Advances in Artifical Intelligence

Research in Computing Science

***RESEARCH IN COMPUTING SCIENCE**, Año 19, Volumen 148, No. 11, Noviembre del 2019, es una publicación mensual, editada por el Instituto Politécnico Nacional, a través del 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, Ciudad de México, Tel. 57 29 60 00, ext. 56571. https://www.rcs.cic.ipn.mx. Editor responsable: Dr. Grigori Sidorov. Reserva de Derechos al Uso Exclusivo del Título No. 04-2019-082310242100-203. ISSN: en trámite, ambos otorgados por el Instituto Nacional del Derecho de Autor. Responsable de la última actualización de este número: el Centro de Investigación en Computación, Dr. Grigori Sidorov, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738. Fecha de última modificación 08 de Noviembre de 2019.*

***RESEARCH IN COMPUTING SCIENCE**, Year 19, Volume 148, No. 11, November 2019, is a monthly publication edited by the National Polytechnic Institute through the Center for Computing Research. Av. Juan de Dios Bátiz S/N, Esq. Miguel Othón de Mendizábal, Nueva Industrial Vallejo, C.P. 07738, Mexico City, Tel. 57 29 60 00, ext. 56571. https://www.rcs.cic.ipn.mx. Editor in charge: Dr. Grigori Sidorov. Reservation of Exclusive Use Rights of Title No. 04-2019-082310242100-203 ISSN: pending, both granted by the National Copyright Institute. Responsible for the latest update of this issue: the Computer Research Center, Dr. Grigori Sidorov, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738. Last modified on November 8, 2019.*

Advances in Artificial Intelligence

María de Lourdes Martínez Villaseñor (ed.)

Instituto Politécnico Nacional,
Centro de Investigación en Computación, México 2019

ISSN: en trámite

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

The intention of this issue of the journal "Research in Computer Science" is the collection of works that apply computational intelligence techniques, addressing problems on topics of intelligent applications, vision and robotics, machine learning, and natural language processing. There are 16 papers presenting intelligent applications for different domains in this issue.

Intelligent solutions like an Internet of Things Platform, crop anlysis, and support for decision making in an echelon supply chain distribution are included in this volume.

Within the topics on computation vision and robotics, works are presented on depth estimation for the NAO robot, collective behaviors in swarms of builder robots, a local image feature approach for RGBD semantic segmentation, and identification of static and dynamic signs of the Mexican Sign Language alphabet.

Research papers on natural language processing are also included in this number: Semantic annotation approach for information search, word embedding as domain ontology enrichment resource, a graph-based tool to produce a model for constraint satisfaction problems, authorship attribution, and a study on how to learn Picture Languages.

Additionally, works are included on diverse applications as a molecular marker for protein electrophoresis using wavelet transform, and a brain-computer system to measure brain activity during a dolphin-assisted therapy.

The papers published here were carefully selected by the editorial committee and revised by at least two external reviewers considering their scientific originality and technical quality. 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).

We would like to thank Mexican Society for Artificial Intelligence (Sociedad Mexicana de Inteligencia Artificial) for their support in the preparation of this volume.

María de Lourdes Martínez Villaseñor
Universidad Panamericana, Mexico
Guest Editor

November 2019

ISSN 1870-4069

Table of Contents

Page

ISSN 1870-4069

Selection of IoT Platform with Multi-Criteria Analysis: Defining Criteria and Experts to Interview

Roberto Contreras-Masse[1], Alberto Ochoa-Zezzatti[1], Vicente García[1], Mayra Elizondo[2]

[1] Universidad Autónoma de Ciudad Juárez, Ciudad Juarez, Mexico
rcontreras@itcj.edu.mx
[2] Universidad Nacional Autónoma de México, México City, Mexico

Abstract. Industry 4.0 is having a great impact in all industries. This is not a unique product, but is composed of several technologies. IoT is a key intelligent factor that allows factories to act intelligently. By adding sensors and actuators to the objects, the object becomes intelligent because it can interact with people, other objects, generate data, generate transactions and react to the environment data. Currently there are very varied implementation options offered by several companies, and this imposes a new challenge to companies that want to implement IoT in their processes. The decision processes that companies must follow should not be free will or by hunches, since this contradicts a methodology and would make the decision process unrepeatable and unjustifiable. Decisions must be supported by methods that consider pros and cons of plural points of view that affect the decision process. With a wide range of IoT platforms, which are not directly comparable to each other, it seems that Multi-Criteria Decision Analysis (MCDA) can be useful to help companies make a decision on what platform to implement, depending on the circumstances prevailing in each company at the time to make the choice. This article shows the complexity of selecting an IoT platform and provides the key decision criteria that must be taken into account when evaluating IoT Platforms alternatives.

Keywords: IoT, platform selection, multi criteria analysis, MCDA, AHP, PROMETHEE.

1 Introduction

Industry 4.0 is having high impact in all industries. This is not a unique product, but is composed of several technologies. Boston Consulting Group has defined nine technological pillars for Industry 4.0: cloud, additive manufacturing, simulation, big data and analysis, autonomous robots, augmented reality, integration of horizontal and vertical systems, cybersecurity and industrial internet of things (IIOT) [22]. IIOT has been used not only in the manufacturing industry, but has expanded to other industries such as health, travel and transportation, energy,

gas and oil, etc. This is one of the main reasons that IIOT is known as the Internet of Things (IoT) [11]. IoT is a key intelligent factor that allows factories to act intelligently. By adding sensors and actuators to objects, the object becomes intelligent because it can interact with people, other objects, generate data, generate transactions and react to environmental data [13,17]. Cities do not ignore this trend, since there is a plan to turn cities into smart cities in certain countries [20].

The decision processes that companies must follow should be supported by methods that consider pros and cons of plural points of view that affect the decision process. Researchers and practitioners have developed over time the techniques that today are part of the domain of Multiple Criteria Decision Analysis (MCDA) which, very simplistically, requires three basic elements: a finite set of actions or alternatives, at least two criteria and at least one decision-making [5]. The MCDA has been the object of study and nowadays there are a lot of methods for decision making in disciplines such as waste management, industrial engineering, strategies, manufacturing, even natural resource management and environmental impact [14].The purpose of this manuscript is precisely to propose a method of MCDA with the corresponding criteria for the selection of an IoT platform, which can serve as a starting point to companies and individuals embarked on implementation projects of Industry 4.0. Our conceptual model to solve the problem is shown in Figure 1.

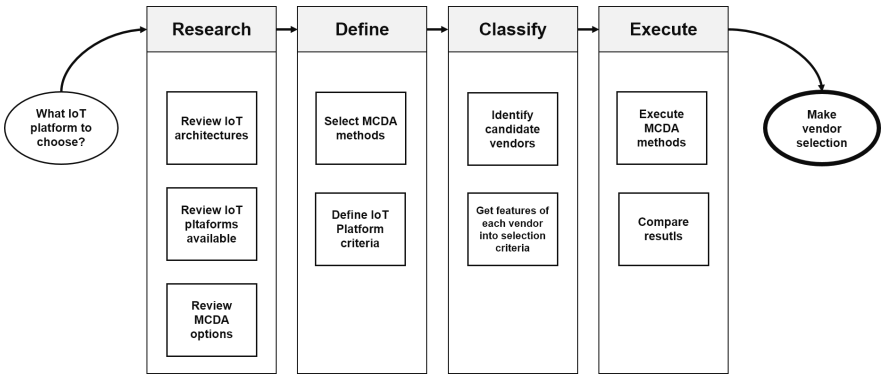


Fig. 1. Conceptual model to select IoT platforms.

This work is organized by sections. Section 2 shows the complexity to compare commercial IoT platforms, including a quick view of the different architectures found in the literature. Section 3 establishes the MCDA reference framework and the methods considered for the selection of platforms, taking into account similar efforts reported in the literature. Section 4 proposes the characteristics to be considered for MCDA. Finally, section 5 discusses the future work to be done and what MCDA methods could fit this kind of decision problem.

2 IoT Architectures and Commercial Platforms

Internet of Things (IoT) continues to evolve. Due to the instrinsic complexity, it is good practice to look at architectural references. IoT have five main requirements in general basis [30]: 1) Enable communication and connectivity between devices and data processing; 2) Establish a mechanism to manage devices, including tasks such as adding or deleting devices, updating software and configurations; 3) Gather all the data produced by the devices and then analyze them to provide a meaningful perspective to the companies or users; 4) Facilitate scalability to handle the increased flow of "data pipes" (hereinafter referred to as data pipelines) and the flow of data, and handle an increasing number of devices; 5) Protect the data by adding the necessary functions to provide privacy and trust between the devices and the users. Table 1 shows the summary of the various multi-layer architectures found in the literature. Technical architecture provides an extreme value to users because it can be

Table 1. IoT Architectures.

Num.	Layers	References
2	Devices and Communication	[28]
3	Devices, Communication and Application	[9,16,21]
4	Devices, Communication, Transport and Application	[4,8,6,18,21,28]
5	Devices, Local processing, Communication, Transport and Applications	[21]
7	Business, Management, Communication, Processing, Acquisition, User interaction and Security	[2,6]
8	Physical devices, Communication, Edge or Fog processing, Data storage, Applications, Collaboration and process, Security	[19]

implemented with different products. Therefore, it is understandable that several companies offer IoT platforms that can be useful for our architectures. Commercial providers aim to flexible options offered, and consumers are responsible for using each component in the best way they consider. The main commercial players identified are, in alphabetical order: Amazon Web Services, Bosch IoT Suite, Google Cloud Platform, IBM Blue Mix (now Watson IoT), Microsoft Azure IoT and Oracle Integrated Cloud [3]. The leading players identified in 2014 by Gartner Group were AWS and Microsoft, but in 2018 Google enters the leaders quadrant. IBM, for its commercial relevance is considered, although it has become a niche player, along with Oracle. Although Bosch IoT does not appear in the panorama detected by Gartner, we include it for being used in several industries. Each of these suppliers has similar characteristics among them and have differentiated within their offer.

3 MCDA as a Tool for Selection of IoT Platform

Making a decision introduce problems to individuals. One of the problems is the integration of heterogeneous data and the uncertainty factor surrounding a decision, and the criteria that usually conflict with each other [14,32]. To carry out a MCDA process, a series of tasks is proposed, based on the three generic steps suggested by [12]: i) identify the objective or goal, ii) select the criteria, parameters, factors, attributes, iii) selection of alternatives, iv) association of attributes with the criteria, v) selection of weight methods to represent the importance of each criterion, and vi) the method of aggregation. [12] included a step that is left out of these proposed tasks, but which should be considered in the discussion before executing the selected action. This step is to understand and compare the preferences of the person making the decision.

The MCDA can be classified according to the basis of the problem, by type, by category or by the methods used to make the analysis. Figure 2 shows a taxonomy adapted from [31]; the methods included in this taxonomy are not exhaustive. The MCDA is a collection of systematic methodologies for comparisons, classification and selection of multiple alternatives, each one with multiple attributes and is dependent on an evaluation matrix. Generally it used to detect and quantify the decisions and considerations from interested parties (stakeholders) about various monetary factors and non-monetary factors to compare alternative course of action [14,31]. The major division that exists in MCDA lies in the category of methodologies. First group considers discrete values with a limited number of known alternatives that involve some compensation or trade-off. This group is called Multiple Attributes Decision Making (MADM). The other group is the Multiple Objectives Decision Making (MODM) and its variable decision values are within a continuous domain with infinite or very numerous options that satisfy the restrictions, preferences or priorities [32]. Also, there is another classification according to the way of adding criteria and it is divided into the American school, which aggregates into a single criterion, and into the European or French school that uses outranking methods. It can be considered a mixture of both schools and they are indirect approaches, such as the Peer Criteria Comparison methods (PCCA) [29].

3.1 Use of MCDA for Selection of IoT Platforms or Technology Platforms: Related Work

When finding the available alternatives of the market, a new question will arise to find the method that helps to select the appropriate option. To answer this last question, a review of the literature is made looking for: a) MCDA methods applied to the selection of IoT platforms and b) knowing the criteria taken into account.

In the literature there is little information on the subject in recent years. Table 2 shows the summary of the work found. The selected methods are focused on AHP, TOPSIS and Fuzzy logic in AHP and TOPSIS. The outranking methods were not implemented, but were considered as an option or for future work by

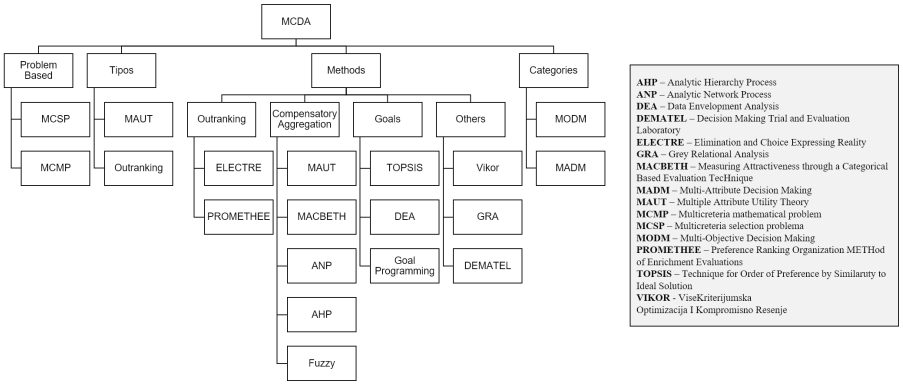


Fig. 2. Taxonomy of MCDA (Adapted from [31]).

some authors [24,26]. The selection of an IoT platform is not dominated by a single criterion, nor is there a single alternative. [15] considered AWS, Azure, Bosch, IBM Watson and Google Cloud within their options, which coincide with some of the alternatives considered in this manuscript. Therefore, it is interesting to review the criteria they included for MCDA, as summarized in Table 2.

Criteria found in literature are purely technical with some hints of economy, and can be found as part of the characteristics of IoT architecture [10]. But when implementing an IoT platform, non-technical aspects should also be considered. As the platform to be considered has its foundation in the cloud, it is valid to review the criteria included in previous MCDA exercises to select a cloud provider, looking for non-technical aspects.

The criteria for selecting a cloud proposed in the CSMIC Framework v 2.1 of 2014 [3] as Index of Measure of Service (SMI) include topics of interest to the organization, financial and usability, together With the technical issues [7]. Some of these criteria can be included to complement the analysis having the technical point of view and the business point of view.

Finally, there is the question about which methods are suitable for this type of problems, noting that the previous work includes AHP, ANP, TOPSIS and Fuzzy Logic, but they leave aside for future research methods such as PROMETHEE and ELECTRE. In the MCDA universe there are many more methods available. Following the decision tree to select an MCDA method written by [29], which considers 56 methods, the number of options can be easily reduced. In the case of selecting an IoT platform that has different criteria, the problem has the characteristics of classification or ranking, ordering the options from best to worst. This technique is useful in real life, since they are hardly conform and

[3] Cloud Services Measurement Initiative Consortium (CSMIC) was created by Carnegie Mellon University to develop Service Measurement Index (SMI). it can be found at `https://spark.adobe.com/page/PN39b/`

Table 2. MCDA related work to select technology.

Yr.	Application	MCDA	Criteria	Ref.
2019	IoT Challenges	AHP, ANP	Communication, Technology, Privacy and security, Legal regulations, Culture	[27]
2018	Cloud service for IoT	FAHP, FTOPSIS	Availability, Privacy, Capacity, Speed, Cost	[25]
2018	Platform IoT	Fuzzy	Security, Device management, Integration level, Processing level, Database functionality, Data collection protocols, Visualization, Analytics variety	[15]
2018	IaaS	TOPSIS	Cost, Computing required, Storage capacity, Operating system	[26]
2018	Distributed IoT Databases	AHP	Usability, Prtability, Support	[1]
2017	IoT Device	AHP	Energy consumption, Implementation time, Difficulty of implementation, Cost, Clock device	[24]
2017	IoT Platform	AHP	Energy, Cost, Computing speed, Data memory, Program memory, device weight	[23]
2013	Ranking cloud services	AHP	Responsibility, Agility, Service assurance, Cost, Performance, Security and privacy, Usability	[7]

subject themselves to a single option, but they have to consider their primary option and another option as backup, assuming that the first option is not viable.

The candidate methods found are COMET, NAIADE II, EVAMIX, MAUT, MAVT, SAW, SMART, TOPSIS, UTA, VIKOR, Fuzzy SAW, Fuzzy TOPSIS, Fuzzy VIKOR, PROMETHEE II, PAMSSEM II, Fuzzy PROMETHEE II, AHP + TOPSIS, AHP + VIKOR, fuzzy AHP + TOPSIS, AHP + Fuzzy TOPSIS, Fuzzy ANP + Fuzzy TOPSIS, AHP, ANP, MACBETH, DEMATEL, REMBRANDT, Fuzzy AHP and Fuzzy ANP.

Of the 29 methods suggested by the decision tree, those used in the literature are included for this type of problem. However, although it would be a very interesting exercise to compare the 29 methods with each other, it is beyond the scope of this article. As the AHP method has been used regularly we suggest to take it as one of the two methods proposed. The other selected method is PRMOETHEE II, which has not been used in previous works, but some authors have considered it for future work.

4 Proposed Criteria and Roles to Participate

In our experience, companies that want to implement IoT show great enthusiasm for the initiative, but on several occasions they have a misconception of what IoT

entails. IoT concepts are technical and of great interest to engineers and systems architects, but the business factors, cost aspects, methods of payment, and commercial conditions, all of them are of great interest for senior management represented by the Chief Officers, referred often as CxO Level. In addition, the wide offer that exists in the market where suppliers have different prices and service schemes make it difficult to compare among each other, or at least difficult to do a linear comparison.

Our proposal identifies and suggests the criteria required for IoT Platform selection for a MCDA exercise with at least two different methods, enabling organizations to compare results and make a well-founded decision. This work does not provide a universal and definitive solution, but rather, it proposes the methodology that any company, be it small or large, can use to decide on the IoT platform that best suits their circumstances and needs. Following the general MCDA process depicted in Figure 3, the decision objective is the selection of an IoT platform.

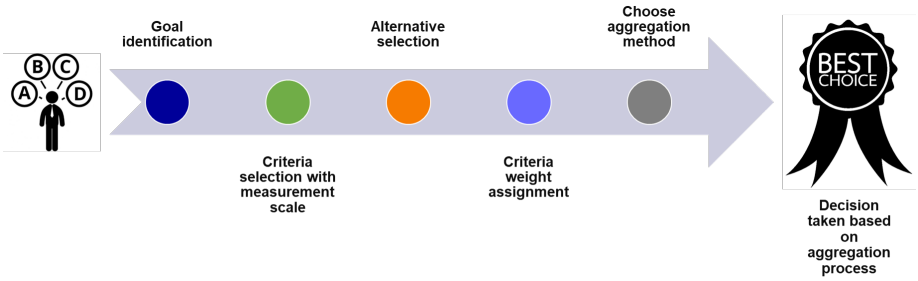


Fig. 3. Conceptual model to select IoT platforms.

The selection of criteria must be consistent with the decision and each criterion must be independent of one another. Each criterion must also be measured on the same scale and applicable to all alternatives. The Table 3 summarizes the criteria to be used together with its definition. Criteria that are qualitative, i.e. based on expert judgement, can be measured with the Saaty scale [24,27]. Criteria that are quantitative should consider equal scenarios, such as the cost of data transmission, which for all alternatives should be calculated with the same number of devices, same message size and same number of messages per day.

The selected criteria are divided into three major areas of interest: technical, economic and social. This is a difference over previous works found in the literature. The selected criteria are also classified as quantitative and qualitative according to their nature, and are summarized in Table **??**.

The existing alternatives for the IoT platform considered in this paper appear in the literature or are widely used in the industry and are recognized as market leaders, in addition to the author's experience with various global suppliers. Thus, the alternatives included in this exercise are: AWS IoT Platform, Microsoft

Azure IoT Platform, Bosch IoT Suite, IBM Watson IoT Platform, Google Cloud IoT Platform, GE Predix IoT Platform, Thingworx (PTC), and SAP Cloud IoT.

Our proposal includes profiles of people who must participate in the expert judgement exercise. It is important that they are not only dedicated to technology in order to enrich the exercise. Table 5 lists the desirable profiles of people who should be involved in a MCDA exercise as experts.

5 Conclusion and Future Work

Selecting an IoT Platform is not an easy task, as it has been found in literature the vast amount of different architectures, vendors and approaches. This work concentrated to find the criteria and labor roles to take a decision on what platform to use.

As a first conclusion, we found MCDA has been used effectively in technological decisions, very close to what we were looking for, but not exactly the problem we faced. AHP and TOPSIS were MCDA methods employed in the past, with good results. However, outranking methods such as PROMETHEE have not been used, but few authors considered it for future work. This opens the opportunity for us to explore these methods in future work. In addition to this, PRMOETHEE I and II are accepted MCDA methods for ranking options, as it was found in literature.

Second, our research suggests the criteria found in previous work are technical oriented, with few criteria considering other business areas, such as economic ones. Most authors that included these kind of criteria, mainly considered cost, which is important but not the only one. Our work is different and provides a more comprehensive criteria, with updated technical aspects based on literature and technical experience; our economic aspects include prestige and market longevity, cost of training, and free tier bonus (offered by most of vendors to compete against others). Another important contribution from our work is the inclusion of social criteria, important for an organization to cover human resources skills. We recommend to consider three main social criteria: Level of Support found in the community, facility to find human resources available in market, and available training programs, offered by vendor, private entities or universities.

As IoT and technology is now part of core business, experts should not come only from IT department, but from different areas of organization. This is our third conclusion, as we suggest the CxO levels that should be considered to provide preferences and expert judgment. One important role is Business Unit Leader, as it deals with daily problems, customers, clients, and details that may not be visible to CxO level.

Our future work will consist to create the mechanism to gather experts' judgment information in a simple and efficient way, and test it against two MCDA methods. One of them will be AHP, widely used in literature, and it could be used as a control method. The second method to use will be PROMETHEE II, as it is a complete outranking method. Also, comparison of those two methods

will help to find effectiveness of methods. Our future work is also considering test our methodology in a large organization and publish our findings.

References

1. Alelaiwi, A.: Evaluating distributed IoT databases for edge/cloud platforms using the analytic hierarchy process. Journal of Parallel and Distributed Computing 124, 41–46 (2019)
2. Contreras-Castillo, J., Zeadally, S., Guerrero Ibáñez, J.A.: A seven-layered model architecture for Internet of Vehicles. Journal of Information and Telecommunication (2017)
3. Dumitru, R.L.: IoT Platforms: Analysis for Building Projects. Informatica Economica (2017)
4. Ferreira, H.G.C., Dias Canedo, E., De Sousa, R.T.: IoT architecture to enable intercommunication through REST API and UPnP using IP, ZigBee and arduino. In: International Conference on Wireless and Mobile Computing, Networking and Communications (2013)
5. Figueira, J., Greco, S., Ehrgott, M.: Multiple criteria decision analysis: state of the art surveys, vol. 78. Springer Science & Business Media (2005)
6. Firdous, F., Mohd Umair, M., Alikhan Siddiqui, D., Mohd Umair, A.: 512 Ms IoT Based Home Automation System over the Cloud. Tech. rep. (2018)
7. Garg, S.K., Versteeg, S., Buyya, R.: A framework for ranking of cloud computing services. Future Generation Computer Systems 29(4), 1012–1023 (2013)
8. Gazis, V., Goertz, M., Huber, M., Leonardi, A., Mathioudakis, K., Wiesmaier, A., Zeiger, F.: Short paper: IoT: Challenges, projects, architectures. In: 2015 18th International Conference on Intelligence in Next Generation Networks. pp. 145–147. IEEE (2015)
9. Gironés, T., Canovas Solbes, J., Parra-Boronat, A.: An Integrated IoT Architecture for Smart Metering. IEEE Communications Magazine 54(12), 50–57 (2016)
10. Guth, J., Breitenbucher, U., Falkenthal, M., Leymann, F., Reinfurt, L.: Comparison of IoT platform architectures: A field study based on a reference architecture. In: 2016 Cloudification of the Internet of Things, CIoT 2016 (2017)
11. Hatzivasilis, G., Fysarakis, K., Soultatos, O., Askoxylakis, I., Papaefstathiou, I., Demetriou, G.: The Industrial Internet of Things as an enabler for a Circular Economy Hy- LP: A novel IIoT protocol, evaluated on a wind park's SDN/NFV-enabled 5G industrial network (2018)
12. Henig, M.I., Buchanan, J.T.: Solving MCDM problems: Process concepts. Journal of Multi-Criteria Decision Analysis 5(1), 3–21 (1996)
13. Höller, J., Tsiatsis, V., Mulligan, C., Karnouskos, S., Avesand, S., Boyle, D.: IoT Architecture – State of the Art. In: From Machine-To-Machine to the Internet of Things (2014)
14. Huang, I.B., Keisler, J., Linkov, I.: Multi-criteria decision analysis in environmental sciences: ten years of applications and trends. Science of the total environment 409(19), 3578–3594 (2011)
15. Kondratenko, Y., Kondratenko, G., Sidenko, I.: Multi-criteria decision making for selecting a rational IoT platform. In: 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT). pp. 147–152. IEEE (2018)

16. Krishnamurthy, R., Cecil, J., Perera, D.: IMECE2017-72293 An Internet of Things (IOT) Based Frameworks for Colloborative Manufacturing. Tech. rep. (2017)
17. Lanotte, R., Merro, M.: A semantic theory of the Internet of Things. Information and Computation 259, 72–101 (2018)
18. Nitti, M., Pilloni, V., Giusto, D., Popescu, V.: IoT Architecture for a sustainable tourism application in a smart city environment. Mobile Information Systems (2017)
19. Rahimi, H., Zibaeenejad, A., Safavi, A.A.: A Novel IoT Architecture based on 5G-IoT and Next Generation Technologies. Tech. rep. (2018)
20. Rathore, M.M., Ahmad, A., Paul, A., Rho, S.: Urban planning and building smart cities based on the Internet of Things using Big Data analytics. Computer Networks 101, 63–80 (2016)
21. Ray, P.P.: A survey on Internet of Things architectures (2018)
22. Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., Harnisch, M.: Industry 4.0: The future of productivity and growth in manufacturing industries. Boston Consulting Group 9(1), 54–89 (2015)
23. Silva, E.M., Agostinho, C., Jardim-Goncalves, R.: A multi-criteria decision model for the selection of a more suitable Internet-of-Things device. In: 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC). pp. 1268–1276. IEEE (2017)
24. Silva, E.M., Jardim-Goncalves, R.: Multi-criteria analysis and decision methodology for the selection of internet-of-things hardware platforms. In: Doctoral Conference on Computing, Electrical and Industrial Systems. pp. 111–121. Springer (2017)
25. Singla, C., Mahajan, N., Kaushal, S., Verma, A., Sangaiah, A.K.: Modelling and Analysis of Multi-objective Service Selection Scheme in IoT-Cloud Environment. In: Cognitive Computing for Big Data Systems Over IoT, pp. 63–77. Springer (2018)
26. Soltani, S., Martin, P., Elgazzar, K.: A hybrid approach to automatic IaaS service selection. Journal of Cloud Computing 7(1), 12 (jul 2018)
27. Uslu, B., Eren, T., Gür, S., Özcan, E.: Evaluation of the Difficulties in the Internet of Things (IoT) with Multi-Criteria Decision-Making. Processes 7(3), 164 (2019)
28. Vasilomanolakis, E., Daubert, J., Luthra, M., Gazis, V., Wiesmaier, A., Kikiras, P.: On the Security and Privacy of Internet of Things Architectures and Systems. In: Proceedings - 2015 International Workshop on Secure Internet of Things, SIoT 2015 (2016)
29. Watrobski, J., Jankowski, J., Pawel, Z., Karczmarczyk, A., Ziolo, M.: Generalised framework for multi-criteria method selection. Omega (2018)
30. Weyrich, M., Ebert, C.: Reference architectures for the internet of things. IEEE Software (1), 112–116 (2016)
31. Whaiduzzaman, M., Gani, A., Anuar, N.B., Shiraz, M., Haque, M.N., Haque, I.T.: Cloud service selection using multicriteria decision analysis. The Scientific World Journal 2014 (2014)
32. Zanakis, S.H., Solomon, A., Wishart, N., Dublish, S.: Multi-attribute decision making: a simulation comparison of select methods. European journal of operational research 107(3), 507–529 (1998)

,,,,

ISSN 1870-4069

Molecular Marker for Protein Electrophoresis by Wavelet Transform

Jorge Arturo Flores-López, Leticia Flores-Pulido,
Lidia Patricia Jaramillo-Quintero

Universidad Autónoma de Tlaxcala,
Facultad de Ciencias Básicas, Ingeniería y Tecnología, Sistemas Inteligentes,
Apizaco, Tlaxcala, Mexico
arturh11flolo@gmail.com,
{leticia.flores.p,lidiapatricia.jaramillo}@uatx.mx

Abstract. Visual Information retrieval is an area where the analysis and recognition is helpful to generate and obtain data from image samples. In this research the image samples are obtained from analysis of electrophoresis that captures protein profiles in tissues. The electrophoresis provides a sample where the proteins are computed to know their molecular weight. The image digital processing takes electrophoresis images and a wavelet transform is applied to the sample that can be fractioned to emphasize the molecular weights that at first sight are not identified. Therefore, with the wavelet transform, it is possible to compute molecular weights of proteins and know their corresponding weight. Electrophoresis is a technique that is used in various analysis such as DNA, medicine, environment and food. All these profiles have different molecular weights, which are known by a marker that is placed in a lane. In this proposal, the wavelet transform is applied to the images of electrophoresis samples, creating the signal of the protein with the approximation coefficients, that achieves to measure a molecular weight. The approximation coefficients are computed at 3 levels of decomposition with the wavelet transform Daubechies. It is then possible to detect a molecular weight of nearly 300 samples of electrophoresis with an accuracy of 97 % in terms of numerical and visual similarity by means of visual information retrieval evaluated with recall and precision metrics.

Keywords: wavelet transform, electrophoresis, protein, molecular marker, visual information retrieval.

1 Introduction

Electrophoresis is the analytical technique used by the professionals of chemistry to achieve the separation of proteins through an electric field. It can be made in silver or coomassi blue substances, this last generates the sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). In this gel, several samples are placed, and once an electric field is induced, the proteins found in the samples placed in each lane, will go down depending on their molecular weight, as light as

they are, they will lower in the gel, when the run in the electric field is finished, the gel is placed in a relativising substance, which after a time allows to visualize the stains of proteins in the gel for molecular measurement. To the above, it is known as the electrophoresis process [1]. It is proposed, then, that the images of electrophoresis are analyzed so that the molecular weight is measured in an automated way.

In the area of digital image processing, the main objective is to emulate the capabilities of human vision, the images are defined as a two-dimensional function $f(x, y)$, where x and y are the intensity coordinates or gray levels. Because the images contain diverse characteristics that affect the optimal functioning of a system that works with them, it is necessary to perform a preprocessing that results in filtering or improvement of contrasts. Once this preprocessing is done, you can classify, segment, recognize objects, or even label them [2].

We have not found so far, research papers that measure the molecular weight of electrophoresis samples by means of wavelet transform in recent publications. However, the state of the art has been divided into three parts: In the first part, we will expose only a couple of works that talk about the recovery of images. In the second part we will talk about those articles that expose various methods for the improvement of electrophoresis images. It should be noted that some research focuses on 2D electrophoresis analysis, but even so, they do not perform molecular weight measurements, which is the main focus of this article. The third part of the state of the art will exhibit two works in which wavelet transform is used for electrophoresis images, however, a measurement of molecular weights is not performed either.

In this work [3] applies two-dimensional gel electrophoresis together with mass spectrography to Identify proteoforms. It is found that about 1000 protein spots are detected in each 2D gel. Because mass spectography does not identify complete protein sequences, when applying a separation in electrophoresis, mass spectography It can be more effective in identification.

The first article is that of [4] where it is mentioned that there are various types of software in which electrophoresis images can be analyzed. In this work it is mentioned that between different software, different images and data contained in them can emerge. In this study it was concluded that the electrophoresis data vary depending on the software, the software considered were TotalLab 120 and LabImage 1D.

In [5], wavelet transform and zero crossing point technique were used in various studies to obtain its signals. With this, they obtained 95 % confidence since. There were no significant differences between the wavelet and other applications. It was tested with daubachies level 8 with decomposition 15 (DB8-15), biorthogonal level 4.4 with decomposition 7 (BIOR4.4-7) and biorthogonal 3.3 with decomposition 4 (BIOR3.3-4) and good calibration results were obtained. Electrophoresis samples are used for protein detection in patients with fibromyalgia. The analysis shows the ability to accept a medication for treatment.

Another way to obtain similarity in the images is that applied in [6] where the proposed algorithm improves distinctive characters of the image using spa-

tial distribution, which is calculated by means of the location histogram. On space, scales and orientations of the local characteristics, in order to achieve a good recovery.

In the work of [7] molecular mass spectrometry is analyzed. Capillary electrophoresis is analyzed by means of a method that combines mass spectrometry with high resolution and rapid separation of the samples. This research offers selective detection with small volumes of examples. The above is done by cell analysis.

There are works with electrophoresis images in which they only focus on the detection or processing of protein stains, primarily, they perform a preprocessing and manage to restore the contrast to more easily dentify the electrophoresis spots [8]. This work focuses on obtaining those spots that do not stand out with the naked eye, followed by that process, use a maximum and minimum algorithm to be able to locate them all.

2 Molecular Marker for Protein Electrophoresis by Wavelet Transform System

The research proposal described in this article (see Figure 1) contains the following stages: initially, the compilation of images of protein electrophoresis in vegetables for human consumption is made. Electrophoresis samples are generated by means of an electrophoresis chamber, which are commonly used in the laboratories of industrial chemical engineers. The samples are adjusted in size to be processed independently by transform wavelet Daubechies Type 1, with decomposition at level 3 (DB1-3). This level is chosen due to the fact that greater relevant characteristics are obtained within the histogram of wavelet coefficients, as well as the possible noise generation of the signal is reduced at minimum. Subsequently, from the wavelet transform Daubechies and the approximation coefficients, an extraction of main characteristics is performed, that is, statistics of the maximum, the minimum, standard deviation, average and median. Each one of these features are obtained to allow the protein computation.

Thus, the analogy between wavelet approximation coefficients and a commercial marker scale of molecular weights is made. The process is the following: (a) the corpus of images are classified in 4 groups (see Figure 3). This profiles are based in molecular weight ranges. Every electrophoresis sample are images with a size of 400 X 700 pixels. Then, the wavelet transform is computed at decomposition 3. The matrix of approximation coefficients is taken for a parametric computation. The main features of parametric computation are: maximum, minimum, standard deviation, mean and median. This process can be observed in Figure 4.

Every feature of the previously mentioned parametric computation is analysed. the values and the ranges. Then, the analysis provides that the *minimum* feature, are the only parameter that discriminate between the 4 groups of molecular weights. So, the values of minimum for each group are:

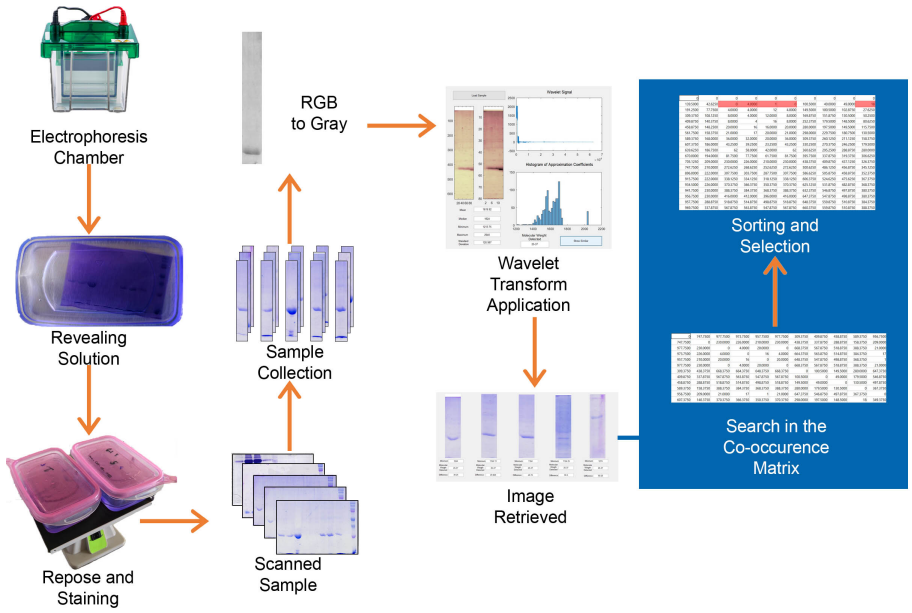


Fig. 1. Visual description of the proposal of molecular markers for protein electrophoresis by wavelet transform.

- First group of molecular weights: this group throws minimum values between 500 and 712. The molecular weight of this group of samples, can be of 10 kD to 15 kD.
- Second group of molecular weights: this group throws minimum values between 713 to 750. The molecular weight of this group of samples, can be of 15 kD to 20 kD.
- Third group of molecular weights: this group throws minimum values between 751 and 1700. The molecular weight of this group of samples, can be of 20 kD to 25 kD.
- Fourth group of molecular weights: this group throws minimum values between 1700 and 2000. The molecular weight of this group of samples, can be of 25 kD to 37 kD.

The molecular weight measurement correspondence and the detail range of values achieved by the minimum are shown in Table 1.

Below is a Equivalences Measurement for Molecular Weight in Table 1, which indicates the analogy on which scales are based for measurements of molecular weights. To create those ranges of measurements, we took into account the molecular weight endorsed by experts, the parametric characteristics indicated, mainly the minimum thrown by the matrix of approximation coefficients of protein profile samples. It should be noted, that this scale can be even more accurate, and throw more measurements. This process will achieve more future work.

Table 1. Minimum equivalence for the molecular weight - scale.

Minimum	Molecular Weight
500 -712	10 kD - 15 kD
713-750	15 kD - 20 kD
751-1700	20 kD - 25 kD
1700-2000	25 kD - 37 kD

Finally, a comparison is made between the elements of the imaging collection that provide molecular weights of proteins already endorsed by experts in the area of molecular metrics and by commercial molecular markers.

The novelty of the proposal is that currently there are not many works that apply wavelet transform to measure or detect the molecular weight of electrophoresis samples. On the other hand, the state of the art regarding the samples for improvement of the resolution of images obtained from samples of electrophoresis protein profiles if it exists, but its reading or detection by automatic means, is still an area of opportunity for the approaches that involve transformed.

3 Methods and Materials

In this section it is explained the foundations on which this research is constituted. The first basic concept is electrophoresis, since it is from this process of industrial chemistry that the samples arise and are created and that was also the source of inspiration for the measurement of molecular weights of proteins. The second basic concept is the wavelet transform. The wavelet transform is widely used in the area of visual information retrieval, since it is an effective method to extract characteristics of a two-dimensional signal. The characteristics in time and frequency are detectable from this type of transformation.

3.1 Electrophoresis

Electrophoresis is an analytical technique used by professionals in the area of industrial chemistry to obtain protein profiles and the analysis of proteins. Its objective is to separate the proteins by means of an electric field. This technique is carried out in polyacrylamide gel, in which the protein samples are placed and then the electric field is applied. Sodium dodecylsufate (SDS) is also used and the natural charge of the proteins can be calculated and separated based on their mass, therefore SDS-PAGE electrophoresis evaluates the purity and estimates the atomic weight of the proteins [1]. The electrophoresis process begins with the placement of the samples in each lane of the gel, then the gel is poured into the electrophoresis chamber which allows the migration of the proteins. The speed of migration of the proteins is proportional to the percentage of the pore of the material and its mass, those that are of greater weight, show resistance

and migrate slowly, those with less weight are those that end up until the end of the, meaning that they will migrate faster. The analysis of electrophoresis uses a standard of measurement of molecular weights, called a marker, its value is around $ 300 US, and its function is precisely to measure the molecular weight of the proteins in question, or the sample in turn. The weight is measured in kiloDaltons (kD); There are different types of markers, the standard is two colors, there is also the five colors that belong to the standards of the brand *kaleidoscope* where the ranges of the markers vary between 10-250 kD. In commercial markers, the colors indicate the range and type of protein contained in the electrophoresis analysis.

3.2 Wavelet Transform

The wavelet transform works with audio signals and images, that is, with one-dimensional or two-dimensional signals. Its function is to decompose the signal into various components, with this it is possible to locate the approximation coefficients and the detail coefficients of the image. The process starts with the original image and then decomposes the image into sub-images until it reaches a low-resolution image. The wavelet transform has several variants, the family of wavelet transform used in this work is the Daubechies transform that was proposed by Ingrid Daubechies as defined in [9]. This type of wavelet was chosen, due to its effectiveness to highlight the coefficients of detail of the type of electrophoresis samples.

Then, a Daubechies transform Type 1 at decomposition level 3 is mapped as $f \mapsto (a^3 \mid d^3)$ where a^3 draws the approximation coefficients and d_3 extracts the detail coefficients. Each value of the signals composed by the coefficients, is formed by a scalar product, being for the approximation matrix $a_m = f \cdot V_m^3$, and for the detail matrix $d_m = f \cdot W_m^3$. Where V_m^3 is the scaled entry signal at level 3 and W_m^3 is the wavelet transform at level 3. The input signal is taken as indicates in the Equation (1):

$$f = (f_1, f_2, ..., f_N). \tag{1}$$

If f_i is the input image, then, Equation (2) represents the calculation of the approximation coefficients:

$$a_m = f_1 \cdot \alpha_1, f_2 \cdot \alpha_2, f_3 \cdot \alpha_3, f_n \cdot \alpha_n, \tag{2}$$

where, a_m is redefined with the substitution of α_i as:

$$a_m = f_1 \frac{1+\sqrt{3}}{4\sqrt{2}}, f_2 \frac{3-\sqrt{3}}{4\sqrt{2}}, f_3 \frac{3+\sqrt{3}}{4\sqrt{2}}, f_4 \frac{1-\sqrt{3}}{4\sqrt{2}}. \tag{3}$$

For the decomposition matrix computation, the Equation (4) represents the computation of detail coefficients:

$$d_m = f_1 \cdot \beta_1, f_2 \cdot \beta_2, f_3 \cdot \beta_3, f_n \cdot \beta_n, \tag{4}$$

where β_i values, can be defined likewise α_i, and Equation 4 can be redefined as:

$$d_m = f_1 \frac{1+\sqrt{3}}{4\sqrt{2}}, f_2 \frac{3-\sqrt{3}}{4\sqrt{2}}, f_3 \frac{3+\sqrt{3}}{4\sqrt{2}}, f_4 \frac{1-\sqrt{3}}{4\sqrt{2}}. \quad (5)$$

The wavelet transforms, in any of its variants, can generate several levels of sub-signals, once the first level is obtained, the same calculations are made, sub signal by sub signal. More information regarding this type of transform can be found in [9].

It is important to mention that the use of Wavelet Daubechies Transformed Type 1 to Decomposition 1 (DB1-L1) was selected since it is known to be one of the simplest wavelet transforms. However, tests were carried out with the Haar Type 1 wavelet transfer to Decomposition 1 (HAAR1-L1) and the Biorthogonal Type 1 wavelet transform to Decomposition 1 (BIOR1-L1). These initial tests are not shown here for reasons of space, however, results almost identical to those obtained by Daubechies in terms of parametric values were obtained. Tests were also performed with other levels of decomposition, that is, apart from testing Daubechies to decomposition 1, it was also tested with Daubechies Type 1 to Decomposition 4 (DB1-L4), obtaining very large parametric variations, which no longer corresponded to the equivalences , nor to the molecular weights of the samples. On the other hand, by testing with Daubechies Level 4 to Decomposition 4 (DB4-L4), the difference between minimum values and molecular weights, was getting bigger and bigger. That is, the lower the decomposition level, the higher the accuracy of the molecular weight calculation.

4 Daubechies Wavelet Transform Implementation

The Toolbox of Matlab R2014a was used to choose the best wavelet transform. The computer equipment was a PC with Windows 10 based on x X64 with intel processor core i7 with 2.60 GHz, NVIDIA GeForce GTX 960M. The selection criterion was based on the clearer visualization of the wavelet coefficients of the electrophoresis samples. The visual selection of the Daubechies transform was visually determining, which is indicated in Figure 2 where the decomposition approximation coefficients level 3 are observed. The collection of electrophoresis images is composed by 64 samples of low molecular weight, 6 samples of the first group, 1 sample of the second group, 22 samples of the third group and 1 sample of the third group.

Tests were performed with the protein samples in Matlab's Toolbox 2-D Wavelet and the representative details of the images were observed. The wavelet transform provides a minimum values that matches with the molecular weights of a commercial marker, visually and numerically speaking. Daubechies Wvelet Transform at 3 decomposition level is chosen to do the work. Also, Haar, or Biorthogonal can be used to do the molecular weight detection as similar results than Daubechies. The approximation and detail matrix of each of the available samples was then carried out. It should be noted that the acquisition of each sample takes approximately 4 working days, so it is not easy or quick to obtain

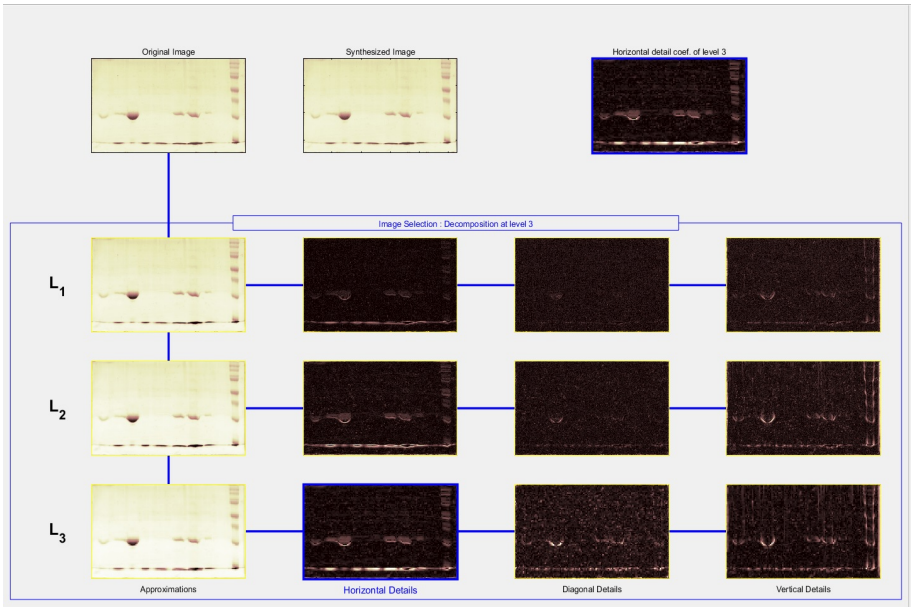


Fig. 2. Daubechies Wavelet Transform in tree structure for level 3 of decomposition for electrophoressi samples ([10]).

each of the samples of protein profiles. Once computed approximation and detail matrices, is possible to make statistics of each one for each sample. The parametric statistics used are the minimum, the maximum, the average and the standard deviation. The behaviour of the system and its operation can be seen in Figure 3.

The performance of the system begins with the loading of a sample of electrophoresis, after which the image synthesized with the approximation coefficients of Daubechies at level 3 is displayed. Each sample is calculated using the parametric statistics already mentioned in the previous section of this article. On the right side of the interface, the obtained wavelet coefficients, both approximation and detail, are displayed. Finally, in the lower graph, the histogram of the most outstanding coefficients of the protein profiles is shown.

The second part of the system, obtains the 5 samples that presented greater similarity with the sample of molecular weights that they want to be within the collection. Similarity was obtained by creating a difference matrix by means of a Euclidean metric [12]. The Euclidean metric is meaningful to establish the differences between sample desired and similar samples and is defined in Expression (6):

$$de_i = \sqrt{(min_a - min_b)^2}, \tag{6}$$

where de_i is the difference between the electrophoresis sample desired and the

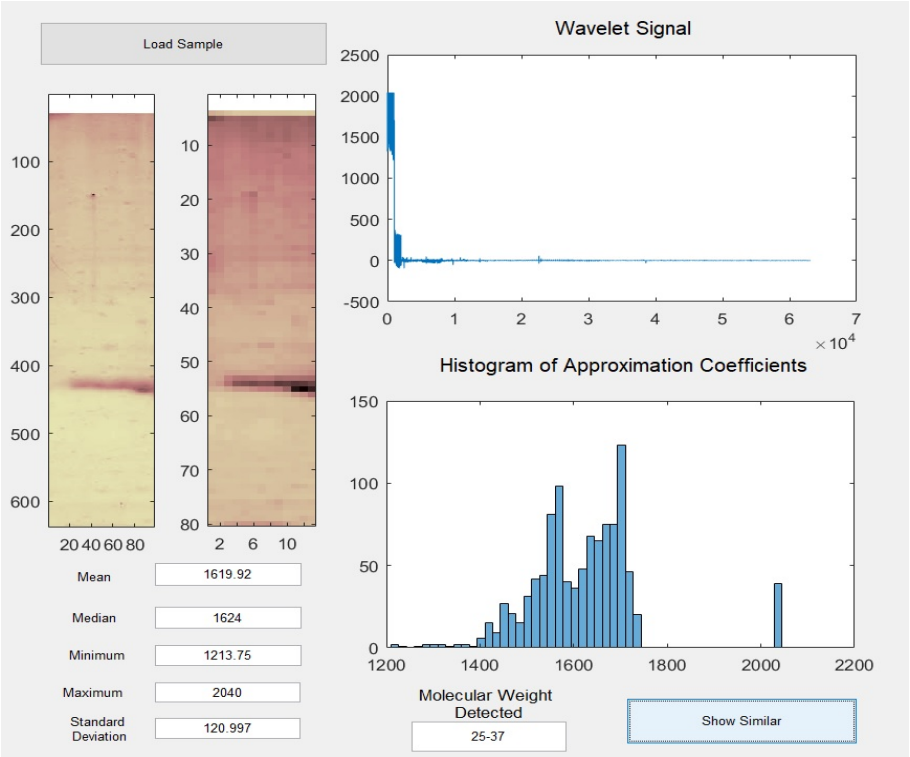


Fig. 3. Example of performance of the molecular weight detection system based on wavelet transform.

electrophoresis similar samples, min_a is the sample desired and min_b is other sample inside the corpus. The objective of euclidean metric is to compute a correlation matrix.

Within the proposed interface, not only the images of greater similarity are displayed, but also the characteristics thrown by the wavelet transform Daubechies. The characteristics are the parametric statistics obtained from the matrix of approximation coefficients, that is, average, median, minimum, maximum and standard deviation. Finally, the difference between the image of the desired molecular weight and the images most similar to that weight within the collection is calculated, in addition to the measurement of the molecular weight of the samples. An example of the performance of the system is shown in Figure 4.

5 Results

On the other hand, in Table 2, 30 random experiments of a corpus of 64 samples of protein profiles are shown, which indicate the desired molecular weight in the

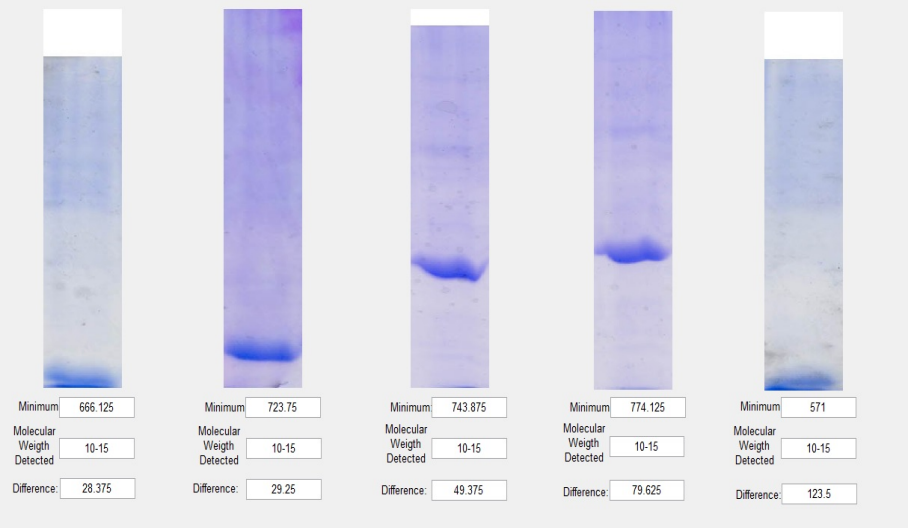


Fig. 4. Example of information retrieval for molecular weight measurement.

sample to be recovered, the minimum value of difference that coincides with the equivalent molecular weight, the number of similar samples recovered by the system, the number of non-similar samples returned by the system, the system relevance values (recall) and the values of precise samples recovered by the system (precision).

The recall and precision metrics aim to evaluate the performance of the visual retrieval system [11]. The recall average of 30 random samples is of 0.22 and precision average is of 0.70. As well as recall achieves 0.5 in each retrieval, the system can be considered as efficient. As well as precision achieves 1.0 in each retrieval, the system can be considered as efficient. The Figure 5 shows the behavior of recall and precision metrics for this visual retrieval system. Details of recall and precision metrics can be consulted in [11] and [12]. Recall is defined in expression (7) and precision is expressed in (8):

$$Recall = \frac{Number \quad of \quad Retrieved \quad Images}{Number \quad of \quad Images \quad in \quad Set}, \tag{7}$$

$$Precision = \frac{Number \quad of \quad Retrieved \quad Images}{Answer \quad Set}, \tag{8}$$

where the *Number of Retrieved Images* imply the similar electrophoresis images that matches with query images, visually spoken and in molecular weight. The value of *Number of Images in Set*, provides the number of images depending on set of query image. Thus, in query image belongs to:

- 1st group: the number of samples availables are 6.
- 2nd group: the number of samples availables are 1.

Table 2. Recall and precision statistics.

Sample	Molecular Weight	Minimum	Similar Samples	No Similar	Recall	Precision
1	10-15	666.13	3	2	0.50	0.60
2	20-25	863.50	1	4	0.17	0.20
3	20-25	967.50	1	4	0.17	0.20
4	10-15	694.50	3	2	0.50	0.60
5	10-15	501.25	4	1	0.67	0.80
6	10-15	571.00	4	1	0.67	0.80
7	20-25	1481.25	4	1	0.18	0.80
8	25-37	1796.00	5	0	0.23	1.00
9	20-25	1284.00	5	0	0.23	1.00
10	15-20	723.75	0	5	0.00	0.00
11	20-25	1244.00	4	1	0.18	0.80
12	20-25	1213.75	1	4	0.05	0.20
13	20-25	944.00	3	2	0.14	0.60
14	15-20	743.88	2	3	0.09	0.40
15	20-25	1164.00	5	0	0.23	1.00
16	20-25	913.50	3	2	0.14	0.60
17	20-25	774.13	2	3	0.09	0.40
18	20-25	1669.00	5	0	0.23	1.00
19	25-37	1713.88	5	0	0.23	1.00
20	25-37	1704.00	5	0	0.23	1.00
21	25-37	1714.00	0	5	0.00	0.00
22	20-25	1094.00	5	0	0.23	1.00
23	20-25	1164.13	5	0	0.23	1.00
24	20-25	1354.13	4	1	0.18	0.80
25	20-25	1051.63	4	1	0.18	0.80
26	20-25	1037.13	4	1	0.18	0.80
27	20-25	1279.25	4	1	0.18	0.80
28	20-25	1300.00	4	1	0.18	0.80
29	20-25	854.00	4	1	0.18	0.80
30	20-25	1165.13	5	0	0.23	1.00
					0.22	**0.70**

- 3rd group: the number of samples availables are 22.
- 4th group: the number of samples availables are 1.

The value of *Answer Set* always will be of 5, because the system and the interface was programmed for to browse 5 retrieved images.

From the technical point of view, it is possible to reproduce the previously detailed experiments. This is possible, through any other collection of scanned electrophoresis images. The images can be produced in an industrial chemistry laboratory. Protein samples can be extracted from food, plants, or tissue samples from various sources. Once obtained, they can be converted to gray levels and subsequently analysed by a transformed wavelet Daubechies. The molecular

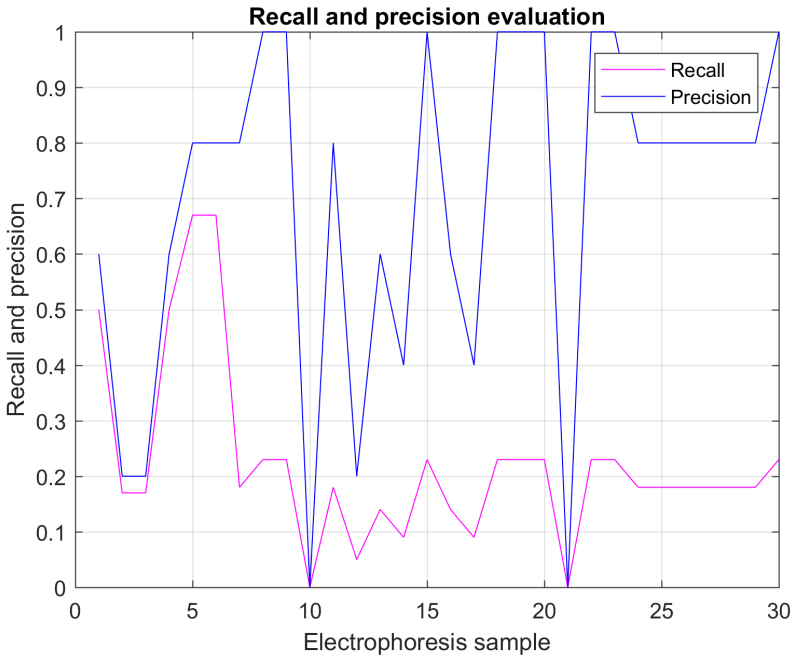


Fig. 5. Evaluation of recall and precision metrics for the weight molecular visual detection.

weight can be calculated using the equivalence values shown in the Table 1 proposed in this article. These equivalences cover the molecular weight readings or ranges comparable to the most commonly used commercial markers.

6 Conclusions

To obtain the electrophoresis samples, a long process is required, because this involves several steps for its realization, the process that requires more time is when the sample is placed in the solution. developer, which helps clean up the analysis and only leaves the areas that contain information on the proteins. This process involves about 96 hours (approximately 4 days), so the obtaining of each one of the 30 samples for the experiments in Table 2, is not a trivial task.

The experimentation stage faced a challenge at the moment of producing the samples, which consisted in that the electric field supply, in many occasions, caused the loss of the same due to the excess voltage. Because of this, the samples must be discarded. Subsequently, the processing part involved the cutting of each of the samples, and standardized each image for analysis, that is, each image had to be cut with a measurement in pixels of 100 pixels wide by 700 pixels long approximately.

To perform the calculation of the wavelet transform, we tried the Haar, the Biorthogonal and the Daubachies variant, achieving similar results, however, only Daubechies results are showed. Visual coefficients of greater importance in the protein signal content highlight the molecular weight score. Then the decomposition levels are analysed, where the wavelet transform Daubechies showed with higher illumination, said protein score. On the other hand, the parametric statistics and the histograms obtained from arrays of approximation coefficients, established an almost direct analogy between the detection of the desired molecular weights and the electrophoresis samples.

By applying the wavelet transform, in the variant Daubechies to a level of decomposition 3, the approximation coefficients of each sample were obtained, from which parametric statistics were obtained such as the mean, the median, the minimum value, maximum and standard deviation. Of these 5 statistics, which are considered as characteristics of the image, only in the case of the minimum value could a correspondence be established between the minimum ranges and the molecular weight marks. The other characteristics did not show much correspondence for the correct detection of molecular weights, because there were no predefined ranges, the readings were spliced or there was definitely no adequate correspondence.

Within the 30 experiments that are observed in the Table 2, it has that in 93 % of the cases, at least one image corresponds effectively to the molecular weight that you want to find, that is, the system recovers statistics and visually at least one molecular weight similar to the one to be found, and the recovery error is 0.06 % Recall and precision statistics can be viewed in Figure 5.

This indicates that it can also be observed that in each of the requests for molecular weights memory is measured (similar images), and precision (images that exactly correspond to the search). In each of the 30 cases taken at random, we can see that while the precision is closer to 1.0 indicates that the search was more accurate. If recall is closer to 0.5, implicates that the number of similar images was higher or more efficient visually and statistically talking.

References

1. Horton, R.: Principios de bioquímica. 4ta. edn. Prentice Hall Person, México (2008)
2. González, R.C., Woods, R.E.: Digital image processing, 4ta. edn. Person, México (2018)
3. Zhan, X., Zhou, T.: Application of Two-dimensional Gel Electrophoresis in Combination with Mass Spectrometry in the Study of Hormone Proteoforms, Chapter Mass Spectrometry in Future Perceptions and Applications. IntechOpen, pp. 1–17 (2018)
4. Kahlenberg, F., Sack, U., Bolt, A.: Impact of Image Analysis Software on Quantitations of 1D Gel Electrophoresis. J. Lab. Med. 36(3), 153–157. Berlin (2012)
5. Reza-Sohrabi, M., Muzabeygi, V., Davallo, M.: Use of Continuous Wavelet Transform Approach for Simultaneous Quantitative Determination of Multicomponent Mixture by UV-Vis Spectrophotometry. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy Vol. 201, pp. 306–314, Elsevier (2018)

6. Liu, P., Miau, Z., Guo, H., Wang, Y., Ai, N.: Adding Spatial Distribution Clue to Aggregated Vector in Image Retrieval. EURASIP Journal on Image and Video Processing, Springer Open, Vol. 9, pp. 1–14 (2018)
7. Zhou, W., Zheng, B., Liu, Y., Wang, C., Sun, W., Li, W., Chen, S.: Advances in Capilarlly Electrophoresis Spectrmetry for Cell Analysis. Trends in Analytical Chemistry, pp. 1–15, Elsevier (2019)
8. Salazar-Centeno, C.A., Niño-Niño, C.A. Diaz-Suarez, R.A.: Detección de bandas de color en una imagen de electroforesis en gel de una dimensión usando un algoritmo de localización basado en máximos y mínimos. ITECKNE 14(2), 12–30 (2017)
9. Walker, J.S.: A primer on wavelet and their scientific applications. 2da. edn. Chapman and Hall/Crc, EUA (2008)
10. Misti, M., Misti, Y., Oppenheim, G., Poggi, J.M.: Wavelet Toolbox for use with Matlab, 2nd edition, Ver. 2.0, Release 5.1, The Mathworks E.U.A. (2000)
11. Chen, Y., Li, J., Wang, J.Z.: Machine Learning and Statistical Modeling Approaches to Image Retrieval. Kluwer Academic Publishers (2004)
12. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press, Addison-Wesley, Edinburgh, England (1999)

ISSN 1870-4069

A Local Image Feature Approach as a Step of a Top-Down RGBD Semantic Segmentation Method

Gerardo Ibarra-Vázquez, Cesar A. Puente-Montejano, José I. Nuñez-Varela

Universidad Autónoma de San Luis Potosí, Facultad de Ingeniería,
San Luis Potosí, Mexico
gerardo.ibarra@alumnos.uaslp.edu.mx, {cesar.puente,jose.nunez}@uaslp.mx

Abstract. Semantic segmentation is a per-pixel class labeling problem, a method to assign a class label from a set of classes to each pixel on an image. Recent works have shown great progress using RGB images. However, indoor environments is still a challenging problem for the state-of-the-art algorithms due to the high variability of the scenarios. Additionally, semantic segmentation architectures based on Convolutional Neural Networks have reported to be vulnerable to adversarial attacks. However, elements of these architectures are similar to hand-crafted features and pipelines used in computer vision. In this paper, we explore the use of local image features by making an analysis and proposing a combination of feature detectors to make a robust classification of these features. This approach is a step of a top-down RGBD semantic segmentation method. Experiments on indoor environments show that the mean classification accuracy of feature descriptors can be improved by up to 3.3% with respect to the performance of a single feature detector. Also, using a balanced dataset an applying a cross-validation technique could improve up to 5.5% of the average of mean accuracy, obtaining better performance than just applying a single feature matching algorithm.

Keywords: local image features, feature detectors, feature descriptor, semantic segmentation.

1 Introduction

Semantic segmentation is a per-pixel class labeling problem, a method to assign a class label from a set of classes to each pixel on a RGB and RGBD (RGB+Depth) image [8]. Computer vision has tackled semantic segmentation on both RGB and RGBD perspectives. RGBD approaches introduced RGBD cameras to assist indoor scene segmentation by increasing the capabilities of getting the shape and spatial information from a depth image [10,11,26,27]. Deep learning has achieved very good results and RGB approaches for semantic segmentation have converged to encoder-decoder architectures based on Convolutional Neural Networks (CNN's) [3,7,17,18]. However, when they have been tested using indoor environments, they have experienced a lack of performance. For example, Badrinarayanan et al. [3] proposed the first encoder-decoder architecture for semantic

segmentation. They reported for an outdoor experiment using Camvid dataset 60.10% of mean Intersection-over-Union (mIoU), an evaluation metric that gives the similarity between the predicted region and the ground-truth. Meanwhile, for indoor environments using the SUN RGBD dataset, they reported 31.84% mIoU. This drop in performance could be explained by the high variability of indoor scenarios. Furthermore, adversarial attacks is a problem recently reported that affect CNN's and encoder-decoder architectures [2,20,31]. Adversarial attacks can be generated through a variety of forms, including making small modifications to the input pixels, using spatial transformations, or by a simple guess and check to find misclassified images.

Hand-crafted representations such as local image features have been widely used in a large variety of computer vision tasks. In specific, they were used before deep learning approaches emerged with excellent results on image classification [23,28]. However, CNN's have outperformed the results made by local image features. Notably local image features and pipelines used in computer vision can be seen as corresponding to layers of a standard CNN. Also, it has been reported in [26] the use of local image feature for an indoor scene segmentation approach using RGBD information. So, the question we address in this paper is whether it is possible to improve the performance of local image features by combining feature detectors and after answering this question build a top-down semantic segmentation method using RGBD images.

2 Local Image Feature Approach

Figure 1 presents an initial proposal of a top-down RGBD semantic segmentation method. Within this proposal, the focus of this work is on the local image feature approach (see Figure 1 yellow rounded box). Local image features are used due to its low computational cost, and robustness to changes on the scale, rotation, viewpoint change, blur and in some cases to lighting conditions. These features are used to influence the perception of the top-down process, recognizing and localizing points over the image. This approach is followed by a region growing segmentation process (currently under development) on the depth image, where the object shapes would be extracted to complete the semantic segmentation method.

Therefore, the contribution of this paper is the analysis of local image features and how a *combination* of feature detectors can improve the classification accuracy of features descriptors (see Figure 1 yellow rounded box). Local image features can be seen as a two-part process: *detection* and *description* (see Figure 1). Feature detection refers to the process of selecting regions or interest points in an image that have unique content, such as edges, corners, ridges or blobs [25]. These interest points can be used for further processing. Feature description involves computing a descriptor, which is typically done on regions centered around the feature detector. A descriptor is a compact vector representation of a local pixel neighborhood around an interest point. A histogram of the image gradients of a region centered on a point is an example of a descriptor. Thus, this

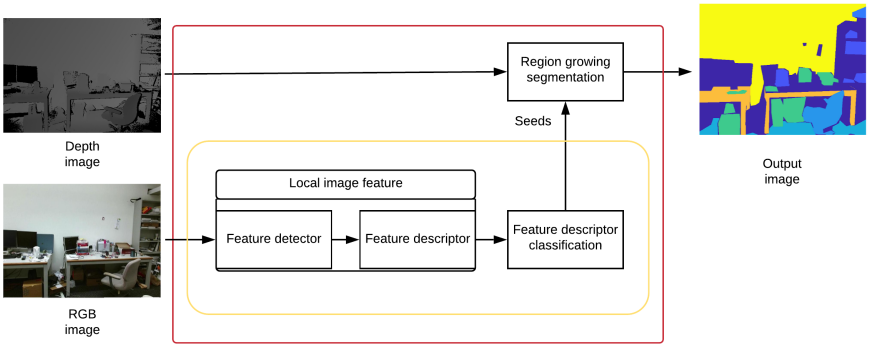


Fig. 1. The semantic segmentation method (red rectangle box), starts by using the local image feature approach by computing the feature detectors on the RGB image and constructing a feature descriptor on each one of the detectors. Afterwards, a classifier predicts a class of the local image features, which in turn becomes the input for the segmentation process (not reported in this paper).

work made an analysis focused on the *combinations* of feature detectors (e.g., regions, points, or corners) which are supposed to have a strong response to a series of filters in either spatial or frequency domains.

2.1 Feature Detectors

Two research works were followed to check their analysis of the current most common feature detectors [21,25]. The first work compares the invariance of feature detectors to the rotation, scale, and affine transformations, as well as some qualities such as repeatability, localization, robustness, and efficiency (see [25] to check such comparison). In terms of invariance, Maximally Stable Extremal Regions (MSER) [6], and Salient Regions (SR) [1], are invariant to all image transformations. The most common used Scale-Invariant Feature Transform (SIFT) [19], Speeded-Up Robust Features (SURF) [4], and Binary Robust Invariant Scalable Keypoints (BRISK) [16] detectors are invariant to rotation and scale. Corner detectors like Harris [13] and Features from Accelerated Segment Test (FAST) [24], are only invariant to rotation. In terms of qualities, MSER has a good performance on all the qualities mentioned above as well as the Harris corner detector. SIFT has a good performance on robustness and localization, while SURF has a good performance on efficiency and localization. Even though SR is invariant to all the transformations, it has poor performance on all the mentioned qualities.

Runtime performance is also taken into consideration. Mikolajczyk et al. [21] analyzed runtime performance and the number of regions for a test image of size 800x600 pixels. The shortest runtime was 0.66 seconds obtained by MSER. The longest was SR with 2013.89 seconds. Harris-Affine achieved the second

fastest time with 1.43 seconds. Hessian-Affine have the third fastest run time with 2.73 seconds. Hence, based on the results reviewed above, we decided to use the following four feature detectors:

- **Maximally Stable Extremal Region (MSER)** [6], extracts from an image a number of co-variant regions, called MSERs. An MSER is a stable connected component of some sets of gray-level pixels of the image.
- **Harris Corner** [13], finds corner points using the Harris-Stephens algorithm which considers the differential of the corner score with respect to direction directly.
- **Features from Accelerated Segment Test (FAST)** [24], is a corner detection method that uses a circle of 16 pixels (a Bresenham circle of radius 3), to classify whether a candidate point is actually a corner.
- **Binary Robust Invariant Scalable Keypoints (BRISK)** [16], is a novel scale-space FAST-based detector in combination with a bit-string descriptor obtained from intensity comparisons retrieved by dedicated sampling of each keypoint neighborhood.

2.2 Feature Descriptors

A descriptor is a compact vector representation of a local pixel neighborhood around an interest point or a region. This vector can be constructed using a feature detector as an input to build a representation of the surrounding pixels. Speeded Up Robust Features (SURF) [4] has a detector and a descriptor part. This paper uses the SURF descriptor in a combination of the feature detectors listed above[1]. SURF descriptor is extracted constructing a square region centered around the interest point (given by a feature detector), and oriented along an assigned orientation. The size of this region is $20s$, where s is the scale in which the interest point was found from a scale-space extrema detection. The orientation is computed with a sliding orientation window that detects the dominant orientation of a Gaussian weighted Haar wavelet.

The following procedure describes how the SURF descriptor is calculated from each one of the four detectors used in this work. A SURF descriptor is computed from MSER using a circle representing the feature with an area proportional to the MSER ellipse area. This area is needed to approximate a scale to construct the descriptor and is computed in terms of the ellipse's axes. The scale value must be greater or equal to 1.6 as is needed by the SURF descriptor [5]. Therefore, the MSER ellipse area is saturated to 1.6. The SURF descriptor orientation uses the MSER orientation directly. Since Harris, FAST and BRISK, generate interest points rather than regions, the minimum scale value of 1.6 was used to construct the square region on which the descriptor would be extracted. Since these interest points do not have an orientation assigned, thus the upright orientation was chosen.

[1] The SURF descriptor was compared against *Binary Robust Invariant Scalable Keypoint* (BRISK), *Fast Retina Keypoint* and *KAZE* descriptors in an experiment not reported in this paper. SURF descriptor showed the best results.

2.3 Classification Algorithms

Classification algorithms are needed to predict a class for each feature descriptor on the local image feature approach. In these experiments four different machine learning classifiers were used and are described next:

- **Probabilistic Neural Network (PNN)**: Is a two-layer network where the first layer computes the distance from the input vector to the training vectors of each class and the second layer produces an output vector of probabilities with the sums of the contributions of each class. The maximum value of these probabilities is chosen as the predicted class. The Euclidean distance computed from the center point of the training vectors is approximated applying a radial basis function using a sigma value.
- **Support Vector Machines (SVM)**: Builds a model that assigns new examples to one category or another. In this paper, a multi-class model for SVM is used that utilizes an Error-Correcting Output Codes (ECOC) model. ECOC reduces the problem of classification with three or more classes to a set of binary classifiers. Additionally, it uses $K(K–1)/2$ binary SVM models using one versus one coding design, where K is the number of classes.
- **Deep Neural Networks (DNN)**: Is an artificial neural network of multiple processing layers that learns representations of data with multiple levels of abstraction.
- **Feed-Forward Neural Network (FFNN)**: Is an artificial neural network composed typically of three layers: an input layer, a hidden layer and an output layer. A specific implementation of this algorithm is used for classification of local image features for semantic segmentation in [26]. It is used as the appearance model for the unary potential function of a Conditional Random Field (CRF) algorithm.

2.4 Dataset

A state-of-the-art dataset was chosen to test the combination of feature detectors and the classification algorithms. The SUN RGBD dataset [29] from Princeton University was selected. This dataset contains 10,335 RGBD images of indoor scenarios such as bedroom, furniture store, office, among others. Four different sensors (Intel RealSense, Microsoft Kinect v1 and v2, and Asus Xtion), were used to capture the images. The dataset contains annotations in 2D and 3D for both, objects and rooms. The dataset is composed of the NYU depth v2 [27], Berkeley B3DO [14], and SUN3D [30] datasets. Also, it provides benchmarks on six important tasks: Scene categorization, semantic segmentation, object detection, object orientation, and room layout estimation.

Furthermore, to complement the experiment of local image features and check the behaviour of the classification algorithms to balanced data. We constructed a balanced set of images from the RGBD Object Dataset [15]. This dataset contains 300 common everyday objects from multiple view angles, totaling 250,000 RGBD images organized into 51 categories. The objects are commonly found in indoor environments, such as homes and offices. As each object category has

different numbers of objects, we selected specific categories and build a balanced dataset.

3 Experiments and Results

The main idea presented in this paper is the combination of feature detectors to improve the classification accuracy of feature descriptors. An example is shown in Figure 2, where it is observed a combination of two feature detectors, MSER and FAST, on a test image. Different feature detectors are observed on the same object, e.g. the tripod which contains detectors of both types. This could allow having different starting points for a segmentation algorithm.

MSER FAST MSER+FAST

Fig. 2. Example showing the combination of two feature detectors (MSER and FAST) using the "cameraman" image.

In terms of the dataset, SUN RGBD provides two ground truth sets for all images, one using 37 classes and another one using 6,590 classes. Handa et al. [12] proposed the ground truth of semantic segmentation for a variety of datasets to standardize benchmarks between datasets. They proposed 13 classes for the SUN RGBD dataset (see Table 1). Rather than using the classes as defined by the SUN RGBD dataset, we followed the classes defined by Handa et al. For our experiments we manually selected a subset of 200 RGBD images for training and 93 images for testing from the SUN RGBD dataset (see Figure 1). All selected images belong to an *office* scenario.

Table 1. 13 classes defined by Handa et al. for the SUN RGBD dataset scenarios [12].

Number	1	2	3	4	5	6	7	8	9	10	11	12	13
Name	Bed	Books	Ceiling	Chair	Floor	Furniture	Objects	Picture	Sofa	Table	TV	Wall	Window

For our experiments, we used a computer with an Intel Xeon E5-1620 processor and 48GB of RAM. The hyperparameters for the classification algorithms were set as follows:

- **PNN**: $\sigma = 0.025$.
- **SVM**: One vs one coding, linear kernel function, kernel scale of 1, polynomial order of 3, and iterations limit of 1×10^6.
- **DNN**: An input layer, 3 fully connected layers of 1000, 500, 50 neurons and a softmax output layer. It was trained using stochastic gradient descent with momentum, an initial learning rate of 0.01, a maximum of 20 epochs and mini-batch size of 250.
- **FFNN**: An input layer, a fully connected layer of 1000 neurons and a softmax output layer. It was trained using stochastic gradient descent with momentum, initial learning rate of 0.01, maximum epochs of 20, and mini batch size of 250.

Table 2. List of feature detectors combinations and the mean accuracy over the 13 classes for each classification algorithm.

Experiment		Mean accuracy			
Feature detectors combinations	Descriptor	PNN	SVM	DNN	FFNN
MSER	SURF	21.5%	28.5%	27.8%	26.7%
MSER + Harris	SURF	21.5%	30.2%	29.4%	28.7%
MSER + FAST	SURF	21.3%	**30.6%**	29.4%	29.9%
MSER + BRISK	SURF	NA	30.4%	**30.2%**	28.6%
MSER + Harris + FAST + BRISK	SURF	NA	**30.6%**	30.1%	**30.0%**

Experiment results and the list of feature detector combinations are shown on Table 2. It is seen that there was no improvement using the PNN algorithm with any feature detector combinations. The mean accuracy obtained over the 13 classes is 21.5% for MSER and MSER + Harris combinations. Whereas for the MSER + FAST combination the mean accuracy is 21.3%. It should be mentioned that it was not possible to process MSER + BRISK and the combination of all detectors because the training data was too large and it was not possible to process using the PNN algorithm. On the other hand, SVM shows an improvement for all feature detectors combinations. MSER + Harris obtained 30.2% of mean accuracy, MSER + BRISK 30.4%, MSER + FAST and the combination of all feature detectors obtained 30.6%. This means an improvement of up to 2.1% from 28.5% of the MSER detector. DNN also obtained an improvement up to 2.4%, and its best result was 30.2% using MSER + BRISK. MSER + Harris + FAST + BRISK combination obtained 30.1%. MSER + Harris and MSER + FAST improved to 29.4% from the 27.8% of

the MSER detector. For the case of FFNN, the result for the combinations was the following: MSER + Harris 28.7%, MSER + BRISK 28.6%, MSER + FAST 29.9% and MSER + Harris + FAST + BRISK 30.0% of mean accuracy. The best result of FFNN showed an improvement of 3.3% from the obtained 26.7% of the MSER detector.

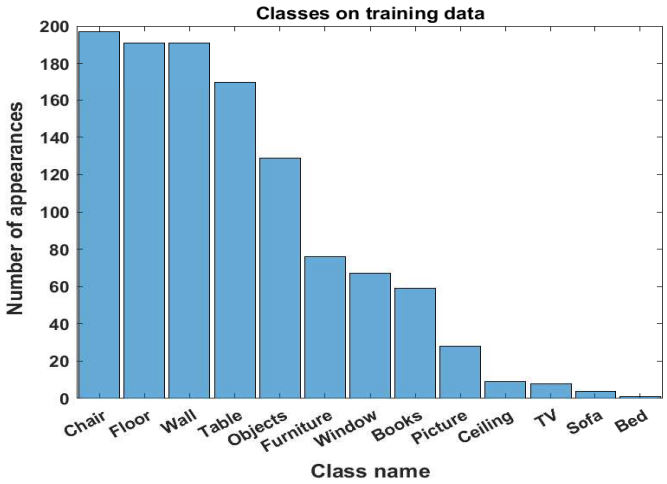


Fig. 3. Histogram of the number of appearances of classes in the training data.

In general, all algorithms presented a low performance. Thus, further analysis of the dataset showed that the training data was unbalanced which it causes the low accuracy of the four classification algorithms. Figure 3 presents the histogram of classes for the selected scenario (*office*), where the objects with most appearances in the training images are *chair*, *floor*, *wall* and *table*. However, other classes are not as recurrent, hence they are more difficult to classify.

Therefore, we defined a dataset for the second experiment choosing five object categories (food bag, food box, notebook, kleenex, and instant noodles) from the RGB-D Object Dataset. From these categories, five objects were selected and 40 images were taken from each object. A total of 1000 images were obtained which were split into five groups of 200 images each (40 images from each object category). We applied a cross-validation technique with four groups for training and one for testing. The same experiment settings listed above were used for this experiment. Additionally, we made a comparison with a feature matching algorithm[22] because we wanted to compare with a traditional local image features technique for object recognition. It consists of a priority search on hierarchical k-means trees for approximate nearest neighbor search in high-dimensional spaces. It is an efficient method for clustering and matching features in large datasets.

Table 3 shows the results for the experiment using a balanced dataset. It is shown that combining feature detectors can significantly increase the average

value of the mean classification accuracy of feature descriptors in some cases up to 5.5% (MSER+BRISK and FFNN case). Also, the best configuration for this experiment was MSER+BRISK detectors obtaining better performance with all the algorithms. It improved for PNN 3.1% of the average value of the mean acurracy, SVM increased 3.04% using the same combination, and DNN 2.03%. The feature matching was the algorithm with the lowest performance. The low performance can be explained by the inability of the algorithm to generalize a class from the feature descriptors. It only focuses on find similarities with the training descriptors. Just in one case the computer could not process the test dataset because it got out of memory, it was using PNN and the sum of all the detectors.

Table 3. List of feature detectors combinations, the average value and the standard deviation using the cross-validation technique of the mean accuracy over the five classes for each classification algorithm.

Experiment		Results									
Feature detectors combinations	Descriptor	PNN		SVM		DNN		FFNN		MATCH	
		avg	std	avg	avg	avg	std	avg	std	avg	std
MSER	SURF	41.80%	7.47%	40.94%	7.42%	38.74%	4.80%	38.96%	7.41%	7.40%	0.9%
MSER + Harris	SURF	34.19%	6.57%	33.61%	6.76%	32.61%	6.77%	33.23%	5.61%	**8.64%**	1.33%
MSER + FAST	SURF	38.40%	6.45%	39.74%	9.10%	37.56%	8.38%	39.27%	8.56%	7.44%	1.20%
MSER + BRISK	SURF	**44.94%**	5.40%	**43.98%**	7.77%	**40.77%**	5.86%	**44.46%**	7.93%	5.11%	0.63%
MSER + Harris + FAST + BRISK	SURF	NA	NA	37.68%	8.52%	37.44%	6.81%	38.96%	8.06%	6.99%	1.11%

Table 4. Time results for each algorithm on the training and testing stages. The average processing time is shown using the cross-correlation technique.

Experiment		Time results (seconds)									
Feature detectors combinations	Descriptor	PNN		SVM		DNN		FFNN		MATCH	
		train	test	train	test	train	test	train	test	train	test
MSER	SURF	0.214	172.3	102.5	**0.102**	157.7	0.343	**61.90**	0.141	0.022	1.61
MSER + Harris	SURF	0.105	747.9	498.2	**0.110**	306.5	0.644	**118.7**	0.258	0.028	3.87
MSER + FAST	SURF	0.107	433.0	379.5	**0.101**	272.1	0.567	**105.1**	0.229	0.029	3.27
MSER + BRISK	SURF	0.117	806.6	508.1	**0.120**	336.37	0.706	**134.7**	0.268	0.034	4.00
MSER + Harris + FAST + BRISK	SURF	0.194	NA	1726.4	**0.205**	600.9	1.249	**232.6**	0.506	0.056	15.42

Table 4 shows the average time performance for training and testing stages over the cross-validation technique using an Intel Xeon E5-1620 processor and 48GB of RAM. Dismissing PNN, which its architecture build the structure of the classifier by simply adjusting some of its parameters during the training process and the feature matching algorithm because we only measure the process of the storage of the training feature descriptors, the best training time performance is FFNN for all the experiments compared to SVM and DNN. It has 40% less time in the training process in the worst case (with MSER and SVM). Although SVM

does not have the best training time, it has the best time performance for the testing stage. It has an average of 0.1276 seconds classifying features descriptors of 200 images (an average of 0.638 milliseconds per image). Regardless of the increment of features, it only increases approximately 202% of its testing time on the worst case. DNN, FNN, and Feature matching increased their testing time approximately 381%, 358%, and 957% respectively when the number of feature detectors becomes larger.

4 Conclusions and Future Work

In this paper, we proposed a combination of feature detectors for feature descriptor classification as a step of a top-down RGBD semantic segmentation method. In the first experiment, local image features were used with four classification algorithms: a Probabilistic Neural Network (PNN), a Support Vector Machine (SVM), a Deep Neural Network (DNN), and a Feed Forward Neural Network (FFNN). This showed that the combination of feature detectors could improve performance on the classification of feature detectors on indoor environments. In particular, mean accuracy over 13 classes could be improved up to 3.3% in comparison to the FFNN implemented in [26]. This experiment showed that the combination of feature detectors of Maximally Stable Extremal Region (MSER) and Binary Robust Invariant Scalable Keypoints (BRISK) improved about 2.4% of mean accuracy using DNN. Support Vector Machine improved its performance by 2.1% using Maximally Stable Extremal Region and Features from Accelerated Segment Test (FAST) detectors, also using the combination of the four detectors (MSER + FAST + BRISK and Harris detector). Probabilistic Neural Network was the only one that could not improve its performance using the combination of feature detectors.

In the second experiment, the same four classification algorithms were compared with a feature matching algorithm. A balanced dataset was defined along with a cross-validation technique. It is showed that using MSER + BRISK combination could improve the average value of the mean classification accuracy by up to 5.5% using the FFNN implementation. The best performance was obtained by PNN with 44.94%, while the worst case was the feature matching algorithm with 5.11% using MSER + BRISK. However, taking into consideration the time performances, SVM has the best time performance for testing data. It takes 0.120 seconds for classifying features detectors of 200 images using the MSER + BRISK combination, approximately 0.6 milliseconds per image with the third best average value of mean classification accuracy of 43.98%. FFNN obtained the best training time with 134.7 seconds and the second testing time of 0.268 seconds with the second best result of 44.46% of the average value of the mean accuracy.

In conclusion, it was observed with these experiments that the best combination of feature detectors could not be defined for the first experiment. Only the sum of all detectors obtained two best performances. However, the second experiment showed that the MSER + BRISK was the best combination

for the classification algorithms, except for the feature matching algorithm. The explanation on the MSER + BRISK is that unlike the other detectors (Harris and FAST), BRISK is invariant to rotation and scaling, so it is more robust. SVM was the algorithm with the best average performance for all the combinations of detectors in both experiments followed by DNN, FFNN, and PNN. The structure of one versus other classes helped SVM to classify better the feature descriptors. We noted that, in general, classification of feature descriptors was not high. One reason could be the generalization problem of local image features. It should be mentioned that we have tackled a part of the problem of semantic segmentation from a specific environment (indoor scenes). Garcia et al. [9] reported that the best result found for SUN RGBD Dataset was 48.10% of mean Intersection-over-Union (mIoU) obtained by Z. Li et al.[17]. So, it is still a challenging problem. As future work, we will analyze the generalization problem in terms of the local image features by exploring several neural network approaches. Furthermore, new combinations of feature detectors and feature descriptors will be made to have a broader perspective on the improvement and robustness of the proposed approach. Finally, this classification will be used for the problem of the semantic segmentation process as it is shown in Figure 1.

References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1597–1604. IEEE (2009)
2. Arnab, A., Miksik, O., Torr, P.H.: On the robustness of semantic segmentation models to adversarial attacks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 888–897 (2018)
3. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39(12), 2481–2495 (2017)
4. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: European Conference on Computer Vision. pp. 404–417. Springer (2006)
5. Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV library. " O'Reilly Media, Inc." (2008)
6. Donoser, M., Bischof, H.: Efficient maximally stable extremal region (mser) tracking. In: null. pp. 553–560. IEEE (2006)
7. Eigen, D., Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2650–2658 (2015)
8. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J.: A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857 (2017)
9. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., Garcia-Rodriguez, J.: A survey on deep learning techniques for image and video semantic segmentation. Applied Soft Computing 70, 41–65 (2018)
10. Gupta, S., Arbeláez, P., Girshick, R., Malik, J.: Indoor scene understanding with rgb-d images: Bottom-up segmentation, object detection and semantic segmentation. International Journal of Computer Vision 112(2), 133–149 (2015)

11. Gupta, S., Girshick, R., Arbeláez, P., Malik, J.: Learning rich features from rgb-d images for object detection and segmentation. In: European Conference on Computer Vision. pp. 345–360. Springer (2014)
12. Handa, A., Pătrăucean, V., Stent, S., Cipolla, R.: Scenenet: an annotated model generator for indoor scene understanding. In: IEEE International Conference on Robotics and automation (ICRA) (2016)
13. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey vision conference. vol. 15, pp. 10–5244. Citeseer (1988)
14. Janoch, A., Karayev, S., Jia, Y., Barron, J.T., Fritz, M., Saenko, K., Darrell, T.: A category-level 3d object dataset: Putting the kinect to work. In: Consumer Depth Cameras for Computer Vision. pp. 141–165. Springer (2013)
15. Lai, K., Bo, L., Ren, X., Fox, D.: A large-scale hierarchical multi-view rgb-d object dataset. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 1817–1824. IEEE (2011)
16. Leutenegger, S., Chli, M., Siegwart, R.Y.: Brisk: Binary robust invariant scalable keypoints. In: IEEE International Conference on Computer Vision. pp. 2548–2555. IEEE (2011)
17. Li, Z., Gan, Y., Liang, X., Yu, Y., Cheng, H., Lin, L.: Lstm-cf: Unifying context modeling and fusion with lstms for rgb-d scene labeling. In: European Conference on Computer Vision. pp. 541–557. Springer (2016)
18. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3431–3440 (2015)
19. Lowe, D.G.: Object recognition from local scale-invariant features. In: The proceedings of the seventh IEEE international conference on Computer vision. vol. 2, pp. 1150–1157. IEEE (1999)
20. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A.: Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083 (2017)
21. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. International Journal of Computer Vision 65(1-2), 43–72 (2005)
22. Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. VISAPP (1) 2(331-340), 2 (2009)
23. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: European conference on computer vision. pp. 143–156. Springer (2010)
24. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: European Conference on Computer Vision. pp. 430–443. Springer (2006)
25. Salahat, E., Qasaimeh, M.: Recent advances in features extraction and description algorithms: A comprehensive survey. In: IEEE International Conference on Industrial Technology. pp. 1059–1063. IEEE (2017)
26. Silberman, N., Fergus, R.: Indoor scene segmentation using a structured light sensor. In: IEEE International Conference on Computer Vision Workshops. pp. 601–608. IEEE (2011)
27. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. Computer Vision–ECCV 2012 pp. 746–760 (2012)
28. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep fisher networks for large-scale image classification. In: Advances in neural information processing systems. pp. 163–171 (2013)

29. Song, S., Lichtenberg, S.P., Xiao, J.: Sun rgb-d: A rgb-d scene understanding benchmark suite. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 567–576 (2015)
30. Xiao, J., Owens, A., Torralba, A.: Sun3d: A database of big spaces reconstructed using sfm and object labels. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1625–1632 (2013)
31. Xie, C., Wang, J., Zhang, Z., Zhou, Y., Xie, L., Yuille, A.: Adversarial examples for semantic segmentation and object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1369–1378 (2017)

ISSN 1870-4069

Depth Estimation Using Optical Flow and CNN for the NAO Robot

Oswualdo Alquisiris-Quecha[1], Jose Martinez-Carranza[1,2]

[1]Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Computer Science Department, Mexico
[2]University of Bristol, Computer Science Department, UK
oswualdoaq@inaoep.mx, carranza@inaoep.mx

Abstract. In the robotics field, one of the main challenges is autonomous navigation using a single camera where camera images are processed in a frame to frame basis. However, getting clear and noise-free images is still a challenge under erratic motion, typical of moving robots. To solve this problem, several works use optical flow techniques to eliminate blurred reference points in RGB images when the robot moves. The NAO robot is an example of a robotic platform that generates an oscillatory movement when walking, thus producing blurred images that may compromise the image processing task. In this work, we focus on the problem of depth estimation in a single image for the NAO robot, which proves useful for autonomous navigation. For the depth estimation, we argue that the erratic movement exhibited by the walking motion of the robot could be exploited to obtain optical flow vectors, which are strongly related to depth observed by the NAO's camera. Thus, we present a real-time system based on a Convolutional Neural Network (CNN) architecture that uses optical flow as input channels in order to estimate depth. To this aim, we present a new dataset that includes optical flow images associated to depth images for training. Our results indicate that optical flow can be exploited in humanoid robots such as NAO, but we are confident that it could be used in other platforms with erratic motion.

Keywords: depth estimation, deep learning, CNN, optical flow, NAO robot.

1 Introduction

One of the main challenges in the field of robotics is the autonomous navigation with a single camera, in which the images are processed frame by frame. However, due to the movement itself that produces a robotic system, such as the case of the NAO robot which generates an oscillatory movement when walking through the environment, it is not possible to obtain clear and noise-free images to be analyzed correctly in subsequent processes, therefore this remains a challenge under the erratic movements of moving robots.

To try to solve this problem, several techniques have been proposed, always trying to compensate the movement of the robot by optical and/or digital

stabilization on RGB images. However, when using this type of images it is possible to obtain blurred reference points when the robot moves, which is why optical flow techniques are used to solve this typical problem in RGB images

In this paper, we focus on the problem of depth estimation in a single image for the NAO robot, which is useful for the task of autonomous navigation. We argue that the erratic movement of a robotic system could be exploited to obtain optical flow vectors, which are strongly related to the depth observed by the camera. With this, it would no longer be necessary to use stabilization systems for the input images, freeing the computational resource that this implies for the robot, which is very useful for computational systems with little computing power like the case of the NAO robot.

Therefore, we present a system based on a CNN architecture that uses optical flow as input channels to estimate depth. For this purpose, we present a new dataset that includes optical flow images associated with depth images for training. Our results indicate that the optical flow can be exploited in humanoid robots like NAO, but we trust that it could be used in other platforms with erratic movement.

2 Related Work

Under the idea of autonomous navigation with the NAO robot, several investigations have been carried out trying to solve the problem of locating the robot within its environment to thereby achieve autonomous navigation. Initially, solutions were proposed by means of navigation based on estimates using environmental marks such as bar codes or landmarks [6,3,13] or using visual memories [1,4]. other works add additional sensors to the robot [15]. However, this considerably reduces the autonomy time of the robot and increases the instability of the system due to the added weight of the sensor—, which is not considered within the robot's kinematics. On the other hand, MonoSLAM based systems have been used for navigation where the robot incrementally constructs a map of an unknown environment in which it is located and in a parallel way estimates your trajectory of displacement in the environment through the use of a single camera as in [14,12].

However, a typical problem of these systems that work with RGB images is the obtaining of images with distortion and defocus due to the erratic movement of the robot when moving through the environment, for this reason, optical flow techniques have been used, which by its principle of operation works by having sequences of moving images.

This approach to employ optical flow techniques in conjunction with deep learning techniques, such as the case of Deep Learning, for the estimation of depth in monocular cameras, increasingly it is becoming a growing area to be used in mobile robot and humanoid as in [7] where they employ techniques that involve the analysis of the optical flow of a monocular image for the estimation of the depth and speed of displacement in an environment, similar to [11] where an architecture based on CNN is proposed for the estimation of distance of an

object in a 3D scene by using visual characteristics of optical flow,the network was trained with optical flow components, which has control signals for the evasion of obstacles in a mobile robot as system output.

3 Methodology

The estimation of depth in an environment using a single monocular image is an important task in autonomous navigation. Therefore, in this section, a general description of the proposed system for depth estimation is made using optical flow images for autonomous navigation tasks in humanoid robots such as the NAO robot. The general architecture is shown in Figure 1, in which the process of training and estimating the depth of the system is described in a general way.

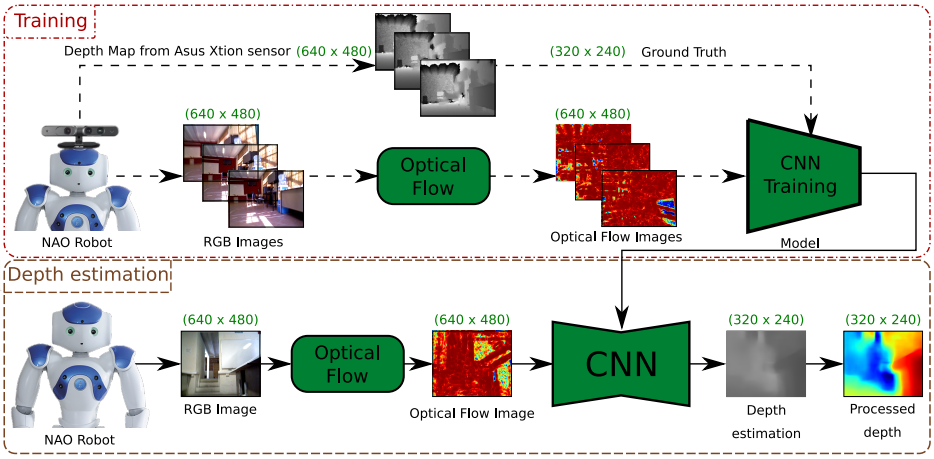


Fig. 1. General methodology.

The idea of using optical flow in depth estimation tasks is due to the fact that the flow vector of an image has a direct relationship with the relative depth of the objects in the image, besides that to obtain these values it is necessary to have sequences of images with movement. Movement that exists in every humanoid robot when moving around its environment, so it is proposed to use it in conjunction with deep learning techniques to generate depth maps of the environment.

The CNN network used is the DenseDepth proposed by [2] which is an encoder-decoder type network for depth estimation in RGB images, for the encoder part the RGB image is encoded in a feature vector using the DenseNet-169 and the decoder is composed of a successive series of ascending sampling layers with connections associated with the encoder without requiring any batch normalization, where the resolution of the input images of the network is 640 x 480 for the RGB and 320 x 240 for the depth maps estimated by it.

4 Description of the Dataset

In order to obtain images that correspond to those that the NAO robot perceives in its environment for the training of the neural network, it was necessary to search for the best alternative to gather the necessary data for the training phase. This is because the NAO robot only has one RGB camera and for the purposes of depth estimation it is necessary to generate a dataset where each RGB image has its corresponding depth map. Therefore, it was decided to use the Asus Xtion depth sensor, which was placed on the head of the NAO robot (See Figure 2) in order to obtain the images that the robot would see, also capturing the erratic movement of the device when moving in the environment.

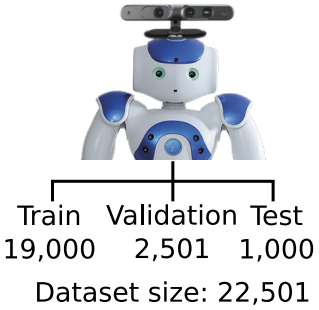


Fig. 2. Dataset from NAO image.

For the generation of the new dataset an implementation was made in C++ language which operates on the ROS system, where to access the sensor data the OpenNI software is used, considering a data synchronization policy of both RGB images and their corresponding depth images.

For RGB images, the developed system takes the I_{i-1} and I_i images from a sequence of images and the Farneback method is applied to calculate the optical flow for each pixel of the image, that is, a flow calculation is performed dense optical on the pair of input images obtaining an output image with the apparent movements of the objects that are inside the scene.

On the other hand, for the depth images corresponding to each calculated optical flow image, a process of removing outliers (negative values or NaN) is performed by assigning the value 0 in each pixel where there is an atypical value. The missing depth values are filled using the painting method proposed by [10], then a process of normalization of the data is carried out at a range of $[0 - 255]$ where the maximum distance corresponding to the depth is set to the 5.5 meters, the above through the equation (3):

$$I_i = (255 * D_i)/5.5, \tag{1}$$

where: I_i is the resulting depth image when applying normalization and D_i is the depth image without outliers.

The dataset is made up of 22,501 elements, with images obtained in interior scenes, of which they are divided into three groups: Training, Validation and Test, with 19,000, 2,501 and 1,000 data respectively (See Figure 2). Each group consists of RGB images, optical flow and depth map, each with a resolution of 640 x 480, which are each acquired at the same time. An example can be seen in Figure 3.

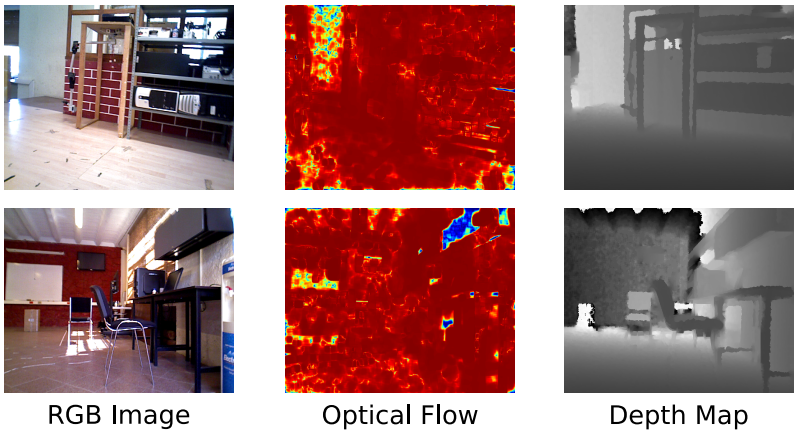


Fig. 3. Structure of the dataset.

5 Training

For the training of the CNN network, the dataset proposed in Section 4 is used, which contains images with interior scenes captured by the Asus Xtion sensor, which was placed on the head of the NAO robot to capture the images that the robot would see , also containing the erratic movement of the robot when navigating its environment. This dataset is composed of pairs of optical flow images and their corresponding depth image, in addition to the technique of data augmentation through transformations of training data whose technique has been shown to generate higher performance and obtain better accuracy. To do this, transformations of horizontal rotations are made to the images with a probability of 0.5, this is because vertical rotations in an image with an interior scene may not provide much information to the learning process and in some cases it could confuse the system because of the similarities of the geometries of floors and ceilings. In the same way different permutations are applied in the color channels with a probability of 0.25, this is considered according to the results obtained by [2] using this parameter.

For the training process, the CNN network is configured using the ADAM optimizer [8], which combines the methodology of Momentum and Root Mean Square Propagation (RMSProp), calculating a linear combination between the gradient and the previous increase, and considers the gradients recently appeared in the updates to maintain different learning rates per variable, with a learning rate value of 0.0001 and values $\beta_1 = 0.9$, $\beta_2 = 0.999$ (default values, optimizer's own), with a batch size of 1 for 50 epochs.

In addition, the loss function proposed by [2] is used, where this function seeks to balance the reconstruction of depth images by minimizing the difference of the ground truth values and at the same time penalizes the high frequency distortions in the image domain of depth. The loss function consists of three terms and is defined as:

$$L(y, \hat{y}) = \lambda L_{depth}(y, \hat{y}) + L_{grad}(y, \hat{y}) + L_{SSIM}(y, \hat{y}), \tag{2}$$

where:

$$\lambda = 0.1, \tag{3}$$

$$L_{depth}(y, \hat{y}) = \frac{1}{n} \sum_{p}^{n} |y_p, \hat{y_p}|, \tag{4}$$

$$L_{grad}(y, \hat{y}) = \frac{1}{n} \sum_{p}^{n} |g_x(y_p, \hat{y_p})| + |g_y(y_p, \hat{y_p})|, \tag{5}$$

$$L_{SSIM}(y_p, \hat{y_p}) = \frac{1 - SSIM(y, \hat{y})}{2}. \tag{6}$$

In the equation (5), g_x y g_y represent the differences in the x and y components for the gradients of the depth image of y and $\hat{y}$

6 Experiments and Results

The network implemented in TensorFlow for depth estimation using optical flow images of the scene was trained using a GeForce GTX 1050 GPU with 640 Cuda Cores, with the pre-trained weights of DenseDepth [2], using the learning transfer technique using the new dataset proposed in this paper.

6.1 Evaluation

For the quantitative evaluation the method is compared with other works of depth estimation that use RGB images as input, the evaluation is done using six evaluation