudas y motivando a continuar con el curso. En el momento de contestar acertadamente el concepto, se le aumentará los puntos que le corresponden.

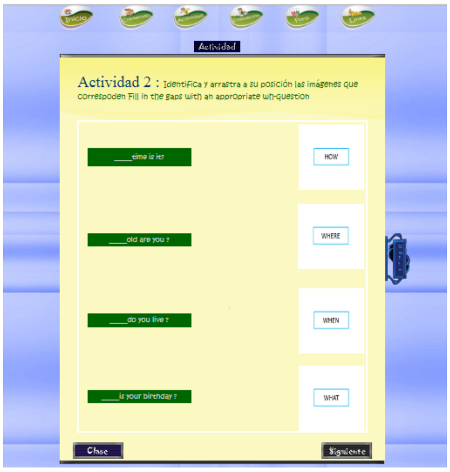


Fig. 5. Pantalla de una actividad de arrastrar el concepto con la imagen.

La información de los movimientos del alumno es almacenada en una bitácora para consultas posteriores, es decir, se guarda un seguimiento del alumno y los movimientos realizados en sus actividades. Al finalizar las actividades, el sistema muestra el nivel con el que terminó y da la opción para borrar su historial. Este nivel está determinado por la cantidad de preguntas que maneja el sistema y los puntos adquiridos con sus respuestas correctas.

Están registrados en la base de datos los nombres de los todos académicos de dicha materia, se le asignó como password su matrícula única, y su primer apellido.

Las funciones que puede desempeñar un profesor en el Sistema educativo multimedia son las siguientes:

- Ver el contenido temático de la materia.
- Ver la lista de sus alumnos con información de avances.

En la elaboración de las preguntas o actividades a realizar por el alumno se considera el proceso intelectual que se toma en cuenta para cada una de ellas, y dependiendo del tipo de aprendizaje que cada una representa. Dentro de este proceso se le otorgó una ponderación a cada tipo de aprendizaje únicamente como ayuda para establecer el nivel en el que el alumno se encuentra, ya sea, alto, medio o bajo. La ponderación mencionada es: al conocimiento indispensable 6 puntos, al conocimiento esencial 4 puntos y al conocimiento antecedente 6 puntos.

4.3. Resultados

El SHA fue piloteado con alumnos adultos de los cursos básicos del instituto en estudio con una muestra n=32, cuyo objetivo fue obtener el diagnóstico inicial y final. Las evaluaciones permitieron revisar el nivel de logro del desarrollo de las competencias en comunicación y comprensión del idioma inglés que adquirieron los alumnos durante el ciclo escolar 2017, siendo el sistema una herramienta de apoyo utilizada con los alumnos de forma presencial durante las sesiones de clases.

A todos, es decir, a los 32 estudiantes se les aplicó la prueba protocolo, al inicio del curso para asegurar igualdad en cuanto a sus estrategias de aprendizaje apegado a las dimensiones manejadas en la Escala de Estrategias de Aprendizaje Contextualizado (ESEAC) (Bernad, 2007).

Las cuatro competencias que se midieron son: Comprender información auditiva en el idioma inglés, Comunicar información oral mediante el idioma inglés, Comprensión lectora del idioma inglés y Manejo de técnicas eficientes donde cada una representa un 25% del total de las competencias y a partir de esta información es como se realiza la evaluación.

Los estudiantes tomaron el curso a lo largo de 3 meses, con los mismos contenidos, y finalmente se evalúo nuevamente la prueba protocolo con el fin de observar el avance en las estrategias de aprendizaje. En la tabla 4 se presentan los resultados iniciales y finales del grupo control.

La competencia que más se desarrolló fue “Comprender información auditiva en el idioma inglés” con un 65% de desarrollo, de 2% a un 65% y la segunda fue “El manejo de técnicas eficientes” con un 60%, logrando alcanzar de un 20% a un 60% y la tercera

“Comunicar información oral mediante el idioma inglés”, logrando alcanzar de un 5% a un 40% , esto conlleva que el uso de un SHA propicia un aprendizaje más significativo en el tema del idioma inglés, como se muestra en la Tabla 4.

Tabla 4. Resultados del Test Inicial, Test Final y aportación del desarrollo para el desarrollo de competencias mediante el SHA.

Competencias	Test Inicial	Test Final	Uso del Sistema
Comprender información auditiva en el idioma inglés	2%	65%	63%
El manejo de técnicas eficientes	20%	60%	40%
Comunicar información oral mediante el idioma inglés	5%	40%	35%

5. Conclusiones

Tomando en cuenta el enfoque de competencias en las nuevas teorías de aprendizaje y considerando los modelos de sistemas de cómputo se logró el diseño y desarrollo de un Sistema Adaptativo Hipermedia para el apoyo del aprendizaje autónomo de idioma inglés.

Se logró promover la comprensión auditiva y oral de este idioma gracias a la potencialidad que nos ofrece la tecnología multimedia.

Considerando las características del estudiante (historial), para adaptar y presentar los contenidos educativos en dicho sistema se lograron adecuar las actividades de aprendizaje convenientes para el desarrollo de las competencias en cuanto al manejo del idioma inglés.

Se logró aumentar la motivación de los estudiantes, y el desarrollo de la habilidad de comprensión auditiva del idioma inglés.

Se sistematizaron estrategias para el logro de un aprendizaje autónomo, y se aplicaron metodologías para atender la diversidad.

En cuanto a la comprensión de información auditiva de los estudiantes según la prueba protocolo, se logró un avance significativo al contrastar los resultados iniciales con los finales en el grupo experimental.

Se puede asegurar que el sistema puede ser utilizado de manera generalizada en todos los grupos de la materia de inglés como refuerzo para el aprendizaje, sin interferir con el estilo de enseñanza.

Se pretende, en un futuro, desarrollar las actividades de aprendizaje de todos los niveles del Marco Común Europeo.

Referencias

1. Antich de León, R.: Metodología de la enseñanza de las Lenguas Extranjeras. Editorial Pueblo y Educación, La Habana (1986)
2. Guijarro, J.: Desarrollo de una metodología para la gestión del proceso de innovación. una perspectiva aplicada, Valencia (2006)

3. Vera, E., López, E.C., Guerrero, J.: Sistema didáctico e-Learning para el apoyo del aprendizaje del idioma inglés (2016)
4. Brusilovsky, P.: Methods and techniques of adaptive hypermedia. User Modeling and User Adapted Interaction. Netherlands, v 6, 2-3, 87-129. ISSN: 0924-1868 (Print) 1573-1391 (Online) (1996)
5. Brusilovsky, P., Millan, P.: User Models for Adaptive Hypermedia y Adaptive Educational Systems. The Adaptive Web, Lecture Notes in Computer Science 4321, Berlin, Springer-Verlag, pp. 3–53. ISSN: 0924-1868 (Print) 1573-1391 (Online) (2007)
6. Praveen, C.: The Eclectic Method- Theory and Its Application to the Learning of English (2013), http://www.ijsrp.org/research-paper-0613/ijsrp-p1844.pdf
7. Aldoobie, N.: ADDIE Model (2015), http://www.aijcrnet.com/journals/Vol_5_No_6_December_2015/10.pdf
8. Brown, H.: Teaching by Principles: An Interactive Approach to Language Pedagogy. Upper Saddle River, New Jersey: Prentice Hall Regents (1994)
9. Weideman, A.: The old and the new: reconsidering eclecticism in language teaching. Unit for Language Skills Development. University of Pretoria. South Africa (2003), http://www.reformationalpublishingproject.com/rpp/docs/6_eclecticism_in_language_07.pdf
10. Amaya, G.: Aprendizaje Autónomo y Competencias. Bogotá (2008)
11. Lara, A.: Aprendizaje Autónomo para estudiantes Universitarios. Colombia: Editorial Universidad de Caldas. Artes y Humanidades (2008)
12. Quesada, C, Sánchez, J.: Calificación y Diagnóstico del Aprendizaje por Computadora. México. Noriega Editores (LIMUSA) (1996)
13. Moreno, A.: Análisis de algunas metodologías para la enseñanza de la Lengua Inglesa en Educación Primaria. Universidad Internacional de la Rioja. Facultad de Educación (2012)
14. Wa'Njogu, K.: Methods of Language Teaching. Yale University (2012)
15. Marcillo, M.A.: La aplicación del Método Ecléctico en el desarrollo de la Expresión Oral del Idioma Inglés. Universidad Central del Ecuador (2017)
16. Aicart, M.: El Método Audio-Lingual. (2015)
17. Rodriguez, M., Rodriguez, J.: Task-Based Language Learning: Old Approach, New Style. A New Lesson to Learn. Vol. 12, No. 2. ISSN 1657-0790. Bogotá, Colombia. Pages 165–178 (2010)

ISSN 1870-4069

DNA Sequence Recognition using Image Representation

Luis A. Santamaría C., Sarahí Zuñiga H., Ivo H. Pineda T., María J. Somodevilla, Mario Rossainz L.

Benémerita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación, Puebla, Mexico

aluloc_kapa@hotmail.com, sarahi.zuhe@gmail.com,
ivopinedatorres@gmail.com, {mariasg,rossainz}@cs.buap.mx

Abstract. In recent years, the field of machine learning has progressed enormously in addressing difficult classification problems. The problem raised in this article is to recognize DNA sequences, recognize the boundaries between exons and introns using a graphic representation of DNA sequences and recent methods of deep learning. The objective of this work is to classify DNA sequences using a convolutional neuronal network (CNN). The set of DNA sequences used for the recognition were 1847 sequences from a database with 4 types of hepatitis C virus (type 1, 2, 3 and 6) taken from the repository available on the ViPR page. The other set of sequences used to recognize limits between exons and introns were sequences from the Molecular database (Splice-junction Gene Sequences) Data Set that has 3190 sequences, available on the ICU page, with three classes of sequences: limit exon-intron, limit intron-exon and none. For the processing of the DNA sequences, a representation method was designed where each nitrogenous base is represented in gray scale to form an image. The generated images were used to train the convolutional neuronal network. The results obtained from the CNN trained with the Hepatitis C virus database suggest that the CNNs are suitable for the classification of the images generated from the DNA sequences. This result led us to perform the experiments for the recognition of exons and introns with the UCI database for the recognition of limits between exons and introns. The results obtained were a training precision of 82%, a validation accuracy of 75% and an evaluation accuracy of 80.8%. It is concluded that it is possible to classify the images of DNA sequences of the databases used.

Keywords. Gene recognition, deep learning, convolutional neuronal networks, coding DNA sequences.

1 Introduction

The automatic learning methods allow to identify characteristics that favor the classification, analysis and recognition of patterns. In the area of biology, the use of machine learning methods facilitates the recognition of DNA sequences. This work recognizes the genes of DNA through images. This article is divided into sections, the first section is the state of the art that is the prior knowledge necessary for the recognition of genes. The second section details the methodology that was used for the analysis of DNA sequences. The third and fourth sections show the results and conclusions obtained.

2 State of the Art

The mechanisms or processes of gene prediction are those that, within the area of computational biology, are used for the algorithmic identification of pieces of sequences, usually genomic DNA [8], and that are biologically functional. This, especially includes genes encoding proteins and regulatory sequences. The identification of genes is one of the first and most important steps to understand the genome of a species once it has been sequenced [11].

Deoxyribonucleic acid (DNA) is composed of four molecules called nucleotides or nitrogenous bases: adenine, thymine, guanine and cytosine [9]. A complete DNA molecule or, in other words, a DNA sequence is composed of an alphabet that contains the letters of the four nitrogenous bases.

$$\begin{aligned} &\Sigma\{ATGC\}, \\ \phi_i = &(V_1, V_2, V_3, \ldots, V_n), \\ &V_i \in \Sigma, \end{aligned} \tag{1}$$

where a string ϕ is a sequence of DNA formed elements of the alphabet Σ (equation 1) can define the characteristics of a living organism, containing all the genetic information in inheritance units called genes. The mechanisms or processes of gene prediction are those that, within the area of computational biology, are used for the algorithmic identification of pieces of sequences, usually genomic DNA [1], and which are biologically functional. This, especially includes genes encoding proteins and regulatory sequences. The identification of genes is one of the first and most important steps to understand the genome of a species once it has been sequenced [2].

Splicing junctions are points in a DNA sequence in which “useless” DNA is removed during the process of creating proteins in higher organisms. The problem posed in this data set is to recognize, given a DNA sequence, the boundaries between the exons (the parts of the DNA sequence retained after splicing) and the introns (the parts of the DNA sequence that are cut).This problem consists of two subtasks: recognition of exon/ intron limits (called EI sites) and recognition of intron exon boundaries (IE sites). In the biological community, the limits of IE refer to the “acceptors” while the boundaries of EI are known as “donors” [7]. Both tasks are complicated since there is no standard sequence to recognize introns and exons, for this reason it is interesting to design tools that help us identify and classify them.

The number of research projects on currently valid genomes is increasing at an accelerated rate, and providing a catalog of genes for these new genomes is a challenge. Obtaining a set of well-characterized genes is a basic requirement in the initial steps of any process of creating a genome. Computational gene search methods can be categorized as based on alignment and sequence composition or a combination of both. Methods based on sequence alignment can be used when trying to predict a gene that encodes a protein for which there is a closely related homologue, this is the approach in GeneWise [5] and PROCRUSTES [4].

Algorithms based on sequence composition (also known as gene search methods) contain a probabilistic model of gene structure based on biological signals (splice sites and translation start / stop sites) and composition properties of functional sequences (exons as coding sequences and introns as intermediate sequences between exons and introns). Unlike alignment-based methods, these algorithms rely on the intrinsic properties of genes to construct predicted genetic structures. Genscan [10] and Geneid [3] are the two examples of this approach and can find known genes and new genes as long as the genes conform to the underlying probabilistic model.

As described a DNA strand is a molecule characterized by four nitrogenous bases represented with numerical or alphabetic values: A (adenine), T (thymine), G (guanine) and C (cytosine) [10]. However, the representation of large amounts of information as DNA sequences do not make their mathematical analysis easy, this creates the need to find new ways of representing information.

In Lapedes [6] and his team trained a neural network to recognize genes in DNA sequences, achieved an accuracy of 91.2% in intron / exon splicing junctions and 92.8% in splicing junctions (exon / intron). What gives a motivation to use of convolutional neural networks to solve this same classification problem. This work consisted of looking for a new way to represent DNA sequences for analysis, as has already been mentioned, there are currently different methods to recognize genes, but these representations complicate their analysis. The proposal we present is to generate images from DNA sequences and subject them to analysis with deep learning techniques, specifically convolutional neuronal networks; used for the classification of images. At the moment a mathematical model has not been found that solves the process of classification by neural networks, but its results get to be so high that they surpass 99% in some cases [1].

2.1 Convolutional Neural Network (CNN)

In recent years, the field of machine learning has made tremendous progress in addressing problems of classification, identification and pattern recognition. In particular, it has been found that a type of model called Convolutional Neural Network CNN, which achieves a reasonable performance in tasks of visual recognition, equaling or exceeding human performance in some domains [11]. A CNN is an algorithm for machine learning in which a model learns to perform classification tasks directly from images, videos or sounds. CNNs are especially useful for locating patterns in images in order to recognize objects, faces and scenes. They learn directly from the image data, using patterns to classify images and eliminate the need for a manual features extraction.

Inception-v3 is one of the tools designed for the Visual Recognition challenge. This is a standard task in artificial vision, where the models try to classify complete images in 1000 ImageNet classes. On the other hand, TensorFlow is a tool for automatic learning. Although, it contains a wide range of functionalities, TensorFlow is designed mainly for models of deep neural networks. Modern models of image recognition have millions of parameters; training them from scratch requires a large amount of tagged training data and a large amount of computing power (hundreds of hours of GPU or more). Transfer learning is a technique that cuts much of this by taking a piece from a model that has already been trained in a related task and reusing it in a new model, in Figure 1 an example of a CNN is shown, the filters are applied to each training image with different resolutions, and the output of each convolved image is used as input for the next layer [2]. Although not as accurate in comparison to the full model training, it is surprisingly effective for many applications, works with moderate amounts of training data (thousands, not millions of tagged images) and can be executed in just thirty minutes in one laptop without a GPU [11].

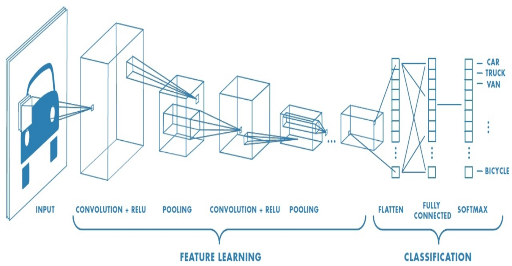


Fig. 1. Example of a Convolutional Neural Network.

3 Methodology

This section describes in detail how images are generated from sequences of DNA and its use to training a convolutional neural network for classification of three classes of sequences. An important appearance that has considered this work is that the CNN are used for the recognition of patterns and classification of images. The sequences of DNA are represented by letters: To use A for the adenine, G for the guanine, C for the cytosine and T for the thymine, however, a CNN is not established to process information under this format, for this reason it has designed a graphic representation of the sequences. The first step was to assign a value of grey to each one of the letters as it shows in Table 1. The scales of grey range from 0 that represents black, to 1 that represents white, in such a way that the resultant intermediate values are tonalities of grey to show a better contrast among them . The second was to do that the sequences could be represented by a specific image to each one. To attain this used a matrix of dimension N X N, where N is the value that coincides with the number of nitrogenous bases (length) of all the sequences of the database employed. Each sequence was planted in the first row and copied in the rest of the rows until having N in total, like this the final result is an image with bars in the scale of grey like which shows in Figure 3, each one of the images obtained is specific for each instance of the database as it observes in figure 2. In total obtained 1847 images of the database of ViPR and 3190 images of the database of the UCI.

Table 1. Color Representation of nitrogenous bases.

Nitrogenous bases	Grey Value
A	0
C	0.3
G	0.7
T	1

1 CCAGCTGCATCACAGGAGGCCAGCGAGCAGGTCTGTTCCAAGGGCCTTCGAGCCAGTCTG
2 CCAGCTGCATCACAGGAGGCCAGCGAGCAGGTCTGTTCCAAGGGCCTTCGAGCCAGTCTG
3 CCAGCTGCATCACAGGAGGCCAGCGAGCAGGTCTGTTCCAAGGGCCTTCGAGCCAGTCTG
.
.
.
60 CCAGCTGCATCACAGGAGGCCAGCGAGCAGGTCTGTTCCAAGGGCCTTCGAGCCAGTCTG

Fig. 2. DNA sequencce for coding.

Fig. 3. Image related.

3.1 Use of CNN to DNA Sequences

This subsection describes how a CNN was trained with the representative images of each sequence. An InceptionV3 CNN was used to which the deep learning transfer was applied to categorize the recognition of three classes of DNA sequences: recognition of exon / intron limits (called EI sites), recognition of intron / exon limits (IE sites) and recognition of neither of the previous two (N).

Once it was possible to represent the DNA sequences as images, a CNN was used and with the software libraryTensorFlow was built a classification model based on a pre-trained convolutional neuronal network. Inception V3 CNNs were used to which the deep learning transfer is applied to categorize the recognition of a database with four classes of DNA sequences:Hepatitis C virus type 1, 2, 3 and 6 and the recognition of another database with three classes of exon / intron limits (called EI sites), recognition of intron / exon boundaries (IE sites) and recognition of none of the previous two (N). To adjust the model to our problem, the last layers of the networks are trained with instances obtained from the databases, both networks were trained in 4000 steps.

First the CNN was trained to classify the 4 types of Hepatitis virus, then a CNN was trained with only 2 classes: EI and IE and finally another CNN was trained with all the classes of the database: EI, IE and N to compare the results of the last two neurons.

4 Results

The classification results for the CNN trained with the database of the four types of Hepatitis C virus show a 95 % evaluation accuracy with 145 images tested and at the end of step (k) 4000 the training precision was 94.5 % and validation accuracy of 95 % as seen in the figure 4. The decreasing behavior of entropy during training, is seen in the figure 5 .

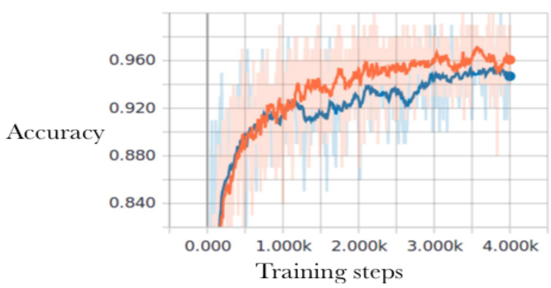


Fig. 4. CNN with classes of Hepatitis C virus type 1, 2, 3 and 6. Orange: training precision. Blue: validation accuracy after 4000 steps (k).

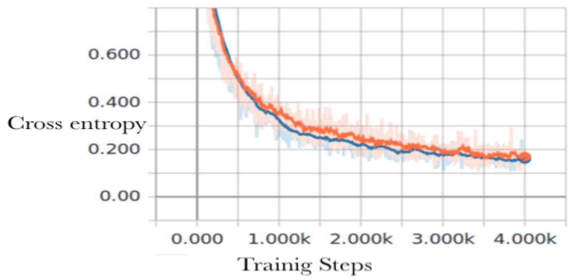


Fig. 5. CNN crossed entropy with classes of Hepatitis C virus type 1, 2, 3 and 6 after 4000 steps (k). Upper stroke: training. Bottom line: validation.

When using a CNN with the classes EI and IE, an evaluation accuracy of 80.8 % is obtained with 177 test images and at the end of step (k) 4000 the training precision is 82 % and the validation accuracy of 75 %. Figure 6 shows how the accuracy of training (orange) and validation (blue) is changing in each step and in Figure 7 shows how entropy decreases with the increase in steps during the training.

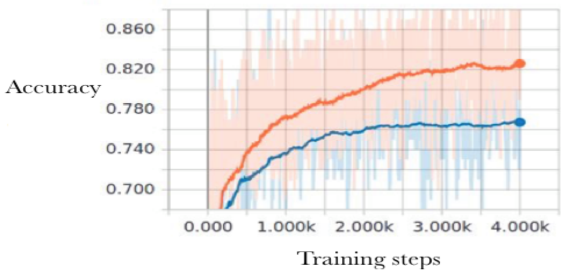


Fig. 6. CNN with classes IE and EI. Orange: training precision. Blue: validation accuracy after 4000 steps (k).

On the other hand, the results of the second CNN where the three classes of the database were used show an evaluation accuracy of 57.5 % with 301 images and at the end of step 4000 the training precision 69 % and the precision of validation with 56 % as shown in Figure 8. Figure 9 shows the changes in entropy at each stage of the training.

sectionConclusions The results obtained from the CNN trained with the Hepatitis C virus database suggest that the automatic learning methodology used in this work is suitable for the classification of the images generated from the DNA sequences, showing important and high percentages of evaluation accuracy, training precision and validation accuracy. These results led us to

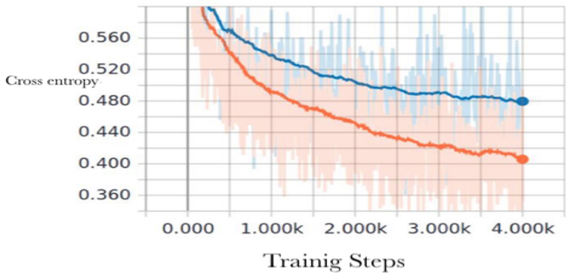


Fig. 7. Cross entropy of CNN with classes IE and EI after 4000 steps (k). Orange: training. Blue: validation.

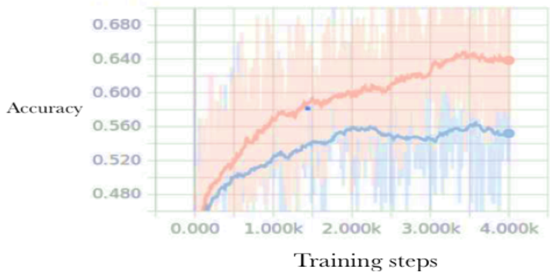


Fig. 8. CNN with classes IE, EI and N. Orange: training precision. Blue: validation accuracy after 4000 steps (k).

carry out the following experiments for the recognition of exons and introns in the following database: For this case, the CNNs show that the percentages of validation accuracy are lower compared to those of a neural network based on the work of Lapedes [6]. The importance of the work is that favorable results are presented to continue exploring the use of convolutional neural networks using the representation of DNA sequences as images, a simple and practical coding method.

In this work, it has been possible to perform DNA sequence classification using a CNN and the results show that CNN are able to perform this classification with an 80.8 % accuracy of evaluation for the experiment with classes IE and EI and 57.5 % for the experiment with classes IE, EI and N. Similar results can be seen in the precision of training and validation of the Figures 6 and 8. In the case of the four types of hepatitis, results of up to 94.5 % of evaluation accuracy are achieved.

The difference between the results obtained for the experiments with two and three classes can be justified that increasing the number of classes increases the entropy Figures 7 and 9. Cross entropy is a metric that can be used to reflect the accuracy of probabilistic forecasts and is closely linked to the maximum likelihood estimation. The crossed entropy is a function that allows to evaluate

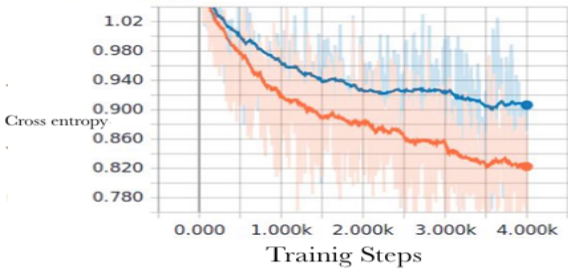


Fig. 9. Cross entropy of CNN with classes IE, EI and N after 4000 steps (k). Orange: training. Blue: validation.

the result of the classification instead of using the metric of the mean square error, the value of the crossed entropy allows to evaluate the progress of the process of learning of the information [1].

On the other hand, there is talk that the transfer of learning is good when there are few images to train the network and that allows to reach acceptable results in most cases, however, it is still possible to further improve validation and training accuracy and decrease entropy if a neural network is trained from scratch,that is, we must have a database of millions of instances and a computer with GPU to train this network but it will surely offer better results than the pre-trained CNN we use for this work.

In conclusion, it can be stated that a convolutional neuronal network of the InceptionV3 model is capable of classifying DNA sequences if the sequence is processed and transformed into an image, however, the accuracy percentages can be improved if a CNN is trained with a larger sequence base.

Acknowledgments

The authors thank the National Council of Science and Technology (CONACyT) of México, the Benemérita Universidad Autónoma de Puebla, which through the Faculty of Computer Science have provided the necessary support for the realization and presentation of this work.

References

1. How to retrain an image classifier for new categories. `https://www.tensorflow.org/tutorials/image_retraining`, accessed: 2018-05-28
2. Mathworks (2018). deep learning. `https://la.mathworks.com/solutions/deep-learning/convolutional-neural-network.html`, accessed: 2018-05-27
3. Almeida, J.S., Vinga, S.: Universal sequence map (usm) of arbitrary discrete sequences. BMC bioinformatics 3(1), 6 (2002)
4. Burge, C., Karlin, S.: Prediction of complete gene structures in human genomic dna1. Journal of molecular biology 268(1), 78–94 (1997)

5. Gelfand, M.S., Mironov, A.A., Pevzner, P.A.: Gene recognition via spliced sequence alignment. Proceedings of the National Academy of Sciences 93(17), 9061–9066 (1996)
6. Lapedes, A., Barnes, C., Burks, C., Farber, R., Sirotkin, K.: Application of neural networks and other machine learning algorithms to dna sequence analysis. Tech. rep., Los Alamos National Lab., NM (USA) (1988)
7. Noordewier, M.O., Towell, G.G., Shavlik, J.W.: Training knowledge-based neural networks to recognize genes in dna sequences. In: Advances in neural information processing systems. pp. 530–536 (1991)
8. Ouzounis, C.A.: Rise and demise of bioinformatics? promise and progress. PLoS computational biology 8(4), e1002487 (2012)
9. Panduro, A.: Biología molecular en la clínica. McGraw-Hill Interamericana (2009)
10. Parra, G., Blanco, E., Guigó, R.: Geneid in drosophila. Genome research 10(4), 511–515 (2000)
11. Salzberg, S., Searls, D., Kasif, S.: Computational gene prediction using neural networks and similarity search. Computational Methods in Molecular Biology 32, 109 (1998)

ISSN 1870-4069

Evolutionary Training of Deep Belief Networks for Handwritten Digit Recognition

Susana Espinoza-Pérez[1], Alfonso Rojas-Domínguez[1], S. Ivvan Valdez-Peña[2], Luis Ernesto Mancilla-Espinoza[1]

[1] Tecnológico Nacional de México - Instituto Tecnológico de León, Mexico

[2] Universidad de Guanajuato, Depto. de Ingeniería Electrónica, Mexico

alfonso.rojas@gmail.com

Abstract. Two of the most representative deep architectures are Deep Convolutional Neural Networks and Deep Belief Networks (DBNs). Both of these can be applied to the problem of pattern classification. Nevertheless, they differ in the training method: while the first is trained by backpropagation of the error through the whole network, the latter is typically pre-trained on a per-layer basis using an unsupervised algorithm known as Contrastive Divergence (CD), and then it is fine-tuned with a gradient descent algorithm. Although metaheuristic algorithms have been widely applied for hyperparameter tuning, little has been published regarding alternative methods to pre-train DBNs. In this work, we substitute the conventional pre-training method with an evolutionary optimization algorithm called the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES). The pretraining is achieved by minimizing the so-called reconstruction error. This proposal is validated on the problem of MNIST digit recognition by training a Deep Belief Network, following the methodology described by Hinton and Salakhutdinov (Science, 2006). It is also compared against the well known Genetic Algorithm (GA). We provide evidence to show that, although the computational cost is significantly highern than CD, the use of CMA-ES leads to a significantly smaller reconstruction error than CD and the GA.

Keywords. Deep learning, deep belief networks, deep autoencoders, metaheuristics, evolutionary algorithms.

1 Introduction

Pattern recognition can be defined as the classification of the data based on the acquired knowledge or statistical information extracted from patterns and its representations [15]. Traditional classifiation methods, such as K-means, Support Vector Machine (SVM), Artificial Neural Networks (ANNs) and linear methods, require the knowledge, contained in a dataset, to be provided to them as features. Therefore, a fundamental step in pattern classification is the selection (and extraction) of features or attributes that best represent the dataset, known as patterns. Typically, an expert does this selection [2]. However, in recent years these experts have been substituted by Deep Neural Networks (DNNs) that are capable of computing and selecting those features during their training process [10].

Despite the fact that deep learning techniques are effectively used as feature learners, some issues arise regarding their training, for instance, since they usually have complex architectures, they are more susceptible to overfitting than simpler models. Depending on the model, they can easily contain hundreds of thousands of parameters that must be calibrated or fine-tuned [22][23]. Fine-tuning the parameters of a machine learning model can be posed as an optimization task, in which the fitness function is the effectiveness of the technique over some validating set, e. g., the best parameters in a DNN are those that show the best classification result for a given problem.

Since about 2006, researchers have been attracted by Restricted Boltzmann Machines (RBMs) due to their simplicity, high level of parallelism and strong representation abilities [22]. RBMs can be interpreted as stochastic neural networks, being mainly used for image reconstruction, in which their hidden layer learns a probability distribution over an input dataset presented to the vsible layer [16]. Few years after their initial proposal, Hinton et al. realized that one can obtain more complex representations by stacking together a few RBMs, thus leading to the so-called Deep Belief Networks (DBNs) [22]. Recently, several works have been published that focus on tuning hyperparameters of DBNs based on nature-inspired metaheuristics, such as Harmony Search [17], Cuckoo Search [21], Artificial Bee Colony Algorithm [9], and Firefly Algorithm [22]. Last year, Papa et. al. [23] employed Quaternion-based Harmony Search to tune hyperparameters of DBNs and evaluated some metaheuristic techniques to tune the so called Infinity Discriminative Restricted Boltzmann Machines in the context of binary images [18].

Besides nature-inspired metaheuristics, there are other types of metaheuristics, such as Evolutionary Algorithms (EAs). The EAs can be classified into different families, for instance: Evolution Strategies (ESs) [5], Genetic Algorithms [12] and Estimation of Distribution Algorithms (EDAs) [24]. ESs were introduced by Rechenberg (1973) and further developed by Schwefel (1981); they are applied to problems in continuous domains [13]. EDAs were introduced in the field of evolutionary computation for the first time by MŰhlenbein and PaaB (1996). All of these are population-based techniques. In general, EDAs possess the advantage that they do not need extra parameters to perform the optimization; instead, a new population is sampled from a probability distribution, which is estimated from the problem database [13]. The *Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES)* [5], uses a multivariate normal distribution to search for the optimum. CMA-ES is an algorithm which, actually, shares characteristics of EAs and EDAs.

Despite the satisfactory results that have been published of metaheuristics used for the optimization of the hyperparameters of DBNs, no studies have been found where the same metaheuristics are utilized for the optimization of the weights of these networks. In this work, the use of the CMA-ES algorithm is proposed for such a task, in the context of training RBMs in an unsupervised manner, which is known as pre-training these to be used in a latter phase of a DBNs designed for classification. The technique is tested on the problem of handwritten digit classification and compared against the typical training procedure of RBMs.

The rest of this paper is organized as follows. Section 2 briefly presents all the techniques that are involved in this work. Section 3 describes our proposal and methodology. Section 4 reports our experimental results and finally, Section 5 offers our conclusions and directions for future work.

2 Materials and Methods

This section introduces all of the required technical background to understand the experimental comparison presented in this work. First the Boltzmann machines and Deep Belief networks are introduced, then the training algorithm known as Contrastive Divergence is briefly explained, and finally, the meta-heuristics compared in this work are reviewed.

2.1 Boltzmann Machines and Deep Belief Nets

Boltzmann Machines are symmetrical, undirected bipartite graphical models [11] containing neuron-like (typically binary) units that perform simple computations with which they are capable of learning internal representations. Their name comes from their sampling distribution being the Boltzmann distribution, popular in statistical mechanics for modelling energy states of a material.

The Boltzmann or Gibbs distribution is shown in Eq. (1). Notice that the maximum (probability) of the Bolztmann is the minimum of $f(x)$. Additionally, the probability mass of x depends on the T value , if T is infinite, the Boltzmann becomes a uniform distribution, on the other hand, if T goes to 0, the Boltzmann becomes a Diract delta:

$$p(x) = \frac{\exp(-f(x)/kT)}{Z}. \tag{1}$$

These properties have been exploited by researchers in computer science, by means of using the Boltzmann distribution for estimating the *minimum* of an energy function $f(x)$. Because such minimum is the most probable state, it would be sampled with a high probability if an adequate T value is chosen. Nevertheless, a direct sampling from the Bolztmann distribution requires dealing with the whole range of $f(x)$, which in turn is an exhaustive search. Fortunately, there are indirect sampling methods whose samples converge to the Boltzmann distribution, one of them is the Gibbs sampler, that takes advantage of neighborhood structures of a graphical model to iteratively generate samples that asymptotically become distributed in agreement with the Boltzmann distribution. If an arbitrary T value is chosen, the sampler could be biased towards a random

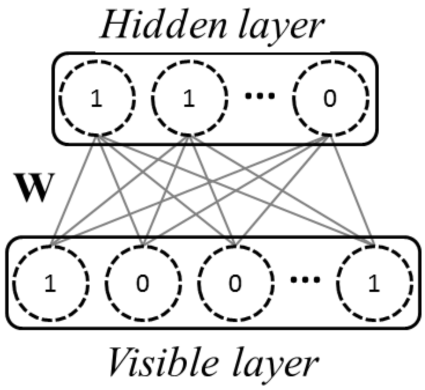


Fig. 1. A Restricted Boltzmann Machine. The matrix **W** represents the weights or connections between the computational binary units of different layers. These connections are undirected.

suboptimal point. To avoid this problem, often a slowly decreasing sequence of values for $\{T\}$ is used. This technique is called a *simulated annealing*.

Because of this, the Boltzmann machines are also classified as "energy based models" and can be seen as the stochastic, generative counterpart of Hopfield networks. In theory, a Boltzmann machine could model the distribution of data presented to it after running for long enough time at a certain energy level. In practice, however, the training time required for the machine to learn grows exponentially with the number of units in the network, and learning becomes impractical. This disadvantage is overcome using an architecture called a "Restricted Boltzmann Machine" (RBM), where the computational units are organized in layers called a visible (input) layer and a hidden layer, and intralayer connections between the units forming the network are prohibited. Thus, in an RBM there are only dependencies between units of different types (see Fig. 1).

Training of one RBM containing two layers (one visible, one hidden) is carried out via a simple algorithm generally known as Contrastive Divergence [6], which will be explained in detail in Section 2.1. Several RBMs can be stacked together one after the other in a daisy-chain fashion, where the hidden layer of one RBM becomes the visible layer of the next RBM. This strategy makes it possible to efficiently train what are called Deep Autoencoders containing many layers for dimensionality reduction and also, by adding a further supervised-learning fine tuning training step, Deep Belief Networks (DBN), which can be used for pattern recognition (e.g. in digital images and video). Each new layer added to a DBN improves the overall generative model much like other models such as multilayer perceptrons (MLPs) or convolutional neural networks (CNNs). Training of a DBN can thus be studied by looking at the training of the stacked RBMs contained within. The propagation function z_i of hidden unit i is:

$$z_i = \sum_j w_{ij} v_j + b_i, \tag{2}$$

where w_{ij} is the symmetric weight on the connection between unit i and unit j in the visible layer, v_j can take the values 0 (the visible unit is OFF) or 1 (the visible unit is ON), and b_i represents a bias added to each of the hidden units. Based on z_i, the hidden unit i will turn on with a probability given by the logistic function in Eq. (3). Although not very common, other types of variables can be used:

$$p(v_i = 1) = \frac{1}{1 + e^{-z_i}}. \tag{3}$$

The same pair of relationships, Eqs. 2 and 3 govern the behaviour of visible units with respect to the hidden units, thus representing the probability that a binary latent variable has a value of 1 during generation (top-down, from hidden to visible units) or inference (bottom-up, from visible to hidden units). If the units in alternating layers are thus sequentially updated, the network will eventually reach a stationary distribution, which is a Boltzmann distribution described in Eq. (4); notice that this is the same distribution as in Eq. (1) for $x = s$, $E(s) = f(x)/T$, and $Z = \sum_q e^{-E(\mathbf{q})}$, which is the normalization constant. In that case, the probability of a state vector $\mathbf{s}$, can be determined solely by the energy of that state vector, normalized by the sum of the energies of all possible binary state vectors $\mathbf{q}$, called the partition function:

$$P(\mathbf{s}) = \frac{e^{-E(s)}}{\sum_q e^{-E(\mathbf{q})}}. \tag{4}$$

A lower energy corresponds to a more desirable configuration of the network, so that training the network means minimizing its energy, defined in Eq. (5):

$$E(\mathbf{s}) = -\sum_i v_i(\mathbf{s}) b_i - \sum_{i<j} w_{ij} v_i(\mathbf{s}) v_j(\mathbf{s}), \tag{5}$$

where $v_i(\mathbf{s})$ represents the binary state assigned to unit i by $\mathbf{s}$. In the next subsection, the necessary steps to perform the training of an RBM by updating its weights are described.

2.2 The Contrastive Divergence Algorithm

During training of a RBM[1], its weights w_{ij} are adjusted via Eq. (6):

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} + \Delta\mathbf{W} = \mathbf{W}^{(t)} + \eta \frac{\partial log(p(v))}{\partial w_{ij}}, \tag{6}$$

where $\mathbf{W}$ represents the matrix of weights w_{ij}, $\Delta\mathbf{W}$ is the adjustment for $\mathbf{W}$ at iteration $(t+1)$ and η is known as the learning rate. This updating rule is an approximation to the maximum likelihood method, where the gradient is given by:

$$\frac{\partial log(p(v))}{\partial w_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}, \tag{7}$$

where v_i represents a *visible* node, h_j represents a *hidden* node and the angled brackets represent the expected value with respect to either the distribution of the data or the model. While computing the distribution with respect to the data is straightforward, the expectation with respect to the model requires the use of extended alternating Gibbs sampling between the visible and the hidden units. The training algorithm developed by Hinton, and known as "Contrastive Divergence" (CD) uses only n steps of Gibbs sampling ($n = 1$ works sufficiently well) [6] . The pseudo-code for CD is shown in Algorithm 1.

Algorithm 1 Pseudo-code of the Contrastive Divergence algorithm

1: Set the values of the training units to a training vector (binary data is requiered).
2: Update the hidden units given the visible units: $p(h_i = 1|\mathbf{V})$ according to Eq. (3).
3: Update the visible units given the hidden units: $p(v_j = 1|\mathbf{H})$ according to Eq. (3).
4: Re-update the hidden units given the newly obtained values for the vissible units, (redo step 2 using the new $\mathbf{V}$).
5: Perform weight adjustment $\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} + \Delta\mathbf{W}$ with Eq. (6).
6: **return** $\mathbf{W}$.

In order to demonstrate the benefits of performing pre-training on the feature extraction module in a DBN, the classification performance of the network with and without the pretraining stage is contrasted in Fig. 2. It can be observed that from the very beginning of the fine-tuning stage, the pre-trained network rapidly achieves a significantly lower error than the network without pre-training. After epoch 10, the networks improve their performance very slowly or practically have converged to their final performance. Applying the fine-tuning for many more epochs does not produce any further improvement.

As can be appreciated, CD is a very simple method for unsupervised training of an RBM. Nevertheless, it is an iterative method that requires a lot of data. In this work we test the hypothesis of whether population-based metaheuristics can perform better than CD. Furthermore, we intend to generate indications about the most adequate metaheuristic to perform the training process of an RBM. In this vein, stochastic algorithms which use descent directions are known as well-performing methods in the context of convolutional neural networks [14]. Based on such evidence, we propose to use the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES), which is a population based metaheuristic that employs a probability distribution where the maximum-variance is oriented towards the direction of steepest descent. In order to test our hypothesis and our proposal, the performance of CD is contrasted against that of the CMA-ES and against the Genetic Algorithm (GA), which is arguably the most widely

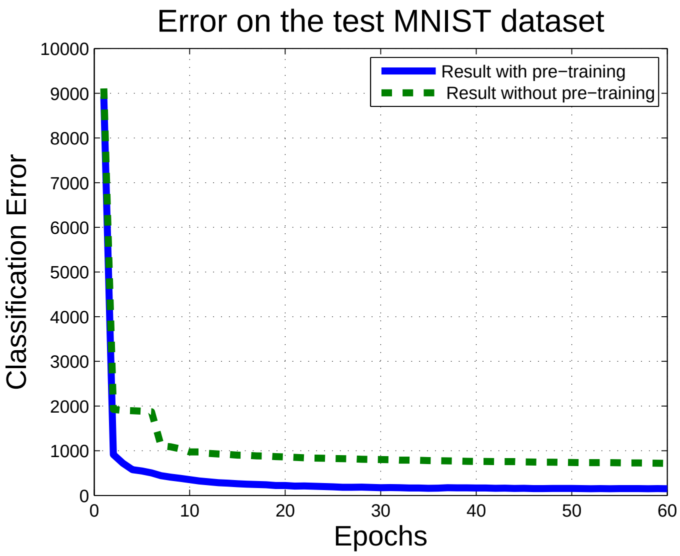


Fig. 2. Effect of employing pre-training against random weight-initialization in the classification performance.

used metaheuristic and, consequently, a suitable basis for comparison. These evolutionary algorithms are described in the following subsection.

2.3 Evolutionary Algorithms

Evolutionary Algorithms (EAs) are based on the principles of biological evolution. They evolve a population of candidate solutions (individuals), by applying reproduction and variation operators on a selected set (parent set), this set is selected with a bias to the best solutions. The intention is to produce fitter offspring for the next generation. The offspring replace the current population, usually, preserving the best solution through generations, this is called elitism. The process is repeated until a stopping criterion is met. The EAs can be classified into different families, for instance: Evolution Strategies (ES) [5], Genetic Algorithms [12] and Estimation of Distribution Algorithms (EDAs) [24], all of them are evolutionary algorithms but each of them has a particular way of working.

The *Genetic Algorithm (GA)* is arguably the most popular of the Evolutionary Algorithms. Initially developed by Holland [8] and popularized through the work of Goldberg [4], GAs are now widely applied in science and engineering as adaptive algorithms for solving practical problems. The GA is also a population-based technique inspired by natural selection and genetics. This technique maintains a population of individuals called chromosomes and formed by "genes" in some D-dimensional search space. Adaptation strategies (selection, crossover and mutation) are used to make the individuals evolve in search for the global optimum. In this way a new population is created and the process is repeated in a new generation. In GA, the selection operator chooses a pair of chromosomes for crossover, combining some of the genes of each of the parents. Only the best of the new chromosomes is passed on to the next generation (elitism). The mutation alters a chromosome, creating a new one. The process above is summarised in Algorithm 2.

Algorithm 2 Genetic Algorithm Pseudo-code

Initialize chromosome population.
Set $P_r \leftarrow$ Probability of reproduction.
Set $P_m \leftarrow$ Probability of mutation.
while stopCrit $\neq$ **true do**
    Evaluate the population.
    Choose two parent chromosomes.
    if P_r **then**
        Apply the crossover operator.
    end if
    if P_m **then**
        Apply the mutation operator.
    end if
    Accept the new solution if its fitness is lower.
end while
return $\mathbf{W}_l$

The *Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES)* [5] uses a multivariate normal distribution to search for the optimum, it is an algorithm which, actually, shares characteristics of EAs and EDAs. In continuous domains, ESs [3], usually, produce the offspring set by generating Normal deviates with mean at a parent position, and a covariance matrix often adapted in two fashions: 1) by using evolutionary operators on the parameters which determine the size and rotation of the covariance matrix, 2) by adapting the covariance matrix using improvement directions. The first is named as the self-adapted-parameter fashion while the second is the covariance matrix adaptation (CMA).

In the CMA, a set of parent solutions is used to determine the size, position and orientation of a normal distribution used to sample the candidate solutions. One of its most important characteristics is that the orientation of the multivariate normal model aligns the search to promising regions, in a kind of descent path.

The CMA-ES pseudo-code is shown in Algorithm 3, steps 1-3 initialize the multivariate Gaussian distribution and the algorithm parameters. In steps 5-7, the new candidate solutions are generated from the multivariate Gaussian distribution and evaluated on the objective function f. In step 9, the population is sorted selecting the best solutions. The previous mean is stored in step 10, then, the new mean, the isotropic path p_σ, and the anisotropic path p_c, are computed in steps 11-13. The isotropic and anisotropic path include information about the historical direction of the search, which has been computed using the best solutions, the first is a kind of covariance-normalized direction while the second is an absolute direction. Lastly, in steps 14-15 the new covariance matrix is computed, and the isotropic and anisotropic paths are used to modify it,in order to guide the search toward the improvement directions, hence, it is expected that the maximum variance direction has a large projection over the maximum improvement direction. Steps 5-15 compose the main iteration of the CMA-ES, these steps are repeated until a stopping criterion is met.

Algorithm 3 Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) Pseudocode

1: Set $\lambda \leftarrow$number of samples per iteration, at least two, generally > 4
2: Initialize $m, \sigma, C = I, p_\sigma = 0, p_c = 0$
3: $t \leftarrow 0$
4: **while** stopCrit $\neq$ **true do**
5:     **for** $i = 1 : \lambda$ **do**
6:         $x_i \leftarrow \mathcal{N}(m, \sigma^2 C)$
7:         $\mathcal{F}_i \leftarrow f(x_i)$
8:     **end for**
9:     $x_{1\ldots\lambda} \leftarrow x_{s(1)\ldots s(\lambda)}$ with $s(i) = argsort(f(1, \ldots, \lambda), i)$
10:     $m' = m$
11:     $m \leftarrow update\_m(x_1, \ldots, x_\lambda)$
12:     $p_\sigma \leftarrow update\_ps(p_\sigma, \sigma^{-1} C^{-1/2}(m - m'), ||p_\sigma||)$
13:     $p_c \leftarrow update\_pc(p_c, \sigma^{-1}(m - m'), ||p_\sigma||)$
14:     $C \leftarrow update\_C(C, p_c, (x_1 - m')/\sigma, \ldots, (x_\lambda - m')/\sigma)$
15:     $\sigma \leftarrow update\_sigma(\sigma, \ldots, ||p_\sigma||)$
16: **end while**
17: **return** x_1

3 Methodology

In this work, we substitute the traditional method of performing the pre-training, with evolutionary optimization algorithms, the CMA-ES and the well known GA, for a direct comparison of their abilities to perform the pre-training of three RBMs stacked together to form a DBN. The network pre-training is achieved by minimizing the reconstruction error, Eq. (8). This proposal is validated on the problem of MNIST digit recognition by training a Deep Belief Network, following the methodology described by Hinton and Salakhutdinov [7] and is also compared against the well known Genetic Algorithm (GA). All of these techniques were described in Section 2.

The MNIST (Modified National Institute of Standards and Technology) database is a moderately large database of handwritten digits, that has become very popular in the field of machine learning and particularly, deep learning. The database contains instances of digits, each in the form of an image of size

Fig. 3. A sample of the MNIST handwritten digits.

28×28 pixels. There are 60,000 training images and 10,000 testing images in the database (Fig. 3).

The problem of classification of MNIST digits consists of building a system that automatically provides the true class of the images in the testing set. This can be accomplished using a set of RBMs stacked one after the other, as explained in Section 2.1, followed by a “completely connected” or “fully connected” (FC) layer with ten neurons (one per digit class) that provide the predictions of the system [20, 7, 19]. The set of RBMs are considered the feature extraction module and are pre-trained using CD, an unsupervised method, and then the whole network (i.e. including the FC layer) is fine-tuned using the supervised method of gradient descent known as backpropagation. In this work we follow the methodology described by Hinton and Salakhutdinov [7], and modify their code, provided as Supporting Online Material where the authors employed the conjugate gradient method as an algorithm for the fine-tuning of the network.

Hinton and Salakhutdinov defined the architecture of their network with five layers of feature extraction of sizes: $(28 \times 28 = 784) \times 500 \times 500 \times 2000 \times 10$ (last is the FC layer), this is illustrated in Fig. 4-a. The authors also used a training procedure based on what is known as mini-batches, where the training dataset is partitioned into relatively small subsets (the mini-batches) that are presented to the system between weight updates in order to accelerate the training (a more typical approach would be to perform an update of the network weights after each individual instance). For reproducibility purposes, and given that they obtained a competitive result corresponding to 1.2% test error, in this work the architecture and hyper-parameters defined by Hinton and Salkhutdinov are used, with one exception described below.

A population-based metaheuristic maintains a set of candidate solutions, known as “individuals” for short, from which better solutions (or at least not-worse than the current best) are produced. Each of the candidate solutions is represented by a vector typically defined in the solution space; i.e., each

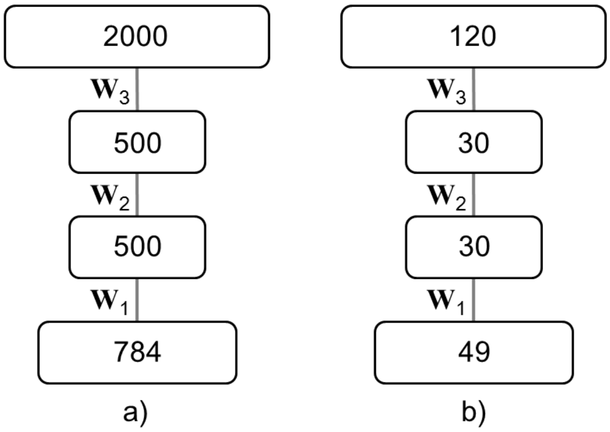


Fig. 4. (a) Architecture used in Hinton and Salakhutdinov. (b) Architecture used in this work.

individual is represented by a vector with as many elements as the dimensionality of the solution space, D. In this work, the solutions that is looked for is a series of weight matrices and bias vectors, one matrix at a time. Each of these matrices contains several thousand elements, for instance, weight matrix $\mathbf{W}_1$ in Fig. 4-a contains $784 \times 500 = 392,000$ elements. This is to say that in order to obtain the optimal $\mathbf{W}_1$, a population of individuals is to be maintained, each individual of size 392,000 elements. The CMA-ES algorithm described in Section 2.3 involves the computation and diagonalization of a covariance matrix of size $D \times D$, and this becomes prohibitively large (the data does not fit in memory) for $D = 392,000$. Although some alternatives (such as hashing algorithms [25]) exist for large-scale and high-dimensional tasks, in this work we use a simpler approach in order to provide a straightforward proof-of-concept of our contribution that can work with other similar problems in smaller dimensions: resizing the training data. By resizing the images in the MNIST dataset to one fourth of their original size, and consequently the dimensions of the network to those illustrated in Fig. 4-b, we found that it is possible for the CMA-ES algorithm to handle the problem directly. Thus, the final architecture used in the experiments is of size: $49 \times 30 \times 30 \times 120 \times 10$. The pseudocode of our experimental methodology is provided in Algorithm 4.

4 Experimental Results

In this section we report the results of performing the pre-training of the feature extraction module for the classification of MNIST digits, comparing the typical training of RBMs against using metaheuristics, in this case the CMA-ES algorithm and the GA. The main two elements for metaheuristics to function are the

Algorithm 4 Methodology for training a DBN

Set $J \leftarrow$ input layer size, $K \leftarrow$ hidden layer size.
For CD: Set $\alpha \leftarrow$ momentum, $\eta \leftarrow$ learning rate.
For CMA-ES: Set $M \leftarrow$ population size.
For GA: Set $M \leftarrow$ population size.
Split dataset, X, into mini-batches, x_i.
while stopCrit $\neq$ **true do**
  for $i = 1 : N$ **do**
    if CD **then**
      Compute E_i with $\mathbf{W}_l^{(t)}$.
      Compute $\Delta\mathbf{W}$, Eq. (7).
      Update $\mathbf{W}_l^{(t+1)} = \mathbf{W}_l^{(t)} + \Delta\mathbf{W}$, Eq. (6).
    else
      if CMA-ES **then**
        Create population $\mathbf{P}^{(t)} = \{\mathbf{W}_l^{(0)}, \mathbf{W}_l^{(1)}, \ldots, \mathbf{W}_l^{(M)}\}$.
        Compute E_i with $\mathbf{P}^{(t)}$.
        Update population with **Algorithm 3**.
      end if
    else
      if GA **then**
        Create population $\mathbf{P}^{(t)} = \{\mathbf{W}_l^{(0)}, \mathbf{W}_l^{(1)}, \ldots, \mathbf{W}_l^{(M)}\}$.
        Compute E_i with $\mathbf{P}^{(t)}$.
        Update population with **Algorithm 2**.
      end if
    end if
  end for
  Compute epoch error: $E = \sum_i E_i$.
end while
return $\mathbf{W}_l$

shape of an individual candidate solution and a fitness function to be optimized (typically, minimized). In this work, the candidate solutions are defined as a concatenation of the weight matrix of the i-th RBM, W_i (flattened), the visible layer biases, b_i, and the hidden layer biases, b_i. All of these are defined as real values. The fitness function used is the reconstruction error, which is the same function guiding the CD algorithm. To compute the reconstruction error of one RBM, input data is given to the RBM, the activation of the hidden layer is computed using Eqs. (2) and (3). Then the reconstruction, $\tilde{v}$ is computed using the same equations, but taking the hidden layer as the input and the visible layer as the output. Thus, the reconstruction error is defined as:

$$E = \sum_j (v_j - \tilde{v}_j)^2. \tag{8}$$

For each of the RBMs, the CD algorithm, the CMA-ES and the GA were executed for 50 iterations, so that the total amount of iterations allocated to train the feature extraction module is 150. The reconstruction error for each iteration was recorded. Since the iterations coincide with the presentation of all the instances of the training data, they are referred to as "epochs".

The control parameters used in the experiments for CD, CMA-ES and GA are reported in Table 1. Notice that three values are reported for the number of visible units in the visible layer, N_{vis}, and the number of units in the hidden layer, N_{hid}. The three values correspond to the three RBMs that compose our system and that are trained sequentially one after another.

Table 1. Control parameters used for the CD, CMA-ES and GA.

Parameter	Value
Contrastive Divergence	
N_{vis}	$\{49, 30, 30\}$
N_{hid}	$\{30, 30, 120\}$
Initial Population, W	$\mathcal{N}(0, 0.1)$
Initial Population, biases	zeros
Learning rate	$\eta = 0.1$
Initial Momentum, α_I	0.5
Momentum α	0.9
CMA-ES	
Problem dimension, D	$N_{vis} \times N_{hid} + N_{vis} + N_{hid}$
Initialization method	$\mathcal{N}(0, 0.5)$
Population size, λ	$4 + \lfloor 3 \log(D) \rfloor$
Recombination points	$\lambda/2$
Genetic Algorithm	
Problem dimension, D	$N_{vis} \times N_{hid} + N_{vis} + N_{hid}$
Initial Population	$\mathcal{N}(0, \sigma_l)$ for each layer l
Standard Deviation	$\sigma_1 = 0.5, \sigma_2 = \sigma_3 = 0.1$
Population size, λ	$\gamma_l(4 + \lfloor 3 \log(D) \rfloor)$, $\gamma_1 = 15, \gamma_2 = 3, \gamma_3 = 3$

The experiment was repeated 30 times and the average result is shown in Fig. 5. The plot shows the reconstruction error obtained by CD (continuous line), CMA-ES (long-dashed line) and the GA (short-dashed line). Vertical bars along the curves are used to show the standard deviation over the 30 trials.

It can be observed that CD exhibits a much smaller variability than CMA-ES. On the other hand, CMA-ES achieves significantly smaller error for the first two RBMs trained. The GA exhibits a much larger variability than CD and comparable to that of CMA-ES, except for the final portion of the epochs, where the population of CMA-ES converges but the one of GA does not. In the end (epoch 150) the mean performance of GA is within a standard deviation of the mean performance of CMA-ES, but in general CMA-ES achieves the lowest error significantly faster than GA; at epochs 85-90, CMA-ES has reached the lowest error recorded. In particular, for the first RBM, GA requires many more epochs

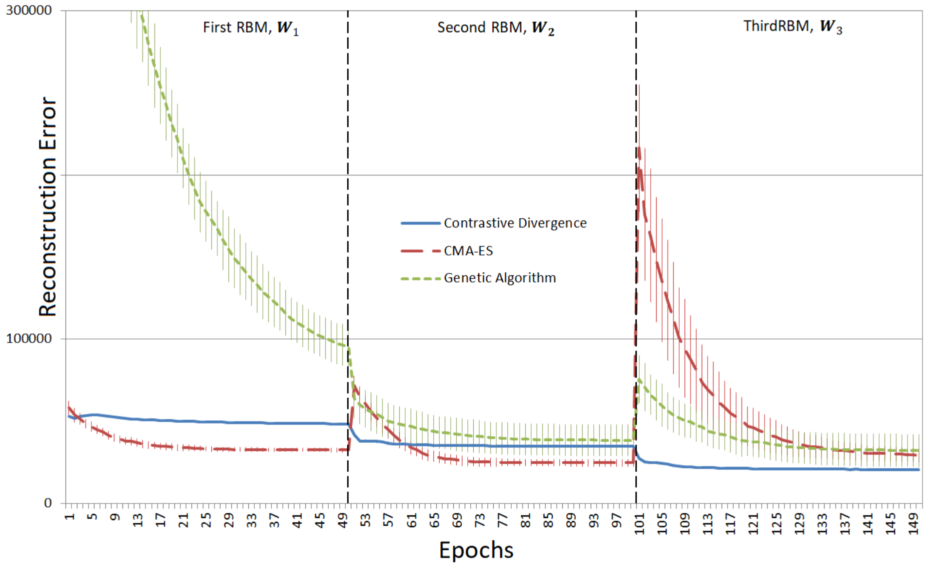


Fig. 5. Error plots comparing CD against CMA-ES and GA.

to significantly reduce the error level, so much so that the beginning of the curve lies outside of the figure. In fact, after 50 epochs one can observe that the curve of the Genetic Algorithm still points significantly downwards, indicating that the GA still has potential to improve the weight optimization beyond the 50 epochs.

Regarding other aspects of the algorithms, it must be noted that in this work the same scheme of training via mini-batches was used for both training strategies. According to this scheme, a training epoch is completed when the network is presented with all of the training data, one mini-batch at a time. While this approach is most beneficial for the Contrastive Divergence algorithm, it may not be the best way to use the training data to guide the CMA-ES and GA algorithms. For instance, one can think of employing a randomized sample of the training data to be used in the computation of an estimate of the reconstruction error, and use that estimate to guide the evolution in these algorithms. This would eliminate a considerably part of the computational cost incurred in having to evaluate all the individuals in the CMA-ES / GA population on all of the training data just to guide the exploration of the solution space. Instead of adjusting the training scheme to benefit each of the algorithms, in this work we employed the same training scheme for both algorithms for the sake of a fair comparison.

5 Conclusion and Future Work

In this work, the proposal of using metaheuristics to perform the pre-training of the feature extraction module in a DBN was put forward and tested using the CMA-ES, which is a population-based metaheuristic. DBNs were pre-trained for the problem of handwritten digit recognition using the MNIST dataset, and their performance was measured by the reconstruction error of the RBMs included in the feature extraction portion of the DBN. The performance of the CMA-ES was compared against that of CD and also against that of the GA. Our results show that CMA-ES performs the best among these algorithms and validate our hypothesis regarding the suitability of CMA-ES to successfully approximate the optimal weights of the DBN. Nevertheless, we have observed that the success of the algorithm strongly depends on the problem dimension. This becomes evident in Fig. 5 where CMA-ES outperforms the other two methods up to the beginning of the third RBM, where it completely loses its advantage and struggles to come back in track. The reason behind this is that the third RBM contains the most units and therefore substantially more weights between its layers. This has a particularly detrimental effect on the CMA-ES algorithm.

It should be noted that in this work our interest was to analyse the performance of metaheuristic techniques for the pre-training of the different RBMs in a DBN designed for handwritten digit recognition but not to observe the end result. This is because the set of stacked RBMs should undergo a second training stage after the pre-training (known as fine-tuning), to improve the classification performance. This second training process is very efficient and may obscure the existing differences between the pre-training achieved by each of the compared techniques. In a future work we will study in detail the effect of the pre-training on the convergence of the fine-tuning step.

As a general conclusion, we have determined that metaheuristics (in particular the CMA-ES), are competitive methods for pre-training a DBN using the reconstruction error. However, metaheuristics are quite demanding in terms of the computational resources (such as memory, operations and objective function evaluations) required to perform the optimization. This drawback are inherit to both metaheuristics tested, but not in the same scale. The CMA-ES requires memory of the order $O(n^2)$ for storing a covariance matrix, in consequence, the size (dimensionality) of the problems that can be tackled by this algorithm are bounded by this computational resource. This disadvantage forced us to apply a scaling to the handwritten digit images in order to reduce their dimensionality. Another disadvantage of employing CMA-ES (and any other population-based algorithm) is the cost of evaluating all the candidate solutions in the population.

In this regard, we can readily identify two main issues to investigate in future work: first, to identify other metaheuristics with a similar or better performance than CMA-ES and with lesser complexity and computational requirements. Second, to employ a different training scheme, more suitable to the particular requirements of the selected metaheuristics (instead of being tailored for CD). In any case, it can be concluded that the combination of metaheuristics and deep learning methods is a promissory line of research.

Acknowledgements. This work is supported by CONACYT, through Scholarship 482124 (S. Espinoza) and Research Grant Cátedras-2598 (A. Rojas). Alfonso Rojas is a CONACYT Research Fellow.

References

1. Ackley, D.H., Hinton, G.E., Sejnowski, T.J.: A learning algorithm for boltzmann machines. Cognitive science 9(1), 147–169 (1985)
2. Andrew R. Webb, K.D.C.: Statistical pattern reconigtion. Wiley, 3 edn. (2011)
3. Beyer, H.G., Schwefel, H.P.: Evolution strategies–a comprehensive introduction. Natural computing 1(1), 3–52 (2002)
4. Goldberg, D.E.: Genetic algorithms in search, optimization, and machine learning. Addison-Wesley (1989)
5. Hansen, N., Niederberger, A., Guzzella, L., Koumoutsakos, P.: A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. IEEE Transactions on Evolutionary Computation 13(1), 180–197 (2009)
6. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural computation 18(7), 1527–1554 (2006)
7. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. science 313(5786), 504–507 (2006)
8. Holland, J.H.: Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press (1992)
9. Horng, M.H.: Fine-tuning parameters of deep belief networks using artificial bee colony algorithm. 2nd International Conference on Artificial Intelligence: Techniques and Applications (AITA 2017) (2017)
10. Huy Nguyen, M.L.N.: A deep neural architecture for sentence-level sentiment classification in twitter social networking. In International Conference of the Pacific Association for Computational Linguistics pp. 15–27 (2017)
11. Jordan, M.I.: Learning in Graphical Models. MIT press, Cambridge Mass (1998)
12. Klanac, A., Jelovica, J.: A concept of omni-optimization for ship structural design. Advancements in Marine Structures, Proceedings of MARSTRUCT pp. 473–481 (2007)
13. Larrañaga, Pedro, Lozano, A, J.: Estimation of distribution algorithms: A new tool for evolutionary computation, vol. 2. Springer Science & Business Media (2001)
14. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature 521(7553), 436 (2015)
15. M. Narasimha Murty, V.S.D.: Pattern Recognition, An Algorithmic Approach. Springer-Verlag London, 1 edn. (1988)
16. Papa, J.P., Rosa, G.H., Marana, A.N., David Cox, W.S.: Model selection for discriminative restricted boltzmann machines through meta-heuristic technique. Journal of Computational Science 9, 14–18 (2015)
17. Papa, J.P., Scheirer, W., Cox, D.: Fine-tuning deep belief networks using harmony search. Applied Soft Computing 46, 875–885 (2016)
18. Passos, L.A., Papa, J.P.: Fine-tuning infinity restricted boltzmann machines. In Graphics, Patterns and Images (SIBGRAPI) pp. 63–70 (2017)
19. Ranzato, M., Huang, F., Boureau, Y., LeCun, Y.: Unsupervised learning of invariant feature hierarchies with applications to object recognition. Proc. of Computer Vision and Pattern Recognition Conference (CVPR 2007) (2007)

20. Revow, M., Williams, C.K., Hinton, G.E.: Using generative models for handwritten digit recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(6), 592–606 (1996)
21. Rodrigues, D., Yang, X.S., Papa, J.P.: Fine-tuning deep belief networks using cuckoo search. Bio-Inspired Computation and Applications in Image Processing pp. 47–59 (2017)
22. de Rosa, G.H., Yang, X.S., Papa, J.P., da Costa, K.A.P.: Learning parameters in deep belief networks through firefly algorithm. In IAPR Workshop on Artificial Neural Networks in Pattern Recognition pp. 138–149 (2016)
23. de Rosa, G.H., Yang, X.S., Papa, J.P., Pereira, D.R.: Quaternion-based deep belief networks fine-tuning. Applied Soft Computing 60, 328–335 (2017)
24. Valdez, S., Hernández, A., Botello, S.: A Boltzmann based estimation of distribution algorithm. Information Sciences 236, 126–137 (2013)
25. Wang, J., Liu, W., Kumar, S., Chang, S.: Learning to hash for indexing big data—a survey. Proceedings of the IEEE 104(1), 34–57 (Jan 2016)

ISSN 1870-4069

Use of Parallel Patterns of Communication between Processes for search of Sequences DNA and RNAi Strings

Mario Rossainz-López[1], Manuel Capel-Tuñón[2], Ivo Pineda-Torres[1], Ivan Olmos-Pineda[1], Arturo Olvera-López[1]

[1]Benemérita Universidad Autónoma de Puebla, Faculty of Computer Science, Puebla, Mexico

[2]University of Granada, Software Engineering Department, College of Informatics and Telecommunications, Granada, Spain

{rossainz,ipineda,iolmos, aolvera}@cs.buap.mx, manuelcapel@ugr.es

Abstract. Within an environment of Parallel Objects, an approach of Structured Parallel Programming and the paradigm of the Orientation to Objects show a programming method based on High Level Parallel Compositions or HLPC to solve two problems of combinatorial optimization: grouping fragments of DNA sequences and the parallel exhaustive search (PES) of RNA strings that help the sequence and assembly of DNAs in the construction of gnomes. The Pipeline and Farm models of parallel patterns of communication between processes are shown as HLPC under the Object Orientation paradigm to solve the cited problems. Each HLPC contains a set of predefined synchronization constraints between processes, as well as the use of synchronous, asynchronous and asynchronous future modes of communication. We show the algorithms that solves the problems, their design and implementation as HLPC and the performance metrics in their parallel execution using multicores and video accelerator card.

Keywords. HLPC, parallel structured programming, parallel objects, pipeline, farm, DNA, RNAi.

1 Introduction

The parallel processing is an alternative to the sequential processing when the limit of performance of a system is reached. In the sequential computation a processor only carries out at the same time an operation, on the contrary of what happens in the calculation parallel, where several processors they can cooperate to solve a given problem, which reduces the time of calculation since several operations can be carried out simultaneously. The present research uses structured parallel programming through a POSIX thread library as a methodological programming proposal based on the pattern of the High-Level Parallel Compositions HLPC [1], [2], the which it is based on the paradigm of Orientation to Objects to solve problems parallelizable using parallel objects.

In this work supply a library of classes that provides the programmer the communication/interaction patterns more commonly used in the parallel programming: the patterns pipeline and farm. With them, problems like the assembly of DNA strings and Parallel Exhaustive Search (PES) of RNAi strings proposed in this paper can be solved. In the case of the first problem mentioned, it is a problem of combinatorial optimization in which diverse heuristics and met heuristics have been proposed to assemble sequences of DNA strings and to provide essential information to understand the species and their mechanisms of life including the human species.

This work shows the implementation of a grouping algorithm that evaluates a set of DNA sequence fragments as a HLPC. The HLPC represents a Farm where worker processes are themselves Pipeline HLPCs. The algorithm determines subgroups of fragments by DNA sequences matching found, which have a high probability of being aligned in an assembly task. Each worker process of HLPC Farm works in parallel with the other worker processes that are generated with a group of fragments of DNA sequences that are internally constructed as graphs represented through the HLPC Pipeline and through an in-depth search the new groups of DNA sequences are generated, which must be processed by some assembly technique to form the contigs of a genome that has been sequenced covering most of its structure but missing a fragment to be completed.

Finally, the design of an experiment is shown through the use of the new HLPC generated called HLPC GraphADN, with genomes of viruses and bacteria available on the web. The pseudo random synthetic readings created to form contigs are shown and the execution performance of this proposal is obtained for eight genomes with an Intel Core i8 processor, a video accelerator card with 1664 CUDA cores and a clock frequency of 1178 MHz. In the case of the second problem a Parallel Exhaustive Search (PES) of RNAi strings is implemented using the inter-process communication pattern called FARM through an HLPC. The authors propose a new HLPC model called HLPC ARNi that is based on the HLPC Farm, which is part of a parallel object library whose details can be consulted in [3], [4].

The HLPC RNAi performs a pre-processing with the input data through its controller object of the Farm, which consists in join in a single string, all the input RNAi strings including the string that contains the characteristics of the organism that is analyzed. The controller then distributes to the farm worker objects within the HLPC RNAi the string constructed so that they, in parallel, perform the Parallel Exhaustive Search, [5], [6], find the matches based on a pre-established substring length and return the search results.

An experiment was designed using the HLPC ARNi with the RNAi strings database located in the Pombase site, referring to the yeast species known as S. pombe., with different lengths of substrings for the search of RNAi strings of the species of yeast mentioned and results of execution times and performance analysis were obtained using 3 to 8 cores of a dual-core Intel server machine.

2 Definition of High Level Parallel Compositions (HLPC)

Using an OO-programming environment, the basic idea is implementing any type of parallel communication patterns between the processes of an application or distributed/parallel algorithm. A HLPC comes from the composition of a set three object types: an object manager (Fig. 1) that represents the HLPC itself and makes an encapsulated abstraction out of it that hides the internal structure [7].

The object manager controls a set of objects references, which address the object collector and several stage objects and represent the HLPC components whose parallel execution is coordinated by the object manager. The objects stage are objects of a specific purpose, in charge of encapsulating a client-server type interface that settles down between the manager and the slave-objects. And a collector object, we can see an object in charge of storing the results received from the stage objects to which is connected, in parallel with other objects of HLPC composition. During a service request the control flow within the stages of a HLPC depends on the implemented communication pattern.

Manager, collector and stages are included in the definition of a PO [8]. POs are active objects, which have intrinsic execution capability. Applications that deploy the PO pattern can exploit the inter-object parallelism as much as the intra-object parallelism. A PO-instance object has a similar structure to that of an object in C++, and additionally defines a scheduling policy that specifies the way in which one or more operations carried out by the instance synchronize. The communication modes used are: The synchronous communication, the asynchronous communication and the asynchronous future. The Synchronization policies are expressed in terms of restrictions; for instance, mutual exclusion in reader/writer processes or the maximum parallelism allowed for writer processes.

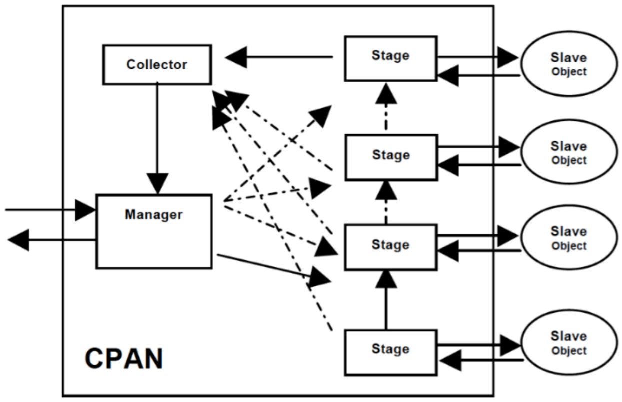


Fig. 1. Internal structure of HLPC.

3 The HLPC Pipeline

We can use this parallel processing technique to solve a problem by splitting it into a series of successive tasks so that data flow in the direction given by the process interconnection structure. Each task can therefore be completed sequentially [9]. In a pipeline, each task is executed by a processor or process as shown in Fig. 2. Using the technique of the pipeline, the idea is to divide the problem in a series of tasks that must be completed, one after another, see Fig. 2. In a pipeline each task can be executed by a process, thread or processor for separate, [9].

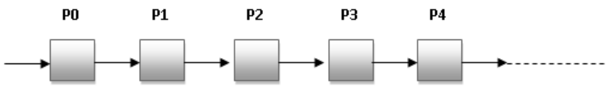


Fig. 2. Internal structure of HLPC.

The parallel pipeline processing technique is presented as a high-level parallel composition for solving a wide variety of problems from several different areas that can already be solved by partially sequential algorithms. In this way, the HLPC pipeline guarantees code parallelization of the resulting algorithm while taking advantage of existing sequential algorithms by using the HLPC processing pattern. The Fig. 3 represents the parallel pattern of communication Pipeline as a HLPC. The details of the implementation can be consulted in [10].

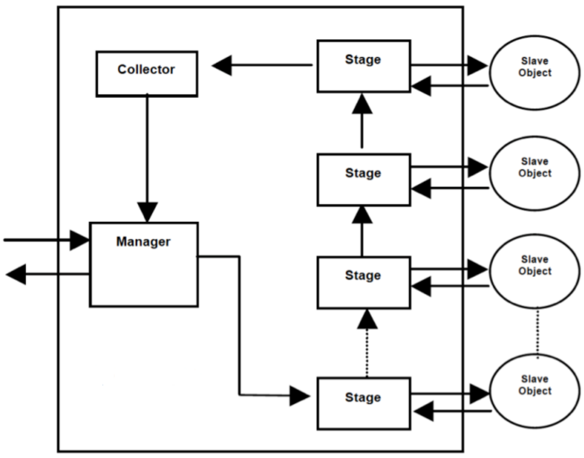


Fig. 3. The HLPC of a Pipeline.

4 The HLPC FARM

The technique of the parallel processing of the farm as a HLPC is shown here. The so named farm parallel pattern of interaction is made up of a set of independent processes, called worker processes, and a process that controls them, called the process controller [11]. The worker processes are executed in parallel until all of them reaches a common objective. The process controller distributes the work and controls the progress of the farm until the solution of the problem is found [12]. Fig. 4 shows the pattern of the farm.

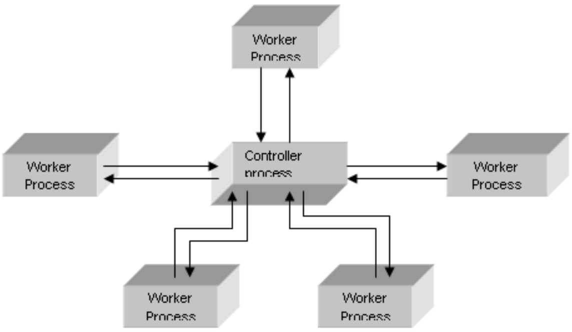


Fig. 4. Farm with a controller and five workers.

The representation of parallel pattern farm as a HLPC is shown in Fig. 5. The details of the implementation can be consulted in [13].

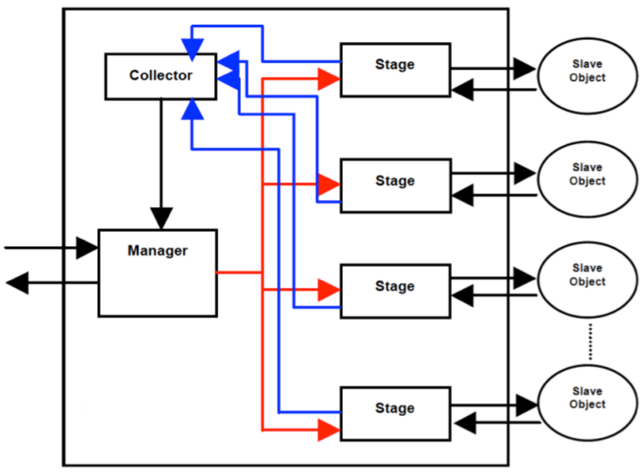


Fig. 5. The HLPC of a farm.

5 Methodology

1. The slave objects, which are executed by the stages and represent instances of the functionality required from the HLPC are created.

2. The sequential algorithm that solves the initial problem is implemented and its instances become methods associated to the slave objects.
3. A list of associations (e.g. slave object, associated method) are created.
4. An instance of the Manager is created to represent the HLPC-manager being constructed. It is then initialized with the list of associations from the previous step. At this moment, the necessary internal stages of the HLPC-manager are automatically created as instances of the Stage and each obtains an association.
5. The initial data to be processed are specified by creating data types and user-defined data as objects, whereas the input data set represents the problem to solve.
6. Application execution is requested to find a solution to the problem.
7. Parallel execution is achieved by transparently using synchronous, asynchronous and future asynchronous communication for the user's main program. The HLPC then creates its Collector object, which is passed as a reference to every stage connected with the Collector object, together with a copy of the data set to be processed. The initial stage works with these data and they are processed by the slave object associated to each stage. The result is then passed on to the next stage which is then created and initialized by the previous stage. In the same way, the second stage processes the data received when its slave object method is executed, and the new result is sent to the next stage, and so on. Finally, the obtained results are received by the Collector object which then compiles, processes and sends them to the Manager object which, in turn, sends the solution outside the HLPC.
8. Finally, the results are shown to the user.

6 Case Study 1: Parallel Exhaustive Search of RNA Strings using HLPC

The search for DNA strings for the construction of genomes of an organism is included within the genomics area and the representation of their information within bioinformatics is done digitally through formats used for this purpose, which are in general flat text formats, relatively simple and easy to analyze (parsing) associated with a search algorithm. The best known and which is used in the present work is the gene sequence representation format called FASTA, [14]. Today there are different approaches for this type of searches, but the most common is to use parallel computing technologies due to the large number of data that are generated and that must be processed in order to achieve an assembly of acceptable DNA sequences. In this work, it is proposed by using the HLPC model to carry out a Parallel Exhaustive Search (PES) of RNA or DNA strings using the communication pattern called FARM. The PES was carried out in plain text files containing a representation of RNA strings of an organism with its respective name. In the HLPC Farm used, the process controller or manager performs the pre-processing of the file extracting the strings written in the FASTA format to create the dictionary formed of the characteristics of the strings and the strings themselves, which is sent to each worker process or stage to perform the exhaustive search using the associated algorithms in the manager object and in the slave objects of the HLPC. This search is carried out in parallel by all the farm worker processes. The HLPC Farm always guarantees a workload balance of these procsesses thanks to the synchronization restriction of the maximum parallelism that its components have guaranteeing the reduction in the execution times of each worker process, but also of the HLPC Farm itself. Then a new model of HLPC called HLPC ARNi is created, which is shown in Fig. 6. The pre-processing of the data is done through a text file as input to the HLPC ARNi. This file contains the name of a text string, as well as its characteristics, which is sent to the Manager of the HLPC or Farm Controller Process. The Manager has the Pre-Processing through a Slave Object, which consists of joining in a single string, all the ARNi strings that are in the TXT input file including the line text that contains the characteristics of the organism in question, then the Manager distributes to each Stage process the corresponding workload defining the limits start and end of the PES for each Stage.

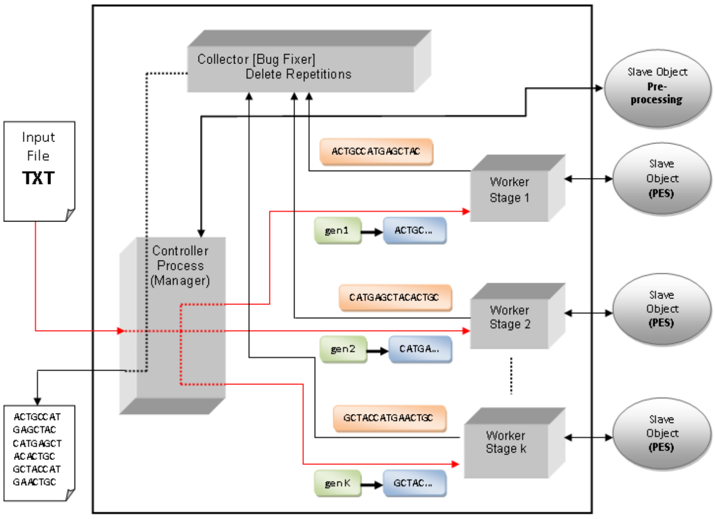
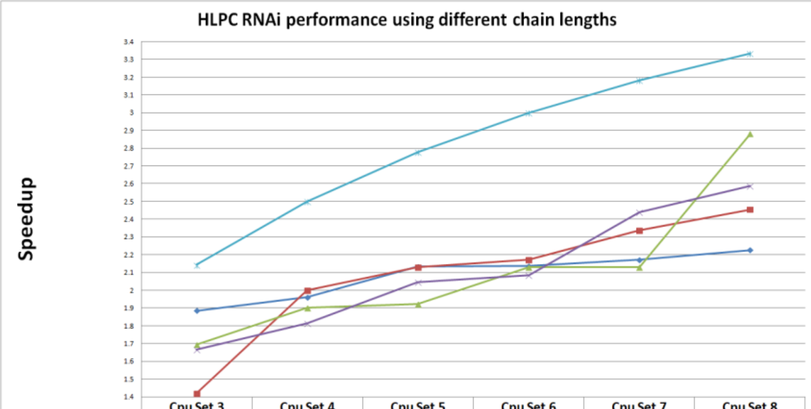


Fig. 6. Model of HLPC ARNi.

The load balance is made using the maximum parallelism in each Stage process (worker) of the Farm which is based on the identifier number of each worker process, [15], [16]. Subsequently the PES is performed in each stage of the HLPC ARNi executing the associated algorithm through the corresponding Slave Object and if a predefined substring is found within the ARNi string it is sent to the Collector object, which receives them in parallel from all the stage processes (workers) connected to it. The Collector bugs fixer eliminates repetitions of strings and the result is sent to the Manager who in turn sends it to an output file to the user.

6.1 Efficiency of HLPC ARNi

An experiment was designed using the HLPC ARNi with the RNAi string database located at the Pombase site (ftp://ftp.ebi.ac.uk/pub/databases/pombase/pombe/Chromosome_Dumps/fasta/). Easts are the agents of fermentation and are found on the surface of plants [17]. The experiment was carried out on a server machine with Intel Xeon processor 2630 2.40 GHz and 8 cores. The operating system installed is Linux Ubuntu 64 bits, with a RAM of 125 Gbytes and a clock frequency of 1200 MHz. Experiments were performed with different number of nodes to determine if the workload was performed correctly. To determine if the length of the strings to be searched has a direct impact on the execution time, experiments with different string lengths were performed for their search, where the execution times increase according to the length of the search string grows, but on the other hand the execution times decrease as the number of nodes (nuclei) that are used in the execution of the search increases. In Fig. 7 shows the scalability of the Speedup found in the HLPC RNAi for different string lengths using 3 to 8 nodes in its execution, showing generally a good acceleration as the number of nodes increases.



	Cpu Set 3	Cpu Set 4	Cpu Set 5	Cpu Set 6	Cpu Set 7	Cpu Set 8
Chain Length = 10	1.885336273	1.961009174	2.132169576	2.1375	2.172808132	2.2265625
Chain Length = 20	1.418485237	2	2.129094412	2.173058014	2.338624339	2.455555556
Chain Length = 30	1.69481654	1.901960784	1.92333113	2.128749086	2.128749086	2.881188119
Chain Length = 40	1.665427509	1.813765182	2.0456621	2.08372093	2.440087146	2.586605081
Amdahl	2.142857143	2.5	2.777777778	3	3.181818182	3.333333333

Fig. 7. Scalability of the magnitude of the Speedup found for the HLPC RNAi in exclusive nodes of 2, 3, 4, 5, 6, 7 and 8 cores.

7 Case Study 2: Assembly of DNA Sequences using HLPC

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule [18]. Knowledge of DNA sequences has become essential to carry out basic biological research. In this regard use of HLPCs for grouping DNA sequence fragments from the parallelization of a clustering algorithm to evaluate a set of fragments are made, which have a high probability of being aligned in an assembly task [19]. The algorithm finds the splices between the fragments using the Myers algorithm and links them in a graph. Then an in-depth search is done in the graph to form the groups and send them as a result. The assembly of DNA strings is proposed as a combinatorial optimization problem and is classified as NP-hard and is based on the paradigm divide-and-conquer using a structure type farm, so that the computational cost of finding the sequence alignments and its splice is substantially reduced with respect to its sequential version. The number of processes required to process the fragments of DNA sequences of a specific genome such as that of a virus or bacteria is determined by the splice of the strings found by the sequential solution algorithm, which looks in parallel for overlaps in the remaining fragments. Two sub-strings of each fragment are taken for comparison with other fragments; and thus, splices are located and associated with the processes (a process for each splice sequence of DNA strings). A splice graph is then generated that shows the relationship between pairs of nodes (processes), as well as the lack of communication among others. The set of nodes (processes) of the graph that are inter-related are grouped together within a worker process pattern farm. Each set of related nodes in the graph are independent and represent the grouping of fragments found. In Fig. 8 is shown the representation of HLPC for grouping DNA sequence fragments.

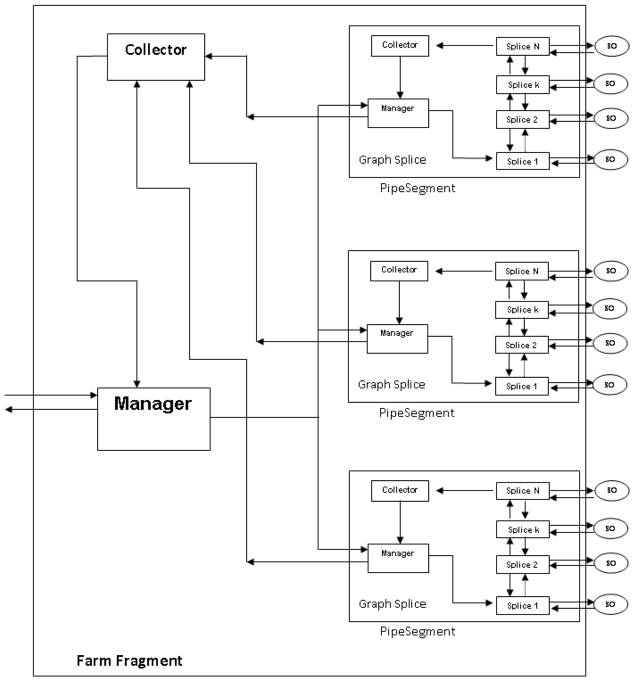


Fig. 8. The HLPC GraphADN.

The new HLPC named HLPC GraphADN is structured as a FARM of n-worker processes, i.e., n-fragments of DNA sequences and each worker process is itself a two directions-communication pipeline HLPC formed by m-stages where each stage of HLPC Pipe represents a splice sequence of DNA strings connected with both, the previous stage as the next stage. The collector object receives the number of formed groups and the elements that belong to each of the formed groups. With the latter information collected, an in-depth search is performed to locate these items and obtain the sequence groups formed by the sequential algorithm assigned to each of the HLPC's slave objects with this result, the user can use an assembly of DNA sequences to try to complete a particular genome or to finish an incomplete sequence of DNA strings of some animal or plant type species.

7.1 Efficiency of HLPC GraphADN

An experiment was designed by using the HLPC GraphADN with genomes of viruses and bacteria available on the web whose data were obtained from European Nucleotide Archive, http://www.ebi.ac.uk. The Scalability of the speedup found in HLPC GraphADN is shown in Fig. 9 to 12. The plot shows the number of processes deployed for the calculation of eight genomes in an experiment conducted on a computer Intel Core i8 processor and using a video accelerator card with 1,664 CUDA cores.

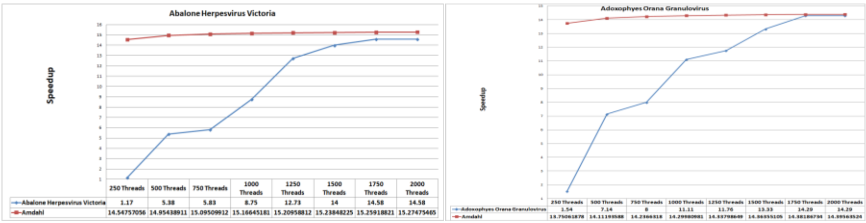


Fig. 9. Speedup found for the HLPC GraphADN for the genome Abalone herpesvirus Victoria and Adoxophyes orana granulovirus.

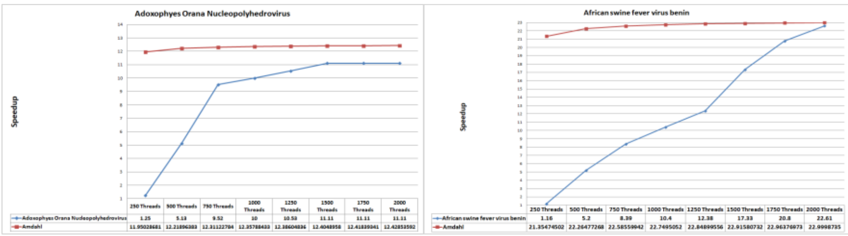


Fig. 10. Speedup found for the HLPC GraphADN for the genome Adoxophyes orana nucleopolyhedrovirus and African swine fever virus benin.

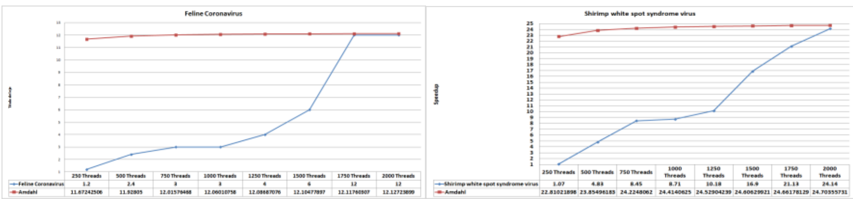


Fig. 11. Speedup found for the HLPC GraphADN for the genome Feline coronavirus and Shirimp white spot syndrome virus.

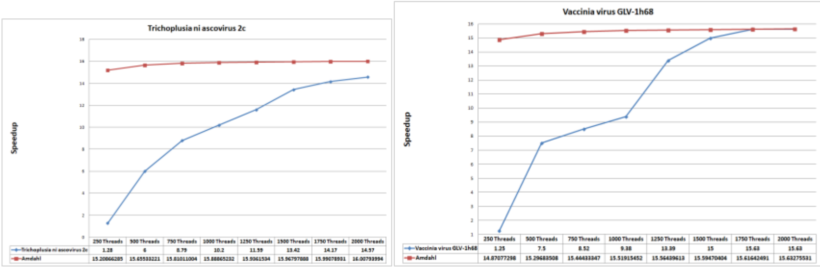


Fig. 12. Speedup found for the HLPC GraphADN for the genome Trichoplusia ni ascovirus 2c and Vaccinia virus GLV-1h68.

8 Comparative Analysis with Other Parallel Approaches

In the literature there are parallel approaches to obtaining DNA sequences that run on parallel machines or computers with multiple cores and that show the obtained accelerations as results. Such are the cases of references [20], [21] and [22]. In [20] the parallelization of two bioinformatic applications used in the analysis of long DNA sequences is shown. Its parallel proposal shows good acceleration results in computers with several processors but not in multicore computers. In [21] the same authors of the reference [20] show the accelerations found in the parallel analysis of DNA sequences using GPU and CUDA. Their results for different sizes of DNA sequences show problems to obtain good acceleration performances and propose to develop new techniques in data processing. Finally, in [22] describe parallel algorithms with different complexities that exhaustively determine all words of size k, k being arbitrarily large, in a source DNA sequence. The results shown that our algorithms achieve a high degree of scalability, allowing the detection of DNA words of 64 nucleotides in only 800 seconds.

Comparing the performance and acceleration of our proposal with those cited, it is concluded that the use of the HLPC in the two case studies presented shows better results than those published in [20] and [21] but not with the results of [22] that are better.

9 Conclusions

The authors have presented a method of design concurrent applications based on the HLPC construction that has been shown susceptible to be used in different platforms. The authors discuss the implementation of HLPCs pipeline and farm as generic and reusable patterns of communication/interaction between processes, which can even be used by inexperienced parallel application programmers to obtain efficient code by only programming the sequential parts of their applications. The HLPC pipeline and HLPC farm have been extracted from a larger library of parallel objects is made up of six sets of classes that constitute the implementation of the parallel patterns farm, pipe and others like treeDV as HLPCs that are part of what will be the library of High Level Parallel Compositions. The authors have presented two case studies: the parallel exhaustive search of RNAi strings through the new HLPC RNAi constructed; referring to the yeast species known as S. pombe, and parallel calculation of the DNA sequences for 8 genomes. In both cases of study, the efficiency of the HLPCs in the solution of the problems has been shown. It has been presented the speedup and low execution times w.r.t. best sequential version of the algorithms that solve these problems. The authors have also obtained good performance in their executions and speedup scalability (compared to Amdahl's law) on the number of processors used to obtain the solution of complex problems.

References

1. Corradi A., Leonardi L.: PO Constraints as tools to synchronize active objects. Journal Object Oriented Programming 10, pp. 42-53 (1991)
2. Danelutto, M.; Orlando, S; et al.: Parallel Programming Models Based on Restricted Computation Structure Approach. Technical Report-Dpt. Informatica, Universitá de Pisa (1995)
3. Rossainz, M., Capel M.: A Parallel Programming Methodology using Communication Patterns named HLPCS or Composition of Parallel Object. In Proceedings 20TH European Modeling & Simulation Symposium.Campora S., Giovanni, Italy (2008)
4. Rossainz, M., Capel M.: Compositions of Parallel Objects to Implement Communication Patterns. In Proceedings XXIII Jornadas de Paralelismo, SARTECO 2012, Elche, España (2012)
5. Blelloch, Guy E.: Programming Parallel Algorithms. Comunications of the ACM, Volume 39, Number 3 (1996)
6. Kumar, Vipin, et al: Introduction to parallel computing. Design and Analysis of Algorithms., USA, Benjamin-Cummings (1994)
7. Rossainz, M.: Una Metodología de Programación Basada en Composiciones Paralelas de Alto Nivel (HLPCs). Universidad de Granada, PhD dissertation (2005)
8. Corradi A, Leonardo L, Zambonelli F.: Experiences toward an Object-Oriented Approach to Structured Parallel Programming. DEIS technical report no. DEIS-LIA-95-007 (1995)
9. De Simone, et al.: Designs Patterns for Parallel Programming. PDPTA International Conference (1997)
10. Rossainz M, Capel M., Domínguez P.: Pipeline as high level parallel composition for the implementation of a sorting algorithm. In Proceedings 27TH European Modeling & Simulation Symposium.Campora Bergeggi, Italy (2015)
11. Roosta, S.: Parallel Processing and Parallel Algorithms. Theory and Computation, Springer (1999)
12. Barry W., Allen M.: Parallel Programming. Techniques and Applications Using Networked Workstations and Parallel Computers, Prentice Hall, ISBN 0-13-671710-1 (1999)
13. Rossainz M., Capel M.: Approach class library of high level parallel compositions to implements communication patterns using structured parallel programming. In Proceedings 26TH European Modeling & Simulation Symposium.Campora Bordeaux, France (2014)
14. Pearson W.R., Lipman D.J.: Improved tools for biological sequence comparison. In Proceedings of the National Academy of Sciences of the United States of America 85, http://fasta.bioch.virginia.edu/fasta_www2/fasta_list2.shtml (1988)
15. Lee R.C.T., Tseng S.S., Chang R.C., Tsai Y.T.: Introducción al diseño y análisis de algoritmos, un enfoque estratégico. Mc Graw Hill (2007)
16. Levitin A.: The Design of Analysis of Algorithms. Wesley (2003)

17. Wood V., Gwilliam R., Rajandream M.A., et al.: The genome sequence of Schizosaccharomyces pombe. Nature (2003)
18. Masoudi-Nejad, A., Narimani, Z. and Hosseinkhan, N.: Next Generation Sequencing and Sequence Assembly. SpringerBriefs in Systems Biology (2013)
19. DanishAli, S. and Farooqui, Z.: Approximate Multiple Pattern String Matching using Bit Parallelism: A Review. International Journal of Computer Applications, Volume 74, No.19, 47–51 (2013)
20. Sanjuan A., Arnau V., Claver J.M.: Análisis Paralelo de Secuencias AND sobre computadores con multiples cores. Actas de las XIX Jornadas de Paralelismo, Castellón, España (2008)
21. Peña A.J., Claver J.M., Sanjuan A., Arnau V.: Análisis Paralelo de Secuencias AND mediante el uso de GPU y CUDA, ResearchGate. https://www.researchgate.net/publication/228857228 (2014)
22. Yang X.Y., Ripoll A., Marin I., Luque E.: Genomic-scale analysis of DNA Words of Arbitrary Length by Parallel Computation. Parallel Computing: Current & Future Issues of High-End Computing, NIC Series, Vol. 33, 623–630 (2008)

ISSN 1870-4069

Automatic Recognition of Abnormal Human Actions with Semi-supervised Training: A Literature Review

Diana-Karina Guevara-Flores, Josefina Guerrero-García, David-Eduardo Pinto-Avendaño

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Puebla, Mexico

diana.guevaraf@alumno.buap.mx

Abstract. In this paper is presented the literature review and the first tests performed for the development of a method consisting of differentiating typical actions in a given environment from those that can be categorized as abnormal or atypical.The objective of this method will be to determine which are the typical situations taking into account temporal and spatial information of a given environment to generate an alarm when a potentially undesirable situation arises. The main difference between this system and those existing in the literature is that it does not seek recognition of pre-established actions, such as running or sitting, and that the system can adapt to different environments. For the development of this research the use of Deep Learning is proposed and due to the complexity of the attributes required by the classifier, the use of a semi-supervised method is proposed.

Keywords. Deep learning, atypical action recognition, semi-supervised training.

1 Introduction

This research work aims to present the theoretical basis for the development and implementation of a model that allows the extraction of features in spatial and temporal dimensions for recognition of atypical situations in uncontrolled environments.

Video analysis with machine learning is a challenge task and there are few methods to capture the movement information of adjacent frames in a long term scheme.

The recognition of human actions in uncontrolled environments has multiple applications, among which are intelligent video surveillance for the analysis of behavior in shopping center customers, students in an educational facility or passengers in an airport. In public spaces it can be very useful for the timely detection of risk situations and in private companies it can allow the optimization of internal processes for the increase of profits.

Simple systems, such as monitoring a home, can be restricted to a single camera, however, more complex tasks require a multi-camera transmission environment.

Classical methods extract the characteristics of fixed images to train classifiers, however, in real applications, even the simplest ones, it is difficult to determine which characteristics are important for a given task.

The task becomes much more complex for video analysis. Most of the proposed methods treat each frame as a fixed image, losing the temporal information. Other methods perform short-term analyzes taking into account only the temporal information of a few frames, which makes the analysis very limited. Finally, the methods that take into account information to long-term are too costly computationally.

Additionally, the classification of atypical actions or situations can be complex even for humans. Actions that should be classified within the same class may seem dramatically different in terms of appearance and movement patterns. In the same way, it is possible that risky actions seem normal either because the individual who perform them pretends that this is the case or because of inherent faults in the system, such as the location of the cameras.

Another problem arises when analyzing highly concurred sites. To perform the analysis, the system must track each of them, for which it will be necessary to identify each individual in the scene. An analysis that includes the authentication of each individual for the analysis of their behavior over extended periods of time, would require limiting the system to a context of fixed users, for example, a company where workers are always the same or a university where students, teachers and administrative staff have relatively fixed routines.

This document is organized as follows: section 2 presents the state of the art; the characteristics of some of the methods used for video processing are presented, as well as the advantages and disadvantages of using each of them. Section 2.1 presents the methods used to evaluate the results obtained by some of the works analyzed in this document. Section 2.2 exposes some of the commercially available systems for video processing. Finally in section 3 the results of this research work are discussed.

2 Literature Review

For the development of the proposed method, an extensive literature review of the existing systems and methods is necessary, as well as the applications for which they have been required and the level of success achieved in their implementation. Also, a search of commercially available systems has been made, some of them dedicated to simple applications, such as monitoring of homes. It have also analyzed complex systems that involve a large number of modules and are capable of monitoring in real time large companies or public sites with a higher level of intelligence.

Models based on deep learning can be trained using supervised, semi-supervised or unsupervised approaches. The supervised approach is the most widely used due to the accuracy of the results obtained but it requires a large amount of tagged data for the training.

There are extensive databases well labeled for supervised training of still images, however, for video tasks there are few data bases and none of them is as widely labeled as images databases. For this reason, some researchers have opted for the realization of synthetic labeling video databases, made from image databases.

Some tasks in which good results are presented with the implementation of Deep Learning techniques are visual recognition of objects, recognition of human actions, natural language processing, audio classification, tracking, image restoration, noise elimination and segmentation, among others. Even some of the results presented in the literature point to a performance superior to that obtained by humans in the same task.

One of the most widely described methods in the literature for the recognition of human actions in video is through convolutional neural networks. Traditionally, this approach allows to treat the video frames as fixed images and perform the recognition of the actions at the level of individual frames, however, in this way the temporal information is not processed and it is not possible to perform movement analysis. This approach is useful for recognizing actions such as sitting or talking on the phone, but does not provide information about how long someone sat or how many different calls an individual made.

To take advantage of temporal information, a proposed approach in the literature is based on 3D convolution to obtain characteristics not only in the spatial dimension [1]. With this method multiple information channels are generated from adjacent video frames, with which it is possible to perform the convolution and sub-sampling in an individual way and then combine the information of all channels.

For the evaluation of the method, a database provided by TRECVID [2] was used, which consists of a 49-hour surveillance video of the London airport using five different cameras with a resolution of 720x576 at 25 fps. This work focused on the recognition of three kinds of actions: talking by cell phone, leaving an object and pointing. In addition, a large number of samples of actions that are not within these classes were generated.

The authors compare the method with a version of the 2D model and four variants of spatial pyramid matching. The results of the cross validation allow to evaluate the performance of each method using three measures: precision, recall and area under the ROC curve. The results of the performance of the proposed method outperform the rest in two of the evaluated actions: talking by cell phone and leaving an object, but are slightly exceeded in the third action: pointing.

In that investigation, because the videos were recorded in a real environment, each frame contains several people, so it was necessary to apply a human detector and tracking. Finally, the developed model was trained using a supervised algorithm, so a large number of labeled samples was required. The number of samples needed can be significantly reduced if a semi-supervised method is used.

Fig. 1. Results for human detection and tracking using TRECVID database [1].

Another method used in the literature for video analysis is deep belief networks, which are probabilistic generative models composed of multiple layers of latent stochastic variables [3]. These latent variables typically have binary values and are called hidden units or feature detectors. This type of networks has generated a lot of interest because they are based on an unsupervised method of learning hierarchical generative models, however the processing of high-dimensional images is limited, so modifications have been proposed for the treatment of images in its real size by means of hybrid networks between deep beliefs and convolutional networks.

A common problem in the treatment of video by means of Deep Learning is the difficulty in handling videos of variable length. This generates degradation in the efficiency of the classification, so that approaches have emerged that allow the representation of video content of arbitrary length, as for example, by means of 3D motion maps [4]. For the classification of actions, this approach is based on the implementation of a generative network to learn the 3D movement map and a discrimination network, based on the learned movements, to classify the actions. The map proposed by this model is a compact and discriminatory representation that eliminates a large amount of redundant information and is capable of capturing distinguishable trajectories around the human body. This approach has been used successfully in action videos, but the quality of the results decreases when the movement is subtle.

Among the free access tools related to this research topic, TensorFlow stands out, which is an open source library capable of offering 93% accuracy, based on neural networks [5]. TensorFlow has been used in projects such as *Deep Dream* [6], which is an image processing algorithm by means of which it is possible to create an alternative dream scene from any image or scenario. Another project of interest created from TensorFlow is *Show and Tell* [7], which is capable of automatically generating precise and highly descriptive captions for any image presented. To carry out the training of this system, millions of manually labeled images were necessary.

2.1 Metrics for Evaluation

The effectiveness of the methods proposed in the literature was evaluated with different metrics. In the case of 3D convolution, the comparison of the results obtained with the results of the analysis with fixed images in a 2D method, spatial pyramid matching (SPM) and spatiotemporal interest points (STIPs) was performed. Other standard assessment metrics based on temporal and spatiotemporal localization are:

– Probability of detection failures,
– False alarm rate,
– Accuracy (average),
– Speed.

2.2 Commercially Available Systems

Below are three commercial systems representative of the technology currently available in the market. The first, developed by the company Panasonic, is a highly sophisticated system capable of analyzing and optimizing the logistics of busy sites such as airports and companies. On the other hand, the so-called Netatmo and Butterfleye systems are low cost and oriented towards home security.

Face recognition Panasonic-FacePRO [8] is one of the most sophisticated commercial systems launched in July 2018. The system is aimed at detecting criminals and thieves in stores or companies and performs recognition and authentication by means of photographs in format jpeg. It has an alarm that notifies the guards of the presence of someone considered dangerous. Suspicious people can be tracked in the sales floor and with the information, it can be create a timeline of their route. The system is robust before:

– Aging,
– Makeup,
– Various facial expressions,
– Image quality.

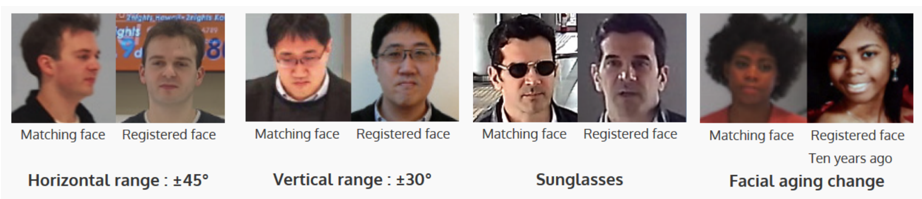


Fig. 2. Facial recognition function of Panasonic-FacePRO.

This system allows authentication even in difficult conditions for conventional technologies, such as faces at an angle greater than 45 degrees to the sides or 30 degrees up or down and partially covered with lenses. In terms of performance, the system achieved the highest level of facial recognition in the world in a comparison test (IJB-A Face Verification Challenge Performance Report/IJBA Face Identification Challenge Performance Report) of NIST. In addition, it is expected to improve performance to achieve face recognition partially covered with face masks.

This technology focuses on obtaining the best possible photograph to make the comparison with their database. The system takes several photographs and sends the best for facial recognition, reducing server and network costs. In a system with 10 or more cameras, costs can be reduced from 40% to 50% (with 10 or more cameras). Conventional systems require a large bandwidth and must be large-scale due to the large amount of storage required for the storage of all captured images. Additionally, the system allows up to 20 cameras in the network connected to a single server. The system dramatically decreases network traffic, as well as the transmission and construction costs of the network.

The algorithm combines Deep learning, machine learning and a similarity calculation method with error suppression. Deep learning technology was developed with the National University of Singapore. The camera adjusts automatically to focus on moving objects, at high speeds and at different light intensities (day and night). For its part, the facial recognition software collects the data of each face it detects and stores it in an easily accessible database. The system allows registering from 10,000 to 30,000 faces. Users can select a face and program an alarm for subsequent detections of this individual or track their movements in chronological order through all the cameras in the system. Additionally it is possible to obtain the count of people and graphic statistics by gender and age.

This system has been successfully tested at Tokyo international airport, in the area of immigration control. The system has been able to compare the photography embedded in the passport chips with a photo taken at the door of facial recognition to verify identity without the need to register travelers or take biometric data.

Netatmo Welcome [9] is an indoor camera that sends alerts to the cell phone when it detects unknown faces. Send photograph of the intruder and record video. For familiar faces it does not record video to protect privacy. The alarm does not activate with pets unless this functionality is activated to monitor them. The video can be stored on a micro SD card, in the personal Dropbox cloud or on the personal FTP server. It has a viewing angle of 130 degrees and full HD resolution. It cost is $200.00 usd. It also has an outdoor version that allows the detection and notification of the presence of people, animals and cars. Its cost is $300 usd.

Butterfleye [10] is a wireless camera with face recognition with 1080p resolution. It has thermal, motion and sound sensors. Some features are available only with monthly plan rent. It is able to recognize family, friends, strangers and pets. It works even if the light or the internet goes out. Stores video from 5 seconds before an event occurs. The system is armed or disarmed according to GPS location. The viewing angle is 120 degrees and the battery life is 2 weeks. Its cost is $200 usd.

3 Discussion

The state of the art was reviewed and an overview of the future work that motivated this investigation was presented. Different methodologies were analyzed for video processing with Deep learning, based on different types of networks and the advantages and disadvantages of each were presented. Additionally, the different evaluation metrics and strategies for training that have been used in similar work are exposed.

In terms of commercial systems, three of the technologies with characteristics more representative of those available in the market were analyzed. Versatility, functionality and cost vary widely depending on whether the application is for home security or for companies and public sites.

References

1. Ji, S., Xu, W., Yang, M., Yu, K.: 3D convolutional neural networks for human action recognition. IEEE transactions on pattern analysis and machine intelligence, 35(1), 221–231 (2013)
2. TREC Video Retrieval Evaluation, https://trecvid.nist.gov/ (2018)
3. Lee, H., Grosse, R., Ranganath, R., Ng, A. Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th annual international conference on machine learning, 609–616 (2009)
4. Sun, Y., Wu, X., Yu, W., Yu, F.: Action recognition with motion map 3D network. Neurocomputing, 297, 33–39 (2018)
5. Home TensorFlow, https://www.tensorflow.org/?hl=es (2018)
6. Deep Dream Generator, https://deepdreamgenerator.com/ (2018)
7. Deep Dream Generator, https://github.com/tensorflow/models/tree/master/... ...research/im2txt (2018)
8. Panasonic Security, https://security.panasonic.com/Face_Recognition (2018)
9. Netatmo oficial, https://www.netatmo.com/es-ES/site/ (2018)
10. Ooma Butterfleye, https://getbutterfleye.com/ (2018)

ISSN 1870-4069

Development of a Multi-user System to Identify the Level of Attention in People

Alfredo Garcia[1], Juan Manuel Gonzalez[1], Amparo Palomino[2]

[1] Benemérita Universidad Autonoma de Puebla,
Facultad de Ciencias de la Computación, Puebla, Mexico

[2] Benemérita Universidad Autonoma de Puebla,
Facultad de Ciencias de la Electrónica, Puebla, Mexico

alfredo_amigo18@hotmail.com,jumagoca78@gmail.com,
ampalomino@gmail.com

Abstract. The level of attention in people is associated with efficiency in their intellectual activities, in their level of understanding and in the development of their creative ability. The physiological variables with greater relevance in this application are the brain waves due to the cognitive relationship that exists between thinking and brain activity, therefore its implication is direct in this context. There are different techniques and systems described in the literature that are used to measure enumerable physiological variables related to the attention states of people. Some current limitations of these systems is that they only use a physiological variable and these systems are of the single-user type. In this paper an analysis of the state of the art is made with respect to the type of variables, techniques, instruments and types of systems used to measure the level of attention in people.

Keywords. Level attention, physiological variables, bio-feedback, brain waves, corporal posture.

1 Introduction

There are several disorders that affect the level of attention of people both in their childhood and adult-hood.

One of the most recognized disorders is the atten-tion deficit / hyperactivity disorder (ADHD) and is usually diagnosed for first time in the childhood; the symptoms persist in the adolescence and in the adulthood.

The ADHD is characterized by lack of attention, impulsivity and hyperactivity. Recently the ADHD has been estimated that affects 3.5% of school-age children around the world and is one of the most common psychiatric disorders among young people. Children with these problems are often unpopular and lack reciprocal friendships, but are not always aware of their own unpopularity. Although these symptoms tend to decrease with age, at least 50% of children with ADHD still have symptoms that de-crease in the adulthood.

Despite the vast literature that supporting the effi-cacy of stimulant medication in the treatment of the attention deficit / hyperactivity disorder (ADHD), several limitations of pharmacological treatments underscore the clear need for effective alternative psychosocial treatments.

To know the degree of affectation that ADHD produces in people, is necessary to have tools that can provide a feedback of the level of attention when the user is executing a specific task.

Currently there is a variety of commercial devices that quantitatively provide the level of concentration, meditation, relaxation and user attention, but generally are type mono-user, using a single physiological variable to obtain the final diagnosis. The performance of these devices is limited because they have restrictions by the manufacturer regarding the software and the hardware implemented.

Recent works describe various techniques to measure the level of attention in people (Digital im-age processing [1], Modification of the frequency range of binaural sounds [2], Wavelet Transform [3], Use of commercial tiaras [4], Classification of brain waves in emotions [5], Classification of text [10], among others, using different physiological variables such as: Facial gesturing [1, 7, 8], Brain waves [2, 3, 4, 5, 6, 9, 36, 37], Body gesturing [8], body posture [8], cardiac pulses [8], respiratory rate [8, 24], way of write a text [10], sight trajectory [25], movement of the head [32], spit [8, 31], sweat [8, 31], temperature [8, 31, 40], glucose in blood [31, 39] and gesticula-tion in the hands [41].

However, there is still a lack of a system that in-volves more than one physiological variable, which generates more hardiness in the system and consequently provides a diagnosis with greater reliability of the level of attention in the user.

In the literature has been verified that the vast majority of reported works are type mono-user, which generates a restriction only to be applied in an isolated context and not in a real social environment such as a classroom.

In this research work is proposed a model to develop a system to measure the level of attention in people, using at least 2 physiological variables, implementing a multi-user system and generating a feedback in the form of a closed loop to carry out some action that helps to restore the user's level of attention.

2 State of the Art Analysis

The variety of techniques applied in the field of research to quantify the level of attention of people leads toward a descriptive analysis that is presented in this work.

2.1 Acquisition Techniques of Physiological Variables Related to the Attention Status of People

Starting from the review in the literature of the variables used to relate the level of attention in people, has been found that the brain waves are the physiological variables with greater relevancy due to the cognitive relationship that exists between think-ing and brain activity. Therefore its implication is direct.

The graph of the Figure 1 shows the relevance that each one of the physiological variables has on the level of attention of the people. This analysis is ob-tained from the state of the art of the related works.

Depending on the physiological variable used, the device is chosen to perform the data acquisition. Within the most devices used in the literature are: the electroencephalogram, WEB cam, motion sensors, gyroscopes, electrodes, mouse, electrocardiogram, electrochemical sensors, keyboard, transducers, cam-eras and optical sensors.

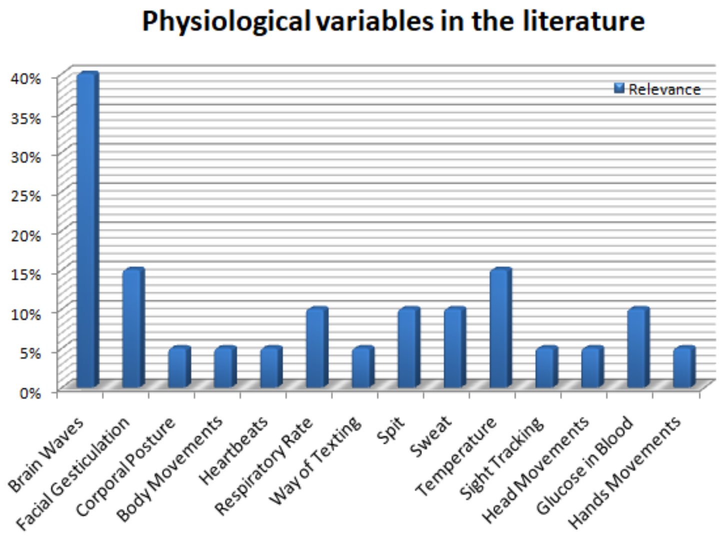


Fig. 1. Relevance of the physiological variables related to the level of attention in the literature.

The Figure 2 shows the relationship between the devices used to obtain the data and the physiological variable sensed.

The techniques applied to obtain the data of the physiological variables also play a fundamental role since the effectiveness of the final diagnosis depends on them. The tables 1 and 2 describe a comparison of various data acquisition techniques and their components as a physiological variable, device and software employees, sampling time, effectiveness and relationship with levels of attention.

2.2 Feedback Techniques Applied on the Attention Status of People

Once the level of attention of the user is known, is necessary to exert some action that gives feedback to the user to stimulate his concentration in the activity or task that he is carrying out.

In the literature has been found that studies related to the classroom showed that environmental factors, such as cognitive assistance technology (CAT), can help people with cognitive disabilities [27].

An example is the battery of the attention training system. This electronically generated response cost system is placed on a student's desk and handled with a remote control that is given to the teacher. It is designed to send comments in order to increase the levels of attention related to tasks. This system was more effective compared to a pre-existing classroom management program that used chip reinforcement [27]. Another example is the watchminder, a vibrat-ing wristwatch. This self-monitoring device aims to increase the task behavior of elementary school children. The results of this study proved effective for two of each three participants.

The graph of the Figure 3 shows the most used feedback systems to stimulate the attention status of people in general. In this comparative analysis is ob-served that visual learning environments are the most common tool in the area of cognitive sciences.

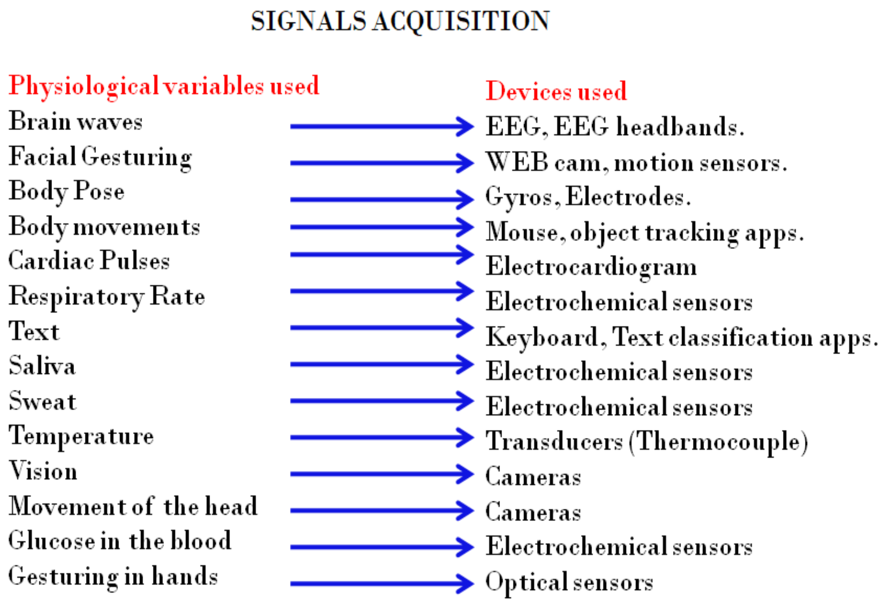


Fig. 2. Devices used to obtain the data of the physiological variable sensed.

3 Construction of the Model for Identifying the Level of Attention in People

In the literature has been found that biofeedback training systems induce a specific mental or physical state in a user through a closed cycle of bio-feedback. These systems gather the physiological state of a person through the detection of hardware, integrate this state into a computer-based interactive system and present the comments so that the user can work to adjust their status [19, 20, 26].

Table 1. Comparison of physiological variables, techniques and relation with the levels of attention.

Reference Number	Physiological variable used	Technique applied	Relationship and impact with the levels of attention
1	Facial Gesticulation	Digital image processing	The automated recognition of emotions can be directly correlated with the levels of attention of a teenager.
2	Binaural waves (auditory waves)	Modification of the frequency range of the incident waves	Binaural waves cause a positive impact on mental states such as active concentration and creative visualization.
3	Brain waves (Alfa, Beta, Delta, Theta y Gamma)	Wavelet Transform	Use of the Mindwave headband for reading brain signals, which are classified in levels of attention
4	Brain waves produced by facial gestures	Suites of EmotivEpoc: *Affective , *Expressiv, *Cognitiv	Use of the EmotivEpoc headband for the reading of brain signals, which are classified in levels of attention.
5	Brain waves (Alfa, Beta, Delta, Theta y Gamma)	Classification of brain waves in emotions through their frequency variations	The automated recognition of emotions can be directly correlated with the levels of attention of a teenager
6	Brain waves (Alfa, Beta, Delta, Theta y Gamma)	Bayesian classification and Hill Climbing search algorithm	The automated recognition of emotions can be directly correlated with the levels of attention of a teenager.
7	Brain waves (Alfa, Beta, Delta, Theta y Gamma)	Digital image processing Affective computing	Automatic feedback can improve levels of adolescent care.
8	Facial gesturing Body movements	Digital image processing Mouse movement	Application of tasks that require cognitive processes such as attention, memory and reasoning.
9	Brain waves Heart waves	Characterization of signal changes Classification of brain and heart waves in emotions through their frequency variations	The automated recognition of emotions can be directly correlated with the levels of attention of a teenager.
10	Text	E-learning (Identification of emotions through the way of writing a text)	The automated recognition of emotions can be directly correlated with the levels of attention of a teenager.

Table 2. Comparison of devices, Software, Sampling time and effectiveness in relation with the levels of attention.

Reference Number	Device used	Software	Sampling time	Effectiveness
1	Web Cam Logitech C170 USB 2.0	Open CV Cmake	1 frame / 64.93 milliseconds	Average detection rate: 84%
2	Headset headband	Visual Studio C++	Not specified	Efficiency of binaural sounds: 80%
3	Bluetooth RFCOMM6 Headband Mindwave	Matlab	Sampling frequency: 512Hz	Not specified
4	Headband Emotive Epoc	Labview 2010	Not specified	60% writing efficiency using the BCI (Brain Computer Interface)
5	Arduino Uno Casco de electrodos	JAVA C#	Not specified	Not specified
6	Headband Emotive Epoc	WEKA Wizard	2048 Hz 56 data / image	Stays correctly classified: 53.7879%
7	WEB CAM of PC used for experiment (variable)	JAVA WEB PHP	Not specified	Accuracy in the recommendation system: 5,757%
8	WEB CAM of PC used for experiment (variable)	Face Tracking SDK Kinect for Windows.	Not specified	Not specified
9	Electroencephalogram Electrocardiogram Prototyping data acquisition card	Prototyping software	Not specified	Not specified
10	PC used for measurement (variable)	Word NET Word NET Affect	Not specified	Not specified

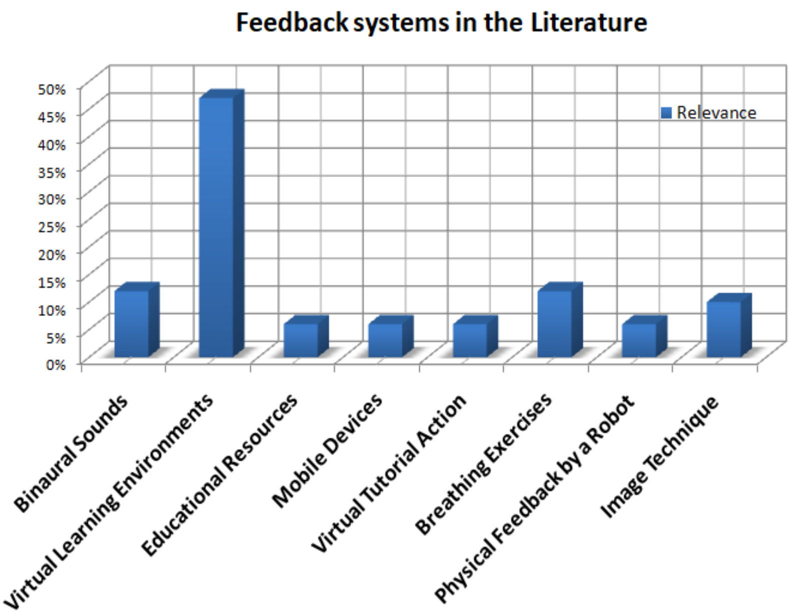


Fig. 3. Relevance of the feedback systems related to the level of attention in the literature.

3.1 Type of Model

For this reason the closed-loop control technique (shown in Figure 4) has been proposed in this work, where the input variables will be the signals of the physiological variables, the output signal will be the feedback that is sent to the user in an automated way, The system floor will be the user whose function is to self-regulate their attention and concentration depending on the feedback received by the system.

3.2 Selected Physiological Variable

According to the relevance of the physiological variables mentioned in the literature, brain waves have been chosen as the main parameter to measure the level of attention of a person.

Body posture has been proposed as a second phy-siological variable, in [8] is placed on its relationship with the state of attention as psycho-educational support in virtual learning environments, this case presenting a development challenge that has not been solved in this context.

3.3 Validation Instruments

These measurement techniques should be available in the language of the participants and should measure the main aspects of cognitive rehabilitation. For example, to evaluate the improvements of attention in longitudinal studies, the following scales and questionnaires could be used: Integrated Visual and Auditory Continuous Performance Test (VAT), Test of Variable Attention (TOVA), among others [20].

3.4 Data Acquisition

According to the physiological variables selected and supported by the reviewed literature, the information collection instruments will be adapted commercial sensors that obtain the measurement of the selected physiological variables, the reading of these signals will be obtained by the system implemented and sent wirelessly to a computer where the processing, classification and feedback of the automated system will be carried out.

4 Limitations and Experimental Advances

This section describes the research work and experi-mental work carried out so far. The work planned in the future is also mentioned.

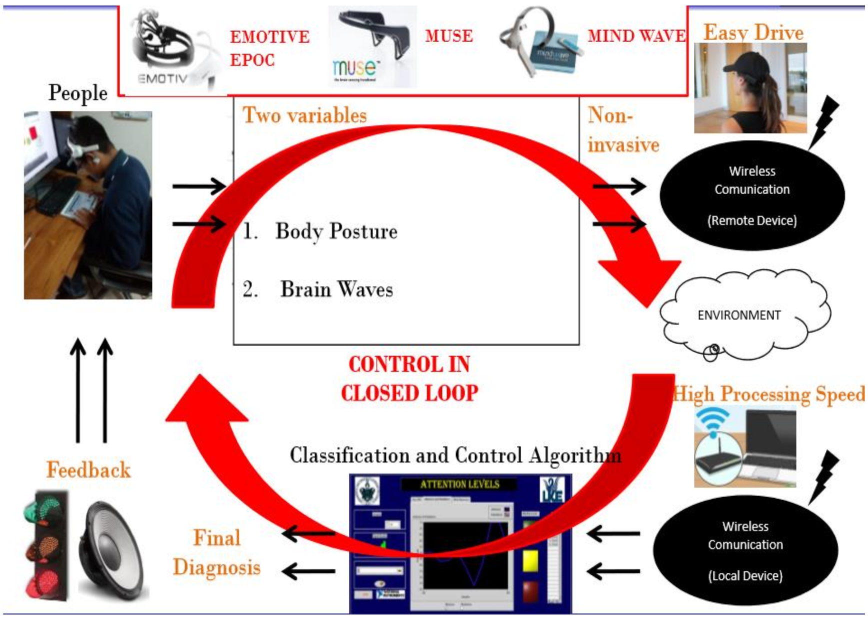


Fig. 4. Model of the closed-loop control technique applied to the attention system.

4.1 Limitations

In the implementation of the works [2, 5, 7, 8, 11, 12, 14, 15, 19, 21, 23, 24, 25, 27, 28, 29, 30] re-viewed so far, systems have been developed of mono-user type, which represents a development challenge towards multi-user systems in this context.

Another limitation that has been found in the literature is the use of a single physiological variable in most of the works [2, 5, 7, 8, 11, 12, 14, 15, 19, 21, 23, 24] presented, which also represents a challenge on the robustness of the systems in this context.

4.2 Experimental advances

In the first instance, experimental tests were car-ried out using the MindWave commercial device of the Neurosky company, to detect the level of atten-tion in first semester students of the computer degree of the BUAP. A sample of 22 students whose ages are between 17 and 22 years old was used.

To obtain the data of the brain signals, a mono-user system was implemented, using the LABVIEW software. Figure 5 illustrates the graphical interface where you can observe the behavior of the brain sig-nals, a traffic light as feedback, a vector where are saved the sampled data and the variation of the user's level of attention.

The test was carried out in two modalities: with tablet and desktop computer. The practical development is shown in figure 6.

With the data obtained an analysis is made and is shown in the graph of Figure 7, which reflects higher concentration reached by users when handling a tab-let than by manipulating the mouse of a desktop computer, which indicates that the device mouse influences as a distractor agent when executing a specific task.

A second experiment was carried out in the labora-tory of the Computation faculty of the BUAP, where 12 undergraduate students were integrated into 3 teams. The activity to be carried out consisted of developing ideas for a science project in an organized way, where each team member participated actively in its elaboration.

While the activity was being carried out, 5 students were randomly monitored, where their attention level was obtained for 10 minutes, the feedback was ob-served in another independent room through a physi-cal traffic light and a laptop where they received the data sensed with the MindWave diadem, which de-scribed the student's level of attention in each sam-pling.

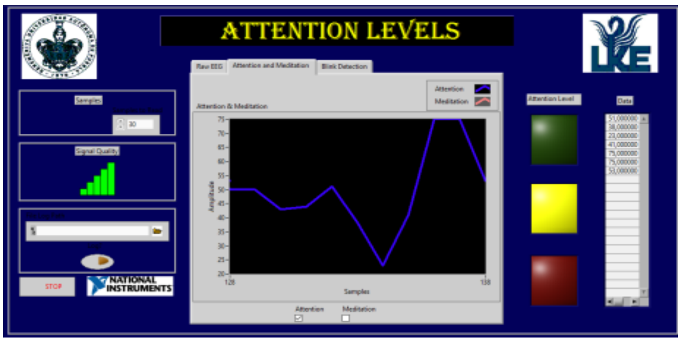


Fig. 5. Graphical interface implemented in LABVIEW.

Fig. 6. Graphical interface implemented in LABVIEW.

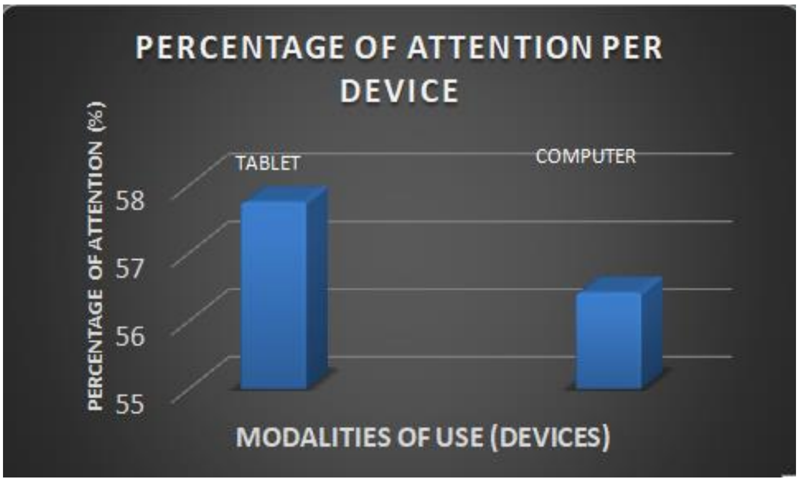


Fig. 7. Percentage of attention per device.

Table 3. Level of attention in the monitored students.

Student	Attention Level (%)
Ruben	57.39
Fernando	45.36
Marcos	60.98
Melani	52.77
Noé	48.35
Global Average	**52.97**

The samples were taken every 500 milliseconds with an instantaneous feedback to the person who monitored the activity in an isolated room. The average of the samples of each monitored student is shown in the Table 3 and in the Figure 8.

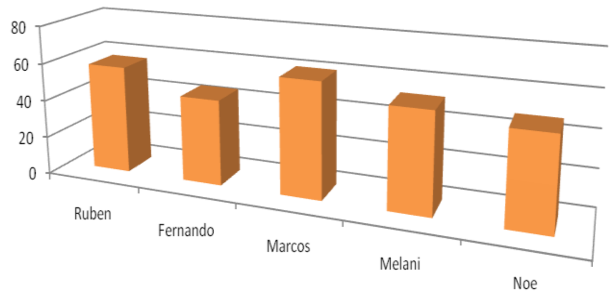


Fig. 8. Graph of the level of attention in the monitored students.

As a continuation of the work, the use of other com-mercial devices was validated in order to compare their performance.

For this motive 3 attention monitors were imple-mented using the three commercial headbands with the highest number of mentions in the literature: Muse, Emotiv Epoc and Neurosky. At first, the soft-ware of each manufacturer is explored and subse-quently a system of transmission and reception of information is implemented to obtain the data read by the sensors, sending them to an own interface to be processed freely.

The software developed by the manufacturer Inte-raxon, provides direct access to the data read by the Muse brain sensor, this feature allows to perform a better interaction with the device and is used to de-velop a mono-user data acquisition interface, shown in figure 9, which has a panel to visualize the level of attention of the user, 1 virtual traffic lights and 1 table where the numerical data of the user's attention level is displayed in an approximate time to the real time.

Fig. 9. Muse attention monitor implemented in Labview.

The software developed by the manufacturer Emotiv Systems, provides direct access to the data read by the brain sensor Emotive Epoc, this feature allows a better interaction with the device and is used to de-velop a single-user data acquisition interface, which It is shown in Figure 10, which has 1 panel to display the user's attention level, 1 virtual traffic lights and 1 table where the numerical data of the user's attention level is displayed in an approximate time to the real time.

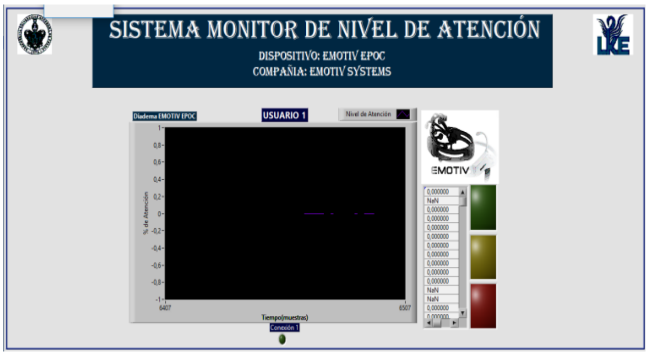


Fig. 10. Emotiv EPOC attention monitor implemented in Labview.

The software developed by the manufacturer Neu-roSky, provides direct access to the data read by the brain sensor MindWave, this feature allows to per-form a better interaction with the device and is used to develop a multi-user data acquisition interface, which is shown in figure 11, which has 2 panels to visualize the level of attention of each user, 2 virtual traffic lights and 2 tables where the numerical data of the level of attention of each user is displayed in a time approximated to the real time.

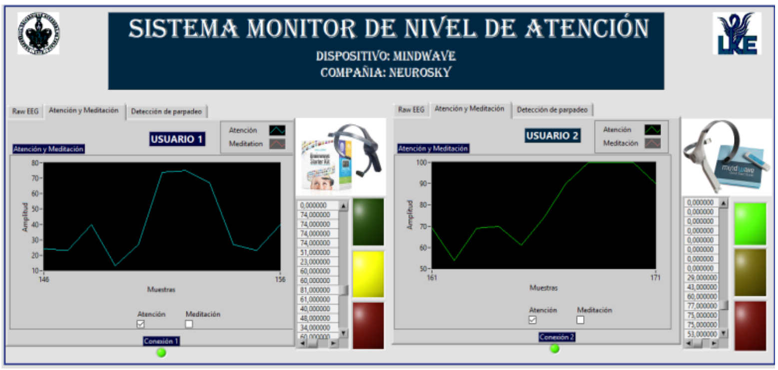


Fig. 11. MindWave attention monitor implemented in Labview.

5 Conclusions

The implementation of single-user systems and the use of a single physiological variable are 2 of the main limitations that have been found in the literature in the context of the systems used to measure the attention of people.

Based on these limitations, the first implementa-tion of a single-user type system was performed, us-ing brain waves as the only physiological variable.

In the second experiment is shown that the level of attention in students when working as a team is on average 53%, only 3% above the expected half, this is due that diverse factors influence a collaborative context, where there is noise, diversity of opinions and other distractors that divert the attention more easily than in a controlled environment where con-centration is stimulated individually.

As a continuation of the work, a data acquisition system is made for 3 different headbands, where a multi-user system with the Mindwave device of the Neurosky company was efficiently implemented.

As a future direction of this research work, has been proposed to correlate a second physiological variable with the attention states and increase the number of users in the system.

References

1. Marín, E.J.: Detección de emociones del usuario. Tesis Pontificia Universidad Católica de Valparaíso, volumen 1, Chile, 1- 67 (2014)
2. Domínguez, C.: Las Ondas Binaurales y sus Efectos. Tesis de Investigación Experimental volumen 1, Ciudad Cooperativa Cruz Azul, 1- 22 (2015)
3. Torres, F., Sánchez, C., Palacio, B.: Adquisición y análisis de señales cerebrales utilizando el dispositivo MindWave. MASKANA, I+D+ingeniería 2014, volumen 1, 1–11 (2014)
4. Rojas, S., Garzón, J., Martínez, D., Escobar, M., Robayo, C.: Lector de ondas cerebrales para implementar un sistema alternativo y aumentativo de comunicación. In Proceedings 10th Latin American and Caribbean Conference for Engineering and Technology, volumen 10, 1–9 (2012)
5. Campazzo, E., Martínez, M., Guzmán, A., Agüero, A.: Desarrollo de interface de detección de emociones para su utilización en redes sociales y entornos virtuales de aprendizaje. In Proceedings XV Workshop de Investigadores en Ciencias de la Computación, volumen 1,Paraná , 1–5 (2013)
6. García, A.E.: Análisis de ondas cerebrales para determinar emociones a partir de estímulos visuales.Universidad Veracruzana Facultad de Estadística e Informática, volumen 1, Xalapa, Veracruz, México, 1–137 (2015)
7. Hernández, A., Vásquez, R., Olivares, B.A., Cortes, G., López, I.: Sistema de detección de emociones para la recomendación de recursos educativos. Programación Matemática y Software, 8(1), 58–66 (2016)
8. Saneiro, M.M.: Apoyo psico-educativo y afectivo en entornos virtuales de aprendizaje. Badajoz, International Journal of Developmental and Educational Psychology, volumen 1, numero 2, España, De INFAD Base de datos, 233–241 (2015)
9. Campazzo, E., Martinez, M., Guzmán, A.E., Agüero, A.: Entornos Virtuales de Aprendizaje integrado a tecnología móvil y detección de emociones. Secretaría de Ciencia y Tecnología/Departamento de Ciencias Exactas Físicas y Naturales/Universidad Nacional de La Rioja, volumen 1, La Rioja, 1–5 (2015)
10. Aballay, L., Aciar, S., Reategui, E.: Propuesta de un Método para Detección de Emociones en E-Learning. In Proceedings ASAI 2015, 16º Simposio Argentino de Inteligencia Artificial, Porto Alegre, Brasil, 121–128 (2015)
11. Garcia, A., Muñoz, M., Suarez, C.: Gestión de emociones en espacios virtuales de formación. Investigación Educativa, 12,(21), Salamanca, España, 45–65 (2008)

12. Aymerich-franch, L.: Los juegos en entornos virtuales como herramientas de aprendizaje: estudio de la respuesta emocional de los participantes. SPHERA PÚBLICA Revista de Ciencias Sociales y de la Comunicación, 12(12), Murcia, España, 183–197 (2012)
13. Hernández, A.M.: Aprendizaje Electrónico Afectivo: un modelo Innovador para Desarrollar una Acción Tutorial Virtual de Naturaleza Inclusiva. Formación Universitaria Volumen 8(2), Granada, España, 19–26 (2015)
14. Desney, S.T., Nijholt, A.: Brain-computer interfaces. Springer, London Dordrecht Heidelberg, volumen 1, New York, USA, 1–275 (2010)
15. Regan, L.M., Shane, D., Kalyn, R.M., Bertram, P.C., Gaetz, M., Doucette, A., Taylor, B.A., Orr, A.P., Keiver, K.: Games as neurofeedback training for children with FASD. In Proceedings IDC '13 Proceedings of the 12th International Conference on Interaction Design and Children, volumen 1, New York, USA, 165–172 (2013)
16. Pascual, M.F., Begoña, Z., Buldian, K.M.: Adaptive cognitive rehabilitation interventions based on serious games for children with ADHD using biofeedback techniques: assessment and evaluation. In Proceedings COMPUTE '10 Proceedings of theThirdAnnual ACM Bangalore Conference, article 29, Bilbao, España, 1–4 (2010)
17. Beaton, R., Merkel, R., Prathipati, J., Weckstein, A., McCricard, S.: Tracking mental engagement: a tool for young people with ADD and ADHD, In Proceedings ASSETS '14 Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility, volumen 1, Virginia, USA, 279–280 (2014)
18. Sonne, T., Jensen, M.M.: Evaluating the ChillFish Biofeedback Game with Children with ADHD. In Proceedings IDC '16 Proceedings of The 15th International Conference on Interaction Design and Children, volumen 1, New York, USA, 529–534 (2016)
19. Sonne, T., Jensen, M.M.: ChillFish: A Respiration Game for Children with ADHD. In Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, volumen 1, New York, USA, 271–278 (2016)
20. Asiry, O., Shen, H., Calder, P.: Extending Attention Span of ADHD Children through an Eye Tracker Directed Adaptive User Interface. In Proceedings ASWEC '15, Vol. II Proceedings of the ASWEC 2015 24th Australasian Software Engineering Conference, volumen 1, Australia, 149–152 (2015)
21. Divia, V.: FOQUS: A Smartwatch Application for Individuals with ADHD and Mental Health Challenges. In Proceedings ASSETS '16 Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, volumen 1, Nevada, USA, 311–312 (2016)
22. Zuckerman, O., Hoffman, G., Rubin, D.K., Klomek, A.B., Shitrit, N., Amsalem, Y., Shlomi, Y.: KIP3: Robotic Companion as an External Cue to Students with ADHD. In Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, volumen 1, New York, USA, 621–626 (2016)
23. Kang, H.W., Zentall, S.S., Burton, T.L.: Use of images in instructional technology for children with attentional difficulties. In Proceedings IDC '07 Proceedings of the 6th international conference on Interaction design and children, volumen 1, New York, USA, 129–132 (2007)
24. McLaren, E.S., Antle, A.N.: Exploring and Evaluating Sound for Helping Children Self-Regulate with a Brain-Computer Application. In Proceedings IDC '17 Proceedings of the 2017 Conference on Interaction Design and Children, volumen 1, California USA, 393–398 (2017)
25. Weisberg, O., Galoz, A., Berkowitz, R., Weiss, N., Peretz, O., Azoulai, S., Rubin D.K., Zuckerman, O.: TangiPlan: designing an assistive technology to enhance executive functioning among children with ADHD. In Proceedings En IDC '14 Proceedings of the 2014 conference on Interaction design and children, volumen 1, New York, USA, 293–296 (2014)

26. Bandodkar, A.J. Wang, J.: Non-invasive wearable electrochemical sensors: a review. Trends in Biotechnology, volumen 32, número 7, San Diego, USA, 363–371 (2014)
27. Maimó, M.F., Mackenzie, I.S., Yee, C.M. Varona, J.: Evaluating Fitts' Law Performance With a Non-ISO Task. In Proceedings of Interacción '17, volumen 1, Cancún, Mexico, 1–8 (2017)
28. Perakakis, M. Potamianos, A.: An Affective Evaluation Tool Using Brain Signals. IUI'13 Companion, volumen 1, USA, 105–106 (2013)
29. Pinto, R.D. Ferreira, H.A.: Development of a Non-invasive Brain Computer Interface for Neurorehabilitation. REHAB '15, volumen 1, Lisbon, Portugal, 1–5 (2015)
30. Eadie, M., Steele, R.J.: Non-invasive Blood Glucose Monitoring and Data Analytics, ICCDA '17, volumen 1, Florida, USA, 138–142 (2017)
31. Blank, M. Sinclair, M.: Demo: Non-invasive and long-term core temperature measurement. SenSys'11, volumen 1, Seattle, USA, 1–2 (2011)
32. Avila, S.O., Rakona, B., Mistry, P.: Non-invasive optical detection of hand gestures. AH '15, volumen 1, Singapur, Singapur, 179–180 (2015)

ISSN 1870-4069

Automatic Identification of Learning-Centered Emotions: Preliminary Study for Data Collection

Yesenia N. González-Meneses, Josefina Guerrero-García

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación, Puebla, Mexico

{yeseniaglez0,joseguga01}@gmail.com

Abstract. The intention of this work is to achieve an automatic identification of emotions in educational environments using machine learning algorithms and physiological and behavioral signal acquisition technologies to identify relations between emotions and learning. Four of the main learning-centered emotions are considered [1]: engagement, boredom, confusion and frustration. It is proposed to make a fusion of data from four signal acquisition technologies with the objective of achieving the identification of emotions in the most precise manner. The development of an appropriate database for the study of emotions is a fundamental task. Therefore, considering the stages of the proposed methodology, the first of them is presented and the design of the experiment that will be executed for data collection with college students during a learning process.

Keywords. Automatic identification of emotions, learning-centered emotions, machine learning.

1 Introduction

From a computational approach, the detection and identification of emotions is a relevant problem within the area of affective computation research, since it is the starting point in the study and development of human-machine interaction systems sensitive to the emotions of human beings. This problem in the human-machine interaction area is known as classification of emotions, automatic detection or identification of emotions, as it is mentioned in [2] and [3].

The complexity of the problems that affective computation deals with lies in the fact it is an interdisciplinary area that encompasses computer science, psychology, and cognitive science [4].

So, to develop models of emotion recognition that generate satisfactory results with an acceptable degree of precision, it will be necessary to study these three areas in depth.

Particularly, in the computational area the challenge is selecting and testing machine learning algorithms that could be integrated in a complete model of recognition of learning-centered emotions. In this sense, two approaches have been used to recognize human emotions [5]: The objective, using sensors or capturing images; as in the recognition of facial expressions, voice, heart rate, body language, body's thermal activity, muscular activity and brain waves. And the subjective, such as contextual analysis through direct observation, surveys, or interrogations of the individuals themselves. The automatic recognition of emotions and the human-machine interaction can be considered, then, from the use of physiological signal sensors that allow the acquisition of data through voice analysis, video images of the face, eyes, head, or body movements of the people [5]. Within the human-machine interfaces that capture physiological signals that can help in the recognition of emotions, as mentioned in [6], the brain wave diadems that send information to an electronic device in the form of an electroencephalogram (EEG) are found.

There are also cardiovascular wristbands, that measure heart rate and provide information in the form of an electrocardiogram (ECG); electrodermal activity sensors, that measure the level of conductivity of the skin through sweat on the hands; thermal cameras that allow the measurement of the temperature change of the human body associated with the different emotional states. There are also devices that measure muscle electrical activity in response to a nerve stimulation of a muscle, in the form of an electromyography (EMG). Regarding devices related to the identification of people's behavior -such as body postures and gestures- there are traditional video cameras, webcams or augment-ed reality that allow the recording of facial expressions and body movements, as well as eye tracking, important to the recognition of emotions. In this type of devices, there are also voice recorders, another medium used to identify emotions, as mentioned in [7].

The data obtained from different devices must be processed and classified based on a specific objective. In this case, they will be used for the recognition of learning-centered emotions and captured in real educational environments when students are executing a learning activity. These activities can be done through an electronic device and using an intelligent tutorial system or MOOC (Massive Open Online Courses), through an educational videogame, using a computer in a traditional way, whether watching a video, studying, investigating or reading.

Therefore, it is expected to contribute to physiological signal processing technics, from their capture, preprocessing, selection of characteristics and, finally, with the identification of emotions centered around learning. For this, we will use machine learning algorithms that will be implemented to process the different data coming from at least four signal acquisition technologies and captured in educational environments, to improve recognition accuracy and obtain a model of emotion recognition with an acceptable level of trust closer to reality within and educational environment.

The objective of identifying emotions during educational activities is to corroborate the relationship between learning-centered emotions and the level of learning obtained by students. This relationship is the base for approaching educational strategies that help to improve the levels of learning and, therefore, the educational level in our country.

2 Problem Description

The problem involving automatic recognition of emotions has been an area of investigation highly active in the last years. Regardless of this, a clear solution, that is within reach of the most people, is still far away. Several drawbacks have influenced the construction of an appropriate solution from a computational point of view. On one hand, a factor that affects the performance of emotional recognizers in real contexts is the difficulty to generate databases with spontaneous emotions. Generally, works are made with actuated databases which provide portraits of emotions representing prototypical and intense emotions that facilitate the search of correlations and the subsequent automatic classification. This kind of databases are usually captured in a controlled environment, which decreases problems in the processing of information (noise, for example). In addition, it can guarantee a balanced number of samples per class. As consequence, there have not been good results when translating the knowledge extracted from these databases to real contexts [8]. In contrast, the data bases with spontaneous records show information with emotional content that does not belong to a single class, but a mixture of them. In other cases, there are samples with a very light emotional charge, close to a neutral emotional state. In addition, databases with spontaneous emotions are usually recorded in noisy environments, such as classrooms, study rooms, entertainment areas, offices, factories or in phone conversations, which leads to the inclusion of noise. Finally, because of the very nature of the problem, it is not possible to ensure a balanced quantity of examples per class.

Another challenge to be solved is the extraction and selection of a set of characteristics that allow recognizing emotions in the data captured spontaneously. Although progress in the area has been important, there is still much to be done in realistic contexts. Therefore, it is necessary to propose and explore other approaches that allow to reach a good performance of the recognition of emotions in real world applications. An evident aspect to consider is the fact that the area of application has an important influence on the accuracy of emotion recognition [7], since this may vary according to the context the age of the participating individuals, the hours of the day in which they work, the comfort with which activities are done, the degree of intrusion of the tools used, the specific type of activity that is being carried out, among others.

Considering the presented problems and the analysis of recent related works, we can identify that educational environments there is a lack of adequate methodologies to recognize the emotional state of students during learning processes through interaction with a computer. Preferably, that integrates diversified characteristics, obtained even from the fusion of data coming from the use of different technologies of acquisition of physiological and behavioral signals based on a model of emotions that allows to become closer to reality and, overall, to the cognitive learning process, which contributes to the analysis of the emotion-learning relation.

2.1 Research Objectives

The objective of this work is to recognize emotions in educational environments using machine learning algorithms and technologies of acquisition of physiological and behavioral signals to identify emotion-learning relationships. To achieve the above, the following activities are proposed:

- Investigate, analyze and select the model of emotions and technologies for the acquisition of physiological data on which the proposal will be based.
- Analyze and select educational environments to make the physiological and behavioral data capture.
- Investigate, study and select machine learning algorithms for the selection of characteristics for the classification of emotions.
- Design a methodology for the identification of learning-centered emotions.

To validate the work done, tests will be carried out with college students and the recognition accuracy will be evaluated with metrics that allow comparing it with related works in the literature. With this we pretend to corroborate how the proposed methodology allows identifying emotions in educational contexts that can contribute to decision making in emotion-learning relationships.

3 Related Works

The analysis of the state of the art that is being carried out consists of a collection of articles focused specifically on the recognition of emotions in learning activities and the identification of the emotion-learning relationship. These investigations in automatic recognition of emotions represent approximately 20%, and the remaining proportion (approximately 80%) are works that recognize basic emotions (happiness, sadness, fear, anger, contempt, disgust and surprise [9]). There is a clear difference in the literature found. There are works that, although they make an identification of emotions, give greater relevance to the analysis of the emotion-learning relationship. Others emphasize especially the algorithms for automatic recognition of emotions and the recognition rate reached and, with less importance, analyze the emotion-learning relationship. Considering these two aspects, a review of the state of the art analyzed so far is presented.

3.1 Works Focused on the Computational Problem of Automatic Recognition of Learning-Centered Emotions

In the literature reviewed, these works represent approximately 63% of the total of papers reviewed that identify learning-centered emotions. Only the most recent research is mentioned here.

In the work of [10] they use a convolutional neuronal network for the recognition of learning-centered emotions. They run tests using three data-bases: RaFD, database of posed facial expressions containing images of 8 basic emotions and two spontaneous databases created by themselves, especially with content related to learning-centered emotions. The emotions they recognize are: Engagement, excitement, boredom and relaxation. The accuracy they achieve when using their databases is 88% and 74%, respectively.

In [11] they propose a computer-assisted method for special school instructors, where they teach students with mental disorders or emotional problems using a system that employs wearable sensor technologies and intelligent recognition of emotions. The emotion recognition module starts with the capture of the signal from brain diadems. The filtered data is sent to the characteristic extraction module. After extracting the characteristics, they are processed with two classifiers: support vector machine and near k-neighbors with a cross-validation of 10 iterations. The emotions they recognize are: happiness, calm, sadness and fear. An expression module shows the instructor suggestions for treatment according to emotional state detected.

In [12] they implement a binary local pattern for the recognition of learning-centered emotions. The purpose of the work is to build a database of spontaneous facial expressions corresponding to affective states in education to be used in different intelligent tutorial systems. The data capture technologies they use are video and EEG diadems (Emotiv-EPOC). The learning-centered emotions that they recognize are: frustration, boredom, engagement and excitement. They use a support vectors machine to do the classification of emotions. After applying the recognizer, its calculated precision was 80%.

The work of [2] explains the construction and validation of a database of facial expressions that they collect by taking pictures with a webcam. Each photograph is labeled with the emotions of the users obtained at that moment from the Emotiv-EPOC device. For the recognition of facial expressions use a technique based on geometry that measures the distances between the central point of the face and 68 other reference points. These measurements are transformed into characteristics to train a support vector machine. They obtain an accuracy, per emotion, of: boredom of 64%, engagement of 64%, excitement of 83% and frustration of 62%. The same model for the recognition of emotions is also used in [13], as part of an affective learning environment based on Web 3.0 to learn how to program in Java.

Finally, in [14] they propose different recommended activities to induce a certain mental state and capture the EEG response for each one of them. They intend to identify the ideal emotional state to learn. These activities are based on psychological research dedicated to measuring the level of attention, concentration and other functions. In the classification process, they use k-means and clustering with a total concentration rate of 96% of 3592 instances.

Of the totality of works reviewed we can conclude that the most popular devices for the capture of data are webcams, followed by EEG Emotiv diadems. From the use of these technologies and other less popular ones, 80% of the works create their own databases to train and test their recognizers. The most used algorithms for classification tasks are artificial neural networks, support vector machine (SVM), clustering techniques, Bayesian classifiers and nearer neighbors (KNN), among others.

Considering this analysis, we will begin the processing of the data collected using the aforementioned algorithms with the purpose of observing their behavior on the signals that we capture. Later we will make a proposal of our own that may include the use of hybrid algorithms.

3.2 Works Focused on the Analysis of the Emotion-Learning Relationship Based on the Automatic Recognition of Learning-Centered Emotions

Up next, the most recent research on the analysis of the emotion-learning relationship using computer techniques for the identification of emotions is listed. These works represent 37% of the literature reviewed so far, in terms of recognition of learning-centered emotions.

In [15], they make a study of the affective states that originate when students learn with technology, using a tutorial to learn the basics of programming in Python. The emotions they can identify are: engagement, confusion, frustration, boredom, curiosity (which were the most frequent affective states), anxiety, happiness, anguish, surprise, disgust, sadness and fear (which were the rarest). With the analysis of their results, they identify the emotion-learning relationship from five different approaches. They conclude that there is no co-occurrence of affective states that can be generalized; they corroborate the affective dynamics model of [1].

In [16] they present the development of an intelligent tutor with recognition and management of emotions for mathematics. And in [17] they develop an affective learning system for algorithmic logic by applying gamification. In both works, a module for the recognition of emotions is integrated. This process is based on the analysis of faces. In the second work they also add the capture of EEG data signals to label the facial images.

The analysis of these works helped us to identify the areas of opportunity regarding our research and to be able to define the design strategies of the methodology and algorithms for the selection of characteristics and classification to be used. This to validate the hypothesis that states that the proposed methodology allows us to identify emotions in educational contexts that contribute to the decision-making in the emotion-learning relationship.

4 Proposed Methodology

General methodology of the investigation is shown graphically in Fig.1. The main stages are:

- Research, analysis and selection of technologies for acquisition of physiological data, emotions model, application context and machine learning algorithms, which will be proposed for the development of the methodology.
- Data processing; it includes the recording and acquisition of physiological and behavioral data to form the data base, the preprocessing required to prepare the data and the implementation of extraction, selection and integration algorithms of relevant characteristics derived from the different signals.
- Identification of emotions; the selected machine learning algorithms will be tested and trained for classifying emotions centered on learning.
- Validation; tests will be carried out to evaluate the complete methodology for the identification of emotions with metrics that measure their execution, precision and accuracy. Subsequently, the results will be interpreted to identify

and analyze emotion-learning relation-ships. Finally, a discussion of the objectives reached will be made.

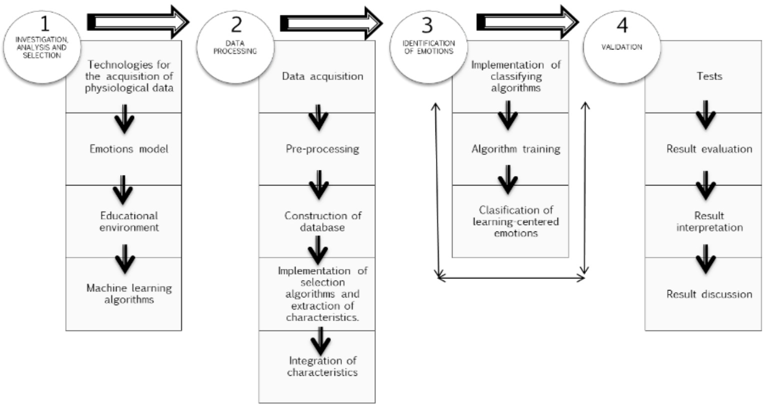


Fig. 1. General Methodology.

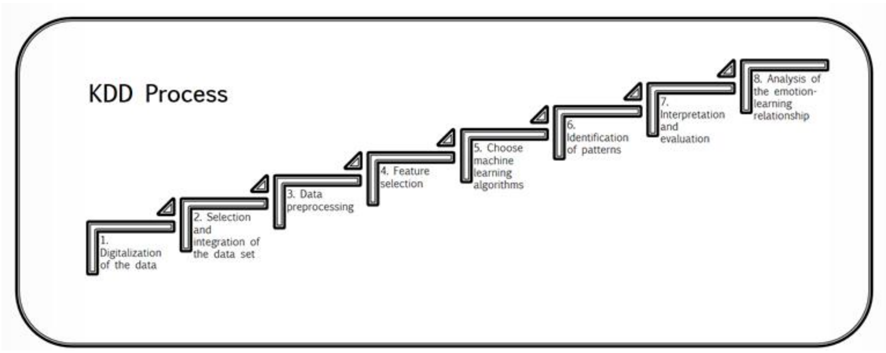


Fig. 2. Knowledge discovery in databases process for automatic identification learning-centered emotions.

Figure 2 show each of the stages of the KDD process (Knowledge Discovery in Databases) [18] which implicitly form part of the computational process for the automatic recognition of learning-centered emotions (these are part of the stages 2 and 3 in figure 1) that propose to implement:

- Recollection of data: correspond to the digitization of data and its storage.
- Data labeling: it consists in selecting and integrating all the data coming from multiple and heterogeneous sources.
- Preprocessing: if necessary, elimination of noise and isolated data. Use of prior knowledge to eliminate inconsistencies and duplicates. Choice and use of strategies to manage the missing information from the data sets.
- Transformation of the data: here the selection of useful characteristics to represent the data, the reduction of dimensionality or methods of transformation and fusion of the data is carried out.

- Data mining: is to choose machine learning algorithms. Choose the task of data mining and mining algorithms that treat all the criteria (classification, regression, clustering or mixed models).
- Identification of patterns: look for patterns of interest in a special way, that is, the recognition emotions.
- Interpretation and evaluation of mine patterns.
- Consolidation of discovered knowledge: discussion and analysis of the relationships of emotions identified with the level of learning.

5 Preliminary Study for Data Collection

The data collection is part of the first stage of the proposed methodology. This includes research, analysis and selection of data acquisition technologies, emotion model, education environment and computer learning algorithms. As part of the preliminary study, the first three elements have been selected. The study, analysis and selection of physio-logical and behavioral data acquisition technologies was carried out by choosing the following: traditional video camera, ICI 9320P thermal camera, heart rate sensor implemented with Arduino and Kinect for Windows. Regarding the model of emotions, we chose to work with discrete and continuous emotions. The learning-centered emotions to recognize will be: engagement, boredom, confusion and frustration. The education environment with which students will interact for the teaching-learning process will be a MOOC of basic algebra using the first topic (duration of 36 minutes).

The experiment for data collection will consist in the capture of behavioral signals and physiological signals in a laboratory of experiments of the Doctorate LKE (Language and Knowledge Engineering) at BUAP (Benemérita Universidad Autónoma de Puebla). We will work with students of the computer faculty at same University, the sample size will be 100 students, who will be asked to participate in a learning activity using the MOOC of basic algebra. The recorded data will be stored, processed and used to recognize the emotions that students present during the learning process. Behavioral signals will correspond to images of facial expressions and movement of the head and hands, stored in JPG and AVI files.

The physiological signals correspond to measurements of the heart rate and temperature of facial areas, stored in TXT, JPG, AVI and XLSX files. The capture of behavioral signals will be done through a video camera and the cameras of the Kinect 360. The capture of physiological signals will be done through a heart rate sensor which will be placed on the ring finger of the left hand. A thermal camera will also be used to capture the facial temperature. Every 10 minutes two tests will be applied to obtain a self-evaluation by the student. In the first test the student must select the emotion he is feeling (engagement, boredom, frustration or confusion). To measure the emotion in a continuous way, the manikin test will be applied, in which the level of each of the three variables involved in the presented emotion will be selected (valence, activation and dominance). In figure 3, photographs of experiments that have been done before starting with formal capture are shown.

Once the database is built, we can proceed to implement the remaining stages proposed in the methodology.

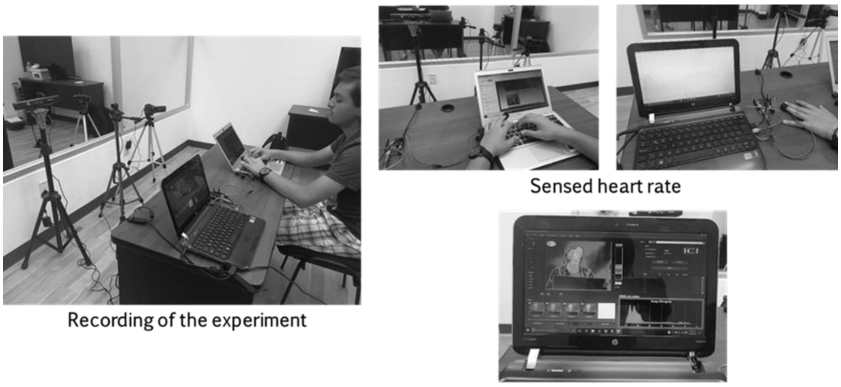


Fig. 3. Pictures of the first experiments.

6 Conclusion

The development of this project will contribute to the identification of gaps in the automatic recognition of emotions centered on learning. It will allow to identify and propose less uncomfortable technologies for students recommended for the collection of data in the identification of emotions centered on learning; we will start with the use of video cameras, thermal cameras, heart rate sensor and Kinect for Windows.

For the development of the proposed methodology, the use of machine learning algorithms is proposed, starting with neural networks and the use of fuzzy techniques. A preliminary analysis of the captured data will allow us to choose the best algorithms according to their characteristics and distribution.

A contribution is expected in the proposal to integrate data from different physiological and behavioral signal acquisition technologies. The data acquisition stage will be carried out through a controlled experiment in which college students will participate using a MOOC to learn algebra. Data of 100 students will be collected. We will try to use a cardiovascular wristband to capture the heart rate that is less obstructive and replace the finger sensor, since we detect discomfort to use the fingers of the left hand in which it is currently placed.

After the data preprocessing will start with the execution of the selected feature extraction and classification algorithms. It is expected to obtain results that contribute to improve the identification of emotions centered in the learning that can be considered for the decision making on pedagogical strategies and learning activities more adapted to each student. The implementation of the process of automatic identification of emotions can be integrated into intelligent tutorials within the student modeling module.

References

1. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learn. Instr., 22(2), 145–157 (2012)
2. Zatarain-Cabada, R., Barron-Estrada, M.L., González-Hernández, F., Rodríguez-Rangel, H.: Building a Face Expression Recognizer and a Face Expression Database for an Intelligent Tutoring System. In: Proceeding IEEE 17th Int. Conf. Adv. Learn. Technol. ICALT 2017, 2161–377X/17, 391–393 (2017)
3. Bosch, N., et al.: Automatic Detection of Learning-Centered Affective States in the Wild. In: Proceeding 20th Int. Conf. Intell. User Interfaces IUI '15, 379–388 (2015)
4. Picard, R.W.: Affective Computing. (1997)
5. Fuentes, C., Herskovic, V., Rodríguez, I., Gerea, C., Marques, M., Rossel, P.O.: A systematic literature re-view about technologies for self-reporting emotional in-formation. J. Ambient Intell. Humaniz. Comput., 1–14 (2016)
6. Zatarain, R., Barrón, M.L., Luis, O.J., Martínez, J.A.: Reconocimiento automático y aspectos éticos de emociones para aplicaciones educativas. Komputer Sapiens, México, 27–31 (2014)
7. Cowie, R., et al.: Emotion recognition in human computer interaction. IEEE Signal Process Mag., 18(1), January, 32–80 (2001)
8. Steidl, S.: Automatic Classification of Emotion-Related User States in Spontaneous Children's Speech. Universität Erlangen-Nürnberg (2009)
9. Ekman, P.: Emotions Revealed. Recognizing Faces and Feelings to Improve Communication and Emotional Life, 1st ed., New York: Henrry Holt and Company (2003)
10. González-Hernández, F., Zatarain-Cabada, R., Barrón-Estrada, M.L., Rodríguez-Rangel, H.: Recognition of learning-centered emotions using a convolutional neural network. Journal Intell. Fuzzy Syst. (2017)
11. Mehmood, R. Lee, H.: Towards Building a Computer Aided Education System for Special Students Using Wearable Sensor Technologies. Sensors, 17(317), 1–22 (2017)
12. Zatarain-Cabada, R., Barrón-Estrada, M.L., González-Hernández, F., Oramas-Bustillos, R., Alor-Hernández, G., Reyes-García, C.A.: Building a Corpus and a Local Binary Pattern Recognizer for Learning-Centered Emotions. Adv. Comput. Intell., II, 524–535 (2017)
13. Zataraín, R., Barrón, M.L., González, F., Reyes-García, C.A.: An Affective and Web 3.0-Based Learning Environment for a Programming Language. Telemat. Informatics (2017)
14. Arana-Llanes, J.Y., González-Serna, G., Pineda-Tapia, R., Olivares-Peregrino, V., Ricarte-Trives, J.J., Latorre-Postigo, J.M.: EEG lecture on recommended activities for the induction of attention and concentration mental states on e-learning students. Journal Intell. Fuzzy Syst. (2017)
15. Bosch, N., D'Mello, S.: The Affective Experience of Novice Computer Programmers. Int. J. Artif. Intell. Educ., 27(1), 181–206 (2015)
16. Barrón, M.L., Zatarain, R., Hernández, Y.: Intelligent Tutor with Emotion Recognition and Student Emotion Management for Math Performance. Rev. Electrónica Investig. Educ., 16, 88–102 (2014)
17. Zatarain-Cabada, R., Barrón-Estrada, M.L., Ríos-Félix, J.M.: Affective Learning System for Algorithmic Logic Applying Gamification. Chapter Lect. Notes Comput. Sci., 576, no. August (2017)
18. Fayyad, U.M., Shapiro, G.P., Padhraic, S.: From data mining to knowledge discovery: An overview. Adv. Knowl. Discov. Data Mining, AAAI Press / MIT Press (1996)

ISSN 1870-4069

Ontology-based Population and Enrichment of Researcher Profiles

Isabel Cruz-Ruiz, Maricela Bravo, José A. Reyes-Ortiz

Autonomous Metropolitan University, Systems Department,
Ciudad de México, Mexico

icruzr2014@hotmail.com, {mcbc,jaro}@azc.uam.mx

Abstract. The representation, management, and exploitation of researcher profiles is an important task that every research institution must achieve. In this paper, we investigate on the use of ontologies as the main solution approach to support the representation of researcher profiles in a given academic environment. We describe the ontology model design, the automatic ontology population processes, and the discovery and enrichment processes of interesting semantic relations between researcher profiles. The functional competency of the enriched ontology is evaluated utilizing a set of inference rules and queries.

Keywords. Ontology population, ontology enrichment, researcher profile.

1 Introduction

Currently, higher education institutions and research institutes have highly specialized human resources who count with high degrees of postgraduate studies. The capacity, expertise and talent accumulated by academic and research staff is one of the most important assets available to institutions. Representing, quantifying and knowing how to better manage these highly specialized human resources is a very important issue; however, it is not an easy task to perform, since it requires the acquisition, representation and intelligent treatment of large volumes of data. A good management of highly specialized human resources can be carried out through the administration of researcher profiles to enable: finding similar profiles to establish new collaborations, looking for specific profiles that allow to integrate a work team with specialists, discovering groups or classes of researchers that address similar topics, discovering groups of researchers that address different problems but that use similar approaches, among other possible applications.

A researcher profile consists of the relevant information regarding previous academic work experience in different research institutions, education and level of studies considering undergraduate, graduate, and specialization studies; an important aspect of a researcher profile is the scientific published articles, chapters, and books, as they represent the researcher topics of interest, and the researcher most active collaborations.

In this paper, we present an ontology solution approach for the acquisition, representation and management of researcher profiles. This ontology solution is evaluated using a set of competency questions through which the functional competency of the proposed solution is evaluated satisfactorily.

This paper reports a contribution in the area of ontology learning. Maedche and Staab [1] define *ontology learning* as the process of automatic or semi-automatic construction, enrichment and adaptation of ontologies. Accordingly, the main tasks involved in ontology learning are ontology enrichment, inconsistency resolution and ontology population.

a) *Ontology enrichment* is the task of extending an existing ontology with additional concepts and semantic relations in the ontology.
b) *Inconsistency resolution* is the task of resolving inconsistencies that appear in an ontology aiming at producing and maintaining a consistent ontology.
c) *Ontology population,* is the task of adding new instances of concepts to the ontology.

The methodological process followed for the construction and evaluation of the proposed solution consist of four phases: ontology design, ontology population, ontology enrichment, and ontology evaluation.

The rest of this paper is organized as follows. Section 2 presents the related work, which is briefly described to compare them with the approach presented in this paper. Section 3 describes the specification of ontology requirements. Section 4 describes the ontology design methods. Section 5 presents the automatic ontology population of each ontology. Section 6 describes the ontology research profile enrichment. Section 7 presents an evaluation based on competency questions. Finally, Section 8 shows the conclusions and future work.

2 Related Work

In this section, we first present the definition of ontology, and describe the related works that address the representation and management of researcher or academic profiles. We analyze their applications and concept coverage.

Over the last decades, different ontology definitions have been presented and discussed. According to Gruber [2] an ontology is an "explicit specification of a conceptualization", an ontology is used to formally define the important concepts of a terminology and the semantic relationships that may exist between concepts. It is frequent that the set of formally defined concepts belong to a specific area of knowledge, and the set of rules and axioms defined are congruent with the particular area of knowledge. In [3] Sowa stated that an ontology represents a catalog of categories to classify entity types that exist in a given domain. In [4] Cámara explained that an ontology can be conceived as an instrument for knowledge representation in a particular topic area, through which knowledge recovery and information retrieval can be executed. Ontologies were selected as the formal representational mechanism as they facilitate reusability, knowledge sharing, and execution of formal reasoning tasks such as satisfiability of concepts, consistency checking, classification and inference.

Concerning researcher profile, Yao, Tang and Li [5] address the problem of researcher profiling by annotating a collection of researcher web pages, and defined a series of difficulties found using this approach. Authors identify tokens in the Web page heuristically, assign tags to each token (Position, Affiliation, Email, Address, Phone, Fax), using the tags, they perform the profiling extraction. In Liu et al [6] authors address the problem of finding experts with required expertise. They describe two ontologies: an *expert ontology*, which defines concepts such as: Person, Publication, Project, and Research Interests; and a *domain ontology* which stores the key concepts (research areas), the attributes of the concepts and the relations between concepts (for example, broader, narrower and part-of). In [7] authors address the problem of automatic extraction of topics of expertise of a person based on the documents accessed by the person through information extraction techniques. They define a user profile using a set of topics with weights determining his level of interest. In [8] authors present a multi-agent paradigm supported by a semantic web architecture to address the challenges of researcher profiling and association. Authors describe an ontological model to represent information such as researcher profiles, conference papers, research centers, etc.

In [9] authors describe ArnetMiner, to address the following questions:

a) How to automatically extract researcher profiles from the Web?
b) How to integrate the extracted information (e.g., researchers' profiles and publications) from different sources?
c) How to model different types of information in a unified approach?
d) How to provide powerful search services based on the constructed network?

In ArnetMiner the schema of a researcher profile was proposed consisting in two main entities: Researcher and Publication. Based on the work reported in [9] in this project we address the same questions and present a solution approach based on the use of ontologies and reasoning tasks.

In [10] authors describe a skill classification ontology model containing skills of research in the area of computer and information science. Their main contributions are:

a) A process to build the skill classification ontology.
b) A methodology to determine expertise of the researcher using the skill classification ontology.
c) A method to retrieve the relevant researchers who may have competency matched to the desired expertise.

Motivated by these related works, we propose an ontology-based solution approach for the acquisition, representation and management of researcher profiles.

3 Specification of Ontology Requirements

The main objective of the ontology model presented in this paper is to facilitate researcher profile processing and reasoning. Considering that every research institution requires the efficient management and dissemination of information relative to the professors and research activities, etc. The ultimate goal of this project is the smart and provisioning of services to researcher communities in which researchers search for specialized publications (such as publications, coauthors, conferences) and are also interested in establishing collaborations with other researchers. Considering this motivation, the following requirements were defined in order to guide the ontology design, construction and evaluation.

3.1 Scope of the Ontology

In order to specify the scope of the ontology, we reviewed the concepts of research profile. Yao, Tang and Li [5] described profiling as the process of obtaining the values associated with the different properties that constitute the person model. Authors define the schema of a researcher profile containing: name, affiliation, position, phone, address, email, research interests, and postgraduate studies. From this definition we consider that the entities (or objects) that constitute a researcher profile are: ***Person*** (for example professors, students, staff, etc.), and ***Publication*** (to extract research interests).

From this initial analysis, we have defined the concept coverage requirements of the ontology and defined the main objective of the ontology, which is to facilitate researcher profile representation and processing in the academic environment.

3.2 Concept coverage

The ontology model should include the following concepts:

a) Data for the identification of persons and researchers such as name, economical number, staff card, etc.
b) Person profile information to represent the user data that is possible to gather from public networks, public Web pages, or public data bases available such as DBLP.
c) Data to represent publications such as: thesis, chapters, journals, etc.

3.3 Competency of the Ontology

Gruninger and Fox [11] proposed six characteristics to evaluate a Business Model. These characteristics were proposed to answer the question of "How can one determine which model is correct for a given task?" To give a guideline on the operation of these characteristics, the authors define the concept of competence of the model as follows: given an appropriately instantiated model and a demonstrator of theorems, the competence of a model is the set of questions that the model can answer. Based on this definition, we may state that

The competence of an ontology model is the set of questions that the ontology can answer.

Evaluation of the competency of an ontology system is crucial to verify that a representational model is complete with respect to a given set of competency questions. During the phase of requirements specification, a team of experts and programmers defined the following set of competency questions for this model:

1. To know how many scientific works does a given researcher has published
2. To find groups of authorship collaborations
3. To know the researcher´s publications from a specialized topic with high degree studies
4. To answer about statistical data of publications
5. To know the scientific productivity of a given department
6. To know the number of female researchers from a given department with published scientific works.

4 Ontology Design

Ontology design is the process of selecting and applying methods, techniques and principles with the objective of producing an ontology model. In this section, the design considerations that were taken into account are described.

A good quality ontology design depends mainly on the selection and incorporation of design principles. Uschold and Grüninger [12] presented their initial ideas and detailed a set of ontology design criteria. For the design and construction of the researcher profile ontology the following design principles were taken in consideration:

- **i.** *Clarity* principle states that an ontology should effectively communicate the intended distinctions. Ambiguity should be minimized, distinctions should be motivated, and examples should be given to understand definitions that lack necessary and sufficient conditions.
- **ii.** *Coherence* design principle specifies that an ontology should be internally consistent. Coherence should also apply to the parts of the definitions that are not axiomatic.
- **iii.** *Extensibility* principle states that an ontology should be designed anticipating possible uses of the shared vocabulary.

Additionally, an initial set of competency questions were used for term elicitation and for final competency evaluation.

4.1 Person Ontology

Person ontology was designed to represent all possible academics, which hold a permanent or temporal position as professor or researcher at the university, such as: academic visitor, full time professor, external sabbaticals, etc. This ontology also represents postgraduate students, and research oriented undergraduate students, among others. Figure 1 shows the main class hierarchy of the ***Person*** ontology. An important characteristic of this ontology is that it uses a unique identifier for every type of person.

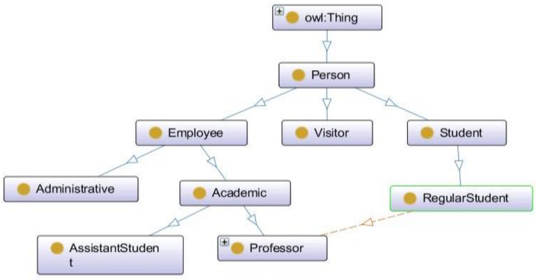


Fig. 1. Person ontology class hierarchy.

The concept ***Person*** is defined as an equivalence through the *hasName* and *hasGender* data properties, indicating that every person individual is obligated to have name and gender to be classified as type of ***Person*** class. The concept ***Employee*** is defined as a sub class of a ***Person*** that *hasEconomicNumber* data property. Whereas the concept ***Student*** is defined as a sub class of ***Person*** that *hasStudentId*. An important concept is a ***Professor*** which is an ***Academic***, is an ***Employee*** and is a ***Person*** that *hasCategory*, *hasDepartment*, and *hasEmail*; and inherits the data property of an ***Academic*** *hasProject*. The class hierarchy of the ***Person*** ontology shows the sub-classification of the class ***Student*** into ***RegularStudent*** and ***AssistantStudent***. This classification addresses the particular need to represent the two types of students that exist in the university where an individual of the ***AssistantStudent*** class is considered to be an ***Academic***, an ***Employee*** and a ***Student***.

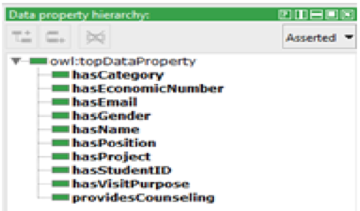


Fig. 2. Data type properties defined for the Person ontology.

The full list of data type properties defined for the ***Person*** ontology are shown in Figure 2.

4.2 Publication Ontology

Scientific published articles, chapters, and books are the most important sources of information in order to integrate a researcher profile. Scientific publications contain the author's topics of interest, conferences and journals of preference, the years of publications and periodicity; also the researcher most active collaborations. In order to build a researcher profile, the design and construction of a publication ontology considered as input the information extracted from the DBLP computer science bibliography on-line reference, extracting the most relevant bibliographic information on major computer science publications.

The ***Publication*** ontology defines the same attributes utilized in DBLP. Figure 3, shows those data type attributes.

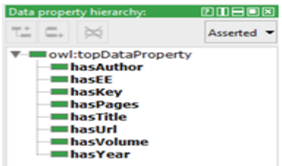


Fig. 3. Data properties of the Publication ontology.

4.3 Researcher Profile Ontology

Researcher Profile ontology was designed to incorporate conceptualizations from ***Person*** and ***Publication*** ontologies. From ***Person*** ontology imports ***Professor*** personal data, such as full name, and economical number; from the ***Publication*** ontology imports publications organized by year, type of publication, among others. Additionally, incorporates ***Department*** and ***Academic Title*** concepts. All these conceptualizations are used to complete the definition of a ***Researcher Profile***, considering the associated publications, the affiliated department, the academic title obtained, and the rest of personal data. Figure 4 shows the main concepts that integrate the ***Researcher Profile*** Ontology.

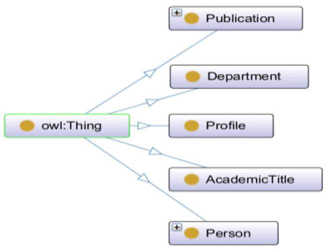


Fig. 4. Researcher Profile ontology class hierarchy.

5 Ontology Population

Ontology population is the process of adding (instantiating) new individuals in the ontology concepts (classes). Automated ontology population is desirable due to the large amount of data that must be extracted and instantiated in the ***Person*** and ***Publication*** ontologies.

5.1 Person Ontology Population

For ***Person*** ontology population, the data source comes from a set of excel files that the management staff of the university uses for different purposes. These excel files contain the information of all academic staff who are affiliated with the university, such as: professor's full name, gender, department, email, economical number, academic projects, and alias.

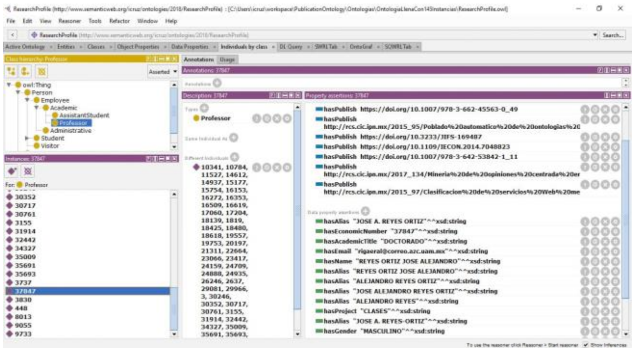


Fig. 5. Person ontology population.

For the automated population of the ***Person*** ontology, two Java modules were developed: a module to *parse and extract* the information from the source files; and another module to *interact* with the ontologies using the Java OWL Application Programming Interface (OWL API) to load and manipulate ontologies, creating new individuals, instantiating object properties and data properties with individuals, and register them in the ontology. Figure 5 shows the values of data type attributes registered for professor "REYES ORTIZ JOSE ALEJANDRO", and the recognized alias names. Alias names are important in order to facilitate the semantic association of the researcher individual with all his publications.

5.2 Publication Ontology Population

For ***Publication*** ontology population, the data was extracted from the DBLP (Digital Bibliography & Library Project) [13], a compressed XML file, which contains more than a million of Computer Science publications. The XML file from the DBLP contains publication title, author names, publication year, volume, EE (a unique publication identifier), URL, and pages (see Figure 6). However, it does not provide the abstract and keywords of publications.

```xml
<article mdate="2018-05-29" key="journals/jifs/Reyes-OrtizB18">
<author>Jos&eacute; A. Reyes-Ortiz</author>
<author>Maricela Bravo</author>
<title>Enhancing patterns with linguistic information for criminal event recognition.</title>
<pages>3027-3036</pages>
<year>2018</year>
<volume>34</volume>
<journal>Journal of Intelligent and Fuzzy Systems</journal>
<number>5</number>
<ee>https://doi.org/10.3233/JIFS-169487</ee>
<url>db/journals/jifs/jifs34.html#Reyes-OrtizB18</url>
</article>
```

Fig. 6. DBLP XML file extract.

A Java module was built to *interact* with the ontologies using the OWL API to load and manipulate ontologies, creating new individuals, instantiating object properties and data properties. Figure 7 shows the instantiation of new ***Publication*** individuals correlated with the identification of the author that published.

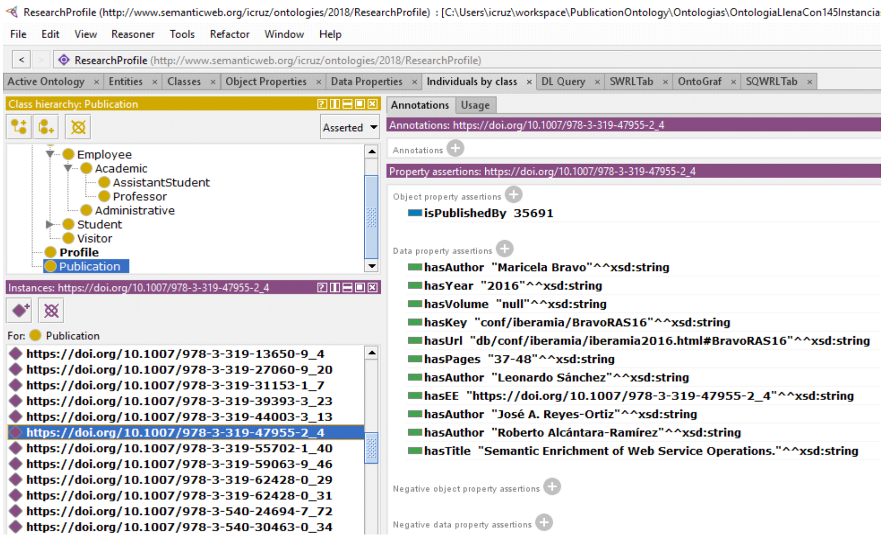


Fig. 7. Publication ontology population.

6 Ontology Enrichment

Ontology enrichment is the automatic process of analyzing the population data values and discovering new interesting semantic relations between individuals. Of particular interest in this enrichment process is the automatic discovery of collaboration relations between authors of publications. For this, the following object properties and inference rules were defined.

collaborateWith is an object property with domain ***Person*** and range ***Person***. This object property was defined to establish semantic relationships between authors of publications.

Figure 8 shows the SWRL rule that was executed to find collaborations between authors of publications is:

```
Publication(?pub) ^ isPublishedBy(?pub, ?prof) ^
isPublishedBy(?pub, ?prof2) ^ differentFrom(?prof, ?prof2) ^
-> collaborateWith(?prof, ?prof2)
```

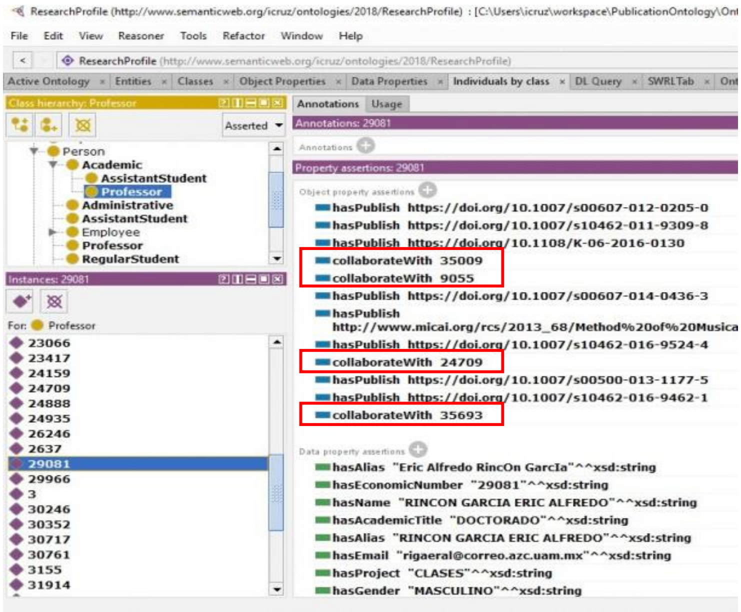


Fig. 8. Discovery of authorship collaborations.

7 Ontology Evaluation

Ontology Evaluation [14] concerns the correct building of the ontology, ensuring that its definitions correctly implement the ontology requirements and competency questions. For evaluation the ***competence of the ontology*** was considered, that is, if it is able to respond to a set of competency questions; and the verification of ***requirements compliance***. The following competency questions were coded in SWRL language and their results were correct.

7.1 Researcher Publications

To know how many scientific works does a given researcher has published, the following rule was defined and executed. Figure 9 shows the result of this competency question.

```
Professor(?prof1) ^ hasPublish(?prof1, ?pub) ^
hasEconomicNumber(?prof1, ?e) ^ swrlb:equal(?e, "14233") ->
sqwrl:count(?pub)
```

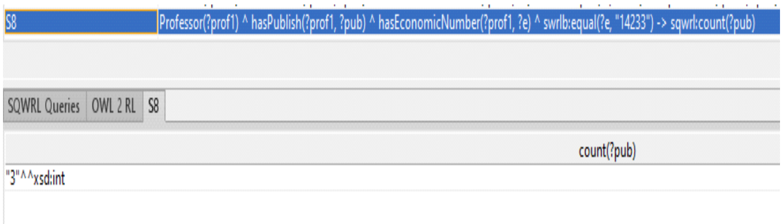


Fig. 9. Discovery of authorship collaborations.

7.2 Collaboration between Researchers

To find groups of authorship collaborations, the following semantic Web rule is defined. The result is shown in Figure 10.

```
Professor(?p1) ^ collaborateWith(?p1, ?p2) ^
hasEconomicNumber(?p1, ?e) ^ swrlb:equal(?e, "14233") ->
sqwrl:select(?p1, ?p2)
```

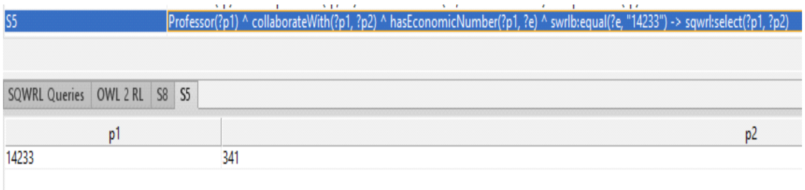


Fig. 10. Discovery of authorship collaborations.

7.3 Qualified and Specialized Researchers

In order to know the researcher´s publications from a specialized topic with high degree studies, the following semantic Web rule was utilized. Figure 11 shows the result of the execution.

```
Professor(?prof1) ^ hasDepartment(?prof1, ?dep) ^ swrlb:contains(?dep,
"ELECTRONICA") ^ hasAcademicTitle(?prof1, ?at) ^ swrlb:equal(?at,
"DOCTORADO") ^ hasPublish(?prof1, ?pub) -> sqwrl:count(?pub)
```

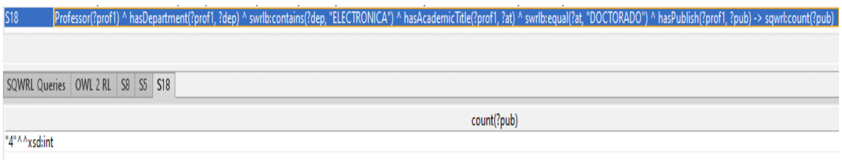


Fig. 11. Highly specialized researchers.

7.4 Publications by Year

The ontology is capable of answering statistical data of publications, for instance: How many publications were there in the year 2017? Figure 12 shows the result of this rule:

```
Publication(?pub) ^
hasYear(?pub, ?y) ^
swrlb:equal(?y, "2017") -> sqwrl:count(?pub)
```

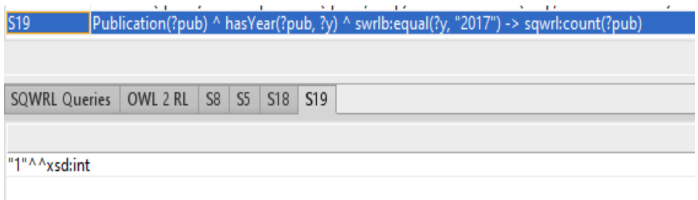


Fig. 12. Publications by year.

7.5 Publications by Department

In order to know the scientific productivity of a given department, the following semantic Web rule is used. Figure 13 shows the result.

```
Professor(?prof1) ^ hasDepartment(?prof1, ?dep) ^ swrlb:contains(?dep,
"SISTEMAS") ^ hasPublish(?prof1, ?pub) -> sqwrl:count(?pub)
```

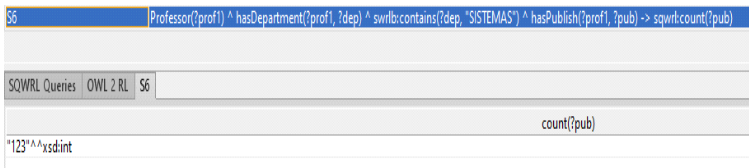


Fig. 13. Publications by department.

7.6 Publications by Gender

To know the number of female researchers from a given department that have published scientific works, the following semantic Web rule was used. Figure 14 shows the results.

```
Professor(?prof) ^
hasGender(?prof, ?gen) ^
swrlb:equal(?gen, "FEMENINO") ^ hasDepartment(?prof, ?dep) ^
swrlb:equal(?dep, "SISTEMAS") ^ hasPublish(?prof, ?pub)
-> sqwrl:count(?prof)
```

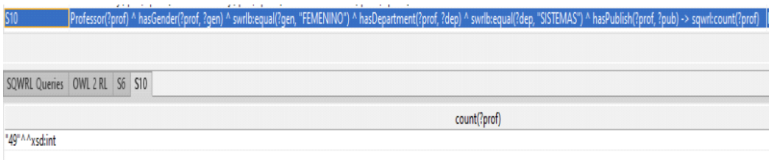


Fig. 14. Publications by gender.

8 Conclusions

In the work reported in this paper, an automated ontology population was used to build researcher profiles. For the population of the ***Person*** ontology, a collection of 373 professors was used, two departments were instantiated in the ***Department*** class, 50 professor individuals are from the Systems ***Department*** and 60 from Electronics. 100% of them were correctly inserted in the ***Person*** ontology, specifically in the ***Professor*** class.

For experimentation purposes two ontologies were generated: one was used for professors of the electronics department and another for professors of the systems department. The systems ontology had a total of 50 researchers in total. Once the universe of professors was divided, the publishing ontology was populated using as a data source the DBLP file, which contains approximately one million articles and more than 56 million lines. The result of the population of the ***Publication*** ontology with professors of the systems department resulted in a total of 135 publications that coincided between the aliases of the professors and the authors indicated within the *<author>* label of the DBLP. At the same time 116 collaboration relationships were found among the professors of the systems department.

The ontology of professors of the electronics department, with a total of 58 professors, was subjected to the same test as the ontology of professors of the systems department and 22 publications were found from the same sample of the DBLP, that is, all the aliases were compared of the researchers of the electronics department against the authors of the million articles of the DBLP. In this ontology 4 collaborative relationships were found, that is, in two publications two or more professors from the same department participated.

As future work, other sources can be considered to continue enriching the ontologies with more semantic relationships, such as ArnetMiner [9], which contains abstracts and keywords of publications in order to enable the semantic relationship between publications, researchers and topics of interest. In this way, the ontology would comply with the characteristic of being scalable and make the profile of each researcher a more complete.

References

1. Maedche, A., Staab, S.: Ontology Learning. In: Handbook on Ontologies (2004)
2. Gruber, Tom. What is an Ontology?, 1993 ksl.stanford.edu/kst/what-is-an-ontology.html

3. Sowa, J. F.: Guided Tour of Ontology, 2001 http://www.jfsowa.com/ontology/guided.htm.
4. Cámara, J. C.: Learning Metadata Standards: Ontologias Barcelona: UPF 2002 http://www.iua.upf.es/ jblat/material/doctorat/ students/jccbis/Ontologias.htm
5. Yao, L., Tang, J., Li, J.: A unified approach to researcher profiling. In: Proceedings of the IEEE International Conference on Web Intelligence, pp. 359–366 (2007)
6. Liu, P., Liu, K., Liu, J.: Ontology-based expertise matching system within academia. In: Proceedings of the IEEE International Conference on Wireless Communications, Networking and Mobile Computing, WiCom, pp. 5431–5434 (2007)
7. Thiagarajan, R., Manjunath, G., Stumptner, M.: Finding experts by semantic matching of user profiles (Doctoral dissertation, CEUR-WS) (2008)
8. Adnan, S., Tahir, A., Basharat, A., de Cesare, S.: Semantic agent oriented architecture for researcher profiling and association (semora). In: Proceedings of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, vol. 3, pp. 559–562 (2009)
9. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: Arnetminer: extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 990–998 (2008)
10. Punnarut, R., Sriharee, G.: A researcher expertise search system using ontology-based data mining. In: Proceedings of the Seventh Asia-Pacific Conference on Conceptual Modelling, vol. 110, pp. 71–78 (2010)
11. Grüninger, M., Fox, M. S.: Methodology for the design and evaluation of ontologies (1995)
12. Uschold, M., King, M.: Towards a methodology for building ontologies Edinburgh. In: Artificial Intelligence Applications Institute, University of Edinburgh, pp. 15–30 (1995)
13. Ley, M.: The DBLP Computer Science Bibliography: Evolution, Research Issues, Perspectives. In Proceedings of the 9th International Symposium on String Processing and Information Retrieval, pp. 1–10 (2002)
14. Gómez-Pérez, A.: Towards a framework to verify knowledge sharing technology. Expert Systems with Applications, vol. 11, no. 4, pp. 519–529 (1996)

ISSN 1870-4069

Towards a Natural Language-based Dialog Model as an Assistant Tool for Dementia Therapy

María J. Somodevilla, Víctor Manuel Morales-de-Jesús

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias de la Computación, Puebla, Mexico

mariajsomodevilla@gmail.com, victor.morales@cs.buap.mx

Abstract. Reminiscence therapy is a treatment of dementia that mitigates the unstable psychological states of a person with dementia, through conversations between patient and caregiver that evoke past experiences focusing on the patient's long-term memory. It is recommended that a person with dementia receive this type of therapy constantly. In this work, we propose the design of a dialog model that generates personalized conversations between a conversational system and patients with Alzheimer's disease, with the aim to assist their caregivers in providing this type of therapy. The proposed model will be capable of getting a semantic interpretation about what the user is saying, identifying the type of dialog, also it will use information from the patient's life history and lifestyle to generate the conversations. During the dialog between the system and the patient, the model searches for relevant information related to the patient's history through the use of ontologies to enrich and develop a further conversation.

Keywords. Dialog model, dementia, conversational systems, reminiscence therapy.

1 Introduction

Dementia is a neurodegenerative and progressive condition characterized by alterations in cognitive processes, behavior, mood, and the ability to perform everyday activities [7]. Alzheimer's disease (AD) is the most common type of dementia in the world and is also one of the main reasons for disability in elderly people after the age of 65, generating dependence on those who suffer this affection. By the year 2010, there were 35.5 million people with dementia worldwide and it is predicted that by 2030 this number will increase to 65.7 million and 115.4 million by 2050 [17]. Therefore, the social cost of dementia is already enormous.

Following this trend, accelerated growth in the population of older adults is expected within the next 20 years in Mexico. According to some estimations, it is expected that during the 2000-2030 period the population in Mexico aged 0 to 14 will have a decrease of 14%, population aged 15 to 64, present a growth of 47%, while population aged over 64 will present a disproportionate growth of 300% [4]. Thus, a significant number of Mexican people will be elderly people, for example, while in 1970 there were 5 to 7 older adults versus 100 young people, it is estimated that by 2050 there will be 139 adults aged over 65 versus 100 young people. The analysis of this data is important because it reveals a need for the transformation of the care requirements for this population group due to its accelerated rate of growth. A population from which almost one third will be elderly people with a diverse level of dependence [7], and as we mentioned above, the main cause of dependency in old age is precisely dementia.

This situation is disturbing because it has been estimated that Mexico is one of the Latin American countries with the highest prevalence of the AD with 7.3 % [11]. Which means that from the 13 million adults aged over 60 who currently live in our country, approximately 1.5 million suffer the AD and it is estimated that by 2050 if not before, the number of Mexicans affected by the AD will increase more than 3.5 million [7].

Although currently, dementia is not a curable disease, it is needed that people suffering from some type of dementia such as Alzheimer's disease can preserve a good quality of life. In this sense, there are both pharmacological and non-pharmacological treatments focused on the reduction of the psychological, behavioral and cognitive impairment symptoms that a person with AD presents. It is preferable that dementia treatments begin with a non-pharmacological intervention, because of this type of treatments promote the use of different methods and techniques to provide physical and emotional safety to the patient without the side effects of pharmacological treatments [8]. Through the implementation of appropriate environments, stimulating tasks and therapies applied according to the specific needs of each patient, the effect of this type of intervention can be maximized [5].

Reminiscence therapy is among this type of intervention, in which a conversation is established between the patient and their caregiver about activities and experiences from the past, where the patient is guided chronologically through their life experiences [19]. This type of therapy is based on the patient's long-term memory using information on their life history, that is, information that is familiar to the patient and relatively easy to remember, since during the mild and moderate phase of the AD the short-term memory is affected mainly. In different studies, the benefits of subjecting patients with dementia to this type of therapies have been studied and it was observed that carrying out constantly this type of conversations supports to enhance the unstable and emotional psychological states of the patients, increasing their sociability and self-confidence [9,19,6].

However, for most people with the AD, it is unreasonable to consult a professional doctor or therapist every day. Also, even for caregivers, the time and effort spent on the counseling are quite limited since they have many others supports to do [14]. In this sense, the objective of this work focuses on the creation of a dialog model capable of generating personalized conversations between a conversational system and patients with AD, based on information about the patient's life history to support caregivers and therapists to provide reminiscence therapy constantly with the purpose of mitigating emotional and psychological symptoms such as anxiety and depression generated by the AD and help to preserve the quality of life of the patient.

2 Related work

A brief review of the literature is presented, as well as a comparison with those proposals considered similar to this work, highlighting the constraints and differences detected with the proposed approach.

It is remarkable to mention that a large number of research proposals focus on the diagnosis of dementia through the classification of symptoms and identification of distinctive features for early detection of dementia. Tanaka et al. [16] proposed a computer avatar with spoken dialog functionalities that produces spoken queries based on the mini-mental state examination (MMS), the Wechsler memory scale-revised, and other related neuropsychological questions. They recorded the interactive data of spoken dialogues and extract different audiovisual features, then two machine learning algorithms were used achieving a 0.93 detection performance rate. Chinaei et al. [2] analyzed several linguistics features that are verbal indicators of confusion in AD like vocabulary richness, parse tree structures, and acoustic cues. They applied several machine learning algorithms to identify dialog-relevant confusion from speech with up to 82% accuracy.

On the other hand, there are approaches that promote the use of assistive technology to help people with dementia with activities of daily living (ADLs) [20,13,18,8] and support different kinds of therapy to address psychological, emotional and behavioral symptoms of dementia [1,3,14,15].

Yasuda et al. [20] developed a remote reminiscence conversation and schedule prompter system via videophone to improve psychological stability and to assist individuals with dementia to perform household tasks. They observed that psychological stability of 1 patient persisted for 3 hours after remote conversations. Futhermore, motivational promter videos were added and they found that the task completion rate afforded by the revised schedule prompter system was 82%. This type of proposals are compared against other instructions strategies in [10]. Ruszicz et al. [13] built a mobile robot, call ED, intented to assist with ADLs through visual monitoring and verbal prompts. They analyzed speech-based interactions between ED and each of 10 older adults with AD as the latter complete daily tasks in a simulated home environment. Their analysis reveals that across the verbal behaviors that indicate confusion, patients with AD are very likely to simply ignore the robot, which accounts for over 40% of all such behaviors.

Wolters et al. [18] analysed the interaction between people with dementia and a simulated intelligent cognitive assistant (ICA) that support people with dementia who need performing everyday tasks by detecting when problems occur and providing tailored and context-sensitive assistance. They conducted three focus groups with people with dementia, carers, and older people without a diagnosis of dementia. Analysis of the focus data showed that voice and interaction style should be chosen based on the preferences of the patients and it is crucial that the user can personalise the ICA themselves.

Similarly, Navarro et al. [8] implemented an assisted cognition system and analyzed the results on the use an adoption of the system to support occupational therapy and evaluated the adoption and effectiveness of the system to ameliorate challenging behaviors and reducing caregiver burden. They conducted an in situ evaluation with two caregiver-patinent dyads and it was observed that intervention personalization and touch-based interface encouraged the adoption of the system, helping reduce challenging behaviors in people with dementia and caregiver burden.

Casey et al. [1] performed a qualitative study wich describes how people with dementia and other key stakeholders could help to design a companion robot (MARIO). In their research they identified elements of importance to end users that make a companion robot acceptable to them and more likely that they would engage and interact with it. At the end, a relevant aspect they found is that people with dementia would like MARIO to have access to their life history, their interests, and hobbies and use this information to foster conversations and reminiscence about events that they could remember more easily. Cruz-Sandoval et al. [3] designed a semi-autonomous agent, called Eva. The agent is capable of participating in simple conversations with the coordination of a human operator to be perceived as an engaging speaker. Eva implements non-pharmacological interventions such as cognitive games and music therapy. They conducted a preliminary evaluation with a group of caregivers (n=8) obtaining quantitative and qualitative results that were used to asses the social abilities of the robot.

As was above-mentioned, a person with dementia should receive these interventions as often as possible. In this sense, virtual agent technology is considered a good alternative to offer daily dementia counseling even at home. Sakakibara et al. [14] proposed a method that dynamically generates personalized dialogs for individual people with dementia. Using the patient's life history the system chooses an appropriate conversation. During the conversation, the system finds new information in linked open data (LOD) relevant to the response and uses it to develop futher conversation. However, their prototype system only implements the personalized dialog for the user's birthplace.

Since language comprehension and production is coparatively well-preserved in people with mild to moderate AD, proposals using dialog systems or voice interfaces to support non pharmacological dementia treatments like cognitive assistance and therapy, have shown be helpful and well adopted. However, in most analyzed proposals, generic dialogs are used, that is, the same conversation is used for all patients, so they do not provide the benefits that have been observed in conversations where the dialog is personalized to each patient as in reminiscence therapy. In this paper, we propose an approach that aims generate personalized dialogs to support reminiscence therapy.

3 Proposed Approach

In this section, we mention our proposal that will be carried out through this work. Firstly, the process of developing the dialog model is described, and then how this dialog model will be integrated in a conversational system is showed.

3.1 Dialog model description

The construction of the dialog model considers different stages according to the particular tasks that need to be performed in each stage. Figure 1 shows the process for the construction of the dialog model.

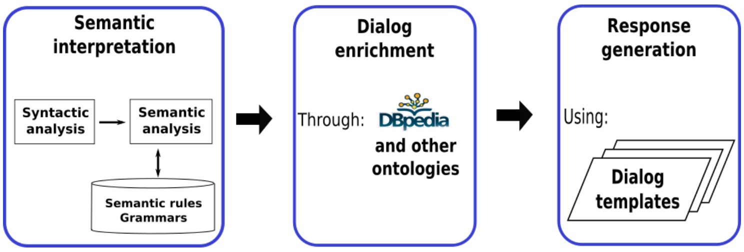


Fig. 1. Development of the dialog model.

Initially, it is necessary that the dialog model will be capable of getting a semantic interpretation about what the user is saying, this is, that the model will understand the type of dialog (question, a request of information, affirmation, etc.), as well as the meaning of what the user said. Therefore, the first thing that is proposed is a syntactic analysis of the sentences with the objective of grammatically labeling each word of the sentence. Subsequently, it is intended to perform a semantic analysis by defining a set of semantic rules according to the patterns detected within the grammar labels added in the syntactic analysis.

As result from the semantic analysis, it is expected to detect some “semantic attributes”, this attributes then will be taken to retrieve relevant information according to the patient’s life history through the use of diverse ontologies such as DBpedia[1]. Finally, the generation of the response is carried out using appropriate dialog templates according to the information retrieved from the ontologies.

Because it is planning that the created model be able to generate a personalized dialog to each patient, it is necessary to have the relevant information of the patient. Thus, the use of the *Center Method* is considered, this method consists in the use of a set of validated forms where information relevant to the family, life history, and lifestyle of the patient with AD is gathered [12]. The forms would be filled semiautomatically in the system by the caregiver according to

[1] https://wiki.dbpedia.org/develop/datasets

the information of the patient. Then the dialog model can use this information to generate personalized conversations according to such information.

3.2 Conversational system

As it was above-mentioned, the dialog model will be integrated into a conversational system prototype. The conversational system uses different modules to perform voice interaction to the user. Figure 2 describes the architecture of the system. First of all, the system must have an input mechanism for communication with the user, in this case, the Google Cloud Speech-to-Text API[2] will be used within the Automatic Speech Recognition (ASR) module to transcript the utterances from the user to text. Subsequently, it is necessary to analyze the text and obtain a semantic interpretation of the information contained in the received message as it was already mentioned. All this process will be done in the Natural Language Understanding (NLU) module using a diverse type of techniques to perform this task. Some of this techniques imply named entity recognition, speech acts classification, etc.

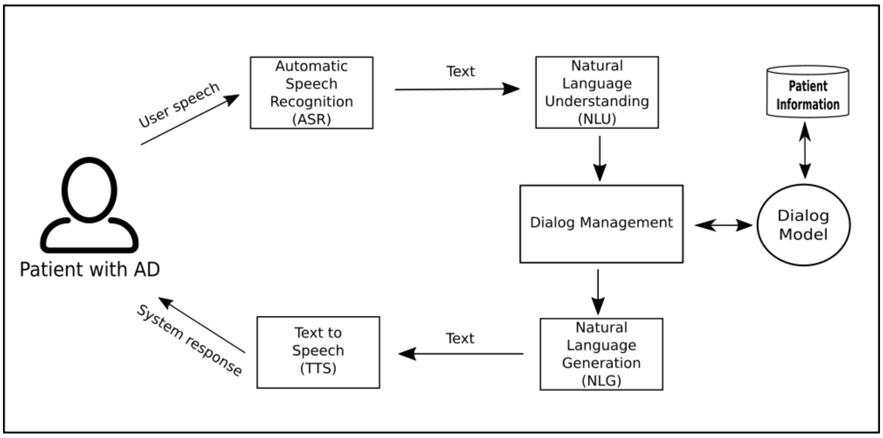


Fig. 2. Architecture of the conversational system proposed.

The Dialog Management module will be in charge of managing the turns of conversation between the user and the system and determining the actions to perform during the interaction according to the current state of the dialog. This module will interact directly with the Dialog Model to interchange semantic information that can be used to select the patient's information relevant for the conversation.

It is necessary to generate a response according to the information provided by the Dialog Model. The response by the system has to be understood by

[2] https://cloud.google.com/speech-to-text/

the user. Hence, to represent in natural language the response provided by the Dialog Management module, diverse techniques will be implemented into the NLG module. Finally, the response will be synthesized through the TTS module using the IBM Watson Text-to-Speech API[3] that offers a more natural-sounding audio.

With the objective to clarify how the system is expected to conduct the dialog with people with AD, the following example in Figure 3 is given. The figure describes a particular conversation that could be performed by the system. First, the system asks the user: "*¿Dónde naciste María? (Where were you born Maria?)*". Suppose that the user María answers "*Yo nací en la ciudad de Puebla (I was born in Puebla city)*".

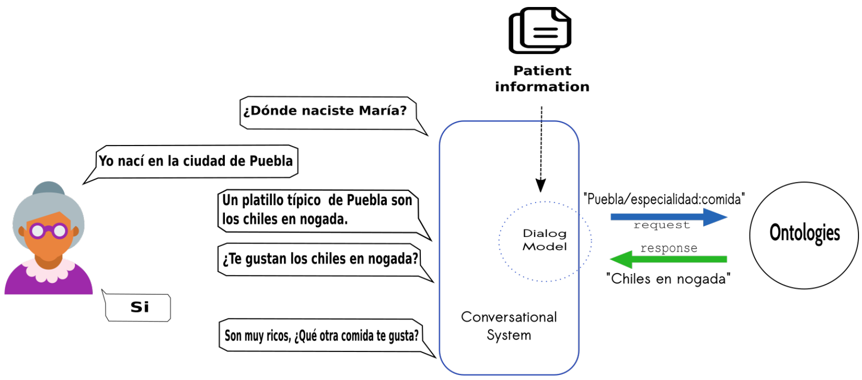


Fig. 3. Example of a dialog generated by the conversational system.

Then, the dialog model would apply some of the semantic rules established to extract the named entity of Puebla and to know that Puebla could be a city or a state. Based on that information a query over different ontologies will be performed to extract relevant information about Puebla. In this case, the query retrieved the traditional food of Puebla where a possible result would be chiles en nogada. Using this information, the model enriches the conversation by mentioning "*Un platillo típico de Puebla son los chiles en nogada (A typical dish of Puebla are chiles en nogada)*", and the system would extend the conversation by asking "*¿Te gustan los chiles en nogada? (Do you like chiles en nogada?)*", where the user could affirm "*Si (Yes)*" and the system would motivate to continue the conversation by saying "*Son muy ricos. ¿Qué otra comida te gusta?(It is delicious. What other kind food do you like too?)*".

[3] https://www.ibm.com/watson/services/text-to-speech/

4 Conclusion

Due to the cognitive impairment that affects people with dementia, in early stage patients present different types of symptoms, one of the most common is short-term memory loss. This situation significantly affects their social behavior towards the people around them and generates unstable psychological states such as anxiety and depression. However, it has been observed that providing patients with non-pharmacological treatments such as reminiscence therapy improve psychological well-being. There are many proposals with the aim of supporting different kinds of therapy to address the psychological, emotional and behavioral symptoms of dementia. Most of them are based on an interaction with the patient through the use of generic dialogs used for all patients, thus the personalization in the intervention is lost.

In this paper, we propose the creation of a dialog model that could be integrated into a conversational system that generates individually personalized dialogs for each patient with AD according to their life history and preferences. Similarly, the model will enrich the dialogs through the use of ontologies that allow the extraction of relevant content associated with the information collected from the user.

In future work, it is initially contemplated to implement diverse natural language understanding techniques that allow us to realize a semantic analysis of the patient's utterances. As well, the selection of useful resources for dynamic dialog generation in real time by the system. Finally, it is considered to perform an experimental evaluation to asses the effectiveness and adoption of the system by people in the early stage of the AD.

References

1. Casey, D., Felzmann, H., Pegman, G., Kouroupetroglou, C., Murphy, K., Koumpis, A., Whelan, S.: What people with dementia want: designing mario an acceptable robot companion. In: International Conference on Computers Helping People with Special Needs. pp. 318–325. Springer (2016)
2. Chinaei, H., Currie, L.C., Danks, A., Lin, H., Mehta, T., Rudzicz, F.: Identifying and avoiding confusion in dialogue with people with alzheimer's disease. Computational Linguistics 43(2), 377–406 (2017)
3. Cruz-Sandoval, D., Favela, J.: Semi-autonomous conversational robot to deal with problematic behaviors from people with dementia. In: Ochoa, S.F., Singh, P., Bravo, J. (eds.) Ubiquitous Computing and Ambient Intelligence. pp. 677–688. Springer International Publishing, Cham (2017)
4. Fernández, P.H., Velarde, S.I., Hernández, M.F., Murguía, V.: Dinámica demográfica 1990-2010 y proyecciones de población 2010-2030. Consejo Nacional de Población (CONAPO) (2014), `http://www.conapo.gob.mx/work/models/CONAPO/Proyecciones/Cuadernos/15_Cuadernillo_Mexico.pdf`
5. Gitlin, L.N., Hodgson, N., Jutkowitz, E., Pizzi, L.: The Cost-Effectiveness of a Nonpharmacologic Intervention for Individuals With Dementia and Family Caregivers: The Tailored Activity Program. The American Journal of Geriatric Psychiatry 18(6), 510–519 (jun 2010)

6. González-Arévalo, K.A.: Terapia de reminiscencia y sus efectos en los pacientes mayores con demencia. Psicogeriatría 5, 101–11 (2015)
7. Gutierrez, L.: Demencias en México. In: Demencias, una visión panorámica, pp. 9–20. Universitaria Potosina (2014)
8. Navarro, R.F., Rodríguez, M.D., Favela, J.: Use and adoption of an assisted cognition system to support therapies for people with dementia. Computational and Mathematical Methods in Medicine 2016 (2016)
9. Neal, M., Barton Wright, P.: Validation therapy for dementia. The Cochrane Library (2003), art. No.: CD001394. DOI: 10.1002/14651858.CD001394
10. Perilli, V., Lancioni, G.E., Hoogeveen, F., Caffó, A., Singh, N., O'Reilly, M., Sigafoos, J., Cassano, G., Oliva, D.: Video prompting versus other instruction strategies for persons with alzheimer's disease. American Journal of Alzheimer's Disease and other Dementias 28(4), 393–402 (jun 2013), `http://journals.sagepub.com/doi/10.1177/1533317513488913`
11. Rodriguez, J.J.L., Ferri, C.P., Acosta, D., Guerra, M., Huang, Y., Jacob, K., Krishnamoorthy, E., Salas, A., Sosa, A.L., Acosta, I., et al.: Prevalence of dementia in latin america, india, and china: a population-based cross-sectional survey. The Lancet 372(9637), 464–474 (2008)
12. Rokkaku, R.: Care management sheet pack for the elderly with dementia: The center method ver. 03. Nihon Ronen Igakkai zasshi. Japanese journal of geriatrics 42(3), 318–319 (2005)
13. Rudzicz, F., Wang, R., Begum, M., Mihailidis, A.: Speech Interaction with Personal Assistive Robots Supporting Aging at Home for Individuals with Alzheimer's Disease. ACM Transactions on Accessible Computing 7(2), 1–22 (may 2015), `http://dl.acm.org/citation.cfm?doid=2785580.2744206`
14. Sakakibara, S., Saiki, S., Nakamura, M., Yasuda, K.: Generating personalized dialogue towards daily counseling system for home dementia care. In: Duffy, V.G. (ed.) Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Health and Safety. pp. 161–172. Springer International Publishing, Cham (2017)
15. Sarabia, M., Young, N., Canavan, K., Edginton, T., Demiris, Y., Vizcaychipi, M.P.: Assistive Robotic Technology to Combat Social Isolation in Acute Hospital Settings. International Journal of Social Robotics (2018), `http://link.springer.com/10.1007/s12369-017-0421-z`
16. Tanaka, H., Adachi, H., Ukita, N., Ikeda, M., Kazui, H., Kudo, T., Nakamura, S.: Detecting dementia through interactive computer avatars. IEEE journal of translational engineering in health and medicine 5, 1–11 (2017)
17. Wimo, A., Prince, M.: World Alzheimer Report 2010. Tech. rep., Alzheimer's Disease International (2010)
18. Wolters, M.K., Kelly, F., Kilgour, J.: Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia. Health Informatics Journal 22(4), 854–866 (2016)
19. Woods, B., Spector, A.E., Jones, C.A., Orrell, M., Davies, S.P.: Reminiscence therapy for dementia. The Cochrane Library (2005)
20. Yasuda, K., Kuwahara, N., Kuwabara, K., Morimoto, K., Tetsutani, N.: Daily assistance for individuals with dementia via videophone. American Journal of Alzheimer's Disease and other Dementias 28(5), 508–516 (aug 2013)

ISSN 1870-4069

Data Acquisition Software for Sign Language Recognition

Manuel Eleazar Martínez-Gutiérrez, José Rafael Rojano-Cáceres, Edgard Benítez-Guerrero, Héctor Eduardo Sánchez-Barrera

Universidad Veracruzana, Facultad de Estadística e informática, Mexico

eleazar.mg.2000@gmail.com, {rrojano, edbenitez}@uv.mx, eduarsb4@gmail.com

Abstract. The research of sign language recognition is a topic of interest in the recent years in the fields of computing vision, artificial intelligence and human-computer interaction. The sign language recognition has at least three steps such as data acquisition, data classification and results. This paper introduces a software capable of facilitating the data acquisition step with the Intel Realsense camera. This software digitize 22 points from the hand of the user in coordinates (x,y,z) and save this data in a comma separated values (CSV) file. Finally, in this paper we tested the software with 19 users recording the alphabet of the Mexican sign language, with the dataset from the software, we trained an artificial neural network to obtain 80.1192% of precision. While this percentage can be increased with a variety of techniques, the test shows that users can create the dataset using this software and that those values can be used in a classification algorithm with acceptable results.

Keywords. Sign language, data acquisition, pattern recognition.

1 Introduction

Sign language has the characteristic of being a visual language, this means that the medium of communication is the sight and the way to express a message is the use of gestures from the hands (including the form of the hands, the movement and the space where it is made), face and the combination of both [1].

The characteristics of the sign language allow the use of different types of research to try to digitize it, there are research with the context of sign language recognition, more specifically in the areas of: computer vision, artificial intelligence and human-computer interaction. In computer vision the researchers develop algorithms to digitize an image or video, more precisely to obtain data of the sign in the image, this can be done with filters [2,3]. The filters can find the edge of the hand, eliminate the background noise from the picture, find the person doing the hand sign, obtain the pixels where the hand is, etc. (this depends on which filter is used). The artificial intelligence takes the data digitized from the computer vision as a dataset, this dataset is taken to a pre-process step to clean the data for the classification algorithm. The purpose of the algorithm is to train from the dataset to be able create a classification model capable of classifying another dataset with a high success rate [4-6]. Finally, the human-computer interaction takes the challenge of incorporating the model from the artificial intelligence into a computer system, the computer system needs to design interfaces to allow sign language users to interact with the software, to do so it is needed an evaluation of the software with sign language users to measure the usability of the system. The system should be able to communicate with the user using the sign language, to perform this, the system should take the filters from computer vision to digitize the input form the user into a dataset to be classified by the classification model, the output of the model will be the input form the user.

In this work it is proposed the use of a software capable of digitize the data of the hand with the use of the Intel Realsense f200, the software can detect 22 points from the hand and saving them in a CSV file, this file can be used as a dataset to work with an algorithm from artificial intelligence.

The paper is organized as follow: Section 2 cover the process of data acquisition. Section 3 is about the software for data acquisition and is explained how the software takes the work capturing the data. Section 4 use the dataset from the software to create a model capable of classifying the signs. Section 5 brings a conclusion to this work.

2 Data Acquisition

As mentioned in the introduction the data acquisition is one step of the sign language recognition, this step consists of digitize the information from the user to make it usable to classification algorithms, the information from the user can be acquired in different ways depending on the tools used, such as visual based like cameras or wearable like the use of electromyogram, sensors in a glove, etc. Those tools have different advantages such as low-cost, hardware and software compatibility, user detection, etc., in the same way those tools have disadvantages such as occlusion, cost, trouble with the users, etc. Depending on the scope of the research the tools to obtain information from the user can change.

2.1 Visual Base Acquisition

The process of data acquisition can be seen as a technical work with images, as in the work [2] where they used a Sobel edge filter to eliminate the background noise from the image and obtain the outline of the hand as a binary image, then a neuronal network is used to generate tokens from the outline of the hand, this tokens represent the points of the shape of the hand, finally the set of tokens from the combination of the images is the dataset used in a classification algorithm. This process can start since the image acquisition as it can be seen in [3] he begins explaining about the resolution from the camera, this is important because the more pixels the image has it is needed more processing power and time to digitize the information, the environmental noises like the light, the position of the cameras and the background of the images. Once the set of images are ready, he smoothed the images with a spatial averaging filter of mask 3x3, then he uses morphological operations of dilation and erosion, next he uses a spatial difference to obtain segments of the hand, then to improve the contrast of the segment image he used the adaptive histogram based contrast enhancement, finally he uses an elliptical Fourier descriptors to create the dataset used in a neuronal network.

The process is similar from the previous works using a special camera like Kinect, the advantages of depth cameras are the ease of digitize the visual data and include it into a software, as it can be seen in [7] the data acquisition with Kinect may require the use of affine transformation to cancel the noise of the signer position and rotation. So, with that in mind, even if the depth camera can obtain the digitize data it may require a few steps to prepare the data.

2.2 Wearable Tool Data Acquisition

The wearables tools have the characteristic of being intrusive, so it can be bothersome to the user, however these tools can approach problems like occlusion or can be used to test different kind of sensors in data acquisition for sign language recognition. For example, in [8] they used the TMS porti electromyogram as a data acquisition tool, unlike with the use of a camera, this one obtains signals that needs to be extracted to be used as a dataset, in this case it is used the mean absolute value and moving variance.

As mentioned before, the use of wearable tools can be because of different reasons like the use of colored gloves to make the hand easier to detect or to evade problems like occlusion that the use of cameras present, in the work [9] they utilize a data glove with 15 sensors divided 3 per finger, with the use of this glove they can digitize the information from the user and evade the problem of occlusion because of the use of sensors the data does not change even if the hands overlap with each other.

3 Sign Language Data Acquisition Software

As seen in the previous section of this paper, the data acquisition is a step in sign language recognition that depending on the tool used it can involve the use of filters or preparations of the data, similarly the choice of the tool depends of its advantages like to evade occlusion, easier hand detection, usability, etc.

In this paper, we propose the use of a software with Intel Realsense f200 depth camera as a data acquisition tool, this camera has the advantages of digitize 22 points from each hand and is capable of facial recognition, the range of capture is between 20-120cm so it is recommended to use this tool for one user at a time, while this is a small range of capture to use in an open space for full body capture like with Kinect, it is optimal for a context when the user is seated and close with the computer.

The specifications to work with the Intel Realsense f200 camera are: 4th generation intel core processor or higher, Microsoft windows 8.1 or higher, 4gb of RAM, 4gb of hard disk space and a USB 3.0 port. The software proposed in this paper works with JAVA and the Intel Realsense SDK, the scope of the software is to capture 22 points of the hand in coordinates (x,y,z), this coordinates are saved in a CSV file with values measured in meters, while this software can capture the face and hands, in this paper we will focus in the use of only one hand.

One of the advantage of the software is the ease of acquiring the hand data from the user, this easiness is compared with the process of data acquisition from the perspective of using a web camera which involves the use of filters and transformation on the image to identify the hand and digitize it to obtain data. Another advantage of the software is the simple design so that the user is capable of digitize his information without the need of an expert to help him.

The software shows the user using the depth camera as shown in figure 1, when the software detects the hand from the user, an icon will appear in the bottom right corner of the interface and the 22 points from the hand will be drawn in colors, the red dot is the center of the palm, the blue dots are the tip of the fingers and the green dots are the rest of points that are the base of the fingers and the wrist. Simultaneously the software draws a line through the points of each finger to help the user represent the finger, this is useful when the user is making a sign and it is confusing to know if the points are in the correct position.

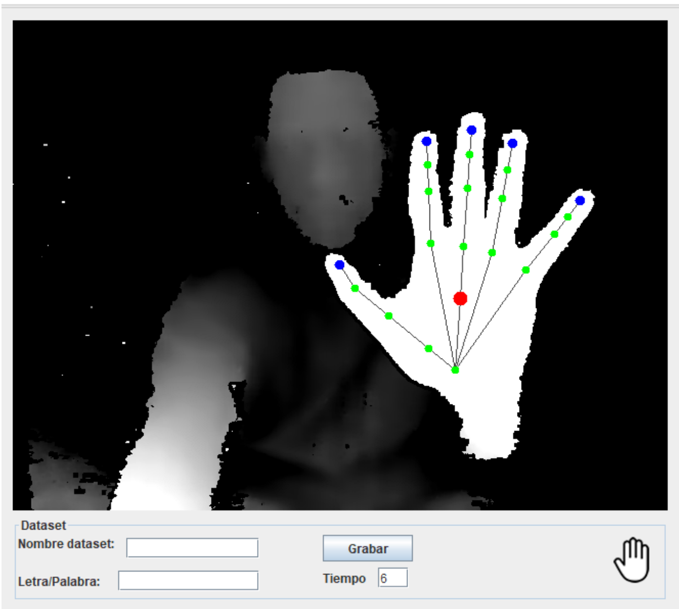


Fig. 1. Data acquisition software with Intel Realsense.

The way to use the software as shown in figure 2 is to fill the data before pressing the button with the name of the CSV file, the letter or word that the user is going to sign and the time that the software is going to record. This inputs are important to create a dataset, in this software if the name of the CSV file is the same of a previous one, the new records are going to be in the end of the previous ones, this is recommended to take control of the number of records per sign and to evade the noise when the user is changing from sign to sign, the letter or word represent the expected value in a classify algorithm so depending on the algorithm it is possible to use a number, the value of the time is because of the number of records that the software is going to write in the CSV file, the camera take 30 frame per second so if the time is in 6 seconds the software is going to write 180 records.

Finally, when the button is used, it will start recording and change its label to a timer to let the user know how much time left to finish the record, simultaneously an icon REC will appear at the bottom right corner of the interface and it will disappear when the time is up.

Fig. 2. Recording the letter “a” from the Mexican sign language.

4 Evaluation of the Software and the Dataset

So, to evaluate the software in this paper we recorded the alphabet of the Mexican sign language with a group of 19 people with basic knowledge of the Mexican sign language, the process was as follow:

1. Equip a place with two computers with the Realsense camera and the software.
2. Schedule the users so only two will use the software at the same time, one in each computer.
3. Prepare a consent document to use the digitize data from the user.
4. Explain the user how to use the software.
5. For the static signs, the user will sign then it will start the record until the time is up.
6. For the signs with movement, the user will repeat the sign until the time is up.

7. The record will be the alphabet twice per user.
8. Combine all the CSV files into one file.
9. Eliminate null records.

The dataset is conformed of 90,305 of records after eliminating null records, those happens when the hand was not recognized while the camera was recording. The dataset is composed of 45 variables, 44 of them are the coordinates (x,y) of the 22 points acquired from the camera and one variable representing the letter of the alphabet, with this dataset we trained an artificial neural network (ANN) multilayer perceptron and backpropagation algorithm.

To train the ANN we split the dataset in 70% to train it and 30% to test it, the structure of the ANN was of 44 neurons in the input layer, 34 neurons in the hidden layer and 27 neurons (representing each letter of the alphabet) in the output layer. To test the classification model, we used cross validation with 30% of the dataset, the results of the test were 81.1192% of correctly classified instances and kappa statistic of 0.8039 which is represented as substantial [10].

5 Conclusion

Based in the state of art we can see the process of data acquisition which involves different approaches, tools, scopes, filters and results but the purpose of them is to take the digitize information from the data acquisition to train a classifier algorithm. In this paper we proposed a software that helps in the data acquisition step with the Intel Realsense camera, this software allow the user to digitize the hand in 22 points of the hand in coordinates data with the flexibility of using one button to record and saving the data in a CSV.

Based in the test of the data presented in this paper, the precision of the ANN was above the 80% only using the data form the software, this percentage can be increased with a variety of techniques such as translation, selection of significative attributes, rotation, etc.

Finally the purpose of the software was a success since in the test, the users were capable of using it and it was possible to create a ANN with the dataset from this test, so in the future with this kind of interaction it will be possible to create applications centered in deaf people with the ease of acquiring the digitize data from the user in the acquisition phase or in the application itself.

References

1. Aldrete, M.: Grámatica de la lengua de señas mexicana. El colegio de México (2008)
2. Raheja, J., Singhal, A., Chaudhary, A.: Android based Portable Hand Sign Recognition System. arXiv Preprint arXiv:1503.03614 (2015) http://doi.org/10.15579/gcsr.vol3.ch1
3. Kishore, P.V.V., Prasad, M.V.D., Prasad, C. R., Rahul, R.: 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN. In Signal Processing and Communication Engineering Systems (SPACES), 2015 International Conference on IEEE, 34–38 (2015)

4. Mohandes, M., Deriche, M., Johar, U., Ilyas, S.: A signer-independent Arabic Sign Language recognition system using face detection, geometric features, and a Hidden Markov Model. Computers and Electrical Engineering, 38(2), 422–433 (2012) http://doi.org/10.1016/j.compeleceng.2011.10.013.
5. Bedregal, B.C., Costa, A.C., Dimuro, G.P.: Fuzzy rule-based hand gesture recognition. In IFIP International Conference on Artificial Intelligence in Theory and Practice, Springer, Boston, MA., 285-294 (2006, August)
6. Islam, M.M., Siddiqua, S., Afnan, J.: Real time hand gesture recognition using different algorithms based on American sign language. In Imaging, Vision & Pattern Recognition (icIVPR), 2017 IEEE International Conference on IEEE, 1-6 (2017)
7. Kumar, P., Saini, R., Roy, P. P., Dogra, D. P.: A position and rotation invariant framework for sign language recognition (SLR) using Kinect. Multimedia Tools and Applications, 77(7), 8823–8846 (2018)
8. Amatanon, V., Chanhang, S., Naiyanetr, P., Thongpang, S.: Sign language-thai alphabet conversion based on electromyogram (EMG). In Proceedings Biomedical Engineering International Conference (BMEiCON), 2014 7th, IEEE, 1–4 (2014)
9. Bedregal, B.C., Costa, A.C., Dimuro, G.P.: Fuzzy rule-based hand gesture recognition. In Proceedings IFIP International Conference on Artificial Intelligence in Theory and Practice, Springer, Boston, MA., 285–294 (2006)
10. Abraira, V.: El índice kappa. Semergen: revista española de medicina de familia, (5), 247–249 (2001)

ISSN 1870-4069

Self-regulation from Emotive Feedback a Catalyst for Creative Task Appraisals in Design Education

T. Prasad-Mukhopadhyay, J. Ruiz-Pinales, R. Guzmán-Cabrera

Universidad de Guanajuato campus Irapuato-Salamanca,
DICIS, Guanajuato, Mexico

Abstract. We examined self-reports on performance of a group with untutored, unconstrained assignments (without external feedback) and contrasted it with 'control' performers in tutored sessions (with external feedback) to check if emotive self-regulation, based on participants' auto-feedback perceptions generate a better product trajectory. Does positive self-feedback work well and influence creative strategy in a positive way, just as proponents of 'emotional intelligence' advocate? Thus, could emotive self-feedback either reinforce or supplant conventional tutoring with external feedback signals in design schools. A brainwave entrainment was carried out to measure energy output and compare with self-reports during given intervals of the assignment. Results indicate presence of a reliable self-regulatory mechanism in which emotive-cognitive reflexes produces sufficient feedforward for task appraisal.

Keywords. Self regulation, emotive feedback, creative task.

1 Introduction

Research on self-feedback and regulation has generally derived from two directions: first in line with precepts of motivational studies of Rogers and Freiberg [1] and Csikszentmihalyi [2]. For Rogers and Freiberg, as well as Cszikszentmihalyi, the creative process uniquely elicits heightened attention and flow experiences – which together endorse the effects of emotive self-regulation [3, 4] that we are studying here, namely the psychological process of feedback and its intimate relation to self-regulatory resolutions of task.

Any design act will probably require emotional satisfaction as much as a recognition of successful implementation of techniques [5,6]. Subliminal (or intrinsic) feedback is shown also to act like a psychological rudder [7]. This feedback system was demonstrated in terms of feedback for individual learners that specifically promotes "self-regulation'' [7, 8]. A combination of emotional feedback and 'self-regulation' may generate sufficient impetus for problem-solving moves and for finishing the task at hand [9].

The general relation between emotions and creativity have to be recognized before defining the role of self-induced motivations by activation and harnessing the creative person's mood states. Self-induced emotive regulation of performance, especially in the context of creativity, has been discussed in terms of intrinsic versus environmental factors in development of artistic abilities [10]. It is believed that the creative person has inner strengths and potential of creativity. For some investigators a basically causal paradigm which relates emotive satisfaction to performance explains creativity [11]. This explains the kind of achievement-oriented behavioralism that evidently guides independently working artists [12].

Although Freud and later theorists like John-Stern [12] explained creative output more in terms of unconscious drives, it is now understood that such drives could be quantified. Following Stern's foundational thoughts on dynamic achievement consciousness as a guide for realizing creative goals, Renzulli suggested that multiple behavioral factors caused creative output and full productive lives for individuals [13].

Creativity has since by explained in terms of motivation and self-reflexive emotional control that suggested a further modification of Renzulli's dynamic causality thesis for a more multi-factor behavioral model which most importantly included a recognition of the importance and function of emotions and mood congruent situations for creative output.

In this paper, feedback experiences are compared for two groups of participants, 'experimental' and 'control' with the purpose of understanding how emotional appraisals of one's performances causes direct behavioral changes in creative people, especially media artists. Research has shown how emotion is known to impact behavior in both direct and mediated ways: but especially by retrospective appraisal of actions and by stimulating further actions in a conscious manner. Given the fact that automatic and controlled behaviors [14] are both impacted by emotive feedback, or specifically mood appraisals, we chose to identify if self-induced mood appraisals are sufficient for maintaining or enhancing (or diminishing) creative performance, especially if the feedback was not derived externally.

In classical terms this is also known as the 'nature' versus the 'nurture' debate [14], the search for the fact if creativity and innovation could be more a matter of self-recognition and 'nature'. The proposition has given birth to what Wei-Chen and colleagues [14] identifies as 'achievement-oriented' theories of creative behavior. On the whole the debate on intrinsic abilities may be studied in this classical context. At least some influential thinkers like Langer and Terman have emphasized on creative intelligence and IQ [11]. Facts show that individual creativity may be enhanced if the right environmental factors are available [15]. Giftedness may be a matter of potential but on the whole self-induced mechanisms of acquisition and realization of creative performances need to be discovered and studied [16]. What precisely are the results of an affective self-feedback mechanism that inspires creative people? How could self-assessment, conscious or otherwise, lead to increased levels of performance and satisfaction in creative performers?

2 Related Work

Yet, affect and emotions are always involved and integrated in the task – like a simultaneous circuit at work. Whether visual simulation (as also audio) depends *solely* on learning by means of prototypes or model emulation as mentioned above [17], or *also* on mood-congruent situations [18], is of the essence in this paper, although a larger number of studies already indicate in favor of recognizing the latter [19]. It has been observed generally that positive emotions benefit learning processes and outcomes [20].

Negative emotions such as anxiety have been interrelated with a lack of or discontinued motivation [21] and positive emotions such as pleasure and self-esteem were observed with high level of motivation [18]. There is an overt need of feeling an impulse, as it guides the designer to imagine and represent newer forms and entities. Therefore, in creative tasks, as opposed to logical reasoning, feeling and motivation, and consequently the sense of uninhibited freedom - or freedom from constraints - is a key factor. Students of arts and design, like their counterparts in other creative thinking processes like conceptual learning, also need a broader positive energy to perform better and achieve goals [18, 22]. In more market-situated contexts apprentices develop their own skills by monitoring what they are learning and by comparing it to existing models that offer aural or visual precepts [23]. Van Moer proposes so far as to evolve an experience-based visual arts learning process, like John Dewey's classic thesis that learning art should ideally depend on individual experience, rather than mandatory advisory intervention [24]. There is a growing consensus on the role of 'self-regulated' learning for students [25], which lays emphasis on the way budding artists are inspired on their own to learn and improvise [23]. There are contrary studies highlighting the facilitatory effect of advisor interactions on students' learning. But a changing digital culture scenario implies more openness toward disinhibited factors of learning. Self-regulated design learning could be related to a student's inner experiences and may constitute a pillar of educative processes. Consider the ideal scenario for a digital arts student. Any learner needs to be free, driven (or motivated), and creative but not uninformed or misguided. As one tries to evolve a certain design one gets to be receptive to pre-existing patterns and images. But that is not all. A learner's resolve is also bound in by her 'feelings', which include subliminal (intrinsic) resources of motivation. Informal freedom of spirit and self-initiated motivation should thus be a key ingredient for this course of simulations, as has been claimed [22]. Again, studies have indicated how positively-valenced emotion facilitates conceptual learning [19]. Also, how "emotional intelligence'' facilitates learning has been already studied [26]. In other cases, researchers have identified neuro-affective integration as the key factor for effective, wholistic learning [27]. But detailed analysis on how emotions may motivate students of digital design is therefore studied here. In digital art emotion may be assumed to play a very crucial role as both design and multimedia expressions employ things like imagination, feelings and desirability [28].

3 Methodology

In non-art contexts this is known as the development of "self-efficacy" in education, now for a long time [22]. We however wished to study the probable success of self-efficacy in context of self-regulated design, in untutored or minimally interventional training scenarios. This is because we wanted to find out if such contexts of creative expression eliminate external feedback and allows participants to rely on their own intuitions and sense of motivation associated with it. Innovation and performance are priorities for any design student. But how could student-interns perform well in market-oriented economies where demand and collective preference compromises individual innovations? It has been suggested that freedom from residual classroom pedagogies and use of game like interactive strategies could contribute to efficient, better and more enjoyable art education. In the experiment conducted here we tried to analyze this for the digital arts education scenario.

The experiment was based on the flowchart, see Figure 1, for a more self-motivated design education of the kind promoted by the more fundamental studies. The fact that emotive intelligence is involved in design strategies was thus adopted as a requisite for understanding the role of emotive feedback on creative practices.

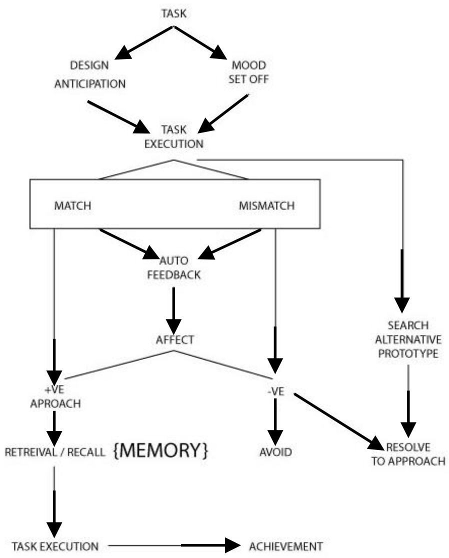


Fig. 1. Flowchart for a dualistic affective cognitive design learning activity based on pre-existing research on creative processes and neuroscientific evidences of the role of emotions.

The experiment was conducted with aspiring artists, mostly students in digital arts and disciplines related to design (n = 20). Participants were in the age range of 19 to 23, and all had prior experience in digital arts creation. Participants were beginner-designers and were selected on the basis of knowing basic software use skills. Also, digital arts curricula generally include learning or doing art projects on a variety software. Secondly, students were selected for a variety of software preferences.

Participants were invited to draw, or photoshop images or illustrations of their choice. Brainwave BCI equipment was fitted to their heads to accomplish the given task. Whereas participants in the Experimental group were free to act on their own mood feedbacks, generated in the creative process, control participants were subjected to normal teaching scenarios with feedback coming from a tutor or other expert in the field, thus conditioning their own sense of emotive satisfaction with task appraisals during the period of the experiment.

The first participants (n = 9) were tested with the EPOC-EEG hardware. The rest (11), were maintained as “control’ and allowed to proceed only with external feedback or instructions as would be appropriate in normalized scenarios of design pedagogy. A simultaneous brainwave graph was recorded during the performance of said tasks. Brainwave activity for the same participants were checked against findings based on conscious semantic responses to the level of ‘emotive satisfaction’ and the sense of regulatory ‘ability’ experienced by participants.

To obtain cognitive data, a Brain-Computer Interface (BCI) device was used to capture brain signals during performance of tasks. The Emotiv-Epoc headset (Figure 4-a) has 14 EEG channels (electrodes). The distribution of sensors in the headset is based on the international 10-20 electrode placement system with two sensors as reference for proper placements on the head. Data mainly from 4 channels were used: AF3, F7, FC6 and AF4 (Figure 4-b). These 4 channels were chosen on the basis of their recognized association with high concentration and activity for creative or problem-solving tasks.

Electrical brain signals are obtained through noninvasive measures. Signal acquisition methods are mainly employed to observe spontaneous brain activity based on the electrical activity of signals.

Fig. 2. (a).

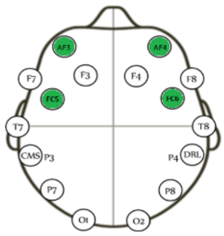


Fig. 2. (b).

Fig. 2. Emotiv-Epoc: (a) headset; (b) headset electrode arrangement.

Once the signals are captured, they are filtered with the purpose of separating them into beta and gamma.

When signals are acquired, generally, they are contaminated by noise and artifacts. Several techniques can be employed to remove noise and artifacts and identify the true signal. Feature extraction is done after noise is removed from the raw signal.

Feature extraction techniques emphasize essential characteristics of the signal and its relation to biomedical events. In our case we used the Empirical Mode Decomposition (EMD). EMD is proposed as the fundamental principle of the Hilbert–Huang transform (HHT). The Hilbert Huang transform is carried out, so to speak, in 2 stages. First, we use the EMD algorithm: in the second stage the instantaneous frequency spectrum of the initial sequence is obtained by applying the Hilbert transform to the results of the previous step.

The HHT allows us to obtain the instantaneous frequency spectrum of nonlinear and non-stationary sequences. These sequences can consequently also be dealt with by using EMD. The signal obtained is not continuous in time. 7680 discrete values are recorded per minute at a sampling frequency of 128 Hz. The signal is broken down into 22 signals (modes), which added together allow us to recover the original signal. This decomposition helps us calculate the energy in each time interval.

To separate the signals in the ranges of beta and gamma waves, the Hilbert frequencies are separated on the following basis: Beta 12 to 39 HZ, with an optimal ± 16 Hz ideal range for creativity [3]. Gamma Range is considered to manifest over 40 Hz [3]. The energies are added in each mode for beta and gamma in the corresponding time interval. Simultaneous presence of higher beta and gamma should indicate more emotive satisfaction (arousal) and creative flow.

Higher energy output should be consistent with stress and concentration factors, higher gamma with more harmony. Supragamma levels of ±70 indicate emotive finésse in task execution, commonly associated with artistic competence. The beta-gamma distribution for these channels are more significant in so far as they reflect typical creative states.

4 Results and Discussion

In order to have an idea of how these signals are visible and therefore available for interpretation (as in Figure 3), the spectrogram obtained for a participant corresponding to the control group is shown. In this case it is for the band width of the beta signal, which ranges from 12 to 40 Hz ($12 \leq \beta \leq 40$ Hz). Yellow indicates very strong activity, as a co-efficient of power for the frequency. Stronger yellow bands in the 10-20 Hz range indicates ideally high creative behavior, the optimal range being 16Hz as is indicated in the literature.

To obtain the signals in the beta and gamma range from the signals caught on respective pre-frontal sensors of the Brainwave BCI the EMD (empirical mode decomposition) method of extraction is applied and the value of the frequency then obtained from the Hilbert-Huang transform. The result of this process is shown in Figures 3 and 4. The Hilbert frequencies are separated using a β band pass filter and a high pass filter for γ.

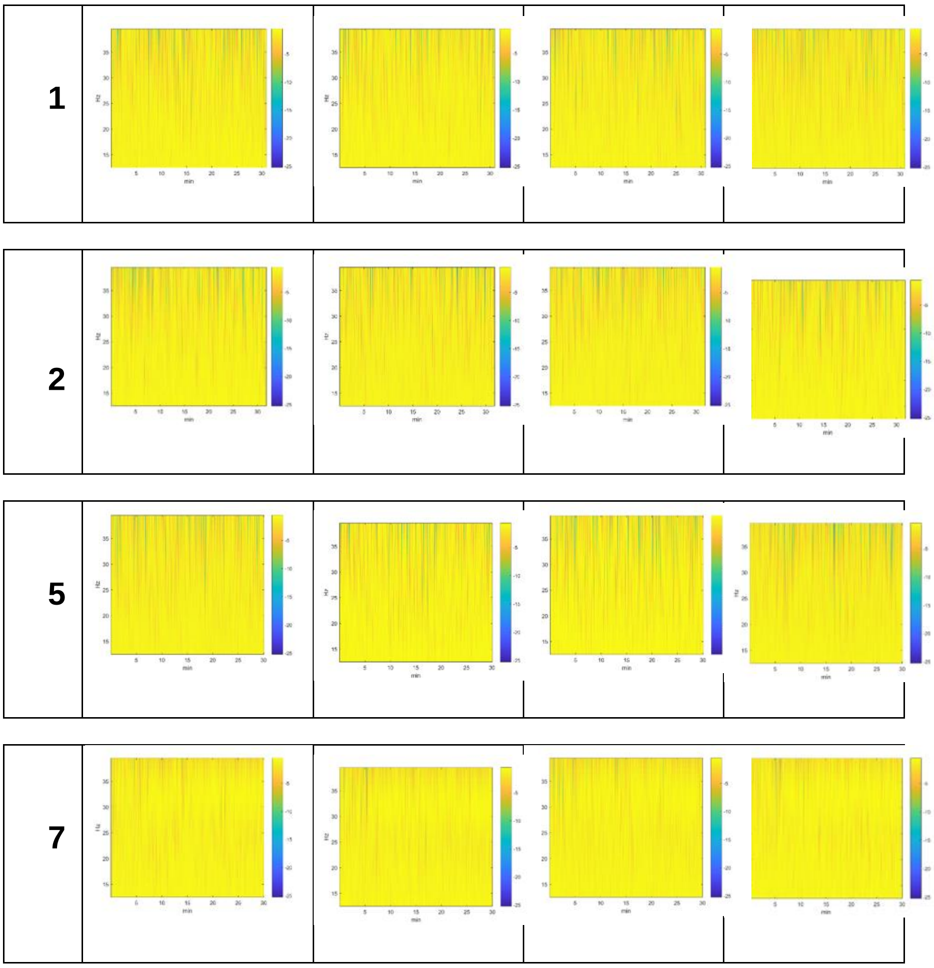


Fig. 3. Beta range signals in sample participant in the Experimental (self-motivated feedback) group. Notice that at least 4 participants (1, 2, 5, 7) demonstrate very strong, i.e. yellow bands in the ±16 Hz range, or beta wave activity for concentration and innovation.

Figure 4 shows, for the same selected participant, the spectrogram corresponding to the gamma signal ($40 \leq \gamma \leq 90$ Hz).

We can see in Figures 3 and 4 a significant variation in color. A higher value in frequency represents greater emotional activity: so, for example 90 Hz is more emotionally proactive with task appraisal scenarios. Figure 4 shows that in the same group of participants as in Figure 3 above there are yellow bands, less prominent, but definitely indicative for a very high frequency close to gamma range frequency. The higher, closer to gamma range indicators, though less regular in the spectrogram, represent achievable peaks in self-motivated participants.

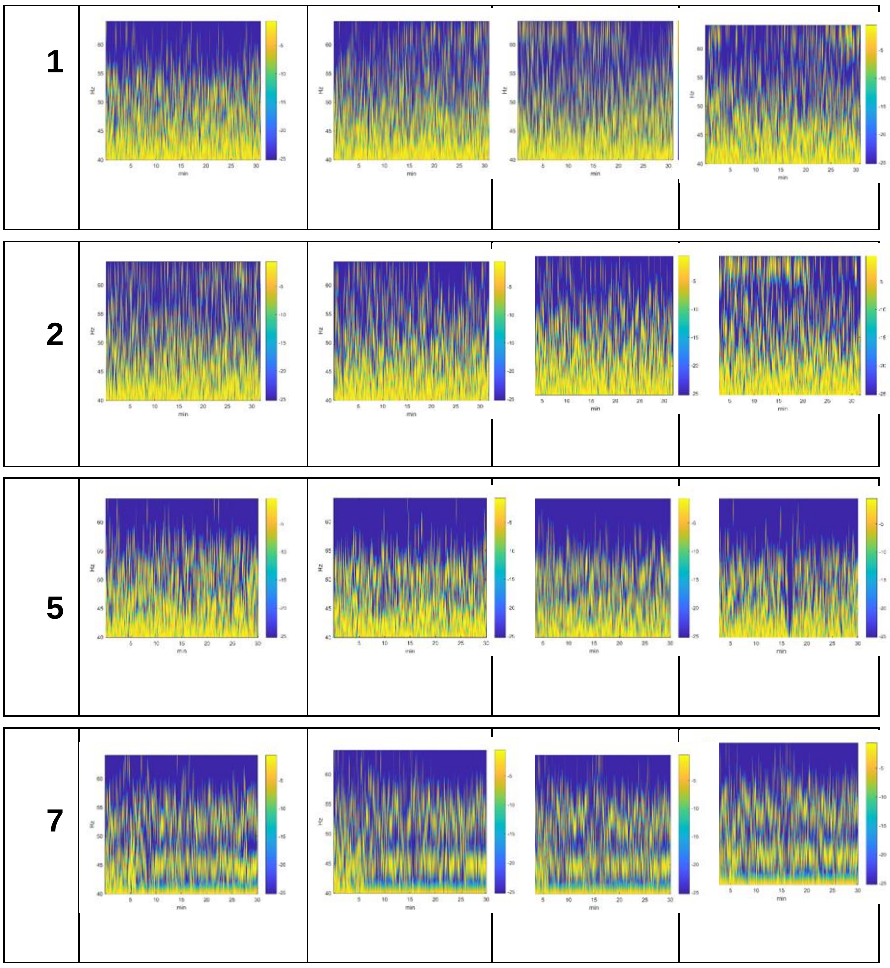


Fig. 4. Gamma signal spectrography for sample participant in same experimental group as in Figure 3 above.

Tables 1 and 2 show the sums of the energies grouped in periods of 5 minutes. The summation helps in analysis of the activity of each participant in the specified intervals during the trajectory of the experiment.

The presence of a possibly normative self-motivational feedback, deriving from successful ‘emotive feedback’ and ‘self-regulation’ may be seen reflected in the Fourier spectrogram analysis of at least 44% of participants in the experimental group as contrasted to only 36 % of participants in the control or external feedback group (see Figure 5 and 6). A sum of the average frequencies recorded for each candidate in each category was considered for statistical analysis. A distinctly high beta wave peak in unconscious Epoc-EEG brainwave BCI was evidently found to be the trend in both experimental and control group participants.

Table 1. Average energy (in picojoules) for all four individual channels usually considered for measuring creativity in participants of experimental group.

EXPERIMENT (Without Training)															
				Beta (AF4+F7+AF3+AF6) (pJ)						Gama (AF4+F7+AF3+AF6) (pJ)					
N	Age	G	Ca	5 min	10 min	15 min	20 min	25 min	30 min	5 min	10 min	15 min	20 min	25 min	30 min
1	21	M	DA	4.1	3.3	21.3	2.8	4.6	2.7	0.6	0.5	0.7	0.7	0.6	0.8
2	22	M	DA	30	1.2	1.6	3.5	6.4	2.7	1.5	0.6	0.7	1.1	1.4	0.6
3	23	F	DA	2.3	50.1	12.2	10.5	6.1	27.1	0.2	18.7	0.4	0.3	0.2	0.6
4	20	F	DA	10.4	28	33.2	23	16	18	0.3	1.1	1.3	0.6	0.6	0.6
5	22	F	M	7.3	11.2	12.1	57.9	34.2	28	0.3	0.5	0.5	0.6	0.3	0.5
6	25	F	DA	9.2	16	18	11	19	9.3	9.3	9.7	9.2	1.8	18.	7.9
7	24	F	DA	33.2	28.7	87.4	87.9	121.4	170.8	5.8	6.4	13.9	15.5	23.5	25

DA: Digital Arts, M: Mechatronics

Table 2. Average energy for all four individual channels usually considered for measuring creativity in participants of control group.

CONTROL (With Training)															
				Beta (AF4+F7+AF3+AF6) (pJ)						Gama (AF4+F7+AF3+AF6) (pJ)					
	Age	G	Ca	5	10	15	20	25	30	5	10	15	20	25	30
	24	F	DA	257.4	31.4	73.5	23.3	48.3	21.4	27.1	10.1	9.7	6	2.2	1.5
	29	M	DA	61.8	14.9	12.4	52.8	14.5	18.8	10.4	6	5.3	4.8	5.3	6.4
3	21	M	DA	22.9	2.8	3	3.6	25	23	0.5	0.2	0.3	0.3	0.3	0.3
	25	F	DA	16.4	275.8	3.1	51.9	101	2.5	1.4	33.2	1	1.7	1.3	0.6
	23	F	DA	9.0	10.7	9.5	13.4	10.1	11.4	0.3	0.3	0.2	0.2	0.2	0.2
	22	M	DA	51.8	4.6	68	4	2.8	9.4	11.7	0.6	17.7	0.5	0.4	0.4
	22	F	DA	291.3	126.3	39	192.4	36.1	16.8	27.3	17.4	9.2	5.9	7.5	2.3
	23	F	DA	6.8	8.3	10	9.2	9.2	8	0.2	0.2	0.3	0.3	0.3	0.3
0	21	M	DA	3.6	4.6	3.9	3.3	4.2	4.7	0.6	0.5	0.4	04	0.5	0.8
1	26	F	EE	14.6	17.6	14.8	15.3	32.1	21.6	0.6	0.9	0.6	0.7	2.3	1.1

DA: Digital Arts, EE: Electronics Engineering

The frequency data was adopted for both experimental and control groups involved in self-regulatory feedback and external tutored feedback for creative trajectories respectively. Of all channels AF4, AF3 F6 and F7 were considered for focus and creative, i.e. problem-solving task appraisals (Jansen 2005; Herrington et al 2005; Gruzlier 2009). Beta radiation studied for pre-frontal problem solving areas indicate activation and level of focus or absorption of participants in their work.

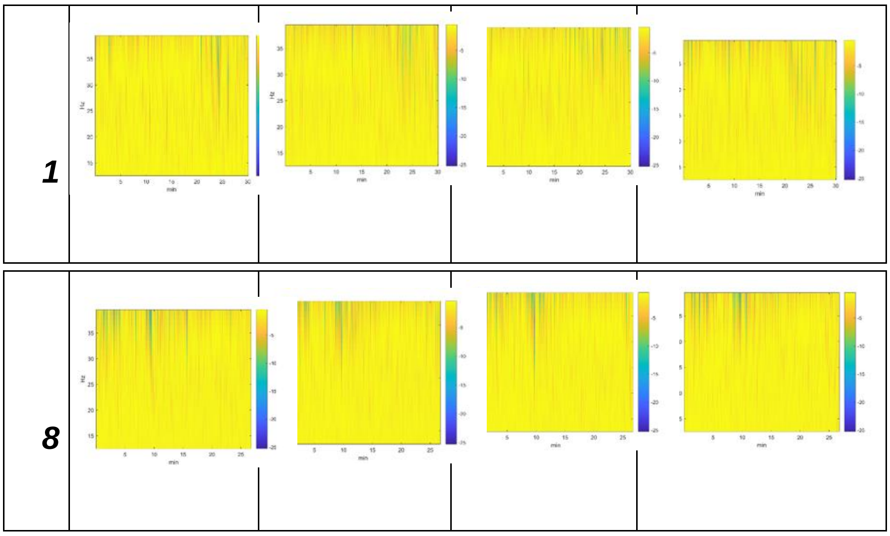


Fig. 5. High yellow band peaks at the ±16 Hz beta frequency range is also visible for Control participants who work on the basis of external feedback. But the number of participants with decidedly thick yellow band (i.e. high energy) output is less than in Experimental group, and also less evenly distributed.

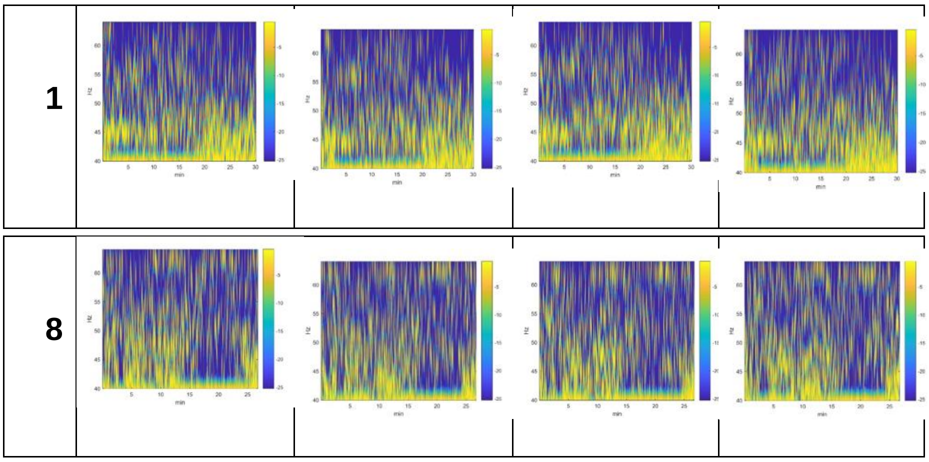


Fig. 6. High yellow band peaks at the ±90 Hz Gamma frequency range is also visible for Control participants who work on the basis of external feedback. But the number of participants, as in Figure 5 above is less than the number of participants in the experimental group and also less evenly distributed.

Of the 44% of higher Beta and Gamma activity as recorded for the self-regulated experimental group participants 1, 2, 7 in the experimental group demonstrate high Gamma wave activity in all PFC terminals, notably AF4, F6 and F7 in the total 30-minute segment. (Figure 3). Participant 2 reflects exclusively high peaks for AF4 and

F7. Participant 5 demonstrates more sporadic creative peaks for both Beta and Gamma.

5 Limitation and Conclusion

On the basis that behavior of the control and experimental groups is the same (and assuming that high energy beta and gamma values would generally reflect the same values in contexts of self-regulatory and external feedback) we observe that performance indicators for brainwave have no significant difference as far as brain activity is concerned (p-value > 0.05).

Contrarily positive self-feedback could indicate towards better flow results for dedicated artists working on self-instruction reliance. There are preponderant high peak outputs, indicative of tranquility and flow for self-motivated and self-regulated creativity, though this may not be a universal rule in creative behavior. Design learning is definitely fostered by self–induced positive emotions.

We know how feedback strategies influence creative and meditative practices (Peper 2012). In this article however we have tried mainly to focus on feedback generated by learner-artists whose work requires imagination, freedom and emotional happiness or satisfaction as distinguished from less emotionally dependent tasks like mathematical problem-solving. Comparing the brainwave functions for the same time-intervals we tend to observe two things. First, that minimally interventional, self-induced motivation provides good impetus for creativity in digital art contexts and is not significantly different from scenarios in which pedagogy, collective inputs, external feedback and limitations imposed from outside are included in the practice. We also observed that instructions and extraneous interventions do not necessarily create a more stable maintenance curve for 'control' participants. The fact that we could have self-sufficient intuitive feeling, especially for highly innovative tasks like creative design, makes us reconsider the necessity of having a free and self-reliant educational environment. The social consequences of implanting such an environment may be desirable in many cases, especially where tutorial intervention is not in emotional and imaginative sync with the performer. In the digital age young learners are often more tech-savvy and advanced in their exposure to a wide variety of media. Assessing the important role played by self-motivated actions in our learning and executive experiences are important and may show us a path out of traditional classroom spaces.

References

1. Rogers, C. R., Freiberg, H. J.: Freedom to learn (1994)
2. Csikszentmihalyi, M.: Flow and the psychology of discovery and invention. New York: Harper Collins (1996)
3. Runco, M. A., Plucker, J. A., Lim, W.: Development and psychometric integrity of a measure of ideational behavior. Creativity Research Journal, 13(3–4), 393–400 (2001)

4. Heo, M. S., Cheon, M. J.: An empirical study on the relationship between role stress and personal creativity: The mediating roles of creative self-efficacy and personal initiative. The Journal of Information Systems, 22(2), 51–83 (2013)
5. Zimmerman, B. J., Kitsantas, A.: Developmental phases in self-regulation: Shifting from process goals to outcome goals. Journal of educational psychology, 89(1), 29 (1997)
6. Hattie, J., Timperley, H.: The power of feedback. Review of educational research, 77(1), 81–112 (2007)
7. Winner, E.: Development in the Arts: Drawing and Music. Handbook of Child Psychology. DOI: 10.1002/9780470147658.chpsy0220 (2007)
8. Russ, S. W.: Affect and creativity: The role of affect and play in the creative process. Psychology Press (1993)
9. Al-kreimeen, R. A.: The Relationship between Individual Creativity and Self-Regulation-From Grade Nine Students Viewpoints in Jordan. International Proceedings of Economics Development and Research, 78, 85 (2014)
10. Heller, K. A., Mönks, F. J., Subotnik, R., Sternberg, R. J. (Eds.).: International handbook of giftedness and talent. Elsevier (2000)
11. Terman, L. M.: The discovery and encouragement of exceptional talent. American psychologist, 9(6), 221 (1954)
12. Stern, S.: The relationship between human resource development and corporate creativity in Japan. Human Resource Development Quarterly, 3(3), 215–234 (1992)
13. Renzulli, J. S.: What is this thing called giftedness, and how do we develop it? A twenty five year perspective. Journal for the Education of the Gifted, 23(1), 3–54 (1999)
14. Wei-Chen, H., Smith, T. J., Smith, M. C.: Design and usability assessment of a dialogue-based cognitive tutoring system to model expert problem solving in research design. British Journal of Educational Technology, 46(1), 82–97 (2015)
15. Gardner, H., Gardner, E.: Art, mind, and brain: A cognitive approach to creativity. Basic Books (2008)
16. Runco, M. A., Yoruk, S.: The neuroscience of divergent thinking. Activitas Nervosa Superior, 56(1-2), 1–16 (2014)
17. Stavridou, F., Kakana, D.: Graphic abilities in relation to mathematical and scientific ability in adolescents. Educational Research,50(1), 75–93 (2008)
18. Kim, C., Pekrun, R.: Emotions and motivation in learning and performance. In Handbook of research on educational communications and technology. pp. 65–75, Springer New York (2014)
19. Endrissata, N., Islamb, G., Noppeneya, C.: Visual organizing: Balancing coordination and creative freedom via mood boards. Journal of Business Research. DOI:10.1016/j.jbusres.2015.10.004
20. Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., Perry, R. P.: Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). Contemporary Educational Psychology, 36(1), 36–48 (2011)
21. Artino, A. R., Stephens, J. M.: Academic motivation and self-regulation: A comparative analysis of undergraduate and graduate students learning online. The Internet and Higher Education, 12(3), 146–151 (2009)
22. Zimmerman, B. J.: Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. American educational research journal, 45(1), 166–183 (2008)
23. Ings, W.: Malleable Thought: The Role of Craft Thinking in Practice-Led Graphic Design, International Journal of Art & Design Education, Volume 34, Issue 2, 180–191, DOI: 10.1111/jade.12013 (2015)
24. Van Moer, E., De Mette, T., Elias, W.: From Obstacle to Growth Dewey's Legacy of Experience-Based Art Education. International Journal of Art & Design Education, 27: 43–52, doi: 10.1111/j.1476-8070.2008.00556.x (2008)

25. Haanstra, F.: Self-Initiated Art Work and School Art. International Journal of Art & Design Education, 29: 271–282. doi: 10.1111/j.1476-8070.2010.01662.x (2010)
26. Spendlove, D.: A Conceptualisation of Emotion within Art and Design Education: A Creative, Learning and Product-Orientated Triadic Schema. International Journal of Art & Design Education, Volume 26, Issue 2, 155–166. DOI: 10.1111/j.1476-8070.2007.00525.x (2007)
27. Clarken, R. H.: Wholistic education: Toward a definition and description. In Annual Meeting of the American Educational Research Association, San Francisco (2006)
28. Silvia, P. J.: Cognitive appraisals and interest in visual art: Exploring an appraisal theory of aesthetic emotions. Empirical studies of the arts, 23(2), 119–133 (2005)

ISSN 1870-4069

A Web-based Tagger for Named Entities Detection

David Muñoz*, Fernando Pérez-Téllez, David Pinto

Computing Department, Institute of Technology Tallaght, Dublin, Ireland

{David.Efrain@postgrad.ittdublin.ie, fernandopt@gmail.com, davideduardopinto@gmail.com}

Abstract. Information Retrieval constitutes an important task for extracting meaningful data in huge amounts of texts since the majority of information in the world exist electronically, and in unstructured and plain text form. Most data-mining applications assimilate only structured information, this means before any predictive model can be applied the data must be prepared in a special way. This paper presents a web-based tagger system for named entities detection in plain text to help users tagging texts resources. The advantage of using this tool is the possibility of tagging any plain text file and with any type of class for the entities, allowing to treat any domain and thus obtaining a wider recognition of named entities. In addition, a small application shows the facility in the training of NER classifier models as an advantage of corpora constructed with the web-based tagger.

Keywords. Named entity recognition, NER, collaborative tagging, named entities, information extraction, information retrieval.

1 Introduction

The increasing use of computers to generate and administrate information causes the information overload. In this sense, Information Retrieval (IR) is the topic commonly associated with online documents. The IR goal is to measure the similarity between an input document and a document collection, finding the best matches in how similar are the results with the input document. In this way, Named Entities (NE) recognition represents an important aspect in the process of natural language understanding, and their effective identification comes to be significant in Information Retrieval related tasks.

The majority of information exists in unstructured and textual data, however for machines is hard to understand the unstructured data and that is why before any learning method -such as predictive models- can be applied the data must be prepared in a special format.

In the task of Named Entities recognition (NER) The Stanford Natural Language Processing Group provides a software implementation in the training of classifier models. However, the training data needs to be structured in a special way of two tabbed

*Corresponding author. Tel.: +52 1 22 26 26 82; E-mail: devasc26@gmail.com.

columns, in which the first column contains the word and the second column contains the tag assigned to the word. Taking into consideration the conversion of unstructured data into structured data is how the proposed system allows constructing corpora in this specific format for any type of plain text files.

In the last years, the task of extracting meaningful data of text has gained the attention in researcher and industry fields. Collaborative tagging systems have emerged as a solution for avoiding noise on information content since every day there is a lot of information generated and the majority is textual data [15].

The problem of highlighting relevant information over the information overload has gained the attention of many researchers, and as an answer, there are some works presenting alternatives as a solution.

Since information extraction is considered as a limited form of full natural language understanding, where the information we are looking for is known beforehand. It includes two fundamental tasks, named entity recognition and relation extraction [10].

In the task of universal semantic tagging, Abzianidze and Bos [8] contributes to better semantic analysis for wide-coverage multilingual text. The authors said that, besides their application in semantic parsing demonstrated in the PMB project, sem-tags can contribute to other NLP tasks, e.g. POS tagging, or research lines rooted in compositional semantics. In their work, the authors have shown that the tags provide semantically fine-grained information, and they are suitable for cross-lingual semantic parsing.

In the task of Named Entity Recognition (NER) in Tweets, Ritter et al [2] proposed a distantly supervised approach which applies Labeled LDA to leverage large amounts of unlabeled data in addition to large dictionaries of entities gathered from Freebase, and combining information about an entity's context across its mentions. This because classifying named entities in tweets is a difficult task since tweets contain a plethora of distinctive named entity types (Companies, Products, Bands, Movies and more), and almost all these types are relatively infrequent. On the other hand, tweets often lack sufficient context to determine an entity's type without the aid of background knowledge.

In the task of named entity classification, Mohamed and Oussalah [11] presented an approach by using the Wikipedia article info boxes where it has significantly reduced the classifier's processing time since the information inside the info box is structured. The proposed approach achieved a classification accuracy of above 97% with 3600 named entities and CoNLL-2003 shared task NER dataset used to validate the classifier's performance.

Font et al [5] proposed a general scheme for building a folksonomy-based tag recommendation system to help users tagging online content resources. They achieved this by using 3 independent steps: 1) Getting candidate tags, selecting a number of candidate tags for every input tag based on a tag-tag similarity matrix derived from a folksonomy, then 2) Aggregating candidate tags, assigning scores to the candidates of step 1 and merging them all in a single list of candidates tags, and finally 3) Selecting which tags to recommend, automatically selecting the candidates that will be part of the final recommendation by determining a threshold and filtering out those candidates whose score is below the threshold.

In the task of recommending items, the classification and prediction have an important role by analyzing data. It is important to know what classification algorithm to use depending on the application to be developed. As noted by Sheshasaayee and Thailambal, Classification is in supervised Learning of Machine Learning where a set of correctly predicted observation is available [3].

Chavaltada et al proposed a framework for automatic product categorization and explained that each classification method is affected in the efficiency of the model since each method have different the parameters [4].

In the task of text mining, Hawizy et al. [9], proposed a tool for semantic text-mining in chemistry where is possible to extract structured scientific data from unstructured scientific literature such as scientific papers or thesis which uses free-flowing natural language combined with domain-specific terminology and numeric phrases. In their work, they also demonstrated that using text mining and natural language processing tools, it is possible to extract both chemical entities and the relationships between those entities, and then making the resulting data available in a machine processable format.

The rest of this paper is organized as follows. First, the fundamentals and techniques employed are explained. Then, the proposed system (web-based tagger) and its architecture are explained. After, the case study of a small demonstration where the Stanford format is used to train the classifier. Finally, the critical evaluation and conclusion are presented.

2 Fundamentals and Techniques

This section introduces the named entities recognition in the process of natural language understanding as well as the use of collaborative systems. It also explains the V-fold cross-validation technique as a method for test the NER model´s ability to predict new named entities in datasets of unknown data.

2.1 Named Entities

NER is an important task in the field of information extraction systems since it aims to locate and classify named entities in raw text into categories previously defined (e.g. Person, Location, Organization). In this way, texts can be represented by their named entities. It is emerged in the Sixth Message Understanding Conference in 1995 [11]. Although, sometimes many of the NEs can be ambiguous to be classified in more than one class, e. g. the automotive company created by Henry Ford in 1903, where “Ford” can be referred to many entities (Name, Company, etc.).

On the other hand, NER systems require a large amount of highly accurate training data to perform well at the task named entities recognition [17]. In this way, excellent training data can be achieved by human feedback, since humans can easily differentiate from one context and another, assigning the correct tag to each named entity in the texts.

2.2 Collaborative Tagging

Collaborative tagging consists of assigning tags by users to a set of information resources such as plain text files. After that, tags can be used for many purposes, including retrieval, browsing, and categorization [7].

Halpin et al [6], claimed that there are three main entities in any tagging system: users, items, and tags. They produced a generative model of collaborative tagging in order to understand the basic dynamics behind tagging. So, they showed how tag co-occurrence networks for a sample domain of tags can be used to analyze the meaning of particular tags given their relationship to others tags. In this way, [1] proposed to model data from collaborative tagging systems with three-mode tensors, in order to capture the three-way correlations between users, tags, and items. He said that by applying multiway analysis, latent correlations are revealed, which help to improve the quality recommendations. He also developed a hybrid scheme that additionally considers content-based information that is extracted from items.

2.3 Cross-validation

Cross-validation methods are widely used in machine learning in order to estimate how accurately a predictive model will perform in practice. Cross-validation is based on the idea of data splitting, as in a prediction model usually is given a dataset of known data. So, part of the data is used for fitting each model, and the rest of the data is used to measure the performance of the model [13].

V-fold Cross-validation.
In practical applications, cross-validation is often computationally expensive. So, in order to achieve a smaller computational cost can be applied the method V-fold cross-validation, which depends on an integer parameter V.

Basically, the value of *V* should be small for complexity reasons, but not too small for decreasing the variability of the algorithm [14]. There are three well-known factors to take into account in order to choose *V*:

- *bias*. When *V* is too small, it leads to underfitting and suboptimal selection.
- *Variability*. The variance is a decreasing function of *V*.
- *Computational complexity*. *V*-fold cross-validation needs to compute at least *V* empirical risk minimizers for each model.

In the least-squares regression setting, *V* has to be chosen largely in order to improve accuracy (by reducing bias and variability); or it will generate computational issued arise when *V* is too big. This is why *V*=5 and *V*=10 are very classical and popular choices [13]. For this work, the value of *V* is 10 because the performance accomplished is much better and the computational power is low allowing to make fast experiments.

The explanation of the *V*-fold cross-validation method is illustrated in Figure 1, where the complete dataset is divided into the test data and the training data with the rest of the dataset. Each time with different elements until the final iteration is complete.

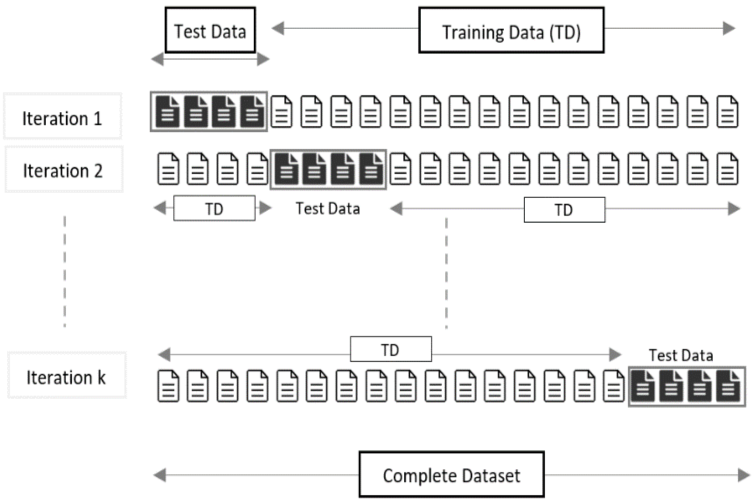


Fig. 1. An illustrative example of V-fold cross-validation.

3 Proposed System

The proposed web-based tagger aims to solve the problem of converting unstructured information to a structured format that can be easily understood by computers, the web-based tool receives plain text files as input and generates output files in a structured format ready to be used in the training of predictive models.

Figure 2 illustrates the role of the web-based tagger in the construction of high quality corpora in three main stages:

- In order to start the corpus construction process, it requires to provide an entry of plain text files to the system.
- In the second stage, the user begins to highlight the most important concepts in each file, assigning to each concept the tag that best identifies it. Since there are many domains in which is possible to tag concepts, the web-based tool allows the users to work in as many domains as necessary. Once all the relevant concepts in the domain have been tagged by the user, it is possible to export the tagged data in one of the three available formats: *Inline XML, keyword context, and tabbed column* format. When the *Inline XML* format is selected it simply find the indexes of each tagged concept enclosing between its corresponding assigned tags. If the *keyword context* format is selected then the concepts highlighted are extracted as well as the words before and after that concept. If the *tabbed column* format is selected then a tokenization step for each plain text file is made.
- After the information in the plain text document has been processed and formatted, it means that the document is now in a structured format and ready to be used in

machine learning algorithms. Finally, the structured data is available to be downloaded and used.

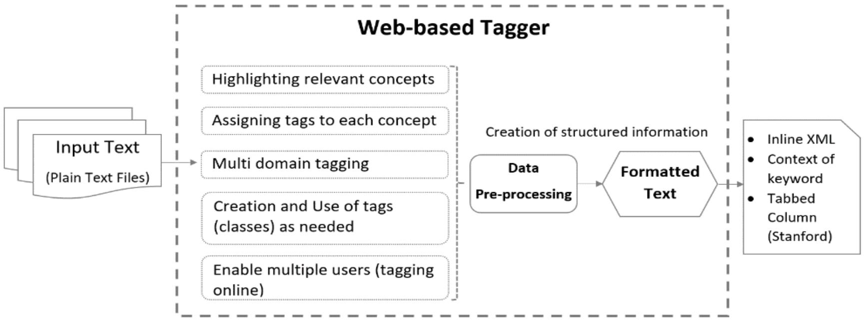


Fig. 2. The role of web-based tagger in the construction of high quality corpora.

The next points in this section describe the system architecture of the web-based tagger and explain more in detail the data output formats generated in order to train each model.

3.1 System Architecture

There are many reasons to have a tagger tool as a web application since it provides access to information from any device with internet connection. Also, it facilitates a huge number of users who can tag different plain text files simultaneously, improving the amount of tagged data.

Since the main task of tagging is to get tagged concepts for users, and in any of the domains the web-based tool needs to allow access to many users and from any place they are. Also, users should download the tagged files when they need it. On the other hand, it needs to facilitate the creation of new tags, since existing a lot of domains in which users can tag concepts.

Following these statements, the system architecture is illustrated in Figure 3. First, users must be authenticated in the Login Screen and depending on whether they are, administrators or taggers, the information, and permissions associated with the user are loaded.

If the authenticated user is an administrator, then its user interface consists of three main functionalities: administration of users, tags, and permissions. If the authenticated user is a tagger, then its user interface is composed of four main functionalities: upload plain text files and after selecting one of them, the user can begin to highlight the important entities until all important concepts have been tagged in the text. Once the named entities in the text have been tagged, the user can download the tagged text in any of the three formats available in the web-based tagger: Inline XML, Tabbed columns, and Keyword context.

The functionalities allowed to each user are explained in the next sections.

Users

As there is a big number of people tagging plain text files, it is necessary to authenticate the user session to know what data is allowed to tag. On the other hand, it facilitates access from any device, it means, once a plain text file was uploaded, the user can use it at any time, in any place, and in any device. There are two types of users:

- Tagger: Taggers are main generators of high quality tagged information since they have a good understanding in finding keywords in textual data. They are allowed to use a lot of different tags, so they are not limited to work in only one or few domains, as well as they are not limited in the number of files. Once they are logged in the system, they are allowed to: upload plain text files, search and select text files, then they can tag keywords in the text and download tagged files. Even, they can download tagged files that belong to other users, so they can use them to generate larger corpora.

  Something important is that each tagger can decide if he makes his file collaborative, this means to be tagged by other users. Thus, the quality of a tagged concept can increase substantially, considering that tags are assigned by different human feedbacks. This is achieved thanks to the system allows reassigning tags to the different concepts that have been tagged.

- The administrator: An administrator is allowed to create, modify and delete users, as well as administrate new tags and assigning tags to each tagger.

  So, tagger users can get tagged texts with high quality in tags and in any domain because there are no class restrictions.

Tags

Since the tags are created by the administrator to be used by taggers in the plain text files, a tag is defined as follows:

- Name. It denotes the class name, e.g. Person, Location, Organization, Skill and Knowledge.
- Id. It denotes a short name reference to the class name, e.g. Per, Loc, Org, Ski, Kno.
- Description. The description is a reference for tagger users, in order to explain what defines a certain class.
- Examples. The examples are provided in order to help users easily identify which concepts belong to a certain type of class.
- Color. The color is designed to highlight the text the concepts that are tagged by users in order to facilitate the reading and identification of each tagged concept.

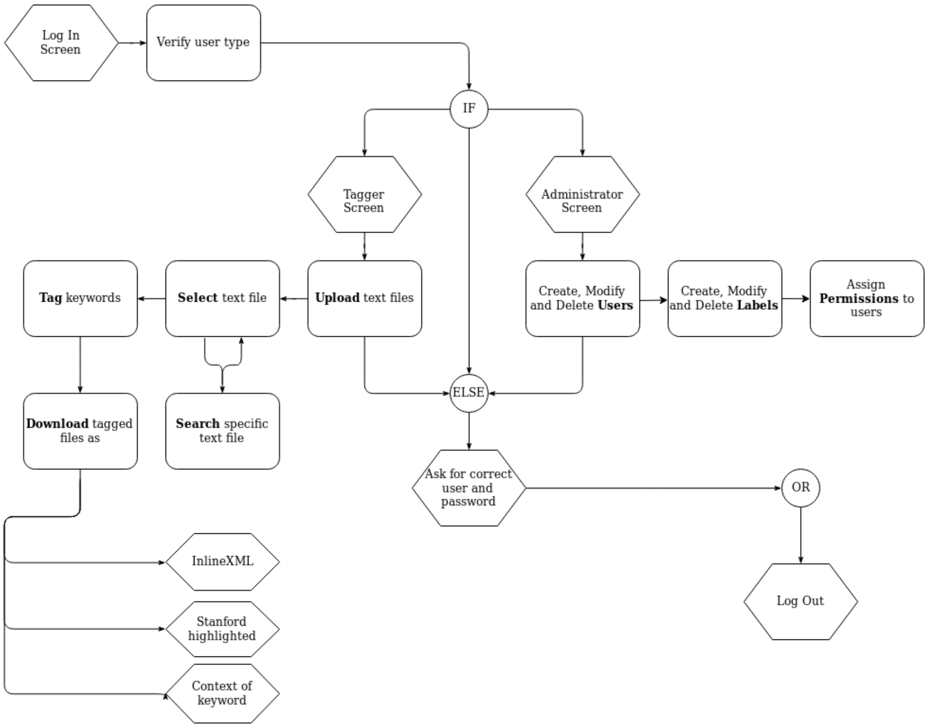


Fig. 3. The system architecture of web-based tagger.

Resources

The web-based tagger allows users to upload plain text files and immediately taggers can start to tag their files. When a file is uploaded a copy of it is generated to differentiate the file that is being displayed in real time from the one that will be tagged internally for later downloading. So, there are three types of formats in which a user can export a tagged text:

– Inline XML. It contains all the original text, with the exception that each item tagged is between the tags assigned to it. Since XML is a well-known and used format because it makes information easily accessible. An illustrative example showing the XML format is presented in Figure 4.

<Organization>Facebook**<Organization>** is an American online social media and social networking service company based in **<Location>**Menlo Park, California**<Location>**. And it was launched by **<Person>**Mark Zuckerberg**<Person>**.

Fig. 4. An example of an Inline XML format.

- Stanford - Tabbed columns. It contains the data prepared to be used in the training of a classifier model provided by Stanford NER [12]. The first column contains the words or tokens of the document and the second column represents the class to which it belongs if it belongs to one, in other cases the value of the second column is 0. An example of the format is presented in Table 1.
- Keyword context. It is a special format in which there are words that precede or proceed the tagged keyword since there are some keywords with different meanings depending in the context they are being used for. It is important to mention that the words that proceed and precede a tagged keyword, are only taken into account if these are not a punctuation marks since a punctuation mark indicates the beginning or end of another idea. An illustrative explanation the Keyword context format is shown in Figure 5.

Facebook is an American online social media and social networking service company based in **Menlo Park**, **California** . And it was launched by **Mark Zuckerberg**.

Fig. 5. An example of Keyword context format showing the importance of punctuation marks indicating the beginning or end of another idea.

Once the plain text files have been tagged any user can download the generated files, even to train immediately a classifier. This is important because in this way users can download datasets generated by other users (supervised training data) and experiment with this corpora as well as modify them if necessary.

The web-based tagger saves a lot of time since it makes the information accessible for everyone. As an advantage, the second format is ready to be used in NER Stanford training. At this point, the corpus can be constructed more quickly and even with more data available to train the classifier.

Table 1. Example of NER Stanford (Tabbed column) format avail-able to download in the web-based tagger.

Token	Class (label)
Facebook	Organization
Is	0
An	0
American	0
online	0
social	0
media	0
and	0
social	0
networking	0
service	0
company	0
based	0
in	0

Menlo	Location
Park	Location
,	0
California	Location
.	0
And	0
it	0
was	0
launched	0
by	0
Mark	Person
Zuckerberg	Person
.	0

3.2 Datasets

The number of examples used in this work is presented in Table 2. The table presents the number of job descriptions employed to train the NER Stanford Classifiers as well as the testing data, it also presents the number of classes and the number of examples per class. For this experiment, 50 different job descriptions related to the computing field were employed, and each one was tagged with at least one of the 3 classes: Knowledge, Skill, and Value.

There are a total of 18,766 tokens of which there exist 2,097 examples with the “Knowledge” tag, 956 examples with the “Skill” tag and 234 examples with the “Value” tag. The remaining 15,477 tokens have no category so in the second column the value for the label is 0.

All of this data is in the format of Stanford – Tabbed Columns provided by the same web-based tagger, and it consists in a tsv file containing the 50 job descriptions in a continuous text with the class labeled for each word, where the first column is for every single word and the second is for the belonging label of each word.

The 50 job descriptions are all related to the IT sector and all of them were obtained from the website: https://www.jobs.ie/. Each one was stored in plain text files.

Table 2. Job descriptions employed and the number of examples belonging to each class.

50 Job descriptions	
Class	Examples
Knowledge	2,097
Skill	956
Value	234
No assigned class	15,477

3.3 Tagger Interface

The main screen of the web-based tagger system is where users can tag texts as well as download them in the corresponding format to train the classifiers. This interface corresponds to the tagger user and a fast explanation of the design is showed in Figure 6.

The first section, allows the user to visualize the current text in the edition and it also shows if the file is enabled to be editable in a collaborative way.

Section two contains an autocomplete field for searching a specific text file in the system as the number of files increase quickly.

Section three shows the allowed tags to be used for the user, these indicate the domain in which a user is working.

In section four, a specific text file can be loaded into the system by selecting from a local computer.

In section five, a list of text files existing in the web-based system is displayed as well as the available formats to download by only selecting the desired file.

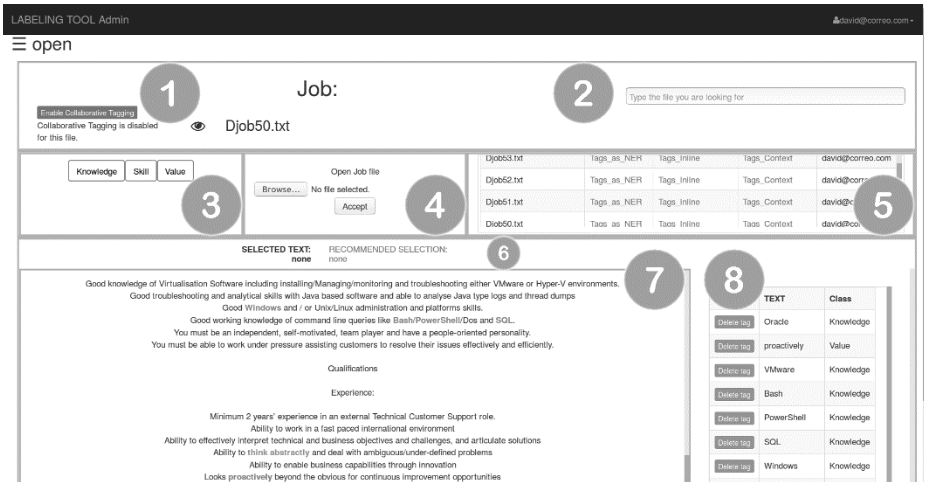


Fig. 6. Tagger interface.

In section 6, appears the selected keyword or concept in the plain text file at the moment of editing, and a recommendation is offered since people can omit one or more characters at the moment of selecting the keyword.

In section 7, the currently selected text file is displayed allowing to observe in real time the tagged keywords by highlighting them in the corresponding color of the tag assigned, it makes easy readable the content for taggers.

Finally, in section 8, a list with the tagged keywords is provided and the option for deleting a tag for a specific keyword.

4 Case Study and Discussion

This section presents an application nugget (small demonstration) of the web-based tagger system in order to show its operation and highlight the advantage to generate more efficient corpus and open access to any domain. The first step is to create a tagged dataset of job descriptions in the IT sector, all of this made by human feedback since humans can differentiate better between two different contexts.

4.1 Data preprocessing

It has been implemented considering the principles of V-fold cross-validation method, in which the original dataset (50 job descriptions) is partitioned in k equal subsamples. Then, of the k subsamples, only one of them is taking as validation data for testing the model and the remaining k-1 subsamples are used as training data (corpus). The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation time.

Table 3. Partitions employed in the training data with V = 10.

50 Job descriptions	
k time	Training data (partitions) 70%
1	f_0-f_{34}
2	f_{15}-f_{49}
3	f_0-f_{19}, f_{46}-f_{49}
4	f_0-f_{19}, f_{21}, f_{23}, f_{25}, f_{27}, f_{29}, f_{31}, f_{33}, f_{35}, f_{37}, f_{39}, f_{41}, f_{43}, f_{45}, f_{47}, f_{49}
5	f_0-f_{20}, f_{22}, f_{24}, f_{26}, f_{28}, f_{30}, f_{32}, f_{34}, f_{36}, f_{38}, f_{40}, f_{42}, f_{44}, f_{46}, f_{48}
6	f_5-f_{39}
7	f_{10}-f_{44}
8	f_0-f_{29}, f_{35}-f_{39}
9	f_0-f_4, f_{10}-f_{14}, f_{20}-f_{24}, f_{30}-f_{49}
10	f_0-f_{24}, f_{26}, f_{28}, f_{30}, f_{32}, f_{34}, f_{36} f_{38}, f_{40}, f_{42}, f_{44}

As explained in fundamentals section, 10 has been chosen as the value of V because it performs better since the value of V is small for complexity reasons but not too much for decreasing the variability of the algorithm.

For this work, the 70% of job descriptions were taken to train the classifier and the remaining 30% to evaluate its performance. This test was executed 10 times by randomly taking 70% each time and using the remaining 30% in each case for its evaluation.

To identify a specific job description, file the notation used is f_x where f means "file" and x is the number of job description used. The first job description is identified with x=0 and the last one with x=49. The partitions used each time in the process are shown in Table 3. The first training data consists of the set of job descriptions from 0 to 34. The second training dataset is composed of job descriptions from 15 to 49.

4.2 Classifier training

When an NER classifier needs to be trained it requires a special tabbed columns format where the first column contains the word and the second column has the tag assigned for that word. The web-based tagger has generated the corpus in this format with the tagged files by users. Since Stanford NER is a Java implementation, the training of a classifier is easy to accomplish with only one command line.

Table 4. Number of iterations and evaluations by each calssifier.

Classifier	Iterations	Evaluations
1	107	125
2	107	126
3	101	118
4	107	125
5	113	126
6	119	134
7	114	126
8	105	115
9	107	122
10	104	116

As the value chosen for V is 10 the number of classifiers is 10 as well, and the training datasets are different every time. Each of this workouts took a different number of iterations and evaluations to each classifier to be ready. The values corresponding to the total number of iterations and evaluations for each classifier are shown in Table 4.

4.3 Results and Discussion

In order to measure the performance of the classifiers, the Stanford parser takes into consideration three measures: Precision, Recall, and F1. However, it is necessary to know some values before calculating these results.

A true positive is a token that the classifier has identified as a specific class that we have also designated with that class. If the classifier designates a token with a class that we have not assigned to that token, then it is seen as a false positive. A false negative is a token that we have identified with a class and the classifier fails to identify as that class. Finally, there exists the true negatives that measure the proportion of actual negatives that are correctly identified.

Precision is calculated by taking the total number of true positives and dividing that number by the combined sum of true positives and false positives.

The Recall is calculated by dividing the number of true positives by the sum total of true positives and false negatives. The F1 is a combination of Precision and Recall.

Table 5. Results of each calssifier.

Execution	Class	Precision	Recall	F1
	K	0.8081	0.5573	0.6596
1	S	0.6615	0.3333	0.4433
	V	0.7857	0.2200	0.3438
	K	0.9864	1.0000	0.9932
2	S	1.0000	0.9907	0.9953
	V	1.0000	1.0000	1.0000
	K	0.5455	0.5415	0.5435
3	S	0.5082	0.3196	0.3924
	V	0.5385	0.1628	0.2500
	K	0.7037	0.6111	0.6541
4	S	0.5667	0.3208	0.4096
	V	0.5556	0.2041	0.2985
	K	0.6549	0.5107	0.5739
5	S	0.6290	0.3250	0.4286

	V	0.6250	0.2273	0.3333
	K	0.7177	0.6224	0.6667
6	S	0.6984	0.3964	0.5057
	V	0.5625	0.2045	0.3000
	K	0.7864	0.6030	0.6826
7	S	0.7636	0.3925	0.5185
	V	0.7500	0.2308	0.3529
	K	0.7448	0.5838	0.6545
8	S	0.6667	0.3750	0.4800
	V	0.6667	0.2273	0.3390
	K	0.5689	0.5556	0.5621
9	S	0.5918	0.2589	0.3602
	V	0.6250	0.2222	0.3279
	K	0.7300	0.5319	0.6154
10	S	0.6508	0.3628	0.4659
	V	0.7273	0.2000	0.3137

Recall and Precision serves as an indicator of the success of our classifiers. In Table 5 the values for Precision, Recall and F1 measures corresponding to each class of the 10 times executed are presented.

For a better understanding of the results obtained in Table 5, a separate chart for each measure is presented in figures 7 to 9.

In Figure 7 the values obtained for Precision can be appreciated, were the Precision measure is the degree to which those tokens identified as Knowledge, Skill or Value by a given classifier are indeed that entity. In the first two iterations the performance of the NER classifier is excellent, however, the performance decrease in iterations 3 to 5, and after that, between iterations 6 to 10 the results obtained show a performance remaining constant around 70% of Precision measure.

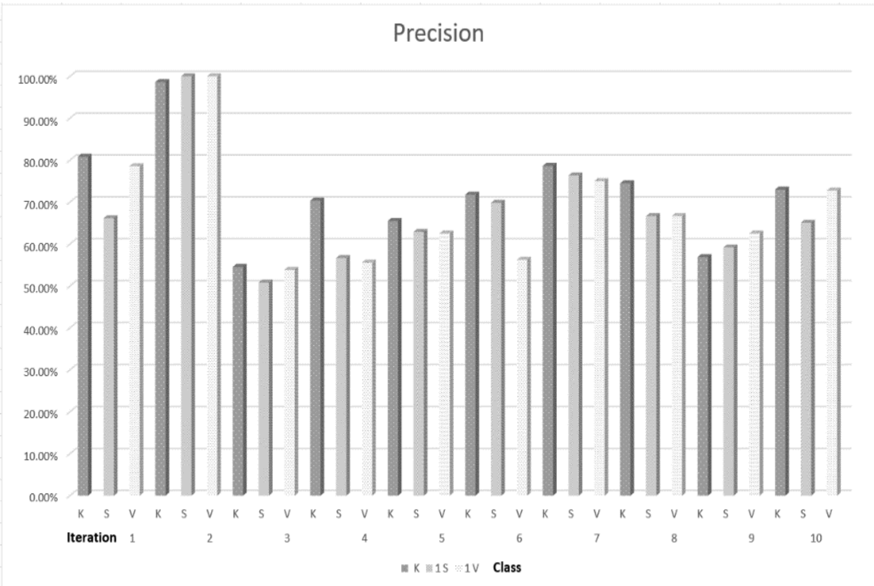


Fig. 7. Precision performance in 10 data tests.

It is also possible to observe the great performance for class Knowledge since most of the keywords reported as Knowledge is in fact Knowledge.

On the other hand, the Recall measure is an indicator of the degree to which a given classifier is able to find all of the tokens that we have identified as Knowledge, Skill or Value. In Figure 8 the values obtained for Recall measure are lower than in the Precision measure, with the exception of iteration 2 that is excellent. It is important to highlight the great performance for Knowledge class. In every iteration, the best performance is always achieved for Knowledge class and it can be possible as a result of more examples provided in this class. However, the results obtained in the Value class are far to be excellent since the examples are only a few pieces provided.

After the Precision and Recall values have been obtained, it is possible to combine those values into an F1 measure which serves as an abstract index of both. In Figure 9 the values obtained for the F1 measure can be appreciated. Again the best results are achieved for Knowledge class in each iteration, while for the Value class the results are low.

After analyzing the results obtained in every iteration and in every measure, it is evident to observe a great performance in precision measurement and always showing excellent results in Knowledge class, this means that most of the keywords identified as Knowledge are in fact Knowledge. Not the same with the Recall measure and Value class, as the number of provided examples is the minimum. The Skill class values are always in the middle of the results, presenting a few changes between a test and another.

The reported accuracy rates presented are good enough, demonstrating an excellent learning in the Knowledge class since wide examples were provided. That why the identified entities as Value in the unknown texts were poor.

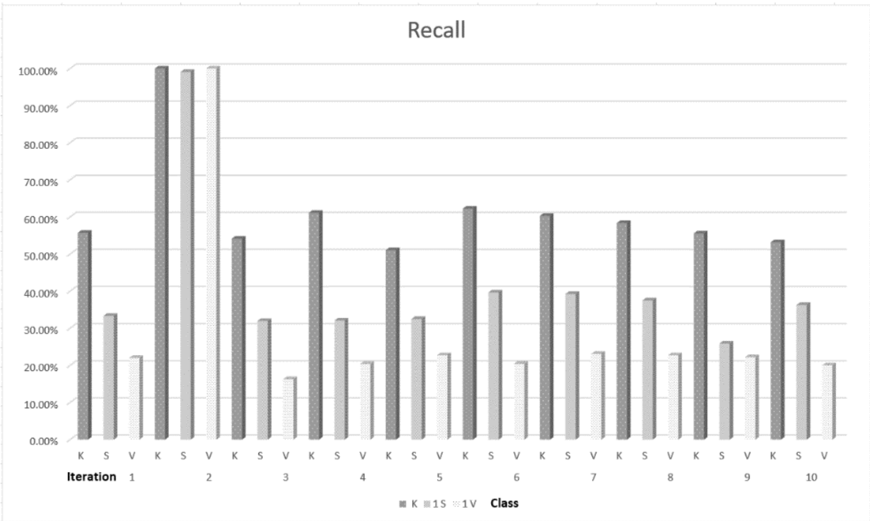


Fig. 8. Recall performance in 10 executions.

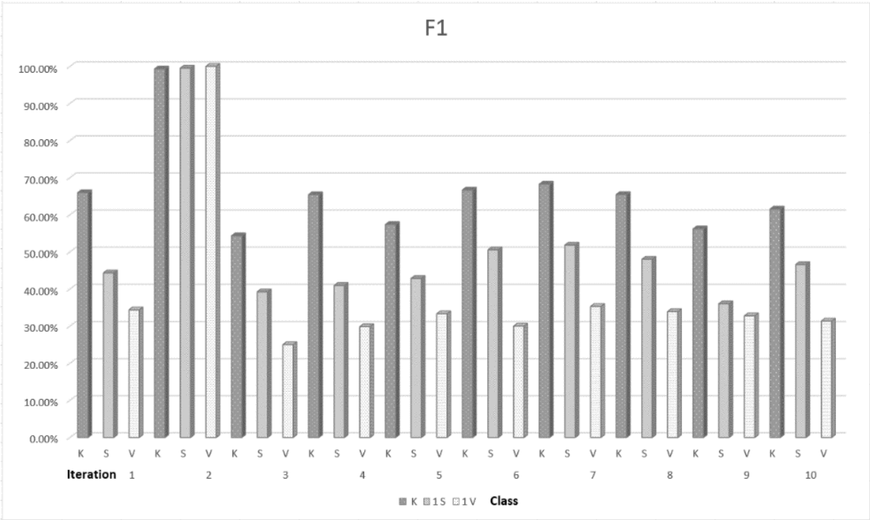


Fig. 9. F1 performance in the complete test.

This experiment shows the advantage of using the CRFClassifier of Stanford in the task of named entities recognizing. And thanks to the different dimensions of examples provided in the three classes it is clear to observe the great performance when a classifier model is trained with a considerable amount of examples for each class.

5 Conclusion

The proposed system allows the construction of high-quality corpora because the texts are tagged with human feedback, in this sense each concept or keyword is tagged under the decision of people which are able to differentiate between one context and another without a problem. The presented work in this paper offers the possibility to generate better corpora in any domain and even faster since these can be constructed in a collaborative way.

As a fast demonstration of the possible applications with the generated corpora in the Tabbed Column Stanford format an application nugget was performed and the performance measures obtained in each class present the excellent use of the CRFClassifier when named entities detection are required to extract relevant information on textual data.

Future experiments will consider the detection of the main characteristics in the domain of job descriptions and Résumés with a huge number of examples for every class since the number of examples are important to improve the results. Identifying the relevant aspects of a new job position in a specific company and the best applicant can reduce the time spent at the moment for finding an applicant.

The web-based tagger can be used to highlight the relevant information in any domain in plain text files. Once the most important data has been tagged the generated corpora can be employed in training classifiers to identify named entities in any kind of text.

References

1. Nanopoulos, A.: Item recommendation in collaborative tagging systems. IEEE transactions on systems, man, and cybernetics, NO. 4, 760–/71 (2011)
2. Ritter, A., Clark, S., Mausam., Etzioni, O.: Named Entity Recognition in Tweets: An Experimental Study, In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 1524–1534 (2011)
3. Sheshasaayee, A., Thailambal, G.: Comparison of Classification Algorithms in Text Mining. International journal of pure and applied mathematics, 425–433 (2017)
4. Chavaltada, C., Pasupa, K., Hardoon, D.: A comparative study of machine learning techniques for automatic product categorization. Springer international publishing, 10–17 (2017)
5. Font, F., Serrà, J., Serra, X.: Folksonomy-based tag recommendation for collaborative tagging systems. Inter-national Journal on Semantic Web and Information Systems, 1–27 (2013)
6. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. WWW 2007, 211–220 (2012)
7. Bischoff, K., Firan, C., Nejdl, W., Paiu, R.: Can all tags be used for search? In: Proceedings of the 17th acm conference on information and knowledge management, 193–202 (2008)
8. Abzianidze, L., Bos, J.: Towards Universal Semantic Tagging (2017)
9. Hawizy, L., Jessop, D.M., Adams, N., Mur-ray-Rust, P.: Chemical Tagger: A tool for semantic text-mining in chemistry. Journal of Cheminformatics 3 (17) (2011)
10. Allahyari, M., Safaei, S., Pouriyeh, S., Trippe, E., Ko-chut, K., Assefi, M., Gutierrez, J.: A brief survey of text mining: classification, clustering and extraction tech-niques. KDD Bigdas (2017)
11. Mohamed, M., Oussalah, M.: Identifying and Extracting Named Entities from Wikipedia Database Using Entity Infoboxes, (IJACSA) International Journal of Advanced Computer Science and Applications 5 (7), 164–169 (2014)
12. NLP. Stanford Named Entity Recognizer. The Stanford natural language processing group. Accessed online on May 2018. https://nlp.stanford.edu/software/CRF-NER.html
13. Arlot, S.: V-fold cross-validation improved: V-fold penalization, 1–17 (2008)
14. Geysser, S.: The predictive sample reuse method with applications. Journal of the American Statistical Association 70 (350), 320–328 (1975)
15. Weiss, S. M., Indurkhya, N., Zhang, T.: Fundamentals of Predictive Text Mining, Springer Verlag, 113–137 (2010)
16. Saquib, S., Siddiqui, J., Ali, R.: Classifications of Recommender Systems: A review. Journal of engineering and technology review, 132–153 (2017)
17. Tardif, S., Curran, J. R., Murphy, T.: Improved Text Categorisation for Wikipedia Named Entities (2006)

ISSN 1870-4069

Methodology for the Analysis of Electrical Consumption

Blanca N. Pérez-Camacho, Juan M. González-Calleros, Iván Olmos-Pineda

Benemérita Universidad Autónoma de Puebla,
Faculty of Computer Science, Puebla, Mexico

{blancanydia.perezc, jumagoca78, ivanoprkl}@gmail.com

Abstract. A demand system is designed in order to analyze and control the way in which electricity consumption occurs in a scenario, this type of system has four objectives to reduce consumption, reduce costs, reduce peak average to ratio and maximize the comfort of the users. The design of these systems can be based on the implementation of IoT. When using IoT for the development of demand systems, a layer of information is integrated which generates a multi-objective optimization problem. Metaheuristics are used to solve multi-objective problems. This paper introduces a proposed methodology to develop demand systems and the first products implemented.

Keywords. Demand system, internet of things, metaheuristic.

1 Introduction

Until now, a methodology for analyzing electric consumption has not been developed from the perspective of the implementation and use of data in an IoT architecture.

Internet of Things (IoT) has been implemented in the way to develop different application areas, like Smart Warehouse, Smart grid, Smart City, Smart meter, Smart Healthcare and Smart Home [1]. There is a common objective to accomplish between the Smart grid and Smart home that is to develop systems that make an efficient electric consumption. This kind of systems is called Demand Systems (DS).

The works focused in the search of the improvement of the electric consumption are about devices or control variables [2-5] buildings consumption [3], houses [6], costs [2,7-10] sensor infrastructure [3] and control algorithms [4,10]. Few works focus on the current challenge of developing solutions supported by IoT.

Demand Systems analyses how the electric energy is used, their objectives could be one or more from these: low Peak Average Ratio (PAR), minimize costs, minimize consumption and maximizer user´s comfort [11]. To develop DS requires: define objectives, variables and what kind of factors needs to be considered. When DS works with consumption data needs to handle a big quantity of data [12] that could use to make decisions [13, 14] and converts into a multi-objective optimization problem [15]. A metaheuristic is used to find a solution to the multi-objective problem.

So, according to the characteristics of the problem, it requires to search for alternative solutions that are appropriate to the context of the problem, which consider the source of the data, the variables and the implementation of a metaheuristic to search the optimal consumption.

Rest of the paper is organized as follows, in Section 2 describes the state-of-the-art work. A description of the layers that make up the methodology proposed to develop projects that implemented Internet of Things is given in Section 3. In Section 4, a brief discussion is carried out. Both conclusions and work to be done in the future are announced in Section 5.

2 Related Work

Many researchers around the world worked to make an optimal electric consumption system in different research lines smart meter, smart grid, neural network, metaheuristic, IoT, Genetic algorithm and big data. Nadeem et al. in [16] considered to develop a DS to reduce consumption under a predefined level, costs and waiting time using hybrid metaheuristics schemes based on Teaching-Learning techniques. The metaheuristics used were Optimization Stopping Rule (OSR), Genetic Algorithm (GA) and Firefly Algorithm (FA); in this work is combined OSR-GA, OSR-TLBO, and OSR-FA. This work considered three electric devices (fridge, dishwasher, and dryer), the variables that were considered for every device: costs average per month, the cost reduction, the priority and the delay per hour per day. It is simulated by the implementation of every technique and their hybrids to each of the devices.

Yao et al. in [17] said that the energy management problem consists in to solve appliance load scheduling and grid power dispatching under a single optimization framework of a utility grid with dynamic costs, a photovoltaic module and the household appliance with three different types: interruptible, uninterruptible and time-varying; in this work was implemented a simulation of a mixed integer linear programming framework. His future consists of two stages, first to implement a genetic algorithm, second to implement a multiobjective optimization framework.

Rahim et al. in [11] described the goal of implementing a demand system, which is reduced to the following to reduce the costs, minimization of Peak to average ratio and maximize comfort. In this work is proposed a generic architecture for a demand system, it models the electric consume in a house, here was considered three algorithms genetic algorithm (GA), binary particles swarm optimization (BPSO) and ant colony optimization (ACO). The GA was more efficient than BPSO and ACO in term of consumption reduction, minimizing PAR while is considered user comfort. In this work was concluded that is still an open problem to Minimize PAR, cost, consumption and maximize comfort.

Javaid et al. in [5] focused over to control electric consumption and maintenance the comfort taking into consideration the user preferences. Was implemented four algorithms: genetic algorithms (GA), teaching-learning base on optimization (TLBO), enhanced differential evolution (EDE) and enhanced differential teaching-learning algorithm (EDTL). The consumption model took two types of devices flexible and inflexible, device, consumption and schedule of use, this model is implemented in a microgrid context.

Jon the other hand, Avaid et al. in [3] presented an electric consumption model that consider houses and devices mount; the kind of devices are interruptible, not interruptible and regular application use. This model is for a demand-side management to reduce peak average to ratio, costs and renewable energy that are considered tariff and time of use. The algorithms that were analyzed are bacterial foraging optimization algorithm (BFOA), genetic algorithm (GA), binary particle swarm optimization (BPSO), wind-driven optimization (WDO). In this work, a genetic binary particle swarm optimization (GBPSO) is proposed. Results showed that GA is better than others in term of PAR reduction and execution time, BPSO is better than the others in term of cost reduction. GBPSO is more efficiently than the others in terms of cost and PAR.

Complementary work, is presented by Hao and Wang where they proposed a demand-side management (DSM) in [10], this DSM reduced cost in terms of tariff and costs, devices identified are interruptible and not interruptible. This DSM implemented a game theory which one can reduce costs and help to increase load demand at the off-peak time. The peak time hours were between 8 to 24 hours. Mohsin et al. developed a demand-side management in [13] with the main purpose of reducing costs, PAR and time, maximize comfort in applying a harmony search algorithm (HSA). This work compared his algorithm with binary particle swarm optimization (BPSO), differential evolution (DE), genetic algorithm (GA), ant colony optimization (ACO), Hybrid Differential Evolution – Harmony Search (DE-HS), immune artificial hybrid algorithm, genetic hybrid algorithm, teaching-learning base on optimization (TLBO) and Shuffled Frog Learning (SFL). Devices considered in this works were defined as interruptible, not interruptible, flexible and inflexible. The results showed that HAS is better in terms of costs, PAR and consumption.

In addition, Huang et al. in [14] where a set of models of demand response was used, using a gradient (PSO) based on particle swarm optimization, the main objective was to schedule the operation of appliances to save energy and reduce cost considering user convenience. The model is simulated and compared to hybrid PSO algorithm and cooperative PSO algorithm. The proposed algorithm shows better results in a real-time application. Similarly, Matei et al.in [18] described an IoT architecture, that consist of four layers sensors, physic, digital and meta through that the data flow. In this work were proposed two times of data process, first in the physic layer and second in meta-layer. The first process to select data and reduce computational cost, and in the second process could be implemented in any algorithm or technique, the choice of it depends on objectives.

Finally, Mohsin et al. concluded in [13] that a deterministic optimization is inefficient and impractical to handle a big problem. A heuristic optimization is better to implement in big data problems. Also, Silva et al. conclude in [19] that the electric consumption varies according to an hour of the day, the day of the weak and season.

Hoon et al. explained in [2], IoT let to monitor, handle and control devices WEB way. The IoT implementation let to get information from data that are generated in real time. There is a challenge for smart meters to develop an intelligent environment (AmI).

Table 1. Goals of systems developed in literature.

Reference	Kind of System	Reduce cost	Reduce Consumption	Comfort	Reduce PAR	Algorithms
Tsai and Lin [20]	DRS	X	X			Sorting genetic algorithm
Yao et al. [17]	DS	X	X			GA
Nadeem et al. [16]	DS	X	X	X		OSR-GA, OSR-TLBO, OSR-FA
Rahim et al. [11]	DRS	X	X	X		GA, BPSO, ACO
Tushar et al. [21]	DRS					Game theory
Javaid et al. [5]	DS	X	X			GA, TLBO, EDE, EDTL
Javaid et al. [3]	DS	X			X	BFOA, GA, BPSO, WDO, GBPSO
Hao and Wang [10]	DS	X				Game theory
Mohsin et al. [13]	DS	X		X	X	HAS
Huang et al. [14]	DRS	X				PSO
Chen et al. [22]	DS	X			X	PSO

Table 2. Kind of appliances in literature.

Reference	Interruptible	Non-Interruptible	Flexible	Not-Flexible
Hao and Wang [10]	X	X		
Mohsin et al. [13]	X	X	X	X
Kakran and Chanana [7]	X	X	X	X
Javaid et al. [3]	X	X	X	
Yao et al. [17]	X	X	X	

In Table 1 is summarized the articles in which demand system and demand-response system (systems that not only take into account the consumer but also the electricity suppliers) developments are presented, identifying for each one of the systems, the implementation objectives to be fulfilled, as well as the metaheuristic algorithms implemented. It is observed in the table that one of the first techniques to be implemented for a demand system is a genetic algorithm and hybrid algorithms. In Table 2 shows the types of appliances in which the household devices can be located.

For the accomplishment of this work were consulted and analyzed 69 articles, that articles were classified according to the areas of development identified; there were 3 articles of smart meter, 29 of smart grid, 1 of nerual network, 5 of metaheuristics, 1 of learning alfgorithm, 12 of IoT, 3 of genetic algorithm, 13 of data minig, 2 of corpus, 3 of bioinspired algorithm and 2 of big data

According to the analysis of 63 articles, the application objectives that could be for the control of processes, big data analysis, behavioral modeling, consumption control, including renewable energy and to form part of a Smart home were identified.

From the reviewed articles it was observed that 73% focus their proposals for the implementation of a demand system from the Smart grid area, 11% from the implementation of a metaheuristic, 4% from the use of the internet of things, 4% implements genetic algorithm and 8% use Smart meters.

3 A methodology for Analysis of Electronic Consumption

A demand system (DS) analyze and control the electric area consumption. The DS objectives [11] are to reduce costs, reduce consumption, reduce PAR and maximizer user comfort. IoT implemented a control and monitoring infrastructure through the web to devices that are in the same red, so it´s possible to develop a DS using this concept. The IoT data flow, its obtained from [18] where describes IoT architecture and [22] describes cyber-physic behavior (see Fig. 1).

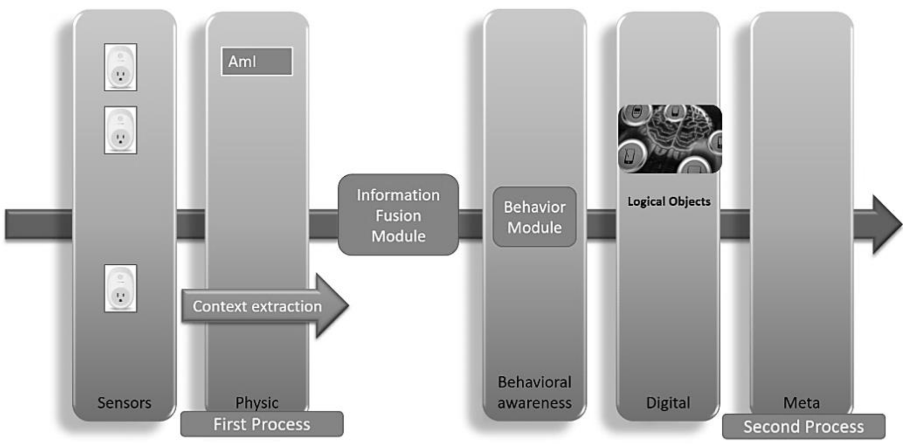


Fig. 1. DS data flow in IoT.

The IoT data flow, flow through five layers called sensor, there is a sensor architecture; physic, here is implemented an ambient intelligent (AmI), AmI extracts the context and in this stage is made a first process whit sensor and user data; the data extracted pass to information module, where all the context data are gathered; behavioral awareness is processed the data and found a behavior; digital, here is created a logical object that represents the real objects; in the meta-layer all the logical objects data are processed according to particular objectives of the problem to be solved.

When is developing a DS and is integrated an information layer, which happens when IoT is implemented, the problems became a multi-objective optimization problem and are solved using metaheuristics [23]. Also, is necessary to identify all variables that determine the electric consumption, to know the types of devices, to identify a meta-heuristic algorithm and to propose an electric consumption model.

The methodology feasibility is going to be illustrated with a case study with a real case scenario of electronic consumption corresponding to every identified device that is frequently used in an office, house, and classroom.

3.1 Consumption and Context of Use

Context of use is a triple composed by the identification of electric consumption device, the environment where the device is used, house, office, classroom, and user's particular needs or preferences. For each device, a set of common characteristics have been identified. Accordingly to the following list:

- *Minimum power* corresponding to the minimum electrical power needed for the device to operate. This value was estimated from public lists of electronic devices publicly available on their websites.
- *Maximum power* corresponding to the minimum electrical power needed for the device to operate. This value was estimated from public lists of electronic devices publicly available on their websites.
- *time of use*. The average daily usage time measured in hours.
- *kind of device.* The devices are classified according to the most common appearance. For example, office supplies, video games, household appliances, electronics, lighting and security devices.
- *The type of devices* corresponding to the a category proposed in related works [4,24,25] that are: interruptible (I), non-interruptible (NI), flexible (F) and not-inflexible (NF). An Interruptible system refers to a system that can be turned on and off without any particular constraint. Interruptible devices are defined like devices that could be used in anytime, and the time of use varies according to the user needs [5, 7,10,13,17]. Not-interruptible devices stop their function once they are finished, their consume could be variable or constant [7,10,13,16,17]. Flexible devices, their functions could be stopped and continue at another time, could be on standby too [3,7,13,17] Not-flexible devices, they could not be turned off because is necessary that they have a constant function [7,13].

3.2 Sensor Layer

Sensor Layer stores sensors raw data, also known as primary data, corresponding to the sensors consumption readings. In a particular context the set of devices may vary and the corresponding arrangement as well, we call this setup the Sensor Architecture. So, from this layer the output is a data base with the consumption readings, consumption meters, of each device that is connected to the sensor architecture of the electricity facility under a specific context.

Data acquisition nowadays could be the result of using IoT devices, or we could even built a no-invasive potentiometer using Arduino Uno and SCT-013 sensors, or any related technology. Data consumption files are obtained and stored in a DB, a sample set is listed as follows in Table 3:

Table 3. Consume list from a laptop.

Current (Amperes)	Power (Watts)
6.4871	1423.1668
3.9958	878.9938
4.4403	976.8803
4.7000	1059.5591
12.5839	2768.4433
9.4061	2009.3459
8.5427	1894.3894
4.9717	1059.7808
3.8485	846.6782
6.0309	1327.0713
18.2753	4020.5761
6.7826	1492.1634
6.0730	1336.0534
6.6512	1409.5651
3.1939	702.6623

This is an input fusion layer and the classifier. The current work just consider data from a specific context which every household device monitored by a non-invasive smart meter that represents a specific consumption that is determined by the occupants' behavior.

In our example, we implemented a database (DB) using MySQL Workbench, DB is in a local server. The DB has 268 registers. Every register is stores: <id_device, device_name, min_ power, max_power, freq_time, shut_down, kind_device>, some registers are showed in Table 4.

Table 4. Consume list from a laptop.

ID device	Device_name	Min power	Max_power	Freq_time (use Hrs/day)	Shut_down	Kind_device
OF1	Cannon (three light)	6.4	400	1.5	Yes	I
OF2	Cannon (three light)	6.4	1000	1.5	Yes	I
OF3	Cannon (one light)	6.4	220	1.5	Yes	I
OF4	Cannon (one light)	6.4	236	1.5	Yes	I
OF5	CPU	14.1	250	4	Yes	F
OF6	CPU iMac 3.06GHz	129.6	365	4	Yes	F
OF9	Digital decoder	55	55	24	Yes	I
OF10	Scanner	65	275	1	Yes	I
OF11	Fax	65	150	1	Yes	F

3.3 Physical Layer

The physical layer corresponds to the composition of the primary data and a particular user data that corresponds to a set of constraints. For this purpose an interactive systems is recommended. a human-computer interface.

The external data is the readings obtained from the room sensors. In a demand system it is identified that the user data have to do with the preferences and times of use of the electrical devices that are monitored through the sensor architecture; and the external data are the data obtained from the environmental conditions (temperature, humidity, presence, etc.), that is, all those data that result in a high degree of importance to be considered for the conditioning of the use of electrical devices within the area you want to control. In our case, we have built a website (see Fig. 2) where users select their devices, graphically, and coherent with the proposal of building better Legos [26].

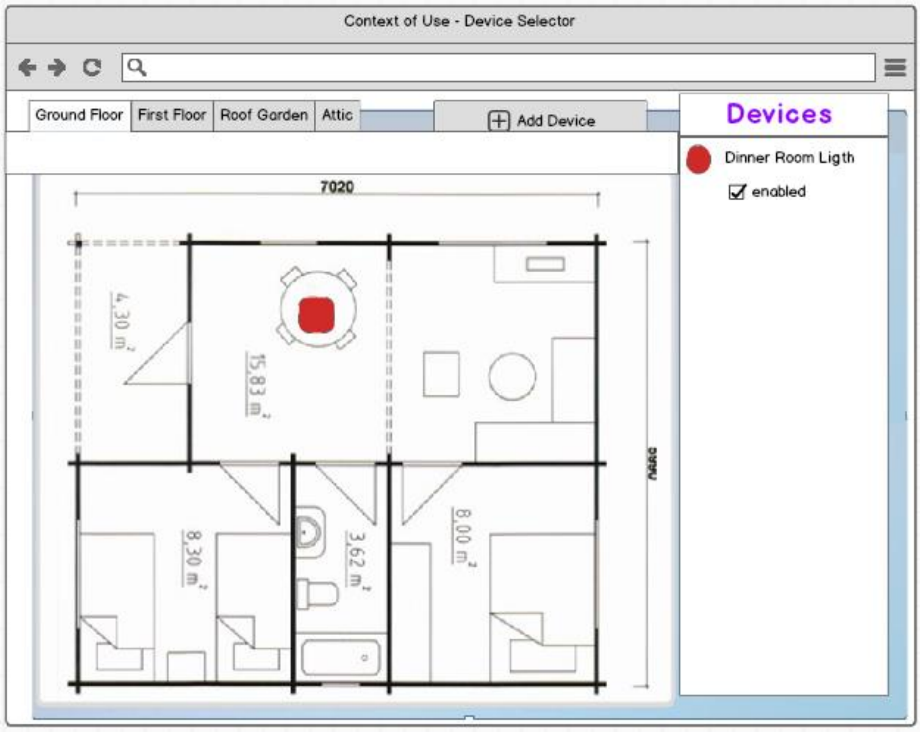


Fig. 2. A web site used to help users to select the physical set of devices available at their facility.

3.4 Fusion Layer

It works as a repository of all the data that must be considered (primary, external and user data) to solve the problem that is being studied; to later be able to group them according to the needs of solution. The output resulting from this module is grouped data.

3.5 Behavioral Layer

A behavioral module is in the behavior layer, in which methods or procedures can be implemented to generate the necessary parameters according to the problem to be solved, this module includes both the user's data (which determines the restrictions to consider) as environmental data (which are the conditions for decision making).

Demand System in the behavior module, a behavioral model can be implemented that identifies the probability and time of use of the devices, giving as output the on and off schedule, and the operating time of each device; and integrate, a module capable of generating the actual consumption data that each device can have. All the resulting data pass in vector form to the next layer.

3.6 Digital Layer

The digital layer is where the data from the previous layer arrive and a space is formed with the vectors that represent the possible behavior scenarios of all the objects that are in the real world. The vector space is sent to the next layer. In a demand system, we have a set of vectors formed from the combinations of possible consumptions that each of the devices that are being monitored may have.

3.7 Target Layer

In the target layer, procedures are implemented that seek proposals for solutions to the objectives that are to be achieved according to the proposed scenario. A demand system is a multi-objective problem and to implement it, a metaheuristic is implemented.

4 Discussion

From the methodology described in the previous section it is possible to identify how the flow of the data is carried out when it is integrated into IoT to solve a problem. In the case of implementation of a demand system consists of selection of devices for reading consumption data, the development of an intelligent environment that serves as an interface to user and serves to read the consumption data to perform a context extraction, develop a module to merge the data received, develop a behavior model that will be implemented in a behavior module, obtain logical objects that emulate the real object, and implement a metaheuristic whose definition of objective function is based on the objectives (objectives of DS system) to meet the original problem.

Has been identified a methodology to identify consumption variable, which consists to identify the devices that are the most frequently are in a house, office, and classroom; the maximum and minimum power: the frequently time of use according to the area in which demand system is going to be implemented; and the classification to which the device belongs (I, NI, F, NF). With the database, devices are possible to simulate N different scenarios with d devices each one.

With the non-invasive data acquisition system, it has been possible to acquire the consumption of devices found in a house (oven, television, phone charger, laptop, charger, washing machine, and refrigerator).

Have been identified metaheuristics to be implemented in a DS to make an efficient consumption, genetic algorithm, particle swarm optimization, teaching-learning based on optimization and optimization stopping-rule.

The data flow in demand system was obtained from two different ideas, one that talks about an architecture that would be considered when is implementing an IoT project, and another one when is talking about a context where is necessary to take in a count the information. The selection of metaheuristic to be implemented is based on what has been identified in the literature, in which the constant of selecting a genetic algorithm is identified against whose results are compared to another metaheuristics implementation.

5 Conclusion and Future Work

This methodology was developed after having analyzed articles that have as objective the development of demand systems, where it is appreciated that the development layers in which they focus only cover one or two of the methodology proposed in the present work. It is appreciated that if the methodology proposed in the present work that uses IoT as part of the implementation process is followed so could be to identify and to comply with every one of the requirements to be met.

The proposed methodology is used as the basis to develop a demand system, where the appliances were identified as their consumption variables too.

As future work, the creation of an intelligent schedule is left to foresee with the probability of use both the schedules and the time of consumption of each device that are contemplated in the scenario. As well as, the implementation of a metaheuristic to analyze and identify the recommendation that is appropriate according to the needs of the user, the reduction of consumption and voltage peaks.

References

1. Parsa, A.: Implementation of Smart Optimal and Automatic Control of Electrical Home Appliances (IoT) (2017)
2. Hoon Kim, T., Ramos, C., Mohammed, S.: Smart City and IoT, Futur. Gener. Comput. Syst., vol. 76, no. July 2014, 159–162 (2017)
3. Javaid, N.: A new heuristically optimized Home Energy Management controller for the smart grid, Sustain. Cities Soc., vol. 34, no. July, 211–227 (2017)
4. Jahn, M., Jentsch, M., Prause, C.R., Pramudianto, F., Al-Akkad, A., Reiners, R.: The energy-aware smart home, 2010 In 5th Int. Conf. Futur. Inf. Technol. (FutureTech), IEEE., pp. 1–8 (2010)
5. Javaid, N.: Demand Side Management in Nearly Zero Energy Buildings Using Heuristic Optimizations, Energies, vol. 10, no. 8, 11–31 (2017)
6. S. R., S. K. R.: Data Mining with Big Data, 2017 In 11th Int. Conf. Intell. Syst. Control, pp. 246–250 (2017)
7. Kakran, S., Chanana, S.: Energy Scheduling of Smart Appliances at Home under the Effect of Dynamic Pricing Schemes and Small Renewable Energy Source, Int. J. Emerg. Electr. Power Syst., pp. 1–12 (2018)
8. Silva, B., Khan, M., Han, K.: Load Balancing Integrated Least Slack Time-Based Appliance Scheduling for Smart Home Energy Management, Sensors, vol. 18, no. 3, p. 685 (2018)
9. Wu, X., Zhu, X., Wu, G.Q., Ding, W.: Data mining with big data, IEEE Trans. Knowl. Data Eng., vol. 26, no. 1, pp. 97–107 (2014)
10. Hao, Y., Wang, W.: Optimal Home Energy Management with PV System in Time of use Tariff Environment, no. 51641702, pp. 5–9
11. Rahim, S.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources, Energy Build., vol. 129, pp. 452–470 (2016)
12. Mocanu, D. C., Mocanu, E., Nguyen, P. H., Gibescu, M., Liotta, A.: Big IoT data mining for real-time energy disaggregation in buildings, 2016 In IEEE Int. Conf. Syst. Man, Cybern. SMC 2016 - Conf. Proc., pp. 3765–3769 (2017)

13. Mohsin, S. M., Javaid, N., Madani, S. A., Abbas, S. K.: Appliance Scheduling in Smart Homes with Harmony Search Algorithm for different Operation Time Intervals Appliance Scheduling in Smart Homes with Harmony Search Algorithm for different Operation Time Intervals, no. February (2018)
14. Huang, Y., Tian, H., Wang, L.: Demand response for home energy management system, Int. J. Electr. Power Energy Syst., vol. 73, pp. 448–455 (2015)
15. Fauvel, C., Claveau, F., Chevrel, P., Fiani, P.: A flexible design methodology to solve energy management problems, Int. J. Electr. Power Energy Syst., vol. 97, no. November 2017, pp. 220–232 (2018)
16. Nadeem, Z., Javaid, N., Malik, A. W., Iqbal, S.: Scheduling appliances with GA, TLBO, FA, OSR and their hybrids using chance constrained optimization for smart homes, Energies, vol. 11, no. 4, pp. 1–30 (2018)
17. Yao, L., Damiran, Z., Lim, W. H.: Energy Management Optimization Scheme for Smart Home Considering Different Types of Appliances (2017)
18. Matei, O., Anton, C., Scholze, S., Cenedese, C.: Multi-layered data mining architecture in the context of Internet of Things, Proc. - 2017 In IEEE 15th Int. Conf. Ind. Informatics, INDIN 2017, pp. 1193–1198 (2017)
19. Silva, B. N., Khan, M., Han, K.: Load balancing integrated least slack time-based appliance scheduling for smart home energy management, Sensors (Switzerland), vol. 18, no. 3 (2018)
20. Tsai, M. S., Lin, Y. H.: Modern development of an Adaptive Non-Intrusive Appliance Load Monitoring system in electricity energy conservation, Appl. Energy, vol. 96, no. December, pp. 55–73 (2012)
21. Tushar, W., Yuen, C., Mohsenian-Rad, H., Saha, T., Poor, H. V., Wood, K. L.: Transforming Energy Networks via Peer to Peer Energy Trading: Potential of Game Theoretic Approaches (2018)
22. Chen, S.: Butler, Not Servant: A Human-Centric Smart Home Energy Management System, IEEE Commun. Mag., vol. 55, no. 2, pp. 27–33 (2017)
23. Fauvel, C., Claveau, F., Chevrel, P., Fiani, P.: A flexible design methodology to solve energy management problems. International journal of Electrical Power and Energy systems, vol. 97, pp. 220-232 (2018)
24. Yao, L., Shen, J. Y., Lim, W. H.: Real-Time Energy Management Optimization for Smart Household, Proc. - 2016 In IEEE Int. Conf. Internet Things; IEEE Green Comput. Commun. IEEE Cyber, Phys. Soc. Comput. IEEE Smart Data, iThings-GreenCom-CPSCom-Smart Data 2016, pp. 20–26 (2017)
25. Alam, M. R., St-Hilaire, M., Kunz, T.: Computational Methods for Residential Energy Cost Optimization in Smart Grids: A Survey, Acm Comput. Surv., vol. 49, no. 1, p. 2 (2016)
26. Jenson, S.: The Future IoT: Building Better Legos. IEEE Computer, 50(2), 68–71 (2017)

ISSN 1870-4069

Ontology: Components and Evaluation, a Review

Cecilia Reyes-Peña, Mireya Tovar-Vidal

Benemérita Universidad Autónoma de Puebla,
Faculty of Computer Science, Puebla, Mexico

reyesp.cecilia@gmail.com, mtovar@cs.buap.mx

Abstract. The ontologies are a powerful tool for representing the knowledge from a particular domain so it is necessary to know its elements in order to guarantee the safety and satisfaction of the task for which was designed and created. This work presents a general review of the elements and evaluation of the ontology in order to offer practical definitions and some ontology application examples.

Keywords: Ontology, ontology's elements, ontology evaluation.

1 Introduction

All days the peoples generate a big information amount to share on the Internet, since a text message until the personal location. Sometimes there are problems when we want to consult, sort or transfer this information because the format is not the same in all sources. The ontology has the purpose of providing knowledge for data structuring by rules mainly on the web, from a particular domain.

We can define the ontology as a formal abstraction of what we wish to represent of a domain, using specific information such as objects, properties, and relationships [11] by a structure normally of hierarchical type; Tello [32] defines ontology as an explicit and formal specification about a shared conceptualization, that has a defined and legible vocabulary to express the main concepts and relationships about a specific domain [24]. The ontology also can be defined as a form of representation about a particular universe of discourse or some part [19], it has a well-defined structured from a set of most representational terms with human-readable text description and its construction methodology depend on clarifying what types of objects are researched in that domain [5].

In general, we can say that the ontology is a theory that specifies a relative vocabulary into a domain in order to help with the semantic interoperability among systems mainly in the web [18]. There are many kinds of ontologies according to their focus, application, creation, the specific domain, and generics [30].

In this work, the elements and evaluation techniques of an ontology are described in offering a general view about semantic knowledge representation and its evaluation from two approaches: validation and verification, and based on criteria.

This work is structured as follows: in section 2, it describes the main elements which forming an ontology; section 3 contains evaluation techniques based on two approaches; section 4 contains the related works about the ontology application and finally, section 5 contains the conclusions and future work.

2 Ontology's Elements

The ontology can be seen as a 5-tuple where its components are: Concepts, relationships, functions, individuals or instances and axioms [32].

$$Ontology =< C, R, F, I, A >, \quad (1)$$

where:

- **Concepts (classes):** are the main formalized elements of the domain [32]. Since the logic, the concepts can be described using specific properties which must be satisfied by them [2].
- **Relationships:** are links between the concepts for representing the ontology structure (taxonomic or not taxonomic).
- **Functions:** are elements with the purpose of calculating information from the other elements.
- **Instances (objects):** are the representation about the main objects within domain according to ontology structure.
- **Axioms:** are the restrictions, rules, logic correspondences definitions [4] which must be accomplished in the relationship between the ontology elements. The axioms can be seen as the smallest unit of knowledge within an ontology [31].

Then, we can see an example of ontology about relevant information about the Master's degree in Computer Science (see Fig.1), where there are three taxonomies: Investigation Line, Person and Subject; a case of taxonomic relation is Student is-a Person [22] [26].

3 Ontology Evaluation

The ontology is a very important tool for the information representation that has become a standard for the knowledge representation [13], so it is necessary to evaluate its main aspects in order to guarantee that representation to be the most real according to the domain. The evaluation of an ontology can be seen from two approaches: verification and validation [1], or evaluation by criteria (see Fig.2).

3.1 Verification

The verification is used to determine if the ontology is correctly constructed or not by the satisfaction of the competency questions which was defined as part of the proposed task, i.e., if the built ontology is suitable for the real world. The verification has the next focus:

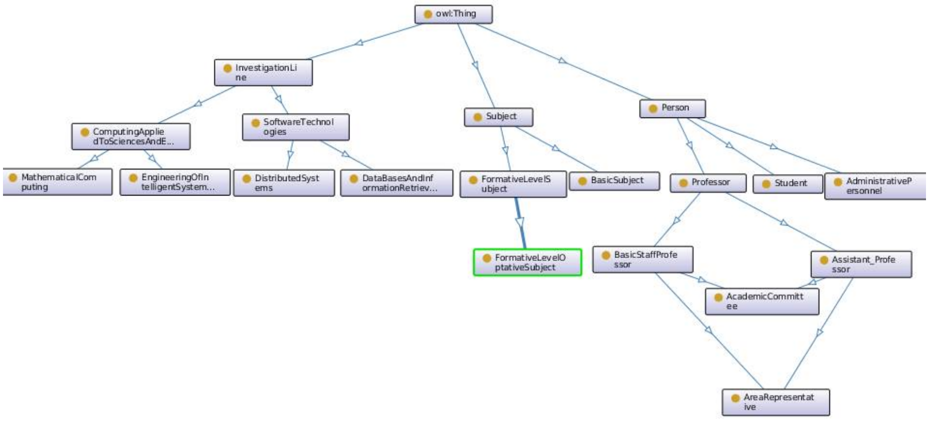


Fig. 1. Ontology Example: Master's degree in Computer Science.

- **Lexical:** the lexical verification is related to the vocabulary used for conceptualizing the domain and is regularly made with precision, recall and F-measure [31] [3]. This verification includes aspects related to the reusability [15].
- **Taxonomic:** specifically, this verification is only focused on is-a and has-a relationships within the ontology.
- **Semantic:** is based on consistency by the semantic features of the ontology [37] considering the meaning and content of these features [15].
- **Context:** can be evaluated by other web ontologies or specific applications [28] which are in the same domain.
- **Syntactic:** this verification is about the coherence in the ontology definitions [28]. For a deep verification, it is necessary using some criteria [15].
- **Structural, architectural and design:** have the purpose to identify the absence of the main concepts of a domain in the structure, loops, concepts with the same definition but different name, among other errors. In this kind of verification does not exist metrics well defined because could be ambiguous [37] and normally the metrics are context-free [8].

3.2 Validation

The validation is about the ontology definitions which should most real possible model and represents a defined domain [24], i.e., indicates if the ontology definitions are a model of the real world [39]. The validation is made by the next techniques:

- **Gold standard:** In this validation, the ontology is compared with a gold ontology that was created by an expert, where both represent the same domain; it assumes the gold ontology is completed and corrected else the validation could have many errors [17] and is commonly used for ontology which was created by semi-automatic and automatic process [40].

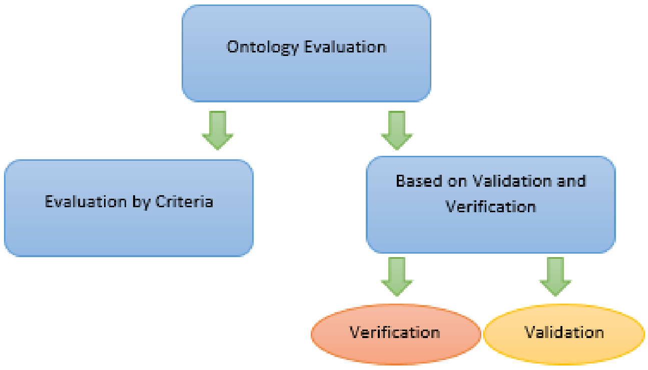


Fig. 2. Ontology Evaluation Approaches.

– **Application based:** the application based validation is a focus for determining if the results of the proposed ontology satisfy or not the task for which it was done [3].
– **Data driven:** for applying this validation is necessary to compare the information of the ontology with existing data about the same domain meanly using a corpus of text documents [3]. A very important aspect in this validation is determining if the ontology has enough elements for representing the domain [15].
– **User based:** basically, this validation is about the experience and perspective around the final user because it becomes suggestive and empiric; for minimizing the error it is advisable the user be an expert in the ontology domain [17].

3.3 Evaluation by Criteria

The evaluation by criteria is a focused technique in aspects or features which can be quantifiable [20] in order to ease the requirements analysis and some of them can be measured by ontology tool as reasoners [40]. In the Table 1 some criteria are described.

3.4 Ontology Evaluation Related Works

Wang et al. [38] developed an ontology using a geographic (spatial and temporal) knowledge of a gazetteer to associate natural hazards news reports by patterns; the main concepts are Happening for describing the processes that occur during

Table 1. Criteria for the ontology evaluation.

Criteria	Description
Lawfulness	About the syntactical error frequency [3]
Richness	Quantifies the most important syntactic features are used in the ontology [3]
Adaptability	How the ontology responses in future uses [20]
Clarity	To communicate about the meaning of the terms independently of the context [20]
Closeness Index	Measures the closeness or structural resemblance among the concepts [14]
Similarity Index	About the proposition correctness into knowledge structure [14]
Accuracy	Measures the representation of knowledge within the ontology in relation to the real world [37]
Coherence	Measure if exist contradictions among the elements of ontology according to logical consistency[37]
Computational Efficiency	Measures the reasoner performance when processing the ontology [15]
Conciseness	About the existence of irrelevant information unnecessary and redundancies [15]
Modularity	Indicates if the ontology was created with an appropriate methodology in order to define if exist reusable components [8]
Tangledness	Measures the distribution of multiple concepts and if exists intersections into the structure [40]
Connectivity	Indicates the most important concepts based on the amount of relationships. [40]
Consistency	Indicates if exists a contradiction among the definitions of the ontological elements [10]
Completeness	About if the content of the ontology is explicit or can be inferred in order to consider it complete [10]
Sensitiveness	Measures how a small change in a definition modifies other elements [10]
Standard Coverage	Identifies elements which are not defined in the ontology [21]
Coupling	About the number of external concepts that are referenced or imported [12]
Coverage	Indicates how well the ontology represents or models the domain [12]

hazard events and object to indicate the involved entities, and the relationships are is-a and part-of type.

In the evaluation ontology area, Ying Shen el at. [29] propose an ontology evaluation approach based on entropy by three elements: data amount, data quality and finally, ontology structure and text visualization, for this, the ontology was seen as a graph where the entropy is measured between two elements and if the value is high it indicates that there is a high redundancy. Another work about evaluation was presented by Djuana [7], he evaluates a folksonomy-based ontology by gold-standard using some important references into state of the art in order to validate the coverage for wider the ontological elements.

There are algorithmic methods for semantic validation of UML (Unified Modeling Language) class diagrams using an ontology as a reference in the domain; the ontology determines if the elements and relationships into diagram are contradictory or not and if are into domain [27].

Barchini et al. [2] propose four dimensions for ontology evaluation since ontology quality approach: descriptive, structural, functional and operative; descriptive dimension refers how well the ontology gives information about its features, structural dimension about how the ontology specifies its elements and definitions, functional dimension about the ontology capacity for giving functions in order to satisfy specific requisites and operational structure determines the ontology capacity for integrating to other physical and logical agents. The ontology proposed by Tovar [33] is a tool for the search information about the social service in a higher level education institution was manually created and validated by the answers of some competency questions.

There is a methodology for ontology evaluation in restricted domains by lexico-syntactic patterns, grouping by formal concept analysis, similarity, latent semantic and dependence graphs using corpora as a reference providing a score based on accuracy measure [36]. Further, the latent semantic analysis is defined as a technique that assumes the words in a common context are semantically related [34].

There are ontology evaluate focus as ontology definitions, the software used for building, share and reusing elements and the documentation about its [9] and some of them are focused to evaluate only ontological elements as semantic relations [35].

4 Ontology Applications Related Works

The representation of the information as the main purpose of an ontology that is why we can many applications about the use of information extraction techniques belonging to NLP (Natural Language Processing) as a tool for ontology designing, creation and instancing. For example, to store the clinical histories of patients, didactic sequence design based on competencies used by an upper middle education professor approach and the creation of management of heterogeneous data system in a university.

Kuna et al. [16] propose an ontology as an extension of an information retrieval system specifically in the query process, this work is designed for scientific document search in the computer sciences domain.

In the medical area, it has proposed to use an ontology for integrating information by heterogeneous way in order to create a repository using a similarity detection algorithm; for creating the mean ontology, it was necessary to mapping the known features to extend the domain [25]. Another application in the same area, it is the use of an ontology for enriching the diabetic patient education in a personalized way [23].

A case of use of an ontology is monitoring of a multi-agent system based on sensors; the ontology saves information about the system process and throws alerts when some sensor does not work on the different physical and logical components of the system [6].

5 Conclusions and Future Work

The ontology is an important standard for information representation, but also when is evaluated can infer knowledge in order to expand itself. We can say the ontology verification is made by measuring features and mathematical functions and the validation is guaranteed by knowledge previously given by an expert in the domain.

In the case of ontology evaluation by criteria, it can use the most appropriated criterion (or a set of them) according the kind of the ontology or its purpose. There is not an only strategy for ontology evaluation that guarantees the best performance, however the evaluation results will depend of a combining between the ontology purpose and the used strategy, in many cases.

Acknowledgement. This work is supported by the Sectoral Research Fund for Education with the CONACyT project CB-257357, and partially supported by the VIEP-BUAP.

References

1. Amith, M.F., He, Z., Bian, J., Lossio-Ventura, J.A., Tao, C.: Assessing the practice of biomedical ontology evaluation: Gaps and opportunities. Journal of biomedical informatics (2018)
2. Barchini, G.E., Álvarez, M.M.: Dimensiones e indicadores de la calidad de una ontología. Avances en Sistemas e Informática 7(1), 29–38 (2010)
3. Brank, J., Grobelnik, M., Mladenić, D.: A survey of ontology evaluation techniques (2005)
4. Chmielewski, M., Paciorkowska, M., Kiedrowicz, M.: A semantic similarity evaluation method and a tool utilised in security applications based on ontology structure and lexicon analysis. In: Mathematics and Computers in Sciences and in Industry (MCSI), 2017 Fourth International Conference on. pp. 224–233. IEEE (2017)
5. Couldry, N., Kallinikos, J.: Ontology (2018)

6. Cruz, A.P.H., Ortega, C.A.P., Granados, L.A.P.: Diseño de una ontología para agentes que monitorean mediciones de sensores. REVISTA COLOMBIANA DE TECNOLOGIAS DE AVANZADA (RCTA) 2(26) (2017)
7. Djuana, E.: Gold-standard evaluation of a folksonomy-based ontology learning model. In: Journal of Physics: Conference Series. vol. 971, p. 012045. IOP Publishing (2018)
8. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: Modelling ontology evaluation and validation. In: European Semantic Web Conference. pp. 140–154. Springer (2006)
9. Gómez-Pérez, A.: Some ideas and examples to evaluate ontologies. In: Artificial Intelligence for Applications, 1995. Proceedings., 11th Conference on. pp. 299–305. IEEE (1995)
10. Gómez-Pérez, A.: Evaluation of ontologies. International Journal of intelligent systems 16(3), 391–409 (2001)
11. Hlomani, H.: Multidimensional Data-driven Ontology Evaluation. Ph.D. thesis (2014)
12. Hlomani, H., Stacey, D.: Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. Semantic Web Journal 1(5), 1–11 (2014)
13. Huitzil, I., Straccia, U., Díaz-Rodríguez, N., Bobillo, F.: Datil: Learning fuzzy ontology datatypes. In: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. pp. 100–112. Springer (2018)
14. Iqbal, R., Murad, M.A.A., Sliman, L., da Silva, C.P.: A mathematical evaluation for measuring correctness of domain ontologies using concept maps. Measurement 118, 73–82 (2018)
15. Jain, S., Meyer, V.: Evaluation and refinement of emergency situation ontology. Int J Inform Educ Technol 8(10), 713–719 (2018)
16. Kuna, H.D., Rey, M., Podkowa, L., Martini, E., Solonezen, L.: Expansión de consultas basada en ontologías para un sistema de recuperación de información. In: XVI Workshop de Investigadores en Ciencias de la Computación (2014)
17. Lourdusamy, R., John, A.: A review on metrics for ontology evaluation. In: 2018 2nd International Conference on Inventive Systems and Control (ICISC). IEEE (2018)
18. Luna, J.A.G., Bonilla, M.L., Torres, I.D.: Metodologías y métodos para la construcción de ontologías. Scientia et technica 2(50), 133–140 (2012)
19. Martinez-Gil, J., Alba, E., Aldana-Montes, J.F.: Optimizing ontology alignments by using genetic algorithms. In: Proceedings of the workshop on nature based reasoning for the semantic Web. Karlsruhe, Germany (2008)
20. McDaniel, M., Storey, V.C., Sugumaran, V.: Assessing the quality of domain ontologies: Metrics and an automated ranking system. Data & Knowledge Engineering 115, 32–47 (2018)
21. Olifer, D.: Evaluation metrics for ontology-based security standards mapping. In: Electrical, Electronic and Information Sciences (eStream), 2015 Open Conference of. pp. 1–4. IEEE (2015)
22. Peña, C.R., Vidal, M.T., González, C.S.V.: Creation of a consulting tool and implementation of an ontology for a master's degree program in computer sciences. Revista Colombiana de Computación 19(1), 29–38 (2018)
23. Quinn, S., Bond, R., Nugent, C.: A two-staged approach to developing and evaluating an ontology for delivering personalized education to diabetic patients. Informatics for Health and Social Care 43(3), 264–279 (2018)

24. Ramos, E., Núñez, H., Casañas, R.: Esquema para evaluar ontologías únicas para un dominio de conocimiento. Enlace 6(1), 57–71 (2009)
25. Ren, S., Lu, X., Wang, T.: Application of ontology in medical heterogeneous data integration. In: Big Data Analysis (ICBDA), 2018 IEEE 3rd International Conference on. pp. 150–155. IEEE (2018)
26. Reyes, C., Tovar, M., Vázquez, S.: Ontology for the description of a masters degree program in computer sciences. In: Proceedings of the XVIII International Conference on Human Computer Interaction, Interacción 2017, Cancun, Mexico, September 25-27, 2017. pp. 12:1–12:4 (2017), http://doi.acm.org/10.1145/3123818.3123836
27. Sadowska, M., Huzar, Z.: Semantic validation of uml class diagrams with the use of domain ontologies expressed in owl 2. In: Software Engineering: Challenges and Solutions, pp. 47–59. Springer (2017)
28. Sekandar, K.: A quality measure for automatic ontology evaluation and improvement. Master's thesis (2018)
29. Shen, Y., Chen, D., Yang, M., Li, Y., Du, N., Lei, K.: Ontology evaluation with path-based text-aware entropy computation. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. pp. 881–884. ACM (2018)
30. Smith, B.: Ontology and information systems, forthcoming in stanford encyclopedia of philosophy. Recuperado de http://ontology. buffalo. edu/ontology (PIC). pdf. Fecha de consulta 20 (2003)
31. Ta, C.D., Thi, T.P.: Automatic evaluation of the computing domain ontology. In: International Conference on Future Data and Security Engineering. pp. 285–295. Springer (2015)
32. Tello, A.L.: Ontologías en la web semántica. España: Universidad De Extremadura (2001)
33. Tovar, M., Flores, J.C., Reyes-Ortiz, J.A.: An ontology for representing information over social service in an educational institution. In: DATA. pp. 391–399 (2017)
34. Tovar, M., Pinto, D., Montes, A., González, G.: An approach based in lsa for evaluation of ontological relations on domain corpora. In: Mexican Conference on Pattern Recognition. pp. 225–233. Springer (2017)
35. Tovar, M., Pinto, D., Montes, A., González, G., Vilarino, D.: Identification of ontological relations in domain corpus using formal concept analysis. Engineering Letters 23(2) (2015)
36. Tovar Vidal, M., Pinto Avendaño, D.E., Montes Rendón, A., González Serna, J.G.: A metric for the evaluation of restricted domain ontologies. Computación y Sistemas 22(1) (2018)
37. Vrandečić, D.: Ontology evaluation. In: Handbook on ontologies, pp. 293–313. Springer (2009)
38. Wang, W., Stewart, K.: Spatiotemporal and semantic information extraction from web news reports about natural hazards. Computers, environment and urban systems 50, 30–40 (2015)
39. YILMAZ, T.K., Durak, U., OĞUZTÜZÜN, H., ÇAĞILTAY, K.: Usability evaluation of a web-based ontology browser: the case of tsont. Turkish Journal of Electrical Engineering & Computer Sciences 26(2), 1115–1128 (2018)
40. Yu, J., Thom, J.A., Tam, A.: Ontology evaluation using wikipedia categories for browsing. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. pp. 223–232. ACM (2007)

ISSN 1870-4069

Robust Algorithm of Clustering for the Detection of Hidden Variables in Bayesian Networks

Niels Martínez-Guevara, Nicandro Cruz-Ramírez, José-Rafael Rojano-Cáceres

Universidad Veracruzana,
Facultad de Estadística e Informática, Mexico

{niemartinez,ncruz,rrojano}@uv.mx
http://www.uv.mx

Abstract. In machine learning there are tree principal areas where the computer is able to learn, the first one is supervised learning, the second one is unsupervised learning and the third one learning by reinforcement. On the first appears something that the literature calls the "learning problem" where they study the most common scenes in relationship to the data and the model, in this paper we purpose a modification of the MS-EM algorithm that estimated a hidden variable from incomplete data in a scene where the model is unknown, and this algorithm is able to get a Bayesian network that explains the presence and relationship of the hidden variable and the dataset.

Keywords. MS-REM, MS-EM, clustering, likelihood, MDL.

1 Background and Introduction

Bayesian Networks or Belief Networks (BN) are graphical representations for probability distribution. They are defined by two components, on the one hand a directed acyclic graph where each node represent a variable of the dataset, and the arc represents a conditional independence between variables. On the other hand, there is a collection of *local interaction models* that describe the conditional probability of each variable [9].

Drawing a Bayesian network from experts can be an expensive process in large applications, however it is also possible to get it from data. Unfortunately the current methods are essentially limited to learn parameters for a fixed network structure [6].

Thus, in order to get a Bayesian Network it is possible to use the Model Selection Expectation-Maximization Algorithm, which also it is capable to obtain hidden values from data. To introduce the concept of hidden values we can take a example given by Heckerman; supposing we have a dataset from a sale cars company, where it is stored information about the last sales, we want to know the economic status of the clients but such information is not implicit in the data. However there are two variables with a big impact to estimated the hidden variable: The price of the car and the monthly payment of the car. Thus, the social economic status of the client can be inferred as better if the price of the car and the monthly payment are high, in other case if the price of the car and monthly payment are low the social economical status is worst [8].

In this paper we analyze the MS-EM algorithm a variant of the clustering algorithm EM that proposes finding an alternative to solve two problems that arise during the process of clustering. The first one is how many values can take a hidden variable? In the special case of EM one of its virtues can also become a big problem, being a soft-clustering algorithm the number of clusters is dynamic and the traditional version of this algorithm obtained his parameters randomly in a range of values with a normal distribution $2 \leq c \leq n$ where c is the number of clusters and n the number of cases in the database, allowing in the first instance the case of duplicate clusters [4].

From a first random experimentation, the algorithm will adjust its values to find the values of the variable, the initial number of clusters c is a random value within a range, the results of the algorithm will be variable and it will be difficult to determine the number of possible values that variable can take since it presents us in the first instance two scenarios, the first, being the latent variable a variable with two possible values ??or the other end being each case of the data set pertaining to a cluster. This can be a big problem because once the initial number of clusters is defined, it can be reduced, however it can not be increased, altering the behavior of the algorithm and generating noise in the execution of the algorithm [11].

The proposal of MS-REM is that the numbers of clusters can be obtained from more to less by adjusting their number until it is adequate, while finding a network that maximizes the value of MDL, implementing an adaptive algorithm such as REM (Robust Expectation-Maximization) which is updating its values? and the number of clusters depending on the iterations of the algorithm, adjusting its probabilities to have a better Log-Likelihood.On the other hand, the original purpose of MS-EM algorithm takes the ability to see the relationship of the variable with the proposed model. In this way, we intend to implement a technique for obtaining the number of clusters and their model.

2 Analysis of the MS-EM Algorithm

Now, the proposal offered by the literature is as follows, if a Bayesian network could be found that explains the data and shows the presence of a hidden variable, as well as the relationship with the observed data, the two problems would be solved. The proposed algorithm is as follows [7]:

```
Procedure MS-EM:
    Choose M(0) and params(0) randomly
        Loop for n=0,1,... until convergence
        Find a Model Mn+1 that maximize Q(:Mn, params n)
        let params n+1 = arg max Q(Mn+1, params : Mn, params n)
```

where M is a Bayesian network and parameters is the probability distribution of the hidden variable. Now the algorithm shows only two steps, so next we will analyze the Expectation function where is represented by the formula 1 and it's the maximum likelihood function [10]:

$$argmaxQ = P(H|M, \theta). \tag{1}$$

Having an initial model M and a random initialization of clusters, we seek to obtain the parameters or the probability θ that each tuple belongs to a cluster H. The traditional version of EM indicates that the number of clusters will be calculated randomly and will be a number between $2 < c < n$ and the probability of each cluster is assigned randomly according to the initial experiment. In each iteration the probabilities of the clusters will be adjusted until there is no significant change, which can be defined in the error variable ε, which is the termination condition of the algorithm [5].

The next step is maximization "Find a Model" Finding a Bayesian network is a problem with high computational complexity, there is a considerable number of possible networks that can represent a database, so to obtain a solution a search algorithm was implemented that was able to find a "better structure" that represent better the relationships between the variables seen in the observe data. To do this we use the Random Mutation Hill-Climbing algorithm (RMHC). And to compare two or more Bayesian structures, the MS-EM algorithm uses the MDL (Minimum Description Length) metric which is describe in the formula 2 [2]:

$$MDL(B_S, D) = -N \cdot H(B_S, D) - \frac{1}{2}K \cdot \log N, \tag{2}$$

where $K = \sum_{i=1}^{n} q_i \cdot (r_i - 1)$ y $H(B_S, D) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} \frac{N_{ijk}}{N} \log \frac{N_{ijk}}{N_{ij}}$.

This formula expresses the following: N is the total number of cases, $H(B_S, D)$ is the Likelihood and $-\frac{1}{2}K \cdot \log N$ is the penalty value where preference is given to networks with fewer arcs and where the value q_i is referring to the number of combinations of the values of variables that can be made between a parent and child node, in the cases of N_{ijk} and N_{ij} referring to the formula are joint frequencies of the parent nodes with respect to the children, which are analysed in the database D, given the structure B_S [2].

Through this formula a negative value will be obtained that the closer it is to zero gets a better is the fitness, following the Ockham's razor principle the better network it's the simples network that represents better the data [3].

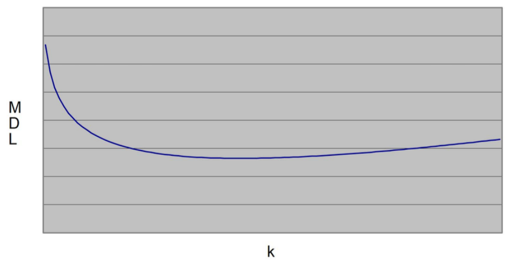


Fig. 1. MDL graph of behavior [3].

With all the formulas and metrics explained, the algorithm can be developed as follows:

1. Step 1: Randomly generate a model and a probability distribution, where the number of clusters is a random value between $2 < c < n$ and the sum of the mean of each cluster is equal to 1.
2. Step 2: Obtain a new network that improves the value of MDL through a RMHC search algorithm.
3. Step 3: Obtain a new probability distribution from the Maximum Likelihood with the new network.
4. Step 4: The two Bayesian networks are compared according to the MDL metric.
5. Step 5: if $(New\_MDL - Current\_MDL! = 0)$ AND $(t < 30)$then
   (a) The new network replaces the current one and returns to Step 2.
6. Else: Return the best distribution of probability and best network.

The previous algorithm is a variant of the EM algorithm where we try to find the "parameters" that we will call Probability Function (Maximum Likelihood) and the structure of the data (Bayesian Network) through the search algorithm.

However, certain limitations that the clustering algorithm has when calculating the number of clusters are still present. One of the proposals to solve this problem is the algorithm called Robust Expectation-Maximization [7].

3 MS-REM Algorithm Construction

During the design phase for the proposal that unites the ideas of an algorithm that focuses on obtaining a categorization from its model and a robust clustering algorithm, it was possible to design the MS-REM algorithm (Model Selection Robust Expectation Maximization) where the characteristics offered by the literature with respect to this algorithm were added, most of these modifications are focused on the calculation of clusters since in their initial configuration, the number of them is no longer a random value between 0 and n, but starts with an initial value of n, where n is the number of cases observables.

The algorithm starts taking into account that it is in the worst case and from it generates a random experiment, where each cluster has a probability of appearing $1/n$ in this first iteration it is possible that some media clusters have a probability of appearing equal to 0 and in other cases they will increase, thus generating larger clusters. This is why the adjustment functions of the algorithm were optimized to eliminate the smallest clusters (the literature suggests eliminating clusters whose mean is less than $1/n$) encouraging that larger clusters have a greater probability of appearing in the next iterations.

Although the iterative process appears to be slower in the first instance, the algorithm starts from the most general to the specific, thus encouraging the data to determine the appropriate groupings for each data set. However, this also has its weaknesses, since in subsequent investigations it was possible to verify in data generated by artificial distributions, that the algorithm allows small clusters, thus generating noise in the classification process cite soor.

With those changes in mind in the MS-REM clustering algorithm is presented as follows:

1. Step 1: Randomly generate a model and a probability distribution, where the **number of clusters is n** and the sum of the mean of each cluster is equal to 1.
2. Step 2: **The number of Clusters is adjusted, eliminating those clusters whose mean is less than 1 / n and the clusters are re-named.**
3. Step 3: Obtain a new network that improves the value of MDL, through the RMHC search algorithm.
4. Step 4: Obtain a new probability distribution from the Maximum Likelihood with the new structure.
5. Step 5: The two Bayesian networks are compared according to the MDL value.
   if $(New_MDL - Current_MDL == 0)$then
   Returns The best probability distribution and best network.
6. Else:
   – if $t < 60$ then: **The network is replaced and returns to Step 2.**
   – Else: **The network is replaced and returns to Step 3.**

When initializing with an initial number of clusters n, it is necessary to make a double adjustment in the algorithm, the first to determine the number of clusters and the second to determine its size.

4 Experimental Results

4.1 Tests with Synthetic Databases

In a first stage of experimentation, two Bayesian networks were defined 3x1x3 (figure 2a) and 3x2x4 (figure 2b) with five random probability distributions and different sample sizes (250,500,1000,2000).

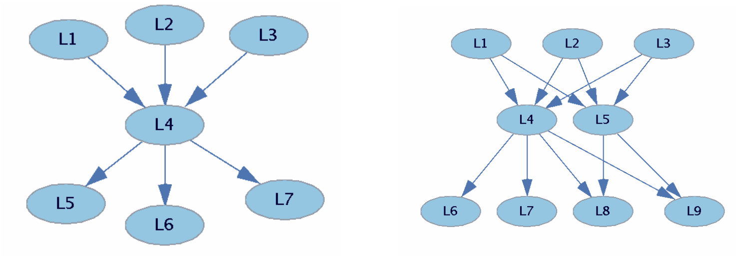


Fig. 2. a. Structure 3x1x3 and b. structure 3x2x4.

For a model, 20 different Dataset were obtained and it is intended to observe the behavior of up to three hidden variables. A total of 120 evaluations were made for all data sets. Subsequently, the values of Log loss were grouped by sample size and number of hidden variables and thus an average was obtained. Then the average Log loss of each structure was plotted by sample size and number of variables.

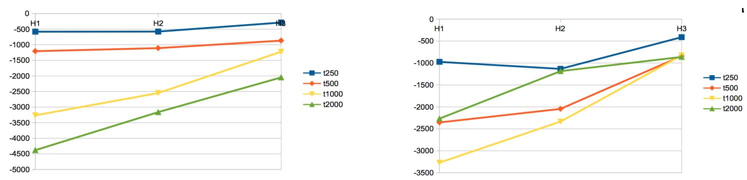


Fig. 3. Log Loss with MS-REM in 3x1x3 and 3x2x4 networks with tree hidden variables.

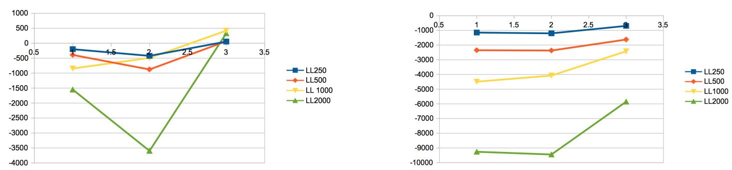


Fig. 4. Log Loss with MS-EM in 3x1x3 and 3x2x4 networks with tree hidden variables.

In these graphs (Figure 3 and Figure 4) we can see if the number of hidden variables is higher is more difficult for the algorithm to find a network with a better MDL and since it starts to find more complex networks or that do not they adequately represent the data, which causes their MDL values to be higher.

This phenomenon can be seen in the graphs in H3, the Log Loss values are closer to zero.

4.2 Performance Statistics

To test the functioning of MS-REM, it was given to different databases that are described in Table 1, all of them belonging to the UCI database, where the class variable is known, this variable was eliminated and 30 independent executions to analyze their results, MS-REM is a random algorithm and it is necessary to perform statistical tests to study its behavior [1].

Table 1. Features of the data sets. Where C represents a given class.

	Iris	B. Cancer	Car	User	Wine
Num. Classes	3	2	4	4	3
Num. Instances	150	286	1728	431	178
Num. Variables	5	10	6	7	13
Dist. Classes	C1(0.33); C2(0.33); C3(0.33)	C1(0.70); C2(0.30)	C1(0.7); C2(0.24); C3(0.03); C4(0.03)	C1(0.12); C2(0.3); C3(0.28); C4(0.3)	C1(0.33); C2(0.39); C3(0.28)

Table 2. Statistical table of the MS-REM algorithm and MS-EM with the Iris database, in relation to MDL and number of clusters (NC).

	MS-REM		MS-EM	
Iris	**MDL**	**NC**	**MDL**	**NC**
Mean	**-693.778**	**10.44**	-735.512	9.967
Median	**-677.550**	**12**	-791.589	13
Best	**-635.196**	**3**	-639.931	3
Worst	**-784.546**	**18**	-811.760	17
SD	**43.905**	**4.183**	44.1580	3.978

From the previous experiment we can see that the MS-REM algorithm presents a better behavior with respect to the original version in databases where the number of values in classes is greater and the distribution of each class is smaller. As with Iris, Wine and User. While in databases such as Car or B. Cancer where a predominant class is found, it performs worse than the original algorithm. Also the size of the database influences, in large databases such as Car the algorithm has a process of adjustment much greater than the original version.

Table 3. Statistical table of the MS-REM algorithm and MS-EM with the Wine database, in relation to MDL and number of clusters (NC).

	MS-REM		MS-EM	
Wine	**MDL**	**NC**	**MDL**	**NC**
Mean	**-3411.179**	**15.5**	-3556.06285	17.2
Median	**-3439.514**	**20**	-3444.62	15
Best	**-3264.215**	**7**	-3427.12	6
Worst	**-3554.198**	**24**	-3746.34	23
SD	**62.614**	**3.900**	139.929	4.209

Table 4. Statistical table of the MS-REM algorithm and MS-EM with the B. Cancer database, in relation to MDL and number of clusters (NC).

	MS-REM		MS-EM	
B. Cancer	**MDL**	**NC**	**MDL**	**NC**
Mean	-4573.242	17.53	**-4472.014**	**14.26**
Median	-4578.216	15	**-4425.009**	**14**
Best	-4365.966	16	**-4256.403**	**8**
Worst	-4958.715	23	**-4601.068**	**14**
SD	164.413	3.381	**126.774**	**2.947**

4.3 Convergence Graphs

Next, we will show the adjustment behavior of the MS-REM algorithm based on the MDL value from the first iteration until reaching convergence. The following graphs represent the average iteration of the performance statistics.

Through these graphs it can be seen how the algorithm starts with a low vale of MDL in all the databases because the algorithm has the worst panorama of clustering as the initial configuration and as it is approached through the adjustment functions to zero. Obtaining a better adjustment of the hidden variable with the observable data and relation with the observable variables through the Bayesian network that is changing at the same time as the means of the clusters until reaching its optimum.

Table 5. Statistical table of the MS-REM algorithm and MS-EM with the User database, in relation to MDL and number of clusters (NC).

	MS-REM		MS-EM	
User	**MDL**	**NC**	**MDL**	**NC**
Mean	**-948.357**	**11.67**	-968.727	11.4
Median	**-946.897**	**11**	-952.361	12
Best	**-906.993**	**7**	-913.339	8
Worst	**-1000.513**	**8**	-1239.14843	14
SD	**19.934**	**2.397**	72.895	2.955

Table 6. Statistical table of the MS-REM algorithm and MS-EM with the Car database, in relation to MDL and number of clusters (NC).

	MS-REM		MS-EM	
Car	**MDL**	**NC**	**MDL**	**NC**
Mean	-21511.723	26.3	**-21446.681**	**21.5**
Median	-20647.463	18	-21796.175	30
Best	-20647.463	18	**-19717.940**	**2**
Worst	-22751.763	34	**-22543.114**	**38**
SD	598.659	4.62	848.772	12.20

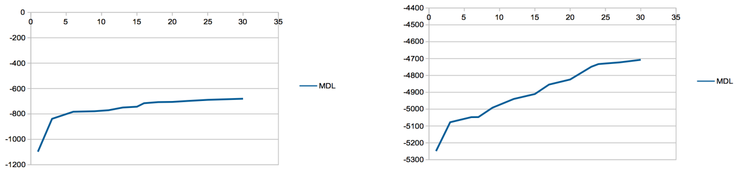


Fig. 5. Convergence graph for Iris and Breast cancer Data sets.

5 Conclusions and Discussion

In the construction of the MS-REM algorithm was presented, different tests were performed to see how to compare the functionality of the original algorithm with the one proposed in this document and see in this way if the modifications of this algorithm are better or didn't show a significant change. In the first place, a test was carried out proposed by the literature where the algorithm was executed with synthetic databases and the distribution of the data was known, we analyzed the loss of information recorded by the two algorithms. In the second place, tests the algorithms with databases of the UCI repository and the clustering process performance was analyzed of the two algorithms.

The MS-REM algorithm initialize in the worst of the cases, this allows its adjustment process to accommodate small clusters, thus causing class divisions to exist in categorical databases. However, what the author proposes is that there may be an infinite number of hidden variables within a database and that the algorithm is only able to find the "most important" at a given a initial

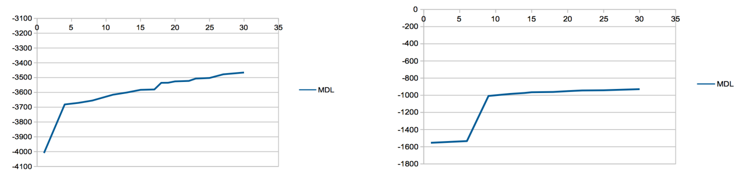


Fig. 6. Convergence graph for Wine and User knowledge modeling Data sets.

configuration and that is why the Bayesian network is changing according to the behavior of the algorithm [7].

Something that shows the literature about the hidden variables is one of its virtues is the fact that it helps us to explain a phenomenon represented in the data [8]. However, as the author mentioned in the original proposal of the algorithm, the stochastic factor within the initial configuration of EM, means that we only find the best variable for a given initial configuration, causing the study of these variables to be complex, because it is difficult to determine that the variable obtained in a certain execution is the most relevant in the database [7].

However, despite this drawback, it is an interesting proposal given that the variables found by the algorithm are obtained entirely from the data and, beyond parameter configuration of the algorithm, the user has no interaction to modify the behavior of the algorithm, thus providing interesting solutions that could explain a phenomenon, offer an alternative of how the data is being categorized or even obtain the values of a variable of great impact on the database.

References

1. Asuncion, A., Newman, D.: Uci machine learning repository (2007)
2. Bouckaert, R.R.: Probabilistic network construction using the minimum description length principle. In: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty. pp. 41–48. Springer (1993)
3. Cruz-Ramírez, N., Acosta-Mesa, H.G., Mezura-Montes, E., Guerra-Hernández, A., de Jesús Hoyos-Rivera, G., Barrientos-Martínez, R.E., Gutiérrez-Fragoso, K., Nava-Fernández, L.A., González-Gaspar, P., Novoa-del Toro, E.M., et al.: How good is crude mdl for solving the bias-variance dilemma? an empirical investigation based on bayesian networks. PloS one 9(3), e92866 (2014)
4. Friedman, J., Hastie, T., Tibshirani, R.: The elements of statistical learning, vol. 1. Springer series in statistics Springer, Berlin (2001)
5. Friedman, N.: The bayesian structural em algorithm. In: Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence. pp. 129–138. Morgan Kaufmann Publishers Inc. (1998)
6. Friedman, N., Goldszmidt, M.: Learning Bayesian networks from data. Morgan Kaufmann (1999)
7. Friedman, N., et al.: Learning belief networks in the presence of missing values and hidden variables. In: ICML. vol. 97, pp. 125–133 (1997)
8. Heckerman, D.: A tutorial on learning with bayesian networks. In: Learning in graphical models, pp. 301–354. Springer (1998)
9. Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Elsevier (2014)
10. Wackerly, D., Mendenhall, W., Scheaffer, R.L.: Mathematical statistics with applications. Nelson Education (2007)
11. Yang, M.S., Lai, C.Y., Lin, C.Y.: A robust em clustering algorithm for gaussian mixture models. Pattern Recognition 45(11), 3950–3961 (2012)

ISSN 1870-4069

Design of an Acquisition System for Monitoring the Current of a GEM Detector

Rabí Soto-Camacho[1], David Pinto Avendaño[1], Sergio Vergara-Limon[2], Maria-Aurora Vargas-Treviño[2], Guy Paic[3]

[1] Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Puebla, México

[2] Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Electrónica, Puebla, México

[3] Universidad Nacional Autónoma de México, Instituto de Ciencias Nucleares, CDMX, México

rabi87_soto22@hotmail.com, dpinto@cs.buap.mx, svergara2@hotmail.com, auroravargast@hotmail.com, Guy.Paic@cern.ch

Abstract. In the development of this work we describe the design proposal of a useful tool for the monitoring of the GEM detector panels in order to measure the current consumption and in this way to know the useful life of the detector, due to in theory the GEM doesn't produce discharges but in real life there are some small discharges (nano and micro amperes) that deteriorate the detector. In this proposal it is also contemplated to design a software to analyze the interaction of electromagnetic fields within a waveguide that represents the pathways within a printed circuit board using the study of electromagnetic theory in the development of a simulator, which will allow to analyze the effects of these fields in the propagation of the signals that are obtained and to determine if the data obtained are adequate to allow their subsequent analysis.

Keywords. Electromagnetic interferences, current monitor, EMI analysis.

1 Introduction

A GEM (gas electron multiplier) is made of a plate (50 microns) of Kapton sheet, a copper coating on each side, surface perforated, each acting as a channel for electron multiplication. Each hole has a bi-conical structure with an outer diameter of 50 microns and a step of 140 microns (see Fig. 1).

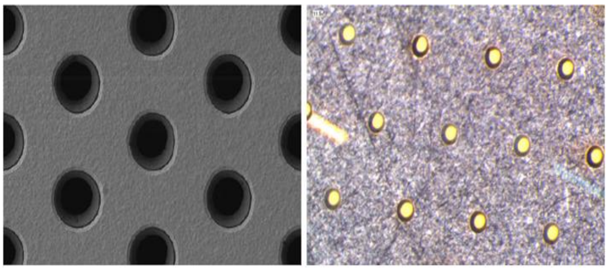


Fig. 1. Microscopic Image of a GEM.

A charged particle or a photon that interacts in the gas of the drift region produces an ionization cluster of positive ions and electrons. Due to the applied electric field, the group of positive ions of the drift goes towards the cathode, while the group of electrons goes towards the holes of the first sheet of GEM.

Here the charge multiplies and while the amplified charge towards the second GEM is frustrated a large percentage of the positive ions produced in the holes is captured by the upper GEM electrode. Thus, the multiplication channel is released quickly and groups of electrons multiply. The gain is proportional to the sum of the GEM voltages (see Fig. 2) [1].

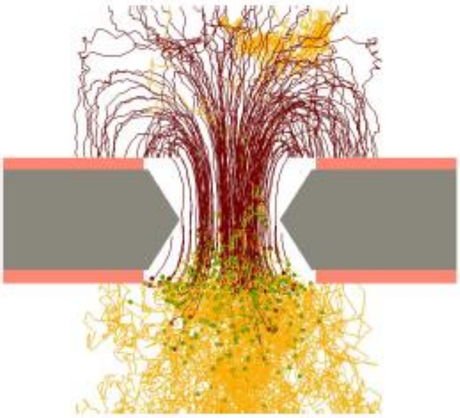


Fig. 2. GEM Operation.

To carry out the measurement of the current consumption of these detectors requires specialized devices for this. This document proposes the development of a current monitor to monitor this consumption, as well as the development of a software to carry out the analysis of electromagnetic fields that can be presented on the monitor.

1.1 Current Monitor

The current monitor is an electronic circuit that serves to monitor current consumption for the system, nowadays there are some devices in the market, but they are not suitable for monitoring each channel of the GEMs and to record the current variations of the detector due to the features of the detector. A prototype was found in the litherature but our proposal in the design is different so the features of the prototype could be useful to compare the performance design [2]. We need to develop a prototype that must be capable to measure the small discharges (nano and micro amperes) and also not interfere in the normal work of the detector. The proposed system is shown in figure 3.

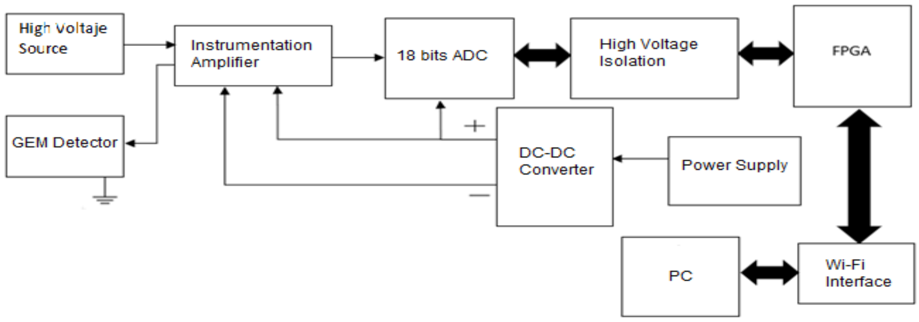


Fig. 3. Current Monitor Block Diagram.

The system must be connected to the high voltage and the GEM detector for which high voltage connectors are required.

The voltage levels necessary to use the current meter card are obtained from the use of a DC-DC converter, in this case in addition to conditioning the card voltage, it allows isolating the high and low voltage supplies. Later a voltage regulator is used in order to condition the operating voltage of the ADC.

Currently there are some software for the analysis of electromagnetic fields, but their costs are very high for this reason a key part of the design of this tool is to provide a tool for analysis at a lower cost. On the other hand, when carrying out the experiments an enormous quantity of data is generated, which must be processed for its study, for this reason the development of this research topic is centered as the first point in the development of a software for analysis of electromagnetic fields to be able to study the effects of said fields on the acquired signals and thus be able to obtain adequate data for their analysis. As a second point of this research project once the analysis is performed and appropriate data are generated, these must be processed to obtain those essential for study and storage [3, 4, 15, 16].

2 Proposal for Developing

For the development of this topic, a block diagram with the work methodology to be followed is presented below:

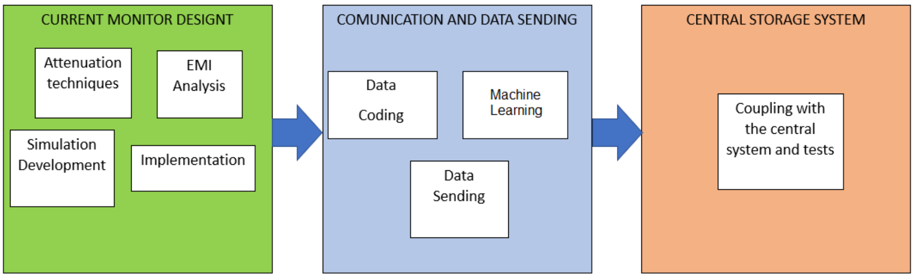


Fig. 4. Proposal Block Diagram.

In the previous diagram, 3 action blocks are presented that define the process to be followed in this work.

A. In the design stage of the Current Monitor various activities are carried out in order to obtain a functional system as a result.

- The study of the techniques for the reduction of electromagnetic interference is done in order to have the adequate knowledge of how electromagnetic fields interact within a medium and thus be able to attenuate them in order to obtain desired signals without noise.
- The analysis of electromagnetic interferences is done to see how the signals are transmitted in different media, as well as to know how they are attenuated in these.
- The software design for the analysis of electromagnetic fields is done in order to determine how these affect the signals and to perform an analysis of it to obtain adequate values of signals for further analysis through the use of data cleansing techniques.
- The instrumentation of the current monitor is performed in order to visualize its operation and observe the interferences that may occur due to the electromagnetic fields inside the circuit and outside.

B. In the Data Processing stage, the following stages are presented:

- In the data coding stage, the 8b / 10b coding is performed in order to have a standard coding pattern with the data.
- he Machine Learning stage is carried out because when making measurements with the system, an important quantity of data is required, therefore it is necessary to use machine learning techniques to classify and find patterns in these.
- The stage of sending data is necessary since it is necessary to store the data once classified so that it can be reused for another type of subsequent analysis.

C. The storage stage is performed in order to have the available data to be able to perform some kind of processing or subsequent analysis.

For the design stage of the current monitor, the study of techniques and theories has been carried out in order to obtain a functional device.

2.1 Propagation of Electromagnetic Waves

The electromagnetic radiation and its propagation are described by the Maxwell's equations with the electric field perpendicular to the magnetic field. These equations determine the effects of transmission, reflection, attenuation and dispersion of the propagated wave, the Maxwell equations are shown in Table 1

Table 1. Maxwell's equations.

Differential form	Integral form	
$\nabla \cdot \vec{D} = \rho_v$	$\oint D \cdot ds = \int_v \rho_v dv$	Gauss Law
$\nabla \cdot \vec{B} = 0$	$\oint B \cdot ds = 0$	Gauss Law (H)
$\nabla x \vec{E} = -\frac{\partial \vec{B}}{\partial t}$	$\int_L E \cdot dl = -\frac{\partial}{\partial t} \int_s B \cdot ds$	Faraday's Law
$\nabla x \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}$	$\int_L H \cdot dl = \int_s \left(J + \frac{\partial D}{\partial t}\right) \cdot ds$	Ampere's Law

where we have:

$\vec{E} =$ Vector electric intensity or electric field [V / m].
$\vec{B} =$ Magnetic flow or magnetic induction $\left[T = {Wb}/{m^2}\right]$.
$\vec{H} =$ Vector of magnetic intensity or magnetic field [A / m].
$\vec{J} =$ Current intensity or conduction [A/m^2].
$\vec{D} =$ Electric flow or displacement current [C/m^2].
q = Electric charge.

The description of this implies that Faraday's law indicates that a magnetic field can produce a current in a closed path if the magnetic flux is changing in time. In relation to Lenz's law, the voltage induced by the magnetic flux has such a polarity that the current established in a closed path gives rise to a flux that opposes the magnetic flux change. Ampere's law states that the magnetic field is generated by both the conduction current and the displacement currents. The first equation of Gauss's law states that the electric flow that passes through any closed surface is equal to the total charge enclosed in the surface. Gauss's law for the magnetic field states that the magnetic flux lines are closed and do not end in a magnetic charge.

The propagation of electromagnetic waves depends on two factors:

- Of the nature of the medium (permittivity or dielectric constant, magnetic permeability and conductivity).
- Of the electromagnetic wave frequency.

The system of equations that relate and describe the electromagnetic propagation with the nature of the medium are:

$$\vec{D} = \varepsilon \vec{E} \quad (1)$$

$$\vec{B} = \mu \vec{H} \quad (2)$$

$$\vec{J} = \sigma \vec{E} \quad (3)$$

$$\varepsilon = \varepsilon_0 \varepsilon_r \quad (4)$$

$$\mu = \mu_0 \mu_r \quad (5)$$

where:

ε= Dielectric permittivity [F/m].

μ= Magnetic permittivity [H/m].

ε_0, μ_0= They correspond to the values of the vacuum.

ε_r, μ_r= They correspond to the values relative to the medium.

σ= Electric conductivity [S/m=1/Ωm].

The dielectric permittivity is a measure of the degree of resistance of the medium to a load flow and is determined by the relationship between the electric displacement and the intensity of the field that produces it.

The Magnetic permeability is the ability of a medium to make way for magnetic field lines and is determined by the relationship between magnetic induction and the external magnetic field [5, 6, 8, 11-14, 17].

2.2 Electromagnetic Interference in Electronic Circuits

Electromagnetic interference (EMI) can be defined as electromagnetic signals that disturb (unintentionally) the normal operation of an electrical or electronic system.

The types of interference, measurement methods and tolerated limits are specified by international standards and are established according to the frequency band of interference.

- Low frequency disturbances: (f <10 KHz), emitted by the network and power supplies.
- Disturbances in band of 10-150 KHz: By impulses of intensity and transitory phenomena of tension switches, etc.
- Disturbances in band of 150 KHz to 30 MHz: By impulses of intensity and transitory phenomena of tension switches, the propagation is by radiation and coupling.
- Disturbances in band of 30 MHz to 300 MHz: Propagation by radiation.
- Disturbances in band of 500 MHz to 18 GHz: It propagates by radiation and usually appears in communications equipment or switching logic circuits.

By an EMI it can be understood as the presence of unwanted voltage or current that can appear in a device as a result of the operation of another device or natural phenomenon.

The coupling between systems consists of the interaction and disturbance of one device depending on another.

The interference can be radiated or conducted, when the noise originates in a source or travels through the air is called EMI radiation, when the noise driven travels through a conductor, as a power line. The noise may have been radiated, deposited in the lines and then driven [6 ,7, 9, 10-13].

3 Conclusions and Future Work

By continuously monitoring the current consumption, the degradation of the GEM can be known since the GEM in initial conditions has a specific current consumption and the more current consumption is visualized, it gives us an idea of how degraded the material is.

It is possible to design and implement a device that allows monitoring the current consumption of the order of nanoamperes using electromagnetic interference attenuation techniques in order to determine if what we are monitoring is noise or a defined signal.

By knowing the techniques and theories about the attenuation of electromagnetic interference it is possible to develop a useful software tool that allows us to know the degree of attenuation of the interferences using certain techniques.

The detection of phenomena of electromagnetic type will allow to improve the state of the art in the design of printed circuits since having a tool that allows us to know the attenuation of electromagnetic interferences can avoid the remaking of circuits due to faults in the medium.

References

1. Santos-Rodrigues, A.T.S.: Gas Electron Multiplier. CERN, Sitio web: https://kt.cern/technologies/gas-electron-multiplier (2016)
2. Schweda K.: Alice TPC GEM High-Voltage System: Engineering Design Review. Sitio Web: https://indico.cern.ch/event/677082/contributions/2772028/attachments/1562708/2464403/hv-edr-schweda.pdf (2017)
3. Software FEKO para análisis de circuitos, disponible en: http://www.indielec.es/feko-cms-4-50-376/ (2017)
4. Software ZUKEN para análisis de circuitos, disponible en: https://www.zuken.com/en (2017)
5. Reitz, J., Milford, F., Christy, R.: Fundamentos de la teoría electromagnética. Tercera edición, Editorial Addison Wesley
6. Resnick, R.: Physics Vol. 2, Tercera edición, Editorial, Continental S.A. de C.V. Mexico.
7. Félix, M.A., Salazar, D., Valenzuela, J.A., Márquez, H.: Análisis de propagación de campos electromagnéticos en guías de onda por medio un método para barrido de parámetro V. SOMI Congreso de Instrumentación, XVIII, disponible en: ftp://www.ece.buap.mx/pub/Secretaria_Academica/ACADEMIA/ASISTENCIA%20A%20EVENTOS%20ACAD%C9MICOS%20Y%20PERMISOS%20TEMPORALES/CONGRESO%20SOMI%20XVIII/PDF/Microondas/71.pdf, 1–11 (2005)

8. Chattopadhyay, S.: Generation of magnetic and electric field measurement and its analysis. Proceedings of the International Confeerence on Electromagrietic Interference and Compatibility, De INCEMIC-99, Sitio Web: https://ieeexplore.ieee.org/document/871631, 219–224 (1999)
9. Wang, C., Chen, Y., Zhang, G., Zhou, Z.: Design of Printed-Circuit Board Rogowski Coil for Highly Accurate Current Measurement. In International Conference on Mechatronics and Automation, Sitio web: http://ieeexplore.ieee.org/document/4304180/ (2007)
10. Sato, T., Matsunaga, T., Maeda, H., Watanabe, M.: Application Examples of Electromagnetic Field Analysis at Fujitsu. FUJITSU Sci. Tech Sitio web: https://www.fujitsu.com/global/documents/about/resources/publications/fstj/archives/ vol 48(4) (2012)
11. Park, H.H., Jang, H.T., Park, H.B.: Simulation and Design of a PCB-Chassis System for Reducing Radiated Emissions. ACES Journal, 26, 679–687 (2011)
12. De Burgh, G.: Making EMI Analysis Part of the Design Process. COTS Journal Sitio web: http://archive.cotsjournalonline.com/articles/view/100226 (2004)
13. Gao, S., Zhang, Z., Bo, R.: Research on the magnetic property of high-speed PCB design. Journal of Chemical and Pharmaceutical Research Sitio web: http://www.jocpr.com/articles/research-on-the-magnetic-property-of-highspeed-pcb-design.pdf (2015)
14. Zumbahlen, H.: Printed Circuit-Board Design Issues. Linear Circuit Design Handbook. sitio web: https://www.oreilly.com/library/view/linear-circuit-design/9780750687034/, 1201–1297 (2008)
15. Antillanca, E.H., Cerda, N.G.: Propuesta Metodológica para el Desarrollo de Software de Investigación. Universidad de Santiago de Chile Sitio web: http://sitios.diinf.usach.cl/kddp/files/2012/10/Paper-Propuesta-de-M%C3%A9todo-para-Sw-de-Investigaci%C3%B3n-Infonor-2010.pdf (2012)
16. Chavarriaga, L.J., Arboleda, J.H.: Modelo de Investigación en Ingeniería del Software: Una propuesta de investigación tecnológica. Universidad San Buenaventura Sitio web: http://www.emn.fr/x-info/harbol07/MIFISIS2004.pdf (2004)
17. Mologni, J.: Compatibilidad electromagnética. Sitio web: http://www.esss.com.br/blog/es/2016/08/compatibilidade-electromagnetica/ (2016)

ISSN 1870-4069

Cyberbullying Detection in Social Networks: A Multi-Stage Approach

Laura P. Del Bosque, Sara E. Garza

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica,
San Nicolás de los Garza, NL, Mexico

{laura.delbosquevg,sara.garzavl}@uanl.edu.mx

Abstract. Cyberbullying, defined as the violent harassment of an individual towards a victim in electronic media, is a serious problem nowadays. If not prevented or mitigated, it can lead to affective disorders, poor academic performance, problems in social relationships, and — ultimately — to suicide attempts in youngsters and children. Because manual supervision in social networks (a space where cyberbullying can naturally occur) is laborious, automated approaches for cyberbullying detection are desirable. However, a considerable number of approaches treat this problem as merely aggressive text message detection, without considering the frequency of the harassment. The approach proposed in this work, in contrast, views cyberbullying detection as a process of three consecutive stages: aggressive text message detection, alleged aggressor and victim detection, and cyberbullying case detection. This approach was tested using a dataset extracted from Twitter; the approach obtained an F-score of 0.947 for the positive (cyberbullying) case.

Keywords. Cyberbullying detection, data mining, text mining, machine learning, sentiment analysis.

1 Introduction

Information technologies allow users to communicate with each other efficiently and to share their ideas and resources mutually. This, at the same time, shortens physical distance and enables collaboration, among other benefits. However, technology also presents a negative face by permitting misbehavior — and even cruelty — to occur more often. A vivid example of the former is given by cases of *cyberbullying*.

Bullying, which was traditionally limited to face-to-face encounters among children and youngsters in school yards, has unfortunately made its entrance into social media under the modality of cyberbullying [37]. Cyberbullying is a malicious, deliberate, repetitive act caused with electronic text messages [22], the main differences with respect to traditional bullying being the use of technology, the capability of reaching a great audience, the lack of time limits in the aggression, and the possibility of the aggressor to feel protected behind the technology. All of this makes cyberbullying more aggressive than traditional bullying [35], and — primarily due to the unsupervised use of technology — gives it the potential to easily get out of control [2], specially in social networks [30], whose rapid growth encourages this pernicious phenomenon [19].

Cyberbullying has been associated with negative experiences, as different studies have shown that victims report a poor academic performance, problems with family and other social relationships, and affective disorders [17, 36, 18]. In the worst case, cyberbullying can lead to suicide attempts when victims cannot cope with emotional distress due to the experienced abuse, aggression, and humiliation [3, 5]. In that sense, cyberbullying can lead to serious consequences that should not be underestimated.

Considering that cyberbullying is *an aggressive intentional act performed by an individual or group of individuals using electronic media to repeatedly contact a defenseless victim* [14], there are three main components in this phenomenon: an aggressive act (object), actors (aggressor-victim), and repetition (a pattern). Even though these three components seem to be present in several definitions [31, 22, 14], works tend to treat cyberbullying detection as a mere problem of aggressive text detection [1, 28, 7, 10, 11] by jumping to the conclusion that, if an aggressive act is taking place, then a harassing pattern has developed. The proposed approach, instead, views cyberbullying detection as a multi-stage *process* and attempts to identify this phenomenon in social networks using several progressive steps: (1) aggressive text message identification, (2) candidate bully and victim identification, and (3) cyberbullying case identification. The first stage relies on a profanity-based unsupervised technique, while the second one uses aspects from graph theory, and the third one is based on outlier detection.

The remainder of this work is organized as follows: Section 2 introduces relevant notions in network theory and Twitter, which is the case study for this work; Section 3 presents related work, Section 4 describes the proposed approach, Section 5 presents experiments and results, and Section 6 discusses conclusions and future work.

2 Background

The current section aims at briefly describing necessary concepts and notation from graph theory and Twitter. While graph theory is essential for social network analysis, understanding basic concepts from Twitter will serve for understanding the case study of this work.

2.1 Graph Theory

Networks are mathematically represented with *graphs*. A graph $G = (N, E)$ is a collection of entities (called *nodes*) and connections (called *edges*). Whenever edges are not bidirectional — that is, the presence of an edge (u, v) does not guarantee the presence of the edge (v, u) — the graph is said to be *directed*. If the edges contain numerical labels, the graph is said to be *weighted* as well. The *degree* of a node is the number of edges attached to this node; in a directed graph, the *in-degree* is the number of edges that enter a node and the *out-degree* is the number of edges that leave a node. A *multigraph* is a graph where multiple edges involving the same pair of nodes are possible.

A *subgraph* $G_s = (N_s, E_s)$ is a portion of a graph where $N_s \subset N$ and $E_s \subset E$. An *edge-induced* subgraph contains all edges in E_s and the nodes attached to these edges.

2.2 Twitter

Twitter[1] is a popular online social networking site where users are available to post short messages termed *tweets*. A user can *follow* other users, this meaning that the user will see the tweets from the followed users in his (her) personal webpage (called *timeline*). Tweets can be *directed* towards a specific set of users by including the addressed user names in the tweet; in Twitter, user names are preceded by the "@" (for example `@ausername`).

3 Related Work

Since the proposed approach covers three successive stages (isolated messages, actors, and cyberbullying cases through message histories), the discussed related work covers either of these stages and is either devoted to cyberbullying or to a similar phenomenon (e.g. pedophilia). For a more comprehensive review on approaches for cyberbullying automated detection, the reader is referred to the work by Salawu et al. [29] (in an *online-first* format to this date).

As mentioned in Section 1, cyberbullying detection has been recast as an aggressive text detection problem. This problem, in turn, has been mostly treated as a profanity detection problem (which is actually a simplification), where profanity is understood as the utilization of foul language. In both scenarios, the detection has been leveraged by machine learning and sentiment analysis techniques. One of the first approaches to address the aggressive text detection problem — outside the context of cyberbullying — was the Smokey system [34], a rule-based system used to identify hostile messages (also called "flames") in e-mail. Other outstanding approaches include the use of classifier ensembles [4], multi-step classifiers [27, 33], and the use of different sets of features — such as statistical, semantic, and linguistic [15]. Within the context of cyberbullying, Al-Garadi et al. [1], Ptaszynski et al. [26], and Yin et al. [38] propose basic machine learning techniques with different features, such as content and sentiment; similarly, Reynolds et al. [28] attempt to extract features to detect cyberbullying. Dinakar et al. [11], in contrast, train different topic-specific classifiers (race, sexuality, religion, intelligence, and physical attributes) to detect the aggression.

[1] Available at `http:\\twitter.com`.

In a later work, Dinakar et al. [10] propose the use of common sense reasoning to treat the case of aggression without the use of profanity.

There are also works that take into account the set of actors involved. For example, Chisholm [6] reports that women with aggressive communication styles tend to exclude their target victim and start conspiring against this victim, while aggressive men tend to use threatening words and phrases more frequently towards their victims. Furthermore, Dadvar et al. [7, 8] report that men and women use foul words with different frequency (men use certain words and women use other words); based on this finding, the authors implement different classifiers based on gender to detect aggressive text. Nahar et al. [19] present a two-step approach for cyberbullying detection, which starts with aggressive message detection and then utilizes the network-based HITS algorithm [16] to detect the actors involved.

Other types of works examine conversations or message histories to detect different types of phenomena. For example, Potha et al. [25] focus on sexual cyberbullying with a methodology based on time-series (a previous work on this approach being given by Potha and Maragoudakis [24]), where histories are represented symbolically and aligned with the aim of detecting a predator pattern. Also, Peersman et al. [23] use a three-stage system to detect cases of pedophilia; other works attempt to characterize pedophile conversations by means of features [12, 21].

4 Approach

The proposed approach aims at identifying cases of cyberbullying in social media by employing a set of successive stages. These stages arise from the definition of *bullying* given by Smith et al. [31]: *An aggressive intentional act performed by an individual or group of individuals repeatedly against a victim.* From this definition, three main components can be highlighted: aggressiveness, actors involved (aggressor-victim), and the repetition of the aggression. In that sense, cyberbullying can be seen as composed by *objects* (aggressive messages), *subjects* (aggressors and victims), and a *pattern*. Taking all of the former into account, the proposed approach consists of three stages:

1. Aggressive message detection
2. Alleged aggressor and victim detection
3. Cyberbullying case detection

The first stage concerns the identification of messages with aggressive content (media other than text remaining as future work) in a *social network*. A social network is considered as a structure composed by individuals that share a cybernetic relationship and have the capability of sending personal messages to each other [19]. At this stage, who sends and who receives the message and with what frequency is irrelevant.

The second stage, which is fed from the results of the first one, concerns the identification of alleged aggressors and victims. The former are defined as sending aggressive messages and the latter, conversely, as receiving aggressive messages. The message histories of these subjects are then more thoroughly analyzed.

The third stage, which takes in the subjects identified at the second stage, concerns the identification of cyberbullying cases by analyzing the frequency and intensity at which alleged aggressors harass alleged victims. This intensity is compared against the intensity of other conversations to search for an abnormal or outlier pattern. If a cyberbullying case is detected, the aggressor, victim, messages, and dates of these messages are returned as output.

The proposed approach uses a sampling technique similar to snowball sampling [20], since a set of messages is extracted from the social network according to several criteria (explained in Section 5); from these messages, a number is selected by the detection algorithm and sampled further according to other criteria (belonging to a certain user or pair of users). This sampling technique, along with the three stages, is explained in the following.

4.1 Aggressive Text Message Detection

The detection of aggressive text messages was performed using the approach proposed by Del Bosque and Garza [9], which concerns an unsupervised lexicon-based, term-counting strategy that identifies profane words. In summary, this approach, given a set M of messages, assigns a score sc_i to each message m_i. Such approach, even though simple, was selected to work around the *imbalanced class* problem; moreover, the aggressiveness score is be necessary for the second and third stages. This approach, in addition, has shown to yield satisfactory results for cyberbullying detection.

4.2 Alleged Aggressor and Victim Detection

In the second stage of the approach, the message sender and receiver become relevant. In that sense, a user that sends messages with a particular frequency and aggressiveness score is considered as an alleged aggressor or bully, while a user that receives messages with a particular frequency and aggressiveness score is considered as an alleged victim. For this case, that particular frequency is two or more messages within M (considering repetition) and that particular aggressiveness score is $sc_i \geq 5$ (considering this is the middle point of the scale used by Del Bosque and Garza [9]).

Formally, if the social network is treated as a directed multigraph (see Figure 1), $E = M$ and N is the subgraph of users induced by E — in other words, each node of the graph is a user of the social network and the users are connected by the messages they direct to each other (only the sample gathered is visible). The weight ω_i of each edge corresponds to its aggressiveness score such that $\omega_i = sc_i$. Let $\deg_{\text{in}}(v, \alpha)$ denote the number of incoming edges to node v where $sc_i > \alpha$, i.e. $\deg_{\text{in}}(v, \alpha) = |\{e_i : sc_i > \alpha, e_i = (u, v)\}|$ and, conversely, let $\deg_{\text{out}}(v, \alpha)$ denote the number of outgoing edges from v where $sc_i > \alpha$, i.e. $\deg_{\text{out}}(v, \alpha) = |\{e_i : sc_i > \alpha, e_i = (v, u)\}|$. Consequently, the set A of alleged aggressors is defined as $A = \{v : \deg_{\text{out}}(v, \alpha) = \beta\}$ and the set V of alleged victims is defined as $V = \{v : \deg_{\text{in}}(v, \alpha) = \beta\}$, where $\alpha = 5$ and $\beta = 2$ as previously stated. Aggressor-victim relationships are then formed by extracting the possible aggressor for each $v \in V$ and the possible victim for each $a \in A$ such that $R = \{(a, v) : a \in A \vee v \in V\}$. Note that, while several of these possible aggressors and victims may not be in the A or V sets, in the third stage a deeper analysis will confirm or deny the existence of a cyberbullying case.

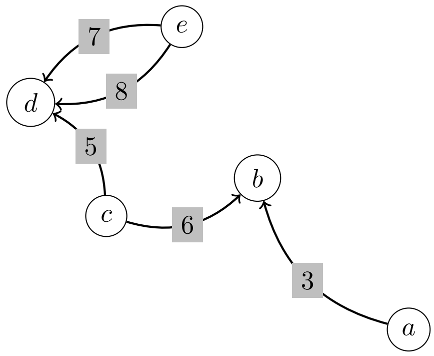


Fig. 1. Directed multigraph where nodes represent users and edges represent message aggressiveness scores. In this case, d is an alleged victim and $\{c, e\}$ comprises a set of alleged aggressors.

4.3 Cyberbullying Case Detection

While the first stage of the proposed approach works at the message level and the second one works at the user level, the third stage works at the historic level. A *message history* H_{au}, in this case, is defined as the set of messages directed from an alleged aggressor a to a user u in date d: $H_{au} = \{m_i : m_i = (a, u)_i, (a, u)_i \in E\}$ (date can be expressed as a function $f_d(m_i) = d$). In the third stage, an aggressive pattern (abnormal situation that spans for a time period) is searched for between an alleged aggressor and an alleged victim. When this pattern exists, a cyberbullying case is confirmed.

To detect an aggressive pattern, first the message histories between the alleged aggressor and each of the aggressor's contacts (including the alleged victim) are fetched for a period of preceding k time units (the selected k being six months, considering this is a reasonable amount of time for harassment to take place and also due to social network API restrictions). Next, an aggressiveness average score is calculated per history:

$$\bar{sc}(a, u) = \frac{\sum_{i=1}^{i=|H|} sc_i}{|H|}, \tag{1}$$

where $H = H_{au}$. Let also $\bar{sc}(a, v)$ correspond to a score where v is a possible victim.

Considering that an aggressor's contacts conform a set C_a, the average score for an alleged aggressor is given by:

$$\bar{sc}(a) = \frac{\sum_{u \in C_a} \bar{sc}(a, u)}{|C_a|}. \tag{2}$$

Then, if $\bar{sc}(a, v) > \bar{sc}(a)$, a case of cyberbullying is detected between a and v where a is the aggressor, v is the victim, and H_{au} contains the messages and dates of the case.

For example, assume that $C_a = \{b, c, v\}$, $H_{av} = \{m_1, m_2, m_3\}$, $sc_1 = 6$, $sc_2 = 8$, and $sc_3 = 9$; then, $\bar{sc}(a, v) = \frac{6+8+9}{3} = 7.66$. Let $\bar{sc}(a, b) = 3.5$ and $\bar{sc}(a, c) = 4$. Then $\bar{sc}(a) = \frac{7.66+3.5+4}{3} = 5.05$. Since $7.66 > 5.05$, a cyberbullying case is reported between a and v with history H_{av}.

5 Experiments and Results

The purpose of the experiments was to detect cyberbullying instances by means of the multi-stage approach and validate these instances using evaluators. Experiments were performed using Twitter, which is a popular social network prone to cyberbullying attacks due to its public, unsupervised nature.

5.1 Setup

A message dataset was collected from Twitter using the API and the methodology described in the works by Del Bosque and Garza [9] and Escalante et al. [13], which in summary consists of obtaining directed messages with seed words (aggressive words, usually) and manually annotating these messages. An extensive list of aggressive words (shown in Table 1) was used to collect messages; these words were gathered with the aid of a native English speaker. A total of 13,313 messages was collected for this dataset. With the proposed approach, it was possible to detect ten alleged aggressors, whose message histories were tracked for six months. With these histories, cyberbullying cases were detected using Eqs. 1 and 2 (an example of a case is presented in Table 2).

Table 1. Aggressive words used to gather messages for cyberbullying cases.

c*nt	wh*re
punk as* b*tch	f*ggot
f*cking f*ggot	f*cking sl*t
f*cking c*nt	motherf*cker

Table 2. Cyberbullying case (bully: @user1, victim: @user2).

Message	Date
@user2 fat as*	May 14
@user2 f*cking fa*g*t sh*t bruh	May 14
@user2 sup wetback	May 1st.
@user2 DUMB*SS WETBACK IM CALLING THE FBI	April 16

A group of these detected cases was presented to a set of nine evaluators (according to Snow et al. [32], a minimum of seven non-expert evaluations are required to emulate the evaluation by an expert); collections of messages not detected by the approach as cyberbullying were also handled to the evaluators, hence the group of cases contained both *positive* (cyberbullying) and *negative* (no cyberbullying) instances. Note that manual annotation is being made *after* the instances are being detected and not *before*, as it is usual in machine learning training data. Therefore, we are validating that what the approach said was positive (case of cyberbullying) is actually marked as positive by a set of evaluators and what the approach did not detect as positive (no cyberbullying) is marked as negative by the evaluators. A total of 26 instances, where 18 were positive and 6 negative, was presented to the evaluators. To obtain the global class of an instance, the majority vote was taken into account; therefore, if an instance was voted as positive by eight evaluators, then the instance was globally classified as a cyberbullying case by the evaluators.

5.2 Results and Discussion

Table 3. Confusion matrix.

		Classified		
		as Positive	as Negative	Total
Class	Positive	18	2	20
	Negative	0	6	6
	Total	18	8	26

Table 3 presents the confusion matrix obtained from the evaluation. As it can be observed from this matrix, the results obtained by the approach closely match the manual evaluation; as a consequence, precision, recall, and F-score (F_1) present considerably high values.

Even though the results look promising, it is also important to keep in mind that this is only a small fraction of a vast social network that receives thousands of messages from thousands of users on a daily basis, thus making this kind of approach like looking for a needle on a haystack. It is possible to miss cases and, consequently, loose recall. However, on the other hand, being able to track

Table 4. Result evaluation.

Class	Precision	Recall	F_1
Positive	0.9	1	0.947
Negative	1	1	1
Average	0.95	1	0.97

these *real* cases on this haystack, regardless of the number of cases, should not be overlooked.

6 Conclusions and Future Work

A multi-stage approach for cyberbullying detection was presented. The first stage identifies aggressive text messages with an unsupervised, profanity-based algorithm; the second stage uses results from the first stage, as well as concepts from graph theory to identify alleged aggressors and victims, and the third stages analyzes alleged aggressor message histories and uses outlier detection to identify cyberbullying cases. The aggressive text message detection algorithm and the multi-stage approach were both validated through a set of experiments; while the algorithm is competitive against different techniques, the multi-stage algorithm is promising for further development. With regard to this, several lines of future work are possible. On one hand, it would be desirable to test the approach against similar methods to have a wider view on its effectiveness. Also, supervised methods could be tested for the aggressive message detection stage (a challenge here would be to address the class imbalance problem), since these methods showed good results. Moreover, techniques such as time series, Big Data, and deep learning could be incorporated into the approach to make it more scalable, robust, and capable of handling large amounts of data. Finally, a series of applications could take advantage from either the overall approach or individual stages; for example, an add-on in social networks could be inserted to warn users before posting an aggressive comment (i.e. the application would ask users if they are sure they want to post the comment, given that a high level of aggressiveness was detected on this comment).

References

1. Al-Garadi, M.A., Varathan, K.D., Ravana, S.D.: Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network. Computers in Human Behavior 63, 433–443 (2016)
2. Alvarez-García, D., Nuñez Perez, J.C., Dobarro González, A., Rodríguez Pérez, C.: Risk factors associated with cybervictimization in adolescence. International Journal of Clinical and Health Psychology 15(3), 226–235 (SEP 2015)
3. Bauman, S., Toomey, R.B., Walker, J.L.: Associations among bullying, cyberbullying, and suicide in high school students. Journal of Adolescence 36(2), 341–350 (2013)

4. Bchir, O., Ismail, B., Maher, M.: Verbal offense detection in social network comments using novel fusion approach. AI Communications 28(4), 765–780 (2015)
5. Campbell, M.A.: Cyber bullying: An old problem in a new guise? Australian Journal of Guidance and Counselling 15(1), 68–76 (2005)
6. Chisholm, J.F.: Cyberspace violence against girls and adolescent females. Annals of the New York Academy of Sciences 1087(1), 74–89 (2006)
7. Dadvar, M., de Jong, F., Ordelman, R., Trieschnigg, R.: Improved cyberbullying detection using gender information. In: Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012). pp. 23–26. University of Ghent, Ghent (2012)
8. Dadvar, M., Trieschnigg, D., Ordelman, R., de Jong, F.: Improving cyberbullying detection with user context. In: Serdyukov, P., Braslavski, P., Kuznetsov, S., Kamps, J., Rüger, S., Agichtein, E., Segalovich, I., Yilmaz, E. (eds.) Advances in Information Retrieval, Lecture Notes in Computer Science, vol. 7814, pp. 693–696. Springer Berlin Heidelberg (2013)
9. Del Bosque, L.P., Garza, S.E.: Aggressive text detection for cyberbullying. In: Gelbukh, A., Espinoza, F.C., Galicia-Haro, S.N. (eds.) Human-Inspired Computing and Its Applications. pp. 221–232. Springer International Publishing, Cham (2014)
10. Dinakar, K., Jones, B., Havasi, C., Lieberman, H., Picard, R.: Common sense reasoning for detection, prevention, and mitigation of cyberbullying. ACM Transactions on Interactive Intelligent Systems (TiiS) 2(3), 18–48 (2012)
11. Dinakar, K., Reichart, R., Lieberman, H.: Modeling the detection of textual cyberbullying. The Social Mobile Web 11(02), 11–17 (2011)
12. Eriksson, G., Karlgren, J.: Features for modelling characteristics of conversations: Notebook for PAN at CLEF 2012. In: CLEF 2012 Evaluation Labs and Workshop, Rome, Italy, 17-20 September 2012 (2012)
13. Escalante, H.J., Villatoro-Tello, E., Garza, S.E., López-Monroy, A.P., y Gómez, M.M., nor Pineda, L.V.: Early detection of deception and aggressiveness using profile-based representations. Expert Systems with Applications 89(Supplement C), 99 – 111 (2017), `http://www.sciencedirect.com/science/article/pii/S0957417417305171`
14. Espelage, D.L., Swearer Napolitano, S.M.: Research on school bullying and victimization: What have we learned and where do we go from here?[mini-series]. School Psychology Review 32(3), 365–383 (2003)
15. Justo, R., Corcoran, T., Lukin, S.M., Walker, M., Torres, M.I.: Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web. Knowledge-Based Systems 69, 124–133 (2014)
16. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM 46(5), 604–632 (Sep 1999)
17. Kowalski, R.M., Limber, S.P.: Psychological, physical, and academic correlates of cyberbullying and traditional bullying. Journal of Adolescent Health 53(1, Supplement), S13–S20 (2013), the Relationship Between Youth Involvement in Bullying and Suicide
18. Machmutow, K., Perren, S., Sticca, F., Alsaker, F.D.: Peer victimisation and depressive symptoms: can specific coping strategies buffer the negative impact of cybervictimisation? Emotional and Behavioural Difficulties 17(3-4), 403–420 (2012)
19. Nahar, V., Xue, L., Chaoyi, P.: An effective approach for cyberbullying detection. Communications in Information Science and Management Engineering 3(5), 238–247 (2012)

20. Newman, M.E.: Networks: An Introduction. Oxford University Press, Oxford (2010)
21. Parapar, J., Losada, D.E., Barreiro, A.: A learning-based approach for the identification of sexual predators in chat logs. In: CLEF (Online Working Notes/Labs/Workshop) (2012)
22. Patchin, J.W., Hinduja, S.: Bullies move beyond the schoolyard: A preliminary look at cyberbullying. Youth violence and juvenile justice 4(2), 148–169 (2006)
23. Peersman, C., Vaassen, F., Van Asch, V., Daelemans, W.: Conversation level constraints on pedophile detection in chat rooms. In: CLEF (Online Working Notes/Labs/Workshop) (2012)
24. Potha, N., Maragoudakis, M.: Cyberbullying detection using time series modeling. In: 2014 IEEE International Conference on Data Mining Workshop. pp. 373–382 (Dec 2014)
25. Potha, N., Maragoudakis, M., Lyras, D.: A biology-inspired, data mining framework for extracting patterns in sexual cyberbullying data. Knowledge-Based Systems 96, 134–155 (MAR 15 2016)
26. Ptaszynski, M., Dybala, P., Matsuba, T., Masui, F., Rzepka, R., Araki, K.: Machine learning and affect analysis against cyber-bullying. In: Proceedings of the 36th AISB. pp. 7–16 (2010)
27. Razavi, A., Inkpen, D., Uritsky, S., Matwin, S.: Offensive language detection using multi-level classification. In: Farzindar, A., Keselj, V. (eds.) Advances in Artificial Intelligence, Lecture Notes in Computer Science, vol. 6085, pp. 16–27. Springer Berlin Heidelberg (2010)
28. Reynolds, K., Kontostathis, A., Edwards, L.: Using machine learning to detect cyberbullying. In: 2011 10th International Conference on Machine Learning and Applications and Workshops. vol. 2, pp. 241–244 (Dec 2011)
29. Salawu, S., He, Y., Lumsden, J.: Approaches to automated detection of cyberbullying: A survey. IEEE Transactions on Affective Computing Pending(Pending), 1 (2018)
30. Slonje, R., Smith, P.K., Frisén, A.: The nature of cyberbullying, and strategies for prevention. Computers in Human Behavior 29(1), 26–32 (2013), including Special Section Youth, Internet, and Wellbeing
31. Smith, P.K., Slee, P., Morita, Y., Catalano, R., Junger-Tas, J., Olweus, D.: The nature of school bullying: A cross-national perspective. Psychology Press, London (1999)
32. Snow, R., O'Connor, B., Jurafsky, D., Ng, A.Y.: Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks. In: Proceedings of the conference on empirical methods in natural language processing. pp. 254–263. ACM, Nueva York, United States (2008)
33. Sood, S.O., Churchill, E.F., Antin, J.: Automatic identification of personal insults on social news sites. Journal of the American Society for Information Science and Technology 63(2), 270–285 (2012)
34. Spertus, E.: Smokey: Automatic recognition of hostile messages. In: AAAI/IAAI. pp. 1058–1065 (1997)
35. Sticca, F., Perren, S.: Is cyberbullying worse than traditional bullying? Examining the differential roles of medium, publicity, and anonymity for the perceived severity of bullying. Journal of youth and adolescence 42(5), 1–12 (2012)
36. Tokunaga, R.S.: Following you home from school: A critical review and synthesis of research on cyberbullying victimization. Computers in Human Behavior 26(3), 277–287 (2010)

37. Walrave, M., Heirman, W.: Cyberbullying: Predicting victimisation and perpetration. Children & Society 25(1), 59–72 (2011)
38. Yin, D., Xue, Z., Hong, L., Davison, B.D., Kontostathis, A., Edwards, L.: Detection of harassment on web 2.0. In: Proceedings of the Content Analysis in the WEB. pp. 1–7 (2009)

ISSN 1870-4069

A Review of Indoor Navigation Systems for a Bipedal Robot: Preliminary Results

Miguel-Angel Ortega-Palacios, Josefina Guerrero-García,
Juan-Manuel González-Calleros

Benemérita Universidad Autónoma de Puebla,
Laboratorio de Ingeniería del Lenguaje y del Conocimiento, Puebla, Mexico

miguel.ortegap@alumno.buap.mx,
[jguerrero,juan.gonzalez]@cs.buap.mx

Abstract. The navigation process of human is quite different from robots. Semantic meaning is what we used to model the environment. One of the most important elements human uses to model the environment is the vision. The keyframes are the elements that human use to model the environment. Therefore, we do not need a specific coordinate to move around. We can only use vision to navigate to anywhere we want to go. To take advantage of the concept that human model the environment, we propose a method which imitates the mapping and navigation process of the human. We will use computer vision to do the mapping of the environment and wayfinding to navigate it. The environment model is created with data collected with an odometer that identifies the obstacles on the route of the robot. Our research proposal is to create an algorithm to avoid obstacles. In this paper, we focus on reviewing existing solutions to address our research problem which is a system mounted on a biped robot navigating indoor environments on its own, without having to use preset routines. We have reviewed existing literature and presented it to help future researchers who had the same goal and could be used as a starting point.

Keywords. Navigation, locomotion, mobile robot, biped robot, indoor environment.

1 Introduction

As more mobile robots start integrating in different areas of society they will need to operate in a wide variety of environments. Often, these environments will be dynamic; objects move and structures physically change. Less dynamic environments can be characterized by physical changes that occur over the course of days, weeks, or months, whereas more dynamic environments involve continuously moving objects, such as, humans or vehicles.

Mobile robots cannot assume the world is static if we expect them to work effectively. In order for a mobile robot to work autonomously in a dynamic setting it must have a way to sense its surroundings. Camera sensors are ubiquitous among modern robots and can provide a great deal of information about the environment. Robotics applications can employ techniques such as computer vision to accomplish tasks such as object recognition, 3D reconstruction, and mapping and localization [1]. Nowadays, most of the robots require moving and performing tasks in a variety of environments which are sometimes even unpredictable. Mobile robot navigation is a challenging problem in the robotics field and numerous studies have been endeavored resulting in a considerable number of solutions. The navigation problem consists of four integral parts namely perception, localization, motion control and path planning.

The path planning is the determination of a collision-free path in a given environment which is often a cluttered environment in a real world situation.

A lot of different path planning approaches have been proposed and tested in various environments with static [2, 3, 4, 5, 6, 7] and dynamic obstacles [8, 9, 10, 11, 12]. These include both classical approaches like potential fields, cell decomposition, bug algorithm, road map and heuristic approaches like neural networks, fuzzy logic, and wavelets [13]. The remaining of this paper is structures as follows: the review of the state of the art is presented in section 2, a review of navigation of mobile robots is presented in section 3, the proposed methodology is presented in section 4, a discussion of the advantages of walking mecnanisms are presented in section 5, the last section concludes with a discussion of the future work.

2 State of the Art

A mobile robot is a device that is capable of locomotion. It has the ability to move around its environment using wheels, tracks, legs, or a combination of them. It may also fly, swim, crawl, or roll. Mobile robots are used for various applications in factories (automated guided vehicles), homes (floor cleaning devices), hospitals (transportation of food and medications), in agriculture (fruit and vegetable picking, fertilization, planting), for military as well as search and rescue operations. They address the demand for flexible material handling, the desire for robots to be able to operate on large structures, and the need for rapid reconfiguration of work areas [14].

Generally, the first step to analyzing the robot gait on any surface is to analyze the gait on flat surfaces. The flat surfaces in this case mean the horizontal plane and the planes with different slope in relation to the horizontal. The ability to detect obstacles by the robot is closely related to sensors that allow to “see” them.

The analyzed type of gait is a statically stable gait. It means that the robot does not fall over if all motors are stopped at any time. For the robot to remain stable, it must have at least three support points [15].

An autonomous robot could define as a device with partial or complete autonomy, which can do specific tasks using one or several algorithms previously programmed.

A fully autonomous device may or may not be controlled by a user machine interface (IMU). The main motivations for the development of robots are based on: unlimited work space, free environment, and the permanent ability to adapt [16].

Robots can play a major role in navigating undulated landmine-infested terrains. Design of mine detection and clearance robots demand special types of locomotion for maneuvering on uneven surfaces. Various modes of locomotion are possible for robots for such special applications, such as legged movement, wheeled movement, track motion, and combinations thereof called hybrid locomotion. Hybrid locomotion uses the advantages of multiple forms of locomotion within a single robot and switches between the various types of locomotion according to the requirements and the terrain conditions. Most hybrid robots move on wheel-leg combinations and can be of two types. In the first type, the wheel is attached to the end of the leg. In this case, the robot is capable of walking using the legs with the wheel movement arrested. In the second case, the wheel and the legs are combined in such a way that they can be operated independently based on the requirement [17]. In the review of the literature, various investigations have been carried out on the navigation of mobile robots in indoor [18, 19, 20, 21, 22, 23, 24, 25, 26] and outdoor environments [20, 26]. Simultaneous Localization and Mapping (SLAM) [18, 19] and topological map [17] are mainly used to obtain an accurate position of the robot, using sensors, lasers [18] or cameras [24] to detect and recognize obstacles, even relying on a computer vision system such as Tensor Flow^{TM} [19, 22]. The projects are completely based on the open source software ROS (Robot Operating System) [19, 24], some authors propose a neural network to compare images [17, 18, 20, 22, 23] by using deep learning techniques [20, 23, 25]. Other works use a local Wi-Fi network to determine the locations of the devices based on the connection data provided by the access points as the devices move through the environment. Finally some authors present a direct method of navigation that uses gradient descending on the difference between two points specified in the space of slow characteristics [27] and an autonomous local navigation and positioning system based on an artificially established magnetic gradient for in-grid or desktop applications [28].

3 Navigation of Mobile Robots

Navigation is essential for a mobile robot and mobile robot navigation is a challenging problem in the robotics field and numerous studies have been endeavored resulting in a considerable number of solutions [31]. It is hard to imagine a mobile robot without the ability of navigation. For example, if a mobile robot wants to search an object in an indoor environment, it has to move around in different rooms without collide into any furniture. Navigation gives a mobile robot more freedom to complete tasks, and makes it more useful and intelligent. If a mobile robot lacks of the ability of navigation, the functions will be highly limited.

The term navigation refers to the guidance of the mobile robot from the starting position to the target position avoiding collisions and unsafe conditions. It consists of three major steps, self-localization, path planning, and map building. Three main anxieties concerning robot navigation problems are efficiency, safety, and accuracy. Path planning is a vital step of the motion control and navigation of the mobile robots. Path planning is categorized as an NP-complete problem and several heuristic based methods have been implemented, such as the application of artificial neural networks, particle swarm optimization (PSO), genetic algorithm (GA), and hybridization between them.

One of the main benefits of heuristic based methods is that it can yield satisfactory results quickly, which is particularly appropriate to solve NP-complete problems.

The path planning is separated into two main fields, global and local path planning. On the first hand, in path planning with local path planning, the calculations of the path are achieved when the mobile robot is in motion; that means, the calculation is fit for generating new paths as the environment changes. On the other hand, with global path planning, the environment should be totally recognized and identified, while the terrain must be static. Path following problems are primarily concerned with the design of control laws that steer an object to reach and to follow a geometric path, while a secondary goal is to force the object moving along the path to satisfy some additional dynamic specifications.

Mobile robots systems are highly dynamic systems that end up inserting several constraints in the control system, such as response speed, precision, mobility constraints, computational cost, computational power, maneuverability, and control stability issues.

As a matter of fact, navigation for robots is complicated. It involves mapping, localization, and path planning. Mapping nowadays highly relies on Simultaneous Localization and Mapping (SLAM), which is used to create high precision metric maps. This kind of map can be presented in 2D or 3D. Metric maps are dedicated to provide rich details and model the environment as precise as possible. Different navigation methods are:

SLAM: A kind of map that robot can use to model the environment is metric map, which is created by SLAM method. Metric map is normally consist of a good number of pixels, and each pixel is marked as an obstacle, a free space or an unknown area. This kind of map tries to be as precise as possible so that robot can do navigation based on it safely.

Semantic Map: One of the most important drawback of a metric map is that it lacks of semantic meaning. Robot can only use this kind of map to model the environment instead of understanding it. Kitchen, living room or restroom cannot be efficiently represented by metric map because a metric map is composed of lots of coordinates. The similar situation is that normally longitude and latitude is not helpful for people who want to reach a place. Therefore, semantic map is a kind of map which tries to store as many semantic meanings as possible.

Topological Map: Another drawback for a metric map is that it needs high computation power because it models the environment precisely. However, topological map is known as its lightweight data usage. This kind of map integrates semantic meaning with space relations together [14].

3.1 Legged Robots

Most mobile robots used wheels or tracks and the limitations of these wheel and track vehicles when on tender ground or tough terrain have been acknowledged ever since they displaced horses and mules. Six-legged robots can be used as search and rescue robots, space robots and discover robots. Legged robots can be used for rescue work after earthquakes and in hazardous places such as the inner of a nuclear reactor, giving biologically stimulated autonomous legged robots terrific potential.

Legged locomotion is fine if there are many depressions and rises that require leg lifting motions to overcome them, given that legged locomotion can move forward by striding over obstacles at how the robot climbs up the steps and its navigation in an unknown environment, observed with the aid of a discussion on the experimental results before concluding [30].

3.2 Walking Mechanisms

Modern humanoid robots can already execute such tasks autonomously, providing the approximate state of the environment is known in advance. However, it is still difficult for modern humanoid robots to perform such tasks without some prior information about the environmental conditions that can be exploited by a programmer to prepare the humanoid robot for the execution of multiple tasks. Integration and continuous sequencing of multiple robot actions remains a problem and some degree of teleoperation is still needed when performing longer task sequences.

There are several reasons why humanoid robots are thought to be interesting:

- Human environments are built for humans, therefore a general-purpose robot designed for human environments, e.g., homes, factories, hospitals, schools, etc., should have a form similar to humans to successfully operate in such environments.
- It is more natural for humans to interact and communicate with robots that look and behave in like humans.
- A humanoid robot can serve as an experimental tool to test the theories about human behavior created by computational neuroscientists, interested in how the human brain operates.

The foremost is the problem of biped locomotion and balance. Unlike other robots, humanoid robots must walk and keep balance during their operation. In the afore mentioned robotics challenge, locomotion turned out to be one of the biggest issues. The basic indicator that describes the balance of a humanoid robot is the concept of zero-moment point, usually abbreviated as ZMP. The concept of ZMP was introduced by Miomir Vukobratovi´c in 1968. It is still the most widely used approach for generating dynamically stable walking movements in which the supporting foot or feet keep contact with the ground surface at all times. This is important to prevent the robot from falling.

Biped Locomotion: It is an important topic in humanoid robotics. Here we focus on walking, which is distinguished from other forms of biped locomotion such as running by the constraint that at least one foot must always be in contact with the ground. Most of the modern humanoid robots exploit the zero-moment point principle to generate stable walking patterns.

Based on the concept defined in Fig. 1 there are two distinct phases in the gait cycle: when both feet are in contact with the ground, the robot is in double support phase. The feet do not move in this phase. Once one of the feet starts moving, the robot transitions from double to single support phase, in which one of the two feet moves. The single support phase is followed by another double support phase once the foot in the swing phase establishes a contact with the ground.

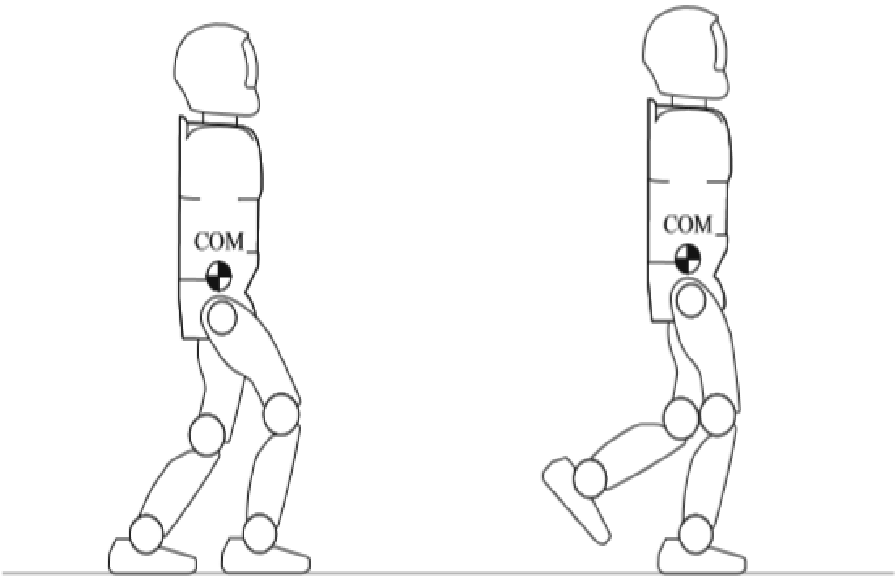


Fig. 1. Single and double support phase. In the double support phase, both feet are in con-tact with the ground and the robot's weight is supported by both legs. In the single support phase, one foot is in motion, whereas the other foot supporting the robot is in contact with the ground [14].

Walking mechanisms are suitable for applications that require movement across rough terrains, especially if compared to conventional wheels. They consist of links and joint, and are intended to simulate walking of human or animal. These linkages can be planar with single degree of freedom, or they can have a more complex motion in 3 dimensional space. Some can have multiple degrees of freedom.

4 Proposed Methodology

The main objective of this work is to develop an autonomous navigation system that integrates elements of the model of bipedal gait, stability and navigation that allows a biped robot to explore indoor environments with a level of autonomy similar to or greater than that reported in the literature. Based on the material provided in the previous section and our experience, the following requirements are defined to the methodology.

a) Select a commercial biped robot, as well as study and analyze its operation.
b) Select and apply a bipedal gait model for the selected robot (Fig. 2).

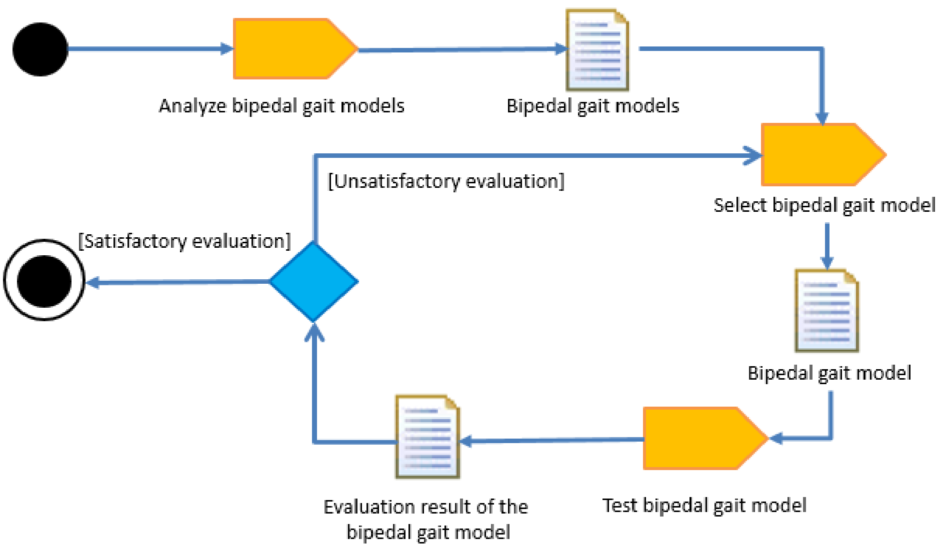


Fig. 2. This diagram shows the process of selection and evaluation of the bipedal gait model for the robot.

c) Select and apply a stability technique for the selected robot (Fig. 3).

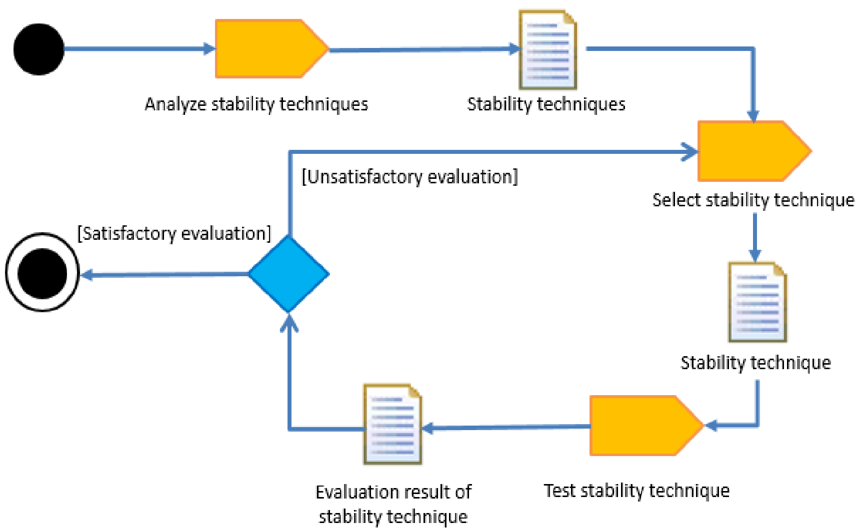


Fig. 3. This diagram shows the process of selection and evaluation of the stability technique for the robot.

d) Select and apply the indoor navigation method for the selected robot (Fig. 4).

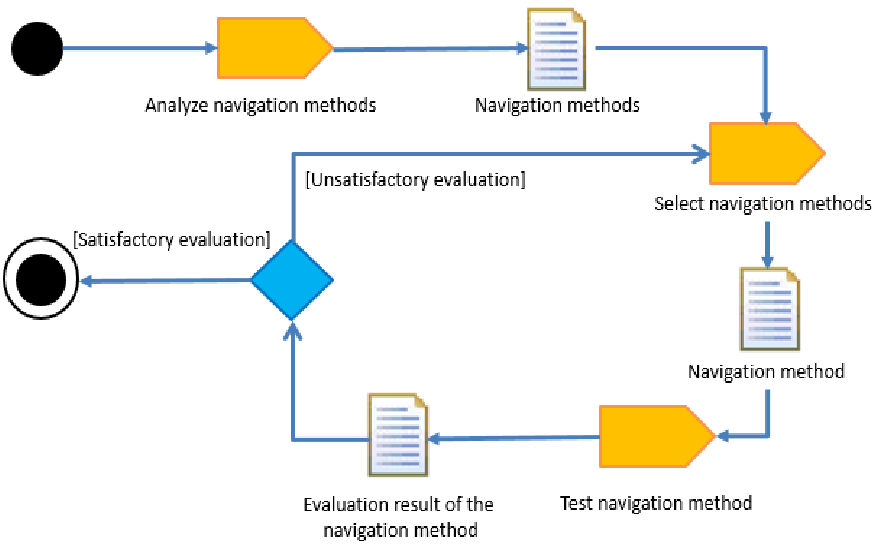


Fig. 4. This diagram shows the process of selection and evaluation of the navigation method for the robot.

e) Evaluate the integration of the bipedal gait model, the stability technique and the indoor navigation method in the selected robot.
f) Select an automatic learning method and implement it in the selected robot.
g) Evaluate the integration of machine learning to clause e).
h) Design the autonomous navigation system for the selected robot.
i) Validate the level of autonomy in the navigation of the selected robot in indoors (Fig. 5).

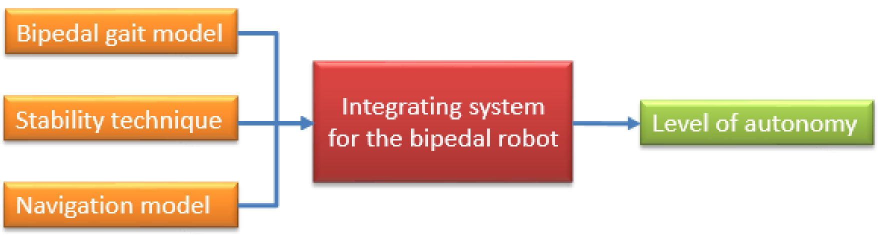


Fig. 5. This diagram shows the integration of the elements of bipedal gait, stability and navigation to obtain the level of autonomy of the bipedal robot.

5 Discussion

In the light of our proposal and based on the literature review we highlight the following elements:

- **Flaws of walking mechanisms**. The shortcomings of the walking mechanisms identified are: a driving member rotates in an unequal speed to obtain a unique speed of robots, or vehicles that are driven by a walking mechanism; the length and height of the steps are fixed; inertial moments and forces cannot be balanced in a satisfied way.
- **Advantages of walking mechanisms.** When it comes to the advantages of walking mechanisms, first of all, it is necessary to mention the movement on rough terrain. It can move on all types of terrains like desert, mountains, snow, rocky. It can be even used in planetary exploration because it has maximum payload to weight ratio. Also it has maximum efficiency for moving. These benefits of using walking mechanisms include higher speed, better fuel economy, greater mobility, better isolation from terrain irregularities, and less environmental damage.
- **Advantages of walking mechanisms when passing an obstacle.** If we compare this mechanism with a common wheel, it is obvious that the walking mechanism will overcome an obstacle much easier by just crossing it. Consider a moving wheel with a constant angular velocity ω, each point on the circumference of the circle will have a velocity ν, in the direction tangent to the circle. It is clear that the leg will have an advantage over the wheel, because it will be easier for it to overcome the obstacle. This way you can reduce the consumed energy [29].
- **Indoor versus outdoor configuration.** Balance is usually not a research problem in wheeled mobile robotics because robots are almost always designed so that all of the wheels are generally in contact with ground. When more than three wheels are used, a suspension system is required to allow all wheels to maintain ground contact when the robot encounters uneven terrain. Therefore, wheeled robot research tends to focus on the problems of traction and stability, maneuverability, and control. This can be further analyzed regarding the environment they are. Regarding outdoor mobile robots, usually researchers need to take into account problems such as the following: higher sensor noise, e.g., in vision systems due to illumination variation; weather variation, such as fog, snow, wet floors, mud, sand; floor irregularities; localization and state estimation issues.

All above-mentioned issues affect directly the mobile robot trajectory tracking control system. Only a few works in literature uses outdoor WMRs with model predictive control that make this an open topic for researchers. In contrast, although indoor mobile robots are often studied, they are always in controllable environments [32].

6 Conclusions

During the review of the literature, it has been validated that the majority of mobile robots are related to image processing and the use of artificial vision to detect and recognize obstacles during navigation, some others use presence sensors, such as ultrasonic or infrared sensors. Even some of these works employ automatic learning techniques, such as reinforcement learning, Q-learning or convolutional neural networks to perform navigation successfully and also be able to memorize the paths made by robot, so that its next journey is more efficient than the previous one. It is important to point out that most of the works related to the navigation of robots use wheeled robots for their simplicity when modeling and controlling their displacement, therefore there is the importance of developing an autonomous navigation system for a biped robot in indoor environments, using any type of sensors or artificial vision for the robot is able to move through a horizontal plane, detect and avoid static or dynamic obstacles.

References

1. Perron, J. M.: Enabling Autonomous Mobile Robots in Dynamic Environments with Computer Vision. Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in the School of Computing Science Faculty of Applied Sciences, B.Sc., York University (2014)
2. Baek, D., Hwang, M., Kim, H., Kwon, D.: Path Planning for Automation of Surgery Robot based on Probabilistic Roadmap and Reinforcement Learning. In 15th Int. Conf. Ubiquitous Robot, 342–347 (2018)
3. Chen, Z., Mendes, A., Yan, Y., Chen, S., Goebel, R.: Intelligent Robotics and Applications. 8102 (2013)
4. Mrudul, K., Mandava, R. K., Vundavilli, P. R.: An Efficient Path Planning Algorithm for Biped Robot using Fast Marching Method. Procedia Comput. Sci. 133, 116–123 (2018)
5. Pattanayak, S., Sahoo, S. C., Choudhury, B. B.: Soft Computing in Data Science, 545 (2015)
6. Rashid, A., Ali, F. R.: Software Implementation of a Static Store System Using the Digital Differential Analyzer Algorithm (2018)
7. Skjetne, R.: Autonomous Path-Planning and - Following for a Marine Surface Robot Mats Håkon Follestad (2017)
8. Chen, J., Kim, P., Cho, Y. K., Ueda, J.: Object-Sensitive Potential Fields for Mobile Robot Navigation and Mapping in Indoor Environments. In 15th Int. Conf. Ubiquitous Robot, 328–333 (2018)
9. Hsu, H. Y., Fu W., Lin Y.: Robot Surveillance System. U. S. Patent Application Publication, Pub. No.: US 2018 / 0181141 A1, (2018)
10. Fink, W., Brooks, A. J. W., Tarbell, M. A.: An Adaptive Hierarchical Approach to Lidar-based Autonomous Robotic Navigation. Micro-Nanotechnol. Sensors. Syst. Appl. X 69 (2018)
11. Graven, O. H., Graven, O. H., Srisuphab, A., Silapachote, P., Sirilertworakul, V., Ampornwathanakun, W., Anekwiroj, P., Maitrichit, N.: An Autonomous Indoor Exploration Robot Rover and 3D Modeling with Photogrammetry. In Int. ECTI North. Sect. Conf. Electr. Electron. Comput. Telecommun. Eng., 44–47 (2018)
12. Roshanineshat, A.: Evolutionary Optimization for Safe Navigation of an Autonomous Robot in Cluttered Dynamic Unknown Environments. Bachelor of Science in Electrical Engineering, University of Zanjan (2018)
13. VInjarapu, A., Gawre, S. K.: A Survey of Autonomous Mobile Robot Path Planning Approaches. In Proceeding International conference on Recent Innovations is Signal Processing and Embedded Systems (RISE-2017), 27–29 (2017)

14. Mihel, M., Bajd, T., Ude, A., Lenarčič, J., Stanovnik, A., Munih, M., Rejc, J., Šlajpah, S.: Robotics. 2nd ed. 2018, 189-191, 209–214 (2018)
15. Globisz, A., Krawczyk, D.: Statically Stable Gait Synthesis Algorithm For 4-legged Walking Robot Based on the Ground Reaction Forces Determination. In 19th International Conference on Research and Education in Mechatronics (REM) (2018)
16. Parada, J. G., Ortega L. E., Ayala L. F., Pérez, F. J., Herrera C. A., Padilla J. A.: A Low-Cost Land Wheeled Autonomous Mini-robot for In-door Surveillance. IEEE Latin America Transaction, 16(5) (2018)
17. Luo, R. C., Shih W.: Autonomous Mobile Robot Intrinsic Navigation Based on Visual Topological Map. In IEEE 27th International Symposium on Industrial Electronics (ISIE), (2018)
18. Guan, X., Cai, C.: A new integrated navigation system for the indoor unmanned aerial vehicles (UAVs) based on the neural network predictive compensation. In 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC) (2018)
19. Grewal, H. S., Jayaprakash, N. T., Matthews, A., Shrivastav, C., George, K.: Autonomous wheelchair navigation in unmapped indoor environments. IEEE International Instrumentation and Measurement Technology Conference (I2MTC) (2018)
20. Cornacchia, M., Kakillioglu, B., Zheng, Y., Velipasalar, S.: Deep Learning Based Obstacle Detection and Classification with Portable Uncalibrated Patterned Light. IEEE Sensors Journal (2018)
21. Catane, A., Feinberg, Y., Dor, O.: Indoor Navigation System, U. S. Patent Application Publication, Pub. No.: US 2013 / 0262223 A1 (2013)
22. Kim, W., Seok, J.: Indoor Semantic Segmentation for Robot Navigating on Mobile. In 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN) 22–25 (2018)
23. Wei, Z., Hong-bin, W., Zhi-ming, Z.: Mobile Robot Applications Grounded in Deep Learning Theories: A Review. Int Rob Auto J, 3(5), 00067 (2017)
24. Baslan, N., Heerklotz, S., Weber, S., Heerklotz, A., Höfig, B., Abu-Khalaf, J.: Navigation and Vision System of a Mobile Robot, In 19th International Conference on Research and Education in Mechatronics (REM), 99–104 (2018)
25. Chen, J., Kim, P., Cho, Y. K., Ueda, J.: Object-sensitive potential fields for mobile robot navigation and mapping in indoor environments. In 2018 15th International Conference on Ubiquitous Robots (UR) (pp. 328-333) (2018).
26. Schtz, M., Dobrev, Y., Carlowitz, C., Vossiek, M.: Wireless Local Positioning with SILO-Based Backscatter Transponders for Autonomous Robot Navigation, IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM) (pp. 1-4) (2018).
27. Qi, X., Chen, L., An, K., Wang, J., Zhang, B., Xu, K., Ye, D., Li C., Ran, L.: Bio-Inspired In-grid Navigation and Positioning Based on an Artificially established Magnetic Gradient, IEEE Transactions on Vehicular Technology PP(99):1-1 (2018).
28. Metka, B., Franzius, M., Bauer-Wersing, U.: Efficient navigation using slow feature gradients. In Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference, 1311–1316 (2017)
29. Vujošević1 V., Mumović1, M., Tomović1, A., Tomović, R.: Robot Based on Walking Jansen Mechanism, IOP Conference Series: Materials Science and Engineering, Volume 393, conference 1 (2018)
30. Bhanuteja, G., Mudkavi, V., Sharma, G.: Design and Construction of Six Legged Robot. International Conference on Design Innovations for 3Cs Compute Communicate Control (2018)
31. Ibraheema, I. K., Ajeilb, F. H., Maherc, R. A.: Path Planning of an Autonomous Mobile Robot in a Dynamic Environment using Modified Bat Swarm Optimization (2018)

32. Nascimento, T. P., Dórea, C. E. T., Gonçalves, L. M. G.: Nonholonomic mobile robots' trajectory tracking model predictive control: a survey. Robotica, 36, 676–696 (2018)

ISSN 1870-4069

Béziau's SP3A Logic and Logic Programming

Jaime Díaz[1], José Luis Carballido[1], Mauricio Osorio[2]

[1] Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación, Mexico

[2] Universidad de las Américas Puebla,
Departamento de Actuaría, Física y Matemáticas, Mexico

{jdiazalt27, jlcarballido7, osoriomauri}@gmail.com

Abstract. We present some three-valued paraconsistent logics, in particular SP3A logic which has been recently introduced as a genuine paraconsistent logic. After discussing the relevance of some rules in classical logic, we motivate the need for paraconsistency. Then we mention areas of applications of non-monotonic reasoning in terms of programming semantics, and mention the relationship that exists between these semantics and logics. In particular we show that logic SP3A belongs to the family of D-elemental logics; these logics can characterize one of the semantics useful in the area of non-monotonic reasoning.

Keywords. Logic programming, paraconsistency, 3-valued logic.

1 Introduction

Consistency is a property that comes from complying that $x \wedge \neg x$ is always false. This means that when a formal system can never produce two or more formulations that contradict each other, then we say that such system is *consistent*. Classical logic is consistent and holds many other desirable properties: de Morgan laws, distributivity, among others and is the right choice whenever a solid and consistent theoretical foundation is needed. Therefore, mathematics, physics, chemistry and engineering take a hold on the basis of classical logic since these disciplines work with instances that generally do not allow contradictions.

This view of the necessity of consistence was strongly defended by Aristotle, and is called *the law on non contradiction (LNC)* (presocratic philosophers did not consider consistency as a matter of necessity) [11]. Logical systems that comply with consistency are called *classical* or *aristotelian*; in aristotelian systems, if we accept a formula that does not comply with the LNC, *explosivity* occurs: the logical system explodes into an infinity of propositions that can be proved as true. This can be expressed as $x \wedge \neg x \to y$ However, the LNC has not been formally proved yet, and even Aristotle failed trying to show the necessity of its compliance [11]. On the other hand, not only presocratics believed that consistency in a classical way is not necesary for a formal system to be valid: as we will see later, logics that do not obey LNC have been developing since the beginning of XXth century. Non classical logics arise then as valid logics even when they do not hold the LNC, and arise too as an answer to situations where contradiction is an inherent feature.

In next section (section 2) we briefly review formal systems and their structure. Section 3, is about paraconsistent logics; in particular we present SP3A and SP3B logics. Section 4 is a brief review of logic programming, where we review some transformations of logic programs in order to avoid inner contradictions. Finally, in the last section we present D-elemental logics and a demonstration that SP3A logic is D-elemental.

2 On Formal Systems, Semantics and Axiomatics

As da Costa remarks in [5], it is difficult to give a precise definition to *semantics*, specially when working with non-classical logics. However, up until now, and given that our pourpose is not a deep philosophical study of semantics, we will understand it as da Costa mentions: *"When first proposed in the fields of logic and formal sciences, the term «semantics» send to present a clear sense. It was supposed to denote that part of an analysis of a language concerned with the determination of the meanings of its (well formed) expressions"* [5]. In this sense, we will then consider that a logic can be defined and analysed when its connectives are defined by presenting their behavior through a truth-table, and well formed formulas (wfs) are presented and evaluated by this same way. This means that semantics allows us to determine if a statement is a tautology, a contradiction, or which is its particular evaluation for any specific interpretation of its variables; moreover, it helps us to determine wether two statements are equivalent to one another. Going back to da Costas's definition of semantics, it must be noted that it implies the determination of the meaning of a wfs. Then, as the semantic approach leads us to elaborate truth-tables that evaluate wfs to defined truth values, at the end we will have a bunch of truth values as the meaning of that specific wfs. Then, it is necessary to understand the meaning of all the truth values in the domain of that particular logic. In classical logic they are ($0 = False$ and $1 = True$), but when it comes to many-valued logics, the matter is not so obvious. Later we will discus about this particular issue.

Mendelson on the other hand, illustrates the axiomatic method by defining a formal theory (a formal system). As Mendelson claims, a formal theory needs 4 conditions to be satisfied:

1. A set of countable symbols to form expressions. An expression is a finite sequence of these symbols.
2. A set of countable expressions that will be called well-formed formulas, and a method to determine if an expression is a wfs.
3. A set of wfs called *axioms*.
4. A finite set of rules of relation between wfs (rules of inference) [8].

Using these rules of inference over axioms, new valid wfs can be obtained. If an obtained expression is a tautology, then it is called a *theorem*.

By this method, we can generate new theorems, and prove things such as logical equivalence or dependence among expressions.

Semantic and axiomatic approaches must not be considered as opposite but as complementary; it is natural to use both in order to perform proofs and calculations.

3 Paraconsistent Logics

In this section we will review some background of paraconsistent logics: What is paraconsistency, forerunners of paraconsistet logic and finally, we will present two paraconsistent logics proposed by Jean-Yves Béziau.

3.1 Non-classical Logics: 3-valued and Paraconsistent Logics

As stated in [3], a *logic* is a set of wfs that satisfies two properties:

1. The set is closed under Modus Ponens: if A and $A \to B$ are in the set, then B is in the set too.
2. The set is closed under substitution: if the formula A is in the logic, then any other formula obtained by replacing all the ocurrences of an atom b in A with other formula B is in the logic too.

Classical logic is the most well known logic, and holds many well defined characteristics; it is 2-valued, obeys LNC, and double negation (the negation of the negation of a formula is equivalent to the original formula $\neg\neg x \to x$); in general, classification of logics is not a precise matter, and it is difficult to claim that a logic is classified in a particular way. However, for us, non-classicism means at least one of two characteristics:

- The logic has 3 or more values.
- The logic does not obey the LNC.

In this sense, our object of interest (SP3A logic) holds both of them, so it will be called a non-classical logic.

Paraconsistency:

There is a group of non-classical logics that are called *paraconsistent logics*. da Costa is one of the main initiators of paraconsistent logics, and his original definition was *"a one-place operator ν is paraconsistent if there is a formula a such that the theory $a, \eta a$ is non-trivial"*, that is, a and ηa are both true simultaneously, with the constraint that this operator should obey all the properties of classical negation (though it is not clear which properties are these) [5].

According to Béziau , *"Paraconsistent logic can be considered a bunch of logical systems in which there is a connective which does not obey the principle of non-contradiction; such a connective is usually called a paraconsistent negation and the main problem is to know if it is legitimate to call such an operator a negation"* [2].

These paraconsistent logics make it possible to deal with contradictions without our formal system exploding into triviality. This property is useful when working with contradictory information (as in non-monotonic reasoning) and makes paraconsistent logics a powerful tool when we need to deal with contradictions; in this work we will focus in just one field of application: Artificial Intelligence.

3-valued logics and the interpretetion of the third value

Among non-classical logics, there is a group of logics that are called *multivalued*; these multivalued logics hold a dominion D with more than two truth values. Logically, there are three-valued logics in which $D = \{0, 1, 2\}$, and where 0 and 2 usually behave classically just as classical 0 and 1 do; but in these logics, the value 1 must be interpreted in a particular way. There are several interpretations for this value, and the interpretation must match with the logic semantics and this interpretation makes the particular logic useful for specific areas of study.

In [6], Coste-Marquis considers multi-valued logics as useful when working with contradictory pieces of information (when working with agents for example); more specifically 3-valued logics, and where the third value (1) is given a specific interpretation: *"Proved both True and False"*. However, this approach does not guarantee that its inferences are consistent. Later we will introduce SP3A and SP3B logics. These logics need an interpretation for their third values. We will propose an interpretation for SP3A.

Although in [6] a particular interpretation for the third value in 3-valued logics is "both true and false", there can be many different ways to understand it. In [4], Ciucci and Dubois mention some of them: possible, unknown, undefined, half-true, irrelevant, inconsistent. These interpretations can be classified in two types: *ontological* (undefined, half-true, irrelevant) and *epistemic* (possible, unknown). Ontological values make reference to a situation where the nature of the third value is not questioned, but understood as an intrinsic feature of the expression; epistemic values on the other hand, are values whose state will eventually change into 0 or 2 in the future.

3.2 Forerunners of Paraconsistency: Łukasiewicz 3-valued Logic, Vasili'év Logic and G_3 Logic

History of paraconsistent logics is long yet almost unknown for most people; Priest states that presocratic philosophers where familiar with logical systems that did not obey the LNC, and it was Christianism that took Aristotelian theories, and the LNC as a dogma for centuries [11]. As a result, classical logic was the only one that developed along this period. In 1910 Jan Łukasiewicz published the book *On the Principle of Contradiction in Aristotle*. This book studies Aristotle's LNC, and concludes that it can not be proved in the sense that *every contradiction is false*: even if Aristotle's arguments were proved, it would prove only that *some contradictions are not true*. Łukasiewicz gives then what he claims is *the only strict and formal proof* for the LNC: the only way to prove it is to assume that contradictory objects are not objects at all, it means they are *nothing*, instead of *something*; this means that anything that is *something* and not *nothing*, does not contain contradictory properties [12].

Łukasiewicz was one of the forerunners on non-classical logics; as he studied the LNC and found out there was no way to prove it valid for every contradiction, he set the precedent for paraconsistent logics, and defined the family of logics $Ł_\omega$.

Vasili'év is also considered a precursor of paraconsistency. Inspired by Lobatchevsky's non-Euclidian geometry, he had some ideas about what he called *imaginary logic* [5]. Łukasiewicz and Vasili'év ignited the development of paraconsistency.

Finally, we will mention a logic defined by Göedel. This logic belongs to the family of multivalued logics G_i, and is the 3-valued logic G_3. The relevance that this logic has for our work is that SP3A logic has no implication connective defined in it and given that we will need it later, we will use G_3 native implication for it.

3.3 Béziau's Paraconsistent Logics SP3A and SP3B

In this work, we present the analysis of some features of SP3A logic. This is a 3-valued paraconsistent logic proposed by Jean-Yves Béziau in *Two Genuine 3-Valued Paraconsistent Logics* [1], where Béziau presents SP3A and SP3B . Each of these 3-valued logics have only three primitive connectives: negation, conjunction and disjunction. In tables 1 and 2, we respectively present the definition of SP3A and SP3B connectives:

Table 1. Truth tables of connectives $\wedge$, $\vee$, and $\neg$ in $SP3A$.

$\wedge$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	2

$\vee$	0	1	2
0	0	1	2
1	1	1	2
2	2	2	2

x	$\neg x$
0	2
1	2
2	0

Table 2. Truth tables of connectives $\wedge$, $\vee$, and $\neg$ in $SP3B$.

$\wedge$	0	1	2
0	0	0	0
1	0	2	1
2	0	1	2

$\vee$	0	1	2
0	0	1	2
1	1	1	2
2	2	2	2

x	$\neg x$
0	2
1	1
2	0

According to the definition of paraconsistent logic given by Béziau , SP3A would be a paraconsistent logic since its negation does not obey the LNC. This can be seen in Table 3

Table 3. Law of Non Contradiction in $SP3A$ and $SP3B$.

x	$\neg x$	$x \wedge \neg x$	$\neg(x \wedge \neg x)$
0	2	0	2
1	2	2	0
2	0	0	2

x	$\neg x$	$x \wedge \neg x$	$\neg x \wedge \neg x$
0	2	0	2
1	1	1	1
2	0	0	2

As can be seen, $\neg(x \wedge \neg x)$ is not a tautology in neither of these logics. That is why Béziau claims these are paraconsistent logics.

Béziau presents his logics and analyses some of their behaviors, mainly in properties related to negation, such as de Morgan laws and double negation; Béziau shows too that substitution theorem does not hold in neither of these logics.

4 Logic Programming

In *"Paraconsistent Logic in a Historical Perspective"* [5], the authors state that the future of Paraconsistent Logic lays in some research lines such as:

- To develop a Paraconsistent Model Theory.
- To develop a paraconsistent Set Theory.
- To develop a Paraconsistent Mathematics.
- To develop further applications to Computer Science, Artificial Intelligence, Law, Everyday Life...

It is in particular the last item in the list above the matter of interest for this work: when working with Logic Programs, Belief Databases, interaction with human beings, sensor fusion, it is common that inconsistencies appear, collapsing the system. There are techniques to avoid it, and using paraconsistent logics to treat the information is one of them. In this section we explore some background of logic programming and belief revision.

Logic programs

Logic programs are formed by 1 or more logic formulas; these programs are designed to make controlled logic inferences in order to demonstrate theorems automatically [3].

According to Schlipf, a *literal* is an atomic formula $R(t_1, t_2...t_n)$ (a *positive* literal) or a negated atomic formula $\neg R(t_1, t_2...t_n)$ (a *negative* literal).

A *logic program* is a finite, or countably infinite, set of *rules*, (implicitly universally quantified) formulas of the form $a \leftarrow b_1 \wedge b_2 \wedge ... \wedge b_n$, where a is a positive literal and the b_i s' are all literals -positive or negative- [13].

Note that in [13], the head of a rule contains just a single literal; however we will use a more general approach where rules are expressions in the form $A \leftarrow B$ where $A = a_1 \vee a_2 \vee ... \vee a_k$ and $B = b_1 \wedge b_2 \wedge ... \wedge b_n \wedge \neg b_{n+1} \wedge ... \wedge \neg b_{n+m}$, with $k \geq 1$ and where $a_i' s$ and $b_i' s$ are atomic formulas [3].

According to Carballido, we can say:

Definition 1. *A logic program (LP) is a* theory*, this is, a set of formulas; a* class of Logic Programs *is a set of LP that satisfies certain property or syntactical limitation. [3].*

Definition 2. *Rules used in Logic Programming hold the following structure: $A \leftarrow B$, where A is called* head of the rule *and B is called* body of the rule. *[3].*

Note that $A \leftarrow B$ is usually used as an alternate form to $B \rightarrow A$ ($\leftarrow$ is a kind of implication [13]).

In order to abbreviate notation, it is common to write a rule in the form: $A \leftarrow B^+ \wedge \neg B^-$, where $B^+ = \{b_1, b_2, ..., b_n\}$ and $B^- = \{b_{n+1}, b_{n+2}, ..., b_{n+m}\}$

Types of programs and rules

Any rule $A \leftarrow B$, where $A = a_1 \vee a2 \vee ... \vee a_k$, $k \geq 1$, and $B = b_1 \wedge b_2 \wedge b_3 \wedge ... \wedge \neg b_n \wedge \neg b_{n+1} \wedge ... \wedge \neg b_{n+m}$, with $n \geq 0$ and $m \geq 0$ is called a *clause or disjunctive rule*. In case that $n = 1$ we will call it a *Normal Clause*. If the clause has no negated atoms, then will call it *Positive*. A program which all of its rules are disjunctive is called a *Disjunctive LP*, and a program which all of its rules are normal is called a *Normal LP*. The size of a disjunctive clause is defined by $k + n + m$ [3]. If there is a program in the form $A \leftarrow$, then it is called a *fact*; an expression in the form $\leftarrow B$ is called a *restriction*.

Stable Model Semantics

In order to explain what the Stable Model Semantics is, we will give two definitions:

Definition 3. *If a logic is stronger than intuitionistic logic, and weaker than classical logic is called* intermediate logic. *An intermediate logic is called* propper *if it is strictly contained in classical logic. Intuitionistic logic is an intermediate logic too [3] [10].*

One intermediate logic to mention is G_3; No one of the intermediate logics is paraconsistent [3].

Definition 4. *If M is a set of atoms of a logical program P, M is a classical model of P, and M is minimal among the classical models of P, then we say that M is a* minimal model *of P.*

Example

Let P be the LP shown:

$a \leftarrow b \wedge \neg c$
$b \leftarrow \neg a$
$b \leftarrow c$

The classical models for P are: $\{a\}$, $\{c, b\}$, $\{a, b\}$, $\{a, b, c\}$.
Minimal classical models are: $\{a\}$, $\{c, b\}$.

In [9], Osorio defines Stable Model Semantics as shown in the following definition:

Definition 5. *Given a disjunctive program P, for any set M of atoms in P, P^M is the program obtained from P by removing:*

1. *Each clause that contains at least one negative literal $\neg b$ in its body, whith $b \in M$.*
2. *All the negative literals in the bodies of remaining clauses.*

Example

Taking the program from the previous example, and considering $M_1 = \{a\}$ and according to the definition of stable, we get P^{M_1} as:

$a \leftarrow b$
$b \leftarrow c$

Models of P^{M_1} are: $\emptyset$, $\{a\}$, $\{a, b\}$ and $\{a, b, c\}$.
Proposing $M_2 = \{b, c\}$, and we can obtain P^{M_2} from applying 5 as:

$b \leftarrow$
$b \leftarrow c$

Models of P^{M_2} are: $\{b\}$ and $\{b, c\}$.

Definition 6. *It is a kwnown fact that any possitive program will always contain at least* 1 *classical minimal model. If M is one of the minimal models of P_M, then we say that M* is a stable model of P *[3] [9].*

Example

From preceding 2 *examples we already know program P, program P^{M_1}, P^{M_2} and the models for each of them. $\{a\} = M$ but it is not a Stable Model because $\{a\}$ is not a minimal model of P^{M_1} ($\emptyset \in \{a\}$).*

Analyzing P^{M_2}, its minimal model is $\{b\} \neq M_2$. Neither M_1 nor M_2 are stable models of P.

In [10], the author shows that the stable semantics can be characterized by modeling in intermadiate logics. We present next another semantics that can also be applied in non-monotonic reasoning, and can be characterized in terms of paraconsistent logics, in particular D-elemental logics. This semantics is called p-stable semantics and offers an alternative to the stable semantics [3].

P-stable Model Semantics

Next, we define another transformation that is helpful to define an alternative semantics to the stable model semantics.

Definition 7. *Let P be a program and M a set of atoms in P. Then, we define $RED(P, M) := \{a \leftarrow B^+ \wedge \neg(B^- \bigcap M) \mid a \leftarrow B^+ \wedge \neg B^- \in P\}$.*

Definition 8. *Let P be a normal program and M a set of atoms in P, then, if $RED(P, M) \Vdash_C M$, we say that M is a p-stable model of P. Where the symbol $\Vdash_C$ means that the formulas on the right can be proved from those on the left in classical logic.*

Example

Continuing with the example, , if we take $M_1 = \{a\}$, then we get $RED(P, M_1)$:

$a \leftarrow b$
$b \leftarrow \neg a$
$b \leftarrow c$

By following the next equivalences: $\neg a \rightarrow b \equiv \neg b \rightarrow \neg\neg a \equiv \neg b \rightarrow a$ so we have $(b \rightarrow a) \wedge (\neg b \rightarrow a)$. From this formula it follows a, and then, M_1 is a p-stable model of P.

4.1 Treating Inconsistency in Artificial Intelligence

When a system interacts with the world, there is a flux of information to take in account; the world changes constantly and new information is generated every instant. Our system has then the necessity to constantly update the information in its data bases. But it is common that new and previous information are opposed to one another. Even more, it may occur that given a set of sensors, some of them give contradictory information at the same instant in time (in sensor fusion for example). If the system interacts with people, the possibility of inconsistent information is even higher: humans are naturally so complex and behave in contradictory ways: a person changes his mind from one moment to another, judges situations in a subjective way and frequently acts in a different way than he thinks or for every enviroment acts in a special manner. If such a system is based on classical logic, these contradictions lead to inconsistencies that could make the system explode so that $(x \wedge \neg x) \to y$, and preventing contradictions becomes a matter of major importance. But if we have a way to deal with contradictions withouth our system exploding, then the problematic situation loses strenghth.

When working with contradictory information there are several approaches that allow us to avoid the trivialization. This is a fundamental problem in A.I.; The problem can be treated by belief revision, belief merging, reasoning from preferred subsets, purification, paraconsistency, etc. The existence of such a wide spectrum of approaches can be explained by the fact that paraconsistency can be achieved in various ways; the particular situation stablishes the needs to be fulfilled and the way to follow in order to achieve this paraconsistency[6].

Paraconsistency taken in a strict sense allows us to deal with inconcistency directly, while the other approaches require some extra-logical information to avoid trivialization; the limitation in these cases is in the fact that this extra information may be too poor or too sophisticated[6].

Belief Revision

Human beings are always changing our minds: new information gets to us and modifies our considerations and knowledge about the world: changes our beliefs. When working with *agents* in Artificial Intelligence leads us to a similar scenery: an *agent* is defined by Russel and Norvind as *"any entity capable of perceiving and acting on the world"* (cited in [7]). When an agent interacts with an environment, fresh information is arriving all the time making it necessary to it to change its «beliefs» (Belief is a relation between an agent and a proposition; however it is difficult to restrict which relations are beliefs and which are not [7]). *Belief Revision* is the discipline that studies the rationality of belief change in agents [7].

Moretto mentions as an example a few agents working in a determined environment.

5 SP3A logic is D-elemental

This last section is dedicated to prove that SP3A logic is D-elemental; definition 9 tells us what does it mean that a logic is D-elemental:

Definition 9. *A multi-valued logic E is* D-Elemental *if its domain D contains three special elements 0, 1 and t, that satisfy the following properties [3]:*

1. *t is a designated value, and 0 is undesignated*
2. *The value assigned to $1 \to 0$ s not designated*
3. *Connectives $\wedge$ and $\vee$ are commutative and associative*
4. *For every value of x, $0 \wedge x = 0$ and $0 \to x \in \{1, t\}$*
5. *For every special element x, $1 \vee x \in \{1, t\}$*
6. *For every special element x, $t \vee x \in \{1, t\}$*
7. *Fragment $\{0, t\}$ matches classical logic (for $\wedge$, $\vee$, $\to$ and $\neg$)*
8. *Fragment $\{0, t\}$ is closed respect to $\wedge$ and $\to$ connectives*
9. *The value assigned to negation of 1 is an element in $\{1, t\}$*
10. *E logic lies between C_ω and C, where C is classical logic. That is $C_\omega \subset E \subset C$*

In the following lines we will prove that SP3A logic is $D - elemental$

The G_3 implication

The logic SP3A does not include any implication as a primitive connective; however, it is necessary to have an implication in SP3A in order to prove it is a $D - elemental$ logic. In order to have an implication for SP3A , now, we are going to define G_3 implication in SP3A terms. G_3 implication is defined as shown in 10

Definition 10. *G_3 implication is defined by:*

$$x \to y = \begin{cases} 1 : x \leq y \\ y : x > y \end{cases}$$

Within SP3A , G_3 implication will be denoted by the symbol "$\to_G$", and it is expressed in terms of the SP3A connectivesby:

$$x \to_G := (\neg x \vee y) \wedge (\neg(x \vee (x \wedge \neg x)) \vee y)$$

The Possitive logic

The *Possitive logic* (Pos) is defined by a set of 8 axiomatic schemes; these schemes are listed below:

For every wfs α, β and γ [3]:

Pos1 $\alpha \to (\beta \to \alpha)$
Pos2 $(\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma))$
Pos3 $\alpha \wedge \beta \to \alpha$
Pos4 $\alpha \wedge \beta \to \beta$
Pos5 $\alpha \to (\beta \to (\alpha \wedge \beta))$
Pos6 $\alpha \to (\alpha \vee \beta)$
Pos7 $\beta \to (\alpha \vee \beta)$
Pos8 $(\alpha \to \gamma) \to ((\beta \to \gamma) \to (\alpha \vee \beta \to \gamma))$

It can be easily proved that SP3A obeys all these axiom schemes using a semantical approach.

The C_ω logic

C_ω logic was proposed by daCosta ; it is defined by the axiom set of Pos, plus $C_\omega 1$ and $C_\omega 2$.

Let α be a well formed formula in C_ω, then

$C_\omega 1$	$\alpha \vee \neg\alpha$
$C_\omega 2$	$\neg\neg\alpha \rightarrow \alpha$

C_ω is a *minimal paraconsistent logic* which means that any other logic E, in which $\neg(A \wedge \neg A)$ is not a theorem must contain every C_ω theorem [3]. Then, as a consequence of this affirmation we can state the next lemma:

Lemma 1. *SP3A logic contains all the theorems in C_ω logic: $C_\omega \subseteq SP3A$*

Proof: As shown in table 3, $\neg(A \wedge \neg A)$ is not a theorem in SP3A. Then, SP3A must contain every C_ω theorem.

SP3A logic is *D-elemental*

Now we proceed to verify that SP3A logic obeys the 10 points that define a *D-elemental* logic:

1. Dominion of variables in SP3A is $D = \{0, 1, 2\}$; in this sense, SP3A logic satisfies this point: 0 and 1 are in its domain, and there is a designated value $t = 2$ while 0 is undesignated
2. In SP3A , $1 \rightarrow_G 0 = 0$; 0 is not designated
3. In SP3A , connectives $\wedge$ and $\vee$ are commutative and associative; it can be proved easily using truth tables
4. For any x in the domain of SP3A , $0 \wedge x = 0$; $0 \rightarrow_G x \in \{1, 2\}$. This occurs because $x \in \{0, 1, 2\}$ and in every case $0 \leq x$ and according to the definition of the G_3 implication, in these cases the evaluation is 1
5. In SP3A , $\vee(x, y) = max(x, y)$. Then, $1 \vee x \in \{1, 2\}$ since $x \in \{0, 1, 2\}$
6. As in SP3A $\vee(x, y) = max(x, y)$, then $t \vee x = 2 \vee x = 2$, and $2 \in \{0, 1, 2\}$
7. In SP3A , fragment $\{0, t\} = \{0, 2\}$ behaves classically as can be seen in table 4
8. let $x \in \{1, 2\}$, then $1 \wedge x \in \{1, 2\}$ and $2 \wedge x = 2$. It occurs too that $1 \rightarrow_G x = 2$ and $2 \rightarrow_G x \in \{1, 2\}$. Then we can say that in SP3A the fragment $\{1, 2\}$ is closed for connectives $\wedge$ and $\rightarrow_G$
9. $\neg 1 = 2$, $2 \in \{1, 2\}$
10. As stated in lemma 1, $C_\omega \subseteq SP3A$; all theorems in SP3A can be proved in C. That is: $C_\omega \subseteq SP3A \subseteq C$

Theorem 1. *SP3A is a* D-elemental *logic.*

Proof: as shown previously.

Paraconsistent logics and logic programming

Finally, we will mention a theorem as stated and proved in [3]:

Theorem 2. *Let P be a disjunctive program. Let M be a set of atoms in P, and let E be a D-elemental logic. $P \bigcup \neg\widetilde{M} \vdash_E M$ iff $RED(P, M) \vdash_E M$.*

Table 4. Truth tables of connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow_G$ for the fragment $\{0, 1, 2\}$ in $SP3A$.

$\wedge$	0	2
0	0	0
2	0	2

$\vee$	0	2
0	0	2
2	2	2

x	$\neg x$
0	2
2	0

$\rightarrow_G$	0	2
0	0	2
2	2	2

References

1. Béziau, J.Y.: Two genuine 3-valued paraconsistent logics. In: Towards Paraconsistent Engineering, pp. 35–47. Springer (2016)
2. Béziau, J.Y.: What is paraconsistent logic? In: In Batens et al. [6. pp. 95–111. Press (2000)
3. Carballido, J.C.: Fundamentos Matemáticos de la Semántica P-estable en Programación Lógica. Ph.D. thesis, Benemérita Universidad Autónoma de Puebla (2009)
4. Ciucci, D., Dubois, D.: A map of dependencies among three-valued logics. Information Sciences 250, 162–177 (2013), `http://www.sciencedirect.com/science/article/pii/S0020025513004775`
5. da Costa N.C.A., B.J., O., B.: Aspects of paraconsistent logic (1995)
6. Coste-Marquis, S., Marquis, P.: Recovering consistency by forgetting inconsistency. Lecture Notes in Computer Science 5293 (2008)
7. M., M.: Belief Revision in Non-Classical Logics. Springer (2013)
8. Mendelson, E.: Introduction to Mathematical Logic. Chapman and Hall (1997)
9. Osorio M., Navarro J.A., A.J., V., B.: Logics with common weak completions. Journal of Logic and Computation 16 (2006)
10. Pearce, D.: Stable inference as intuitionistic validity. The Journal of Logic Programming 38, 79–91 (1999)
11. Priest, G.: Why it's irrational to believe in consistency. In: Brogaard, B., Smith, B. (eds.) Proc. 23 rd International Wittgenstein Symposium. pp. 284–293 (2001)
12. Priest, G.: Torn by reason: łukasiewicz on the principle of non contradiction. Early Analytic Philosophy - New Perspectives on the Tradition (2016)
13. Schlipf, J.: Formalizing a logic for logic programming. Annals of Mathematics and Artificial Intelligence 5, 279–302 (1992)

ISSN 1870-4069

Diseño de prototipo web inclusivo con interfaces naturales para apoyar el examen de admisión de personas con discapacidad visual en educación superior

Carmen Cerón, Etelvina Archundia, Beatriz Beltrán, Migliolo Jair

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Puebla, México

{mceron,etelvina,beltran}@cs.buap.mx,
migliolo.jair@gmail.com

Resumen. En la actualidad, existe una variedad de aplicaciones de las Tecnologías de la Información y la Comunicación para el ámbito educativo. Sin embargo, se requiere atender a estudiantes con necesidades educativas especiales. El objetivo es presentar el diseño, desarrollo e implementación de un prototipo web como herramienta de inclusión para personas con discapacidad visual para apoyar la realización del examen de admisión. La metodología utilizada fue el diseño inclusivo centrado en el usuario y la implementación utiliza JavaScript y un convertidor de Texto a Voz de los archivos del repositorio. Las pruebas de funcionalidad y usabilidad se realizaron en un grupo focal de tres estudiantes con discapacidad visual.

Palabras clave: Educación inclusiva, accesibilidad, discapacidad, interfaces.

Design of Inclusive Web Prototype with Natural Interfaces to Support the Admission Test for People with Visual Disabilities in higher Education

Abstract. Currently there is a diversity of applications of Information and Communication Technologies for the educational field. However, it is necessary to attend students with special educational needs. The purpose is to present the design, development and implementation of web prototype as an alternative tool for the inclusion of students with visual disabilities, as a support for the realization of the undergraduate level admission test. The methodology used was the user-centered inclusive design and the implementation use JavaScript and a Text-to-voice converter of the repository files. Finally, the tests of the usability and functionality tests was realized with a focus group of students with visual disabilities.

Keywords. Inclusive education, accessibility, disability, interfaces.

1. Introducción

La educación en México requiere atender a los estudiantes con capacidades diferentes, lo cual con lleva a generar sistemas educativos incluyentes. Los programas actuales deben ser incluyentes garantizando que los estudiantes puedan tener acceso a la educación obligatoria y superior. La educación inclusiva debe incorporar a los estudiantes con distintas discapacidades en las aulas ordinarias, lo cual implica un esfuerzo permanente.

Según los resultados del Censo de Población y Vivienda realizado por el Instituto Nacional de Estadística y Geografía [1] México tiene una población de 119,530,753 personas, de las cuales 35.2 millones están matriculadas como estudiantes en educación básica escolarizada, lo que representa 73.4% de la matrícula del sistema educativo. En educación media superior, se brinda servicio educativo a 4.4 millones de jóvenes, y en educación superior hay 3.3 millones de alumnos, lo que significa 29.2% de cobertura, distribuidos en siete mil planteles en todo el país [1]. De acuerdo a las estadísticas publicadas por la universidad solo el 10% de los alumnos en edad de ingresar a la educación superior obtiene un lugar y para las personas con discapacidad se reduce aún más en 2%, ya que a pesar de los avances tecnológicos y la apertura a la educación sigue existiendo la falta de inclusión para las personas con discapacidad sin tener los materiales, tecnologías o instalaciones para apoyar el examen de admisión para obtener un lugar en las Instituciones de Educación Superior (IES)

Desde el 2008, la educación inclusiva ha formado parte de la Declaración Mundial de "Educación para Todos", como una alternativa para transformar los sistemas educativos De acuerdo con la UNESCO [2], podemos entender a la educación inclusiva como: *La educación inclusiva puede ser concebida como un proceso que permite abordar y responder a la diversidad de las necesidades de todos los educandos a través de una mayor participación en el aprendizaje, las actividades culturales y comunitarias y reducir la exclusión dentro y fuera del sistema educativo*. Lo anterior implica cambios y modificaciones de contenidos, enfoques, estructuras y estrategias que conlleven a lograr una educación inclusiva para favorecer a todos los actores educativos.

Para los estudiantes con discapacidades, el uso de tecnologías digitales significa principalmente dos ventajas: una mayor posibilidad y acceso a la información y el uso de recursos didácticos para poder apoyar su aprendizaje [3].

El reconocimiento automático de voz es una tecnología, que, está siendo incorporada como la interface idónea para la comunicación entre hombre y la computadora debido a la naturalidad de la comunicación que presentan los sistemas actuales de reconocimiento de voz. Estas herramientas computacionales procesan la señal de voz emitida por el ser humano y reconocen la información contenida en ésta, convirtiéndola en texto o dando órdenes que actúan sobre un proceso mediante el uso de comando.

El propósito de este trabajo es presentar el diseño de un prototipo para apoyar el examen de admisión para estudiantes con discapacidad visual usando interfaces naturales de reconocimiento de voz y el procesamiento del lenguaje natural para facilitar la evaluación. El software EDU-INCLUEXA. permite que el usuario interactúe a través de comandos de voz en cada sección de la evaluación, realizando ejercicios y simulando la Prueba de Actitud Académica, el sistema almacena los resultados de la evaluación y genera un archivo de resultados en formato pdf, los cuales pueden ser consultados posteriormente por el usuario y evaluador.

El artículo está organizado en la siguiente forma: en la sección 2 se presenta la revisión del trabajo relacionado con la investigación. En la sección 3 se define la metodología, el análisis y diseño de del prototipo. En la sección 4 se muestran los resultados de la prueba piloto y la evaluación para estudiantes con discapacidades en el grupo focal a través de un enfoque metodológico cualitativo y la técnica de inspección y exploración. Finalmente, se presentan las conclusiones y el trabajo futuro de la investigación.

2. Trabajo relacionado

En esta sección, se presentan la revisión de algunos trabajos que se encuentran relacionados con procesamiento de voz e interfaces naturales de voz utilizados para apoyar a personas con discapacidad en su vida cotidiana.

Un sistema text-to-speech (TTS) convierte el lenguaje de texto normal en habla; otros sistemas recrean la representación simbólica lingüística como transcripciones fonéticas en habla. La síntesis de texto a voz (TTS), es la conversión automática de un texto a un lenguaje nativo. El TTS, es una tecnología que permite utilizar un n algoritmo llamado motor TTS, que analiza el texto, procesa previamente el texto y sintetiza el texto y finalmente, genera datos de sonido en un formato de audio como salida. La calidad de un sintetizador de voz es juzgada por su similitud con la voz humana y por su capacidad de ser entendido. Un programa inteligible de texto a voz permite apoyar a personas con discapacidad visual o discapacidad lectora para escuchar documentos y poder interactuar.

En [4] la investigación se enfocó en determinar la eficacia del software de convertir "speech-text". Con el objetivo de realizar los diagnósticos y determinar si la carga cognitiva para su capacitación podría disminuir para los participantes que se clasificaron en tres discapacidades categóricas que afectan las áreas de lectura y expresión escrita. Para lo cual analizaron muestras de escritura de los estudiantes con y sin apoyo del software, verificando que los participantes mejoraban en relación a los que utilizaron el software. El software que utilizaron fue aplicaciones en plataformas de iOS y Windows.

En [5] el trabajo presentado se enfoca al uso de software TTS para mejorar las habilidades de lectura para estudiantes con dificultades de rendimiento académico. El software TTS utilizado para este estudio fue el Kurzweil 3000, que fue diseñado como una herramienta compensatoria para personas con discapacidades, incluidas discapacidades de aprendizaje (LD), trastorno por déficit de atención con hiperactividad (TDAH), dificultades de lectura y algunas discapacidades físicas, como la tetraplejia. El software TTS proporciona una presentación visual y auditiva sincronizada de texto. Reproduce texto mediante palabras habladas de documento electrónicos. Finalmente, los resultados demostraron que los participantes que utilizaron el software TTS habían mejorado en la lectura, el vocabulario y comprensión.

En la investigación realizada en [6], diseñaron un software para la conversión de texto a voz para personas con discapacidades visuales. El desarrollo del sintetizador de texto a voz en forma de una aplicación simple, convierte el texto ingresado en voz sintetizada y posteriormente se lee al usuario, el cual puede guardarse como un archivo mp3. Logrando obtener un motor para el idioma ingles y posteriormente para el nigeriano.

Con respecto a las Interfaces de Usuario Natural (NUI), los investigadores se centran en la creación y evaluación de gestos y acciones naturales realizados para una interacción multitáctil.

En [7] afirman que las NUI, son consideradas nuevos métodos para la Interacción Humano Computador (HCI) y el diseño de aplicaciones informáticas basadas en interfaces con las cuales las interacciones se realizan a partir de las acciones naturales de los seres humano, es decir, que permiten reconsiderar las habilidades existentes del usuario. Por lo cual, se consideran bastantes adaptables y accesibles para las personas con alguna discapacidad. Por otra parte, afirman que las características de las NUI son centradas en el usuario (necesidades, deseos y limitaciones), multicanal (habilidades sensoriales y motoras), extendidas, de gran ancho de banda, y de interacción basada en voz (procesamiento de voz y comunicación), basadas en imagen (integran imágenes para la comunicación del usuario), basadas en comportamiento (reconocimiento del comportamiento humano y de expresiones) logrando apoyar a los usuarios en sus necesidades.

El reconocimiento de voz o Automatic Speech Recognition (ASR) [8] es el proceso de convertir una señal de voz a una secuencia de palabras mediante un algoritmo, implementado como un programa de computadora [9]. Las interfaces son mecanismos diseñados para permitir el control de un dispositivo o aplicación mediante la interacción verbal. El reconocimiento de voz es generalmente utilizado como una interfaz humano-computadora usado por un software de aplicación, el cual debe cumplir tres tareas

1. Procesamiento: Convierte la entrada de voz a una forma que el reconocedor pueda procesar, es decir, convertir la señal análoga a digital.
2. Reconocimiento: Identifica lo que se dijo, realizando la traducción.
3. Comunicación: Envía el reconocimiento al software de aplicación.

Tradicionalmente, las tres áreas principales de trabajo en el campo del procesamiento de voz desde un punto de vista del procesamiento de señales son: codificación, síntesis y reconocimiento. Lo cual en [10] y [11] demuestran que la síntesis de texto a voz es una tecnología computacional de rápido crecimiento y desempeña un papel cada vez más importante en la forma en que los usuarios interactúan con el sistema y las interfaces logran integrarse a una variedad de plataformas y adaptarse a las necesidades de los usuarios.

3. Metodología

La metodología que se utilizó fue un enfoque cualitativo, usando el Diseño Centrado en el Usuario Inclusivo (DCUI) y el modelo de Prototipos, el cual se describe en el análisis y diseño del sistema.

3.1. Análisis y diseño del prototipo

El DCU inclusivo se basa en los criterios de usabilidad de heurística [12] y en las líneas de la Guía de Accesibilidad al Contenido en la Web. Por lo cual se aplicaron los criterios de: uso de colores, independencia del dispositivo, la navegación mediante comandos de voz. Así también, se realizó una serie de entrevistas y encuestas con el grupo focal, logrando identificar las necesidades de interacción. Para el análisis y el diseño del sistema se determinaron los diagramas de secuencia, como se muestra en Fig. 1. El sistema permite identificar dos usuarios para el acceso y manipulación del sistema:

- Usuario Aplicador: Puede realizar consultas generales de la evaluación del tema y examen

- Usuario Aspirante con baja visión/Ceguera: siendo el cual elije el tipo de evaluación: Prueba de Aptitud Académica (PAA) o Prueba por Área de Conocimiento (PAC). Además, el software incluye una serie de actividades de ejercicios para cada una de las pruebas como una guía de ayuda antes de aplicar la prueba y finalmente se le muestra su resumen de la evaluación.

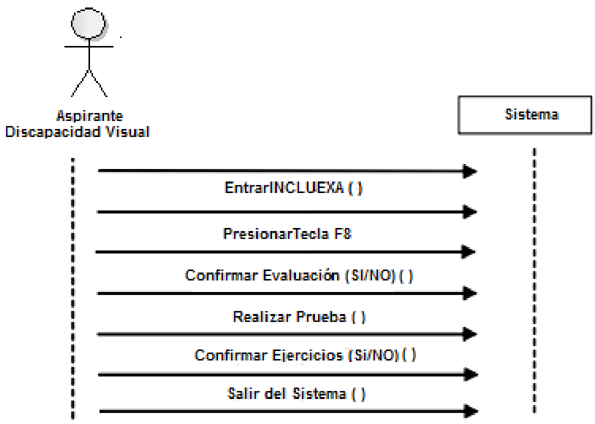


Fig. 1. Diagrama de caso de uso del estudiante.

Así también, se utiliza un dispositivo natural que se encarga de recibir la voz del usuario (la entrada) y posteriormente se encarga de procesar la voz del usuario y ejecutar la función correspondiente (procesamiento de voz) por medio una función lambda que se encarga de interpretar los comandos de voz del usuario, y convertirla en acciones que la API entienda por medio de la libreria Artyom.js en JavaScript. Tal como se muestra en la Fig. 2 las acciones e interacción con el software.

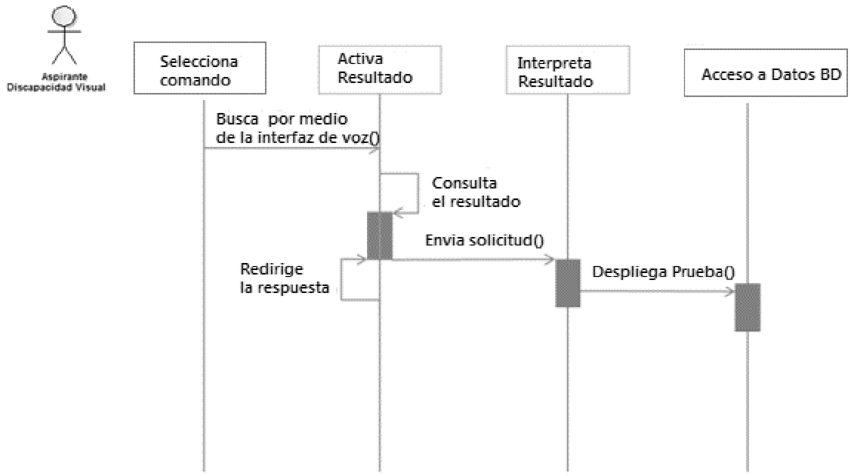


Fig. 2. Diagrama de interacción del usuario con el software.

3.2. Arquitectura y desarrollo del prototipo

El Sistema de reconocimiento de voz, maneja un extractor de características y el clasificador. Cuando se recibe la señal de voz, se pasa a través de un reconocedor que da como resultado la palabra. Después de un procesamiento de lenguaje natural, se realiza una representación semántica y finalmente una acción. Hay dos procesos importantes en la fase de reconocimiento. La extracción de características tiene los siguientes pasos:

- La señal se divide en una colección de segmentos.
- La técnica de procesamiento de señales se utilizará para obtener una representación de las características más distintivas del segmento.
- Basado en sus propias características, se construyen un conjunto de vectores que constituyen la entrada al siguiente módulo.

En cuanto al clasificador probabilístico, tenemos lo siguiente:
Un modelo probabilístico basado en redes neuronales se crea como modelos ocultos de Markov y con la probabilidad de que esto ocurra, se realiza una búsqueda para encontrar la secuencia de segmentos con la mayor probabilidad de ser reconocido.

La estructura del sintetizador de texto a voz se puede dividir en módulos principales:

a) Módulo de Procesamiento de Lenguaje Natural (NLP): Produce una transcripción fonética del texto leído, junto con la fonética.
b) Módulo de procesamiento de señal digital (DSP): transforma la información simbólica que recibe de la PNL en voz audible e inteligible.

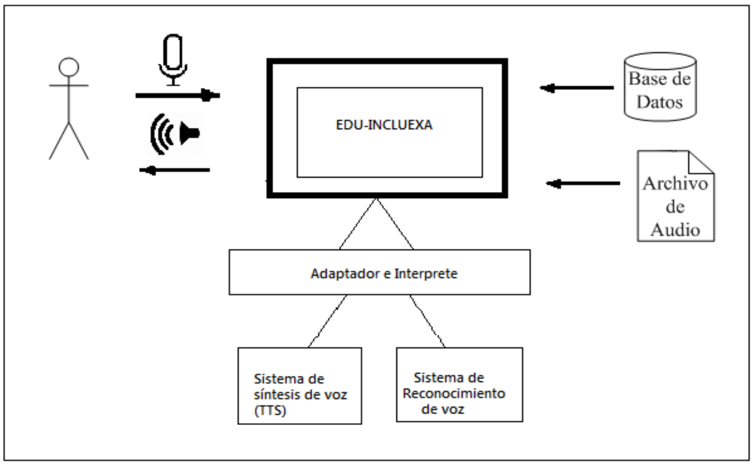


Fig. 3. Arquitectura del sistema.

La biblioteca utilizada en lenguaje JavaScript para el reconocimiento y síntesis de voz fue Artyom.js, que se basa en el "*web-kit-speech-recognition*". Esta biblioteca tiene soporte en español e inglés, por lo que el reconocimiento de voz y el uso de la plataforma se pueden realizar en ambos idiomas. El diseño del prototipo se muestra en la Figura 4, las forma de interactuar, la lista de tareas y configuración de acuerdo a las necesidades del usuario.

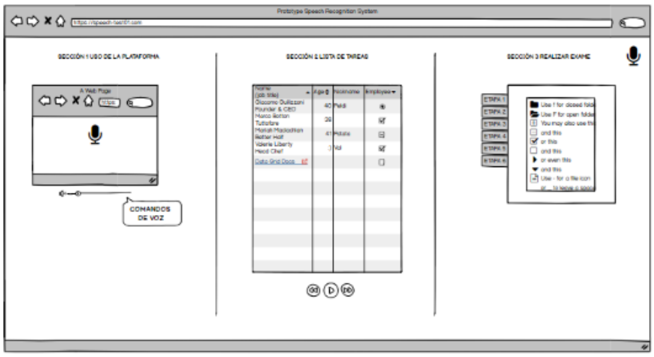


Fig. 4. Prototipo web inclusivo.

El desarrollo del software fue por prototipado, lo cual permitió trabajar con los usuarios. Para la implementación se utilizando HTML, CSS y JavaScript, accediendo directamente a las API de reconocimiento de voz y a la biblioteca Artyom.js. Del mismo modo, se hace uso del reconocimiento de comandos del teclado, de modo que el usuario puede usar algunas combinaciones de teclas para acceder a la plataforma, como se muestra en la Figura 5.

Fig. 5. Ingreso al sistema “EDU-INCLUEXA”.

Después de la pantalla de inicio de sesión, el usuario tiene la posibilidad de navegar por la plataforma, ya sea entre las secciones para la preparación del examen y aplicación de prueba de examen. En la Figura 6 se visualiza la pantalla que presenta las instrucciones para poder contestar las preguntas.

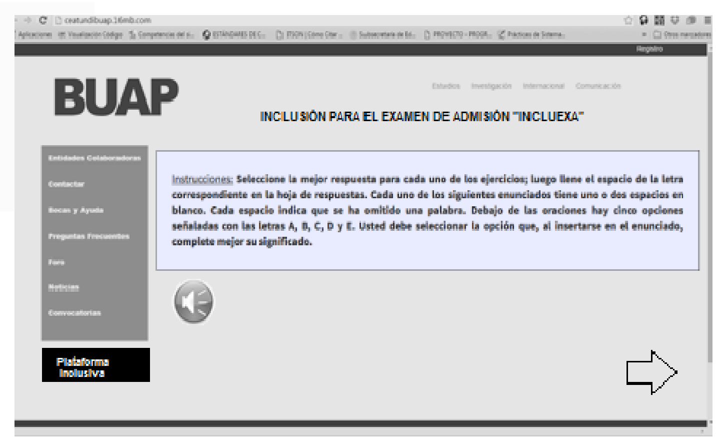


Fig. 6. Pantalla de instrucciones para navegar en la evaluación.

A partir de las instrucciones el usuario con discapacidad visual puede interactuar con voz o teclado para que pueda contestar las preguntas seleccionando la respuesta correcta como se muestra en la Figura 7.

Fig. 7. Pantalla de selección de respuesta por el usuario.

Así también para el seguimiento de su aprendizaje, se le proporciona los resultados en pantalla, se almacena y se genera un archivo en formato pdf, que posteriormente puede consultar el estudiante o evaluador (ver Figura 8).

Fig. 8. Pantalla de resultados del estudiante.

4. Resultados

El prototipo web inclusivo se integró al portal del Centro de Atención a Universitarios con Discapacidad. La prueba funcional del prototipo se llevó a cabo con un grupo focal conformado por dos alumnos con disminución de discapacidad visual, y una alumna con ceguera, los cuales realizaron los ejercicios y pruebas.

Para lo cual, se aplicó la técnica de inspección y exploración bajo tres posibles escenarios que a continuación se describen:

- Situación 1: Al usuario se le dio una breve explicación del uso del sistema y se le acompaño en el sistema en su navegación y ejecución de la evaluación.
- Situación 2: Al usuario se le explicó el uso del sistema y solo se le acompañó al inicio de la evaluación.
- Situación 3: Al usuario se le explico el uso del sistema y no se le acompañó en la navegación ni ejecución de las actividades.

Los usuarios deberán cumplir ciertas tareas para comprobar el funcionamiento del sistema:

- Tarea 1: Acceder al sistema registrando su clave y contraseña, seleccionando el control por voz, comandos o hipervínculos de imágenes.
- Tarea 2: Seleccionar del menú de navegación: Ejercicios de Prueba o Simulador de Aplicación de Prueba de Aptitud Académica con formada por secciones para Lectura, Redacción, Matemáticas e inglés.
- Tarea 3: Recorrer el sistema mediante la interacción seleccionada voz, comandos de teclas o hipervínculos de imágenes.
- Tarea 4: Obtener el resultado de avance de ejercicios y de la aplicación del simulador de la Prueba de Aptitud Académica y generar la evaluación en un archivo.

Los resultados obtenidos de la prueba de inspección se presentan en la Tabla 1. Lo cual refleja que los usuarios con una breve explicación de la Situación 1 su desempeño fue del 87.58% del cumplimiento de las tareas mientras que los usuarios de la Situación 2 al 82.5% y para la situación 3 lograron las tareas solo en un 80.5% esto implica que el software inclusivo apoya la ejercitación para la preparación y simulación del examen de admisión de manera que las interfaces de voz facilitan la interacción con el software a las personas con discapacidad visual.

Tabla 1. Resultados de la Prueba funcional por inspección.

Tarea	Situación 1	Situación 2	Situación 3
1	90%	85%	76%
2	90%	80%	85%
3	85%	85%	80%
4	85%	80%	80%
M=	87.5%	82.5%	80.25%

Esto implica, que el campo del uso de las interfaces de voz debe ser considerado para el diseño de software inclusivo y brindar mayor accesibilidad a las personas que tiene necesidades especiales para la educación.

5. Conclusión y trabajo a futuro

Una de las principales contribuciones, es la aportación de un prototipo de software inclusivo para apoyar a la preparación del examen de admisión a las personas con discapacidad visual logrando integrar las tecnologías de reconocimiento de voz para facilitar la interacción del usuario y el uso de las interfaces naturales.

Este trabajo investigación, cuya finalidad principal fue brindar un apoyo tecnológico para motivar a estudiantes con discapacidad visual para continuar sus estudios a nivel superior y promover la preparación para presentar el examen de admisión para ingresar a las Instituciones de Educación Superior. Dando pauta hacia un nuevo panorama en la educación superior inclusiva apoyado por las tecnologías de voz e interfaces naturales de usuario las cuales brindan una mejor asistencia e interacción con las personas con discapacidad y permiten reducir la brecha digital.

Como perspectiva de este trabajo, se pretende generar más reactivos para la base de datos que apoyen a la Prueba de Aptitud Académica e incorporar la Prueba de Área de Conocimiento. Así como mejorar las interfaces y operatividad del sistema, De tal manera que se pueda constituir una alternativa para presentar el examen de admisión en la educación media superior y superior.

Por otra parte, las pruebas de accesibilidad y usabilidad deben aplicarse a una mayor muestra, para que se obtengan estadísticas sobre el uso del software inclusivo en estudiantes con alguna discapacidad y generar otros apoyos de recursos digitales y herramientas para el aprendizaje de estudiantes en educación superior.

Referencias

1. Instituto Nacional de Estadística y Geografía (INEGI). Censo de Población y Vivienda (2015). Recuperado de http://www.beta.inegi.org.mx/temas/estructura/
2. UNESCO: Directrices sobre políticas de inclusión en la educación. http:// unesdoc.unesco.org/images/0017/001778/177849s.pdf, último acceso 2017/11/09
3. Fernández, J.: Necesidades educativas especiales en el contexto universitario español. Revista de la Educación Superior, 33, 149–162 (2004)
4. McCollum, D., Nation, S., Gunn, S.: The Effects of a Speech-to-Text Software Application on Written Expression for Students with Various Disabilities. National Forum of Special Education Journal, 25, 1–13 (2014)
5. Stoddena, R., Robertsa, K., Takahashia K., Parka, H., Stoddena, J.: Use of text-to-speech software to improve reading skills of high school struggling readers. Procedia Computer Science, 14, 359–362 (2012)
6. Isewon, I., Oyelade, J., Olufunke, O.: Design and Implementation of Text to Speech Conversion for Visually Impaired People, International Journal of Applied Information Systems Foundation of Computer Science FCS, New York, USA, 7(2), 25–30 (2014)
7. Lozada, R, Rivera, L., Molina, F.: Interfaces de Usuario Natural. https://www.researchgate.net/publication/285927377_Interfaces_de_Usuario_Natural,último acceso 2017/10/09
8. Zheng, H., Haiyan, H., Shihong, L., Hong, M.: The Study and Implementation of Text-Speech System for Agricultural Information. In Third IFIP TC 12 International Conference on Computer and Computing Technologies in Agriculture III (CCTA), Oct 2009, Beijing,

China. Springer, IFIP Advances in Information and Communication Technology, AICT-317, 291–296 (2010)

9. Sung, J., Kyun, D., Shin, Y.: Design and Implementation of a Voice Based Navigation for Visually Impaired Persons. International Journal of Bio-Science and Bio-Technology, 5(3),61–68 (2013)
10. Lleida, E: Reconocimiento automático del habla. Centro Politécnico Superior Universidad de Zaragoza. Recuperado de http://www.gtc.cps.unizar.es/~eduardo, último acceso 2017/10/22
11. Newell, A., Gregor, P.: User Sensitive Inclusive Design: in search of a new paradigm. In: CUU First ACM Conference on Universal Usability (2000)
12. Nielsen, J: Ten Usability Heuristics, http://www.useit.com/papers/heuristic/heuristic_list.html, último acceso 2018/01/25

ISSN 1870-4069

Grafos de análisis sintáctico con gramáticas HRGs

Yolanda Moyao[1], Darnes Vilariño[1], Carlos Guillén[2], José de Jesús Lavalle[1]

[1] Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación

[2] Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias Físico Matemáticas

{ymoyao,dvilarino,jlavalle}@cs.buap.mx, cguillen@fcfm.buap.mx

Resumen. En este trabajo se realiza una investigación respecto al estado actual de los trabajos relacionados con la forma en que las Gramáticas de Reemplazo de Hiperaristas (HRGs) pueden ser usadas para definir modelos de lenguajes para transformación de grafos, con el objetivo de resolver problemas en diferentes áreas del ProcesamINTO de Lenguaje Natural (PLN) [5], tales como, análisis sintáctico y semántico, desambiguación del sentido de la palabra, comprensión de texto y resumen, por nombrar algunas. Debido a que el uso de las gramáticas HRGs puede facilitar la comprensión y generación de aplicaciones del Procesamiento de Lenguaje Natural(PLN), se ha incrementado el interés por el estudio del análisis basado en grafos, en partíclar, los grafos se consideran como una herramienta apropiada para representar la estructura semántica dentro del PLN.

Palabras clave: Gramática de reemplazo de hiperarista, procesamiento de lenguaje natural y lenguajes basados en grafos.

Graphs of Syntactic Analysis with Grammars HRGs

Abstract. In this work an investigation is made regarding the current state of the works related to the way in which the Hyperarist Replacement Grammar (HRGs) can be used to define language models for graph transformation, with the objective of solving problems in different Areas of Natural Language Processing (PLN) [5], such as, syntactic and semantic analysis, disambiguation of the sense of the word, text comprehension and summary, to name a few. Because the use of grammars HRGs can facilitate the understanding and generation of applications of Natural Language Processing (PLN), interest in the study of graph-based analysis has increased, in particular, graphs are considered as a tool appropriate to represent the semantic structure within the PLN.

Keywords. Hyperarist replacement grammar, natural language processing and graph-based languages.

1. Introducción

Las gramáticas HRGs, son una generalización de la Gramática Libre de Contexto (CFG) para grafo lenguajes. Un grafo lenguaje es hipergrafo dirigido con hiperaristas etiquetadas sobre marcos semánticos, entidades y roles o argumentos. Una CFG es una gramática que construye cadenas al ir reemplazando símbolos con nuevas subcadenas, mientras que una gramática HRG crea grafos al ir reemplazando aristas con subgrafos.

En los últimos años uno de los temas que ha recibido atención, dentro de la comunidad de Procesamiento de Lenguaje Natural (NLP), son las gramáticas de grafos libres de contexto, ya que éstas pueden ser usadas para generar lenguajes de cadenas sensibles al contexto.

Una de las tareas algorítmicas más importantes en relación con las gramáticas de grafos libre de contexto, así como en las Gramáticas de Reemplazo de Hiperaristas (HRGs), es el análisis sintáctico [5].

Las HRGs son los candidatos formales para la generación y análisis de la representación semántica basada en grafos [18], de hecho, se consideran una herramienta efectiva para resolver problemas de análisis y generación automática de lenguajes libres de contexto, este tipo de gramáticas sirven para resolver cualquier problema dentro del PLN cuya representación pueda llevarse a cabo a través de Grafos Aciclicos Dirigidos (DAG)[10].

Sin embargo, existen problemas dentro del área de PLN que al hacer uso de las HRGs pueden generar lenguajes de grafos libres de contexto para los cuales el análisis semántico es NP-Completo, por tal motivo las investigaciones se enfocan al estudio de técnicas como el árbol de descomposición y el ancho de árbol que hagan a los algoritmos eficientes para algunas clases de grafos restringidos [4].

Existen algunos casos especiales cuya complejidad puede ser polinomial, como el analizador especializado para las HRGs canónicas, las cuales puede ser analizadas de manera eficiente en un tiempo de orden $O(n^c)$, donde c es el número máximo de nodos del lado derecho de la regla presentado por [10]. En este artículo se presentan algunos métodos aplicados por los diferentes autores para las aplicaciones del PLN, a través del uso de las gramáticas HRGs, de la descomposición arbórea y el ancho arbóreo.

El presente artículo está organizado de la siguiente manera. En la sección 2 se presentan los avances en los trabajos de investigación relacionados a la aplicación de los HRGs, en problemas de PLN, en la sección 3 se muestra de la misma manera los avances obtenidos para resolver problemas de PLN con el uso de AMR; por último en las conclusiones se discuten algunas posibles líneas de investigación.

2. HRGs

Los algoritmos para analizar grafos se sabe que de manera general son de complejidad exponencial, es por ello que se está trabajando en diferentes líneas de investigación, enfocadas a la aplicación de teoría de transformación de grafos a NLP, con el objetivo de desarrollar algoritmos que sean eficientes para la generación y análisis de la representación semántica basada en grafos de propósito particular, tal es el caso del estudio y desarrollo de técnicas que se basan en la traslación de las gramáticas HRGs a analizadores de grafos, los cuales ofrecen un gran potencial para aplicaciones del lenguaje natural, en particular, para la generación y comprensión del mismo [2].

Una Gramática de Reemplazo de Hiperarista (HRG) es una tupla $G = (N, T, P, S)$ donde,

- N y T son conjuntos disjuntos finitos de símbolos terminales y no terminales.
- $S \in N$ es el simbolo inicial.
- P es un conjunto finito de producciones de la forma $A \to R$, donde $A \in N$ y R es un grafo fragmentado sobre $N \cup T$.

Un hipergrafo fragmentado es una tupla (V, E, l, X), donde (V, E, l) es un hipergrafo y $X \in V^+$ es una lista de nodos disjuntos llamados nodos externos. Los nodos externos indican como integrar un grafo dentro de otro grafo durante la derivación.

La complejidad en tiempo del algoritmo presentado en [13], es polinomial para el reconocimiento de lenguajes generados por la gramática HRG para grafos conectados de grado acotado, dicho algoritmo está basado en el trabajo presentado por Rozenberg et al. en [17] quiénes mostraron que el análisis sintáctico en gramáticas Controladas por Etiquetas de Nodo (BNLC) acotadas puede realizarse en tiempo polinomial para grafos de grado acotado, árboles, grafos bipartitos completos, grafo outerplanar maximal, grafo con ancho banda $\leq K$, grafos con ancho de corte $\leq K$.

Por otro lado, se tiene el caso del uso de las HRGs libres de contexto que pueden ser representadas por modelos basados en árboles y que además facilitan el uso de herramientas para autómatas arbóreos [12], también, se han presentado otros algoritmos en tiempo polinomial basados en esta técnica, tal es el caso del método para la traducción directa de HRGs a analizadores de grafos combinatorios [14], al igual que el método de [15], en el que utiliza las HRG2s, el árbol de descomposición y el áncho de árbol (tw). Éste juega un papel relevante en teoría de grafos, siendo una característica importante en el algoritmo de árbol de unión del aprendizaje autómatico [15], el cual ha demostrado ser valioso para el proceso de un análisis eficiente. [9]

Por otro lado, [3] describe a detalle un algoritmo más eficiente que el propuesto por [13] para reconocimiento de grafos, esto es factible al considerar que la complejidad computacional de las reglas de reescritura se puede mejorar, debido a que, al llevar a cabo el proceso de binarización de una CFG, que consiste en descomponer una regla de deducción en dos o mas reglas; donde cada una de las nuevas reglas tiene un número más pequeño de variables que la regla original.

También presenta un método de optimización que permite la ejecución en tiempo polinimial del algoritmo de análisis, siempre y cuando el ancho de árbol y el grado del grafo estén acotados; a pesar de esto, aún no se tiene un sistema completo que permita darle solución al problema cuando el tamaño del grafo es muy grande.

En la Fig. 1a se puede observar un grafo cuyo árbol de descomposición se muestra en la Fig. 1b, con un ancho de árbol igual a 3.

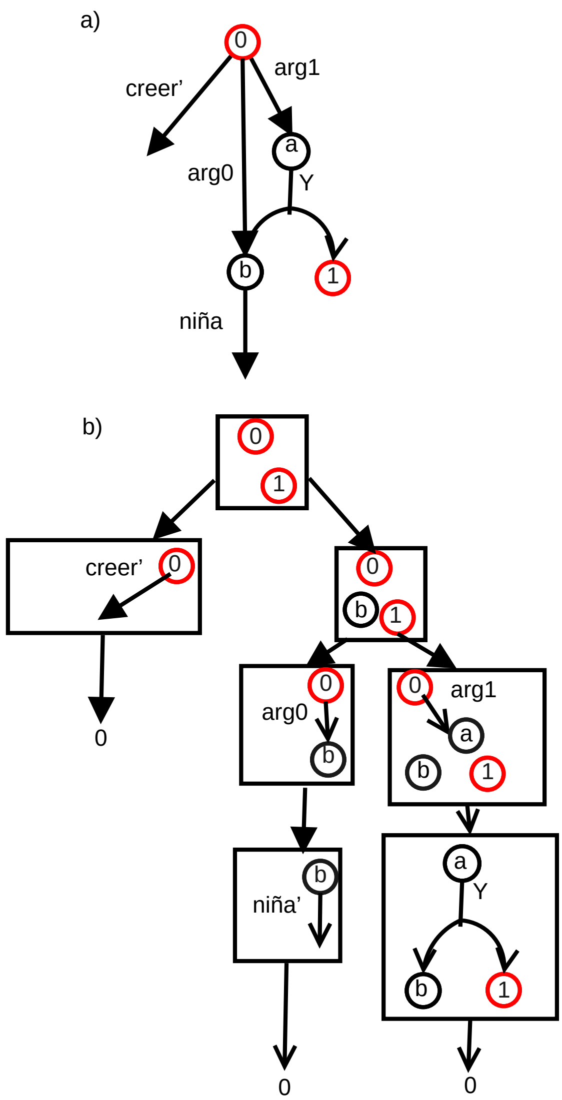


Fig. 1. (a) Parte derecha de la regla, y (b) Una descomposición de árbol.

También, se tienen los modelos generativos de lenguaje de grafos, como el que se introduce en [10], donde se presenta un marco simple para extraer automáticamente HRGs, basado en la definición de árbol de descomposición y luego recorre el árbol para extraer reglas de una forma muy similar a como se extraen reglas de un corpus de árboles RTG (Gramáticas de Árbol Regulares).

Un árbol de descomposición de un grafo $G = (V, E)$ es un tipo de árbol, en el que se tiene un subconjunto de vértices G's para cada nodo. Los nodos de éste árbol T se definen con el conjunto I, y las áristas como el conjunto F. El subconjunto de V asociado con el nodo i de T se denota por X^i, donde $i \in I$ es un subconjunto de V, y el árbol T tiene las siguientes propiedades:

1. Cubierta de vértice: Cada vértice de G está contenido en al menos un nodo del árbol.
2. Cubierta de arista: Para cada arista e del grafo, hay un nodo árbol n tal que cada vértice de $\alpha(e)$ está en n.
3. Manejo de intersección: Dados cualesquiera dos nodos árbol n_0 y n_1, ambos contienen al vértice v, todos los nodos árbol en el único camino de n_0 a n_1 también contienen a v.

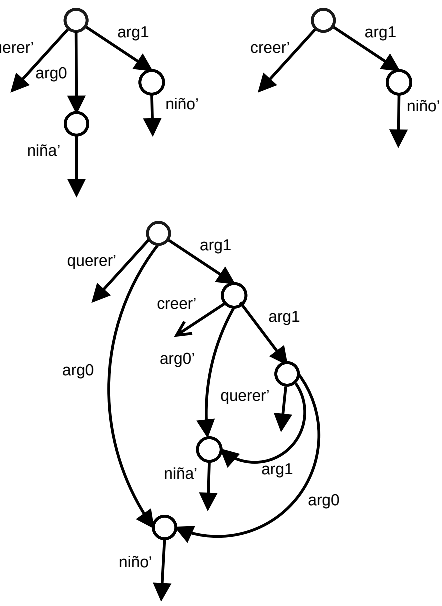


Fig. 2. Algunos elementos de un lenguaje de grafos, representando los significados de (en el sentido de las agujas del reloj desde la esquina superior izquierda): "La niña quiere al niño", "El niño se cree" y "El niño quiere que la niña crea que él la quiere".

Por otro lado, se tiene el caso de la propuesta que hace uso de árbol de descomposición de reglas y grafos de entrada y que trabaja con representaciones basadas en el límite de subgrafos, que ya han sido procesados [5]. Otra de las líneas de investigación, se basa en lenguajes DAG ponderados y que son de interés en NLP, debido a que pueden ser usados para representar la estructura semántica de grafos de forma similar como los de AMR. [6].

A continuación, se muestra un ejemplo con el objetivo de aclarar el concepto de las gramáticas HRGs. Considere un lenguaje de grafo ponderado que involucre solo dos tipos de marcos semánticos (quiero y creo), dos tipos de entidades (niño y niña) y dos roles (arg0 y arg1). La Fig. 2 muestra algunos grafos de este lenguaje. La Fig. 3 muestra cómo derivar uno de estos grafos usando una HRG.

La derivación comienza con un única arista etiquetada con el símbolo no terminal S. El primer paso de re-escritura reemplaza esta arista con un subgrafo, que podemos leer como "El muchacho quiere que algo (X) involucra a sí mismo..El segundo paso de re-escritura reemplaza la arista X con otro subgrafo, que podríamos leer como "El chico quiere que la chica crea algo (Y) involucrando a ambos ". La derivación continúa con un tercer paso de re-escritura, después del cual ya no hay aristas etiquetadas sin elementos terminales.

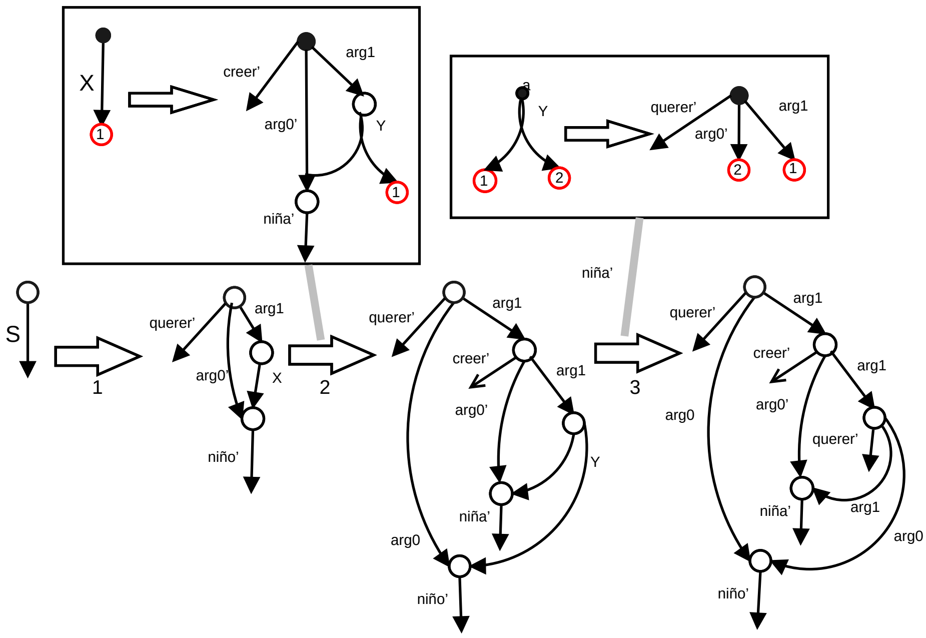


Fig. 3. Derivación de una gramática de sustitución de hiperarista para un grafo que representa el significado de "el chico quiere que la chica crea que él la quiere".

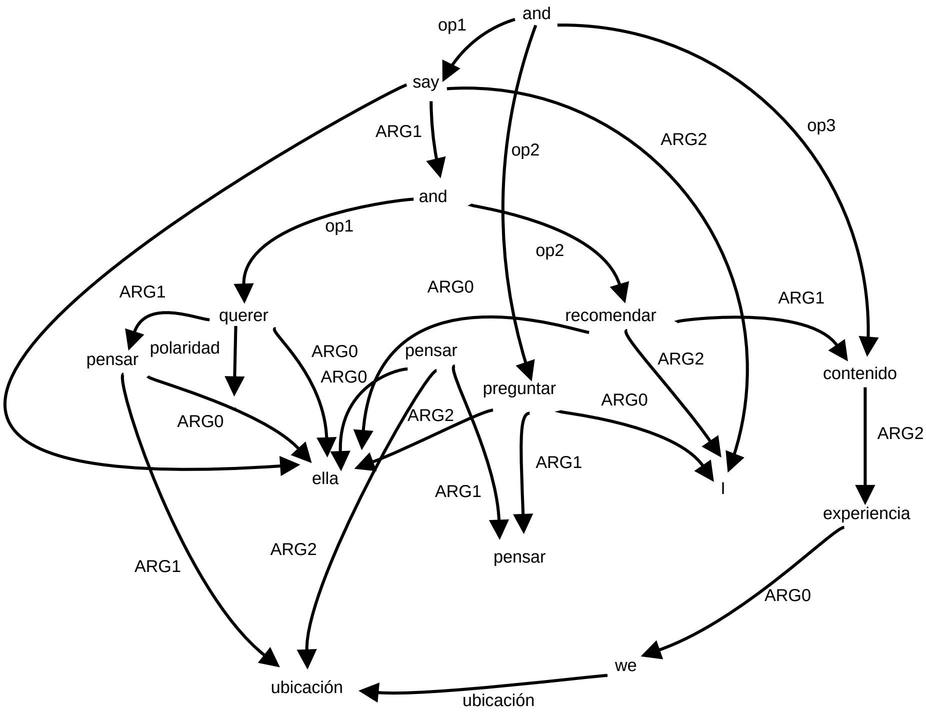


Fig. 4. Una AMR.

3. Grafos para la Representación de Significado Abstracto (AMR)

Las AMR fueron introducidas por [1] como una notación gráfica independiente del dominio para la semántica de significados en lenguaje natural. Las AMR tienen un propósito similar en el ámbito de la representación semántica como lo hace el conocido árbol constituyente para la representación sintáctica. Para este último, se tiene una gran variedad de modelos formales que representan la estructura de la representación correcta de la gramática de árbol regular.[8] y [7].

Un AMR generalmente no es un árbol sino un Grafo Aciclico Dirigido(DAG) como el AMR (algo simplificado y abstracto). Sus vértices son principalmente conceptos de PropBank. [11] conectados por aristas que están etiquetadas por etiquetas de roles, de forma intuitiva, suministrando los conceptos con sus argumentos semánticos.

En este caso, los árboles de dependencia muestran la similitud, pero también la diferencia, entre varios conceptos, que comparten un argumento semántico.

A continuación, se presenta un ejemplo, cuyo grafo se puede observar en la Fig. 4. ”Le pregunté qué pensaba sobre dónde estaríamos y ella dijo ella no quiere pensar en eso, y que debería estar feliz con las experiencias que hemos tenido (el cual Yo soy)."Los nombres de los cuadros de PropBank (las etiquetas de los vértices) se han simplificado en aras de la legibilidad. Se pueden encontrar en [3], que también usa este ejemplo, como nota al margen; uno puede notar que la AMR no especifica si su dicho fue la respuesta a mi pregunta, al revés, o los dos eran independientes. Esto, sin embargo, contribuiría a una discusión sobre los límites de AMR en lugar de modelos de autómatas formales para AMR.

Los AMR pueden ser utilizados para la construcción de Máquinas de Traducción Automática (MT), representación de corpus multilingues, entre otras tareas de PLN.

4. Conclusiones

De acuerdo a los trabajos revisados en este articulo se puede observar que en los últimos años ha ido creciendo el interés por la investigación de aplicaciones de gramáticas modeladas con grafos para resolver problemas dentro del área de PLN. Se pueden mencionar algunos resultados importantes, tales como, el uso de las HRGs que son consideradas como una herramienta efectiva para resolver problemas de comprensión y generación del lenguaje natural, siempre y cuando el problema pueda ser representado por grafos dirigidos y aciclicos.

También se puede concluir que de acuerdo a los trabajos relacionados al uso del árbol de descomposición de un grafo, en general tiene un comportamiento computacional $NP - Completo$, sin embargo, cuando el ancho de árbol para el grafo es acotado entonces, el comportamiento computacional para el algoritmo se puede reducir a un comportamiento de orden polinomial $O(n^c)$.

Dentro de las posibles líneas de investigación relacionadas con la presente revisión, se puede mencionar la necesidad de incorporar otras métricas arbóreas que conduzcan a algoritmos eficientes para clases menos restrictivas de grafos gramáticas, al mismo tiempo la inclusión de un análisis y tratamiento con parámetro fijo tratable más profundo.

Referencias

1. Banarescu, L., Bonial, C., Cai, S., Madalina, Georgescu, Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N.: Abstract meaning representation for sembanking. Proceedings 7th Linguistic Annotation Workshop, ACL 2013, Workshop (2013)
2. Chiang, D., Drewes, F., Gildea, D., Lopez, A., Satta, G.: Weighted DAG automata for semantic graphs. Submitted (2013)
3. Chiang, D., Andreas, J., Bauer, D., Moritz, H.K., Jones, B., and Knight, K.: Parsing graphs with hyperedge replacement grammars. Proceedings of the 51st Meeting of the ACL (2013)
4. Drewes, F., Habel, A., Kreowski, H.: Hyperedge replacement graph grammars. In: Rozenberg, G., editor, Handbook of Graph Grammars and Computing by Graph Transformation, World Scientific, 95–162 (1997)

5. Drewes, F., Hoffmann, B., Minas, M.: Predictive top-down parsing for hyperedge replacement grammars. Proceedings 8th Int'l Conf. on Graph Transformation (ICGT'15), Lecture Notes in Computer Science, Springer (2015)
6. Drewes, F.: Tutorial: Introduction to Graph Transformation Report from Dagstuhl Seminar 15122 Formal Models of Graph Transformation in Natural Language Processing Edited by Drewes F., Knight K., and Kuhlmann M. (2017)
7. Drewes, F.: DAG Automata for Meaning Representation. Proceedings of the 15th Meeting on the Mathematics of Language, London, UK, Association for Computational Linguistics, pages 88–99, (2017)
8. Gécseg, F., Steinby, M.: Tree Automata. Akadémiai Kiadó, Budapest. Online version available under https://arxiv.org/ abs/1509.06233 (1984)
9. Gildea, D.: Grammar factorization by tree decomposition. Computational Linguistics, Association for Computational Linguistics, 37(1), 231–248 (2011)
10. Jones, B., Andreas, J., Bauer, D., Hermann K., Knigh, K.: Semantics-Based Machine Translation with Hyperedge Replacement Grammars. Proceedings of COLING 2012: Technical Papers, pages 1359–1376, COLING 2012, Mumbai, December (2012)
11. Kingsbury, P., Palmer, M.: From Treebank to propbank. Proc. 3rd Intl. Conf. on Language Resources and Evaluation (LREC 2002) (2002)
12. Knight, K., Graehl, J.: An overview of probabilistic tree transducers for natural language processing. Proceedings of the 6th International Conference on Intelligent Text Processing and Computational Linguistics (2005)
13. Lautemann, C.: The complexity of graph languages generated by hyperedge replacement. Acta Informatica, 27, 399–421 (1990)
14. Mazanek, S., Minas, M.: Parsing of hyperedge replacement grammars with graph parser combinators. Proc. 7th International Workshop on Graph Transformation and Visual Modeling Techniques (2008)
15. Moot, R.: Lambek grammars, tree adjoining grammars and hyperedge replacement grammars. Proc. TAG+9, pages 65–72 (2008)
16. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco, CA, 2nd edition (1988)
17. Rozenberg, G., Welzl, E.: Boundary NLC graph grammars basic definitions, normal forms, and complexity. Information and Control, 69, 136–167 (1986)
18. Teichmann, C. Drewes, F.: Efficient HRG-Parsing for the NLP Domain. Report from Dagstuhl Seminar 15122 Formal Models of Graph Transformation in Natural Language Processing Edited by Drewes F., Knight K., and Kuhlmann M. (2015)

Electronic edition
Available online: http://www.rcs.cic.ipn.mx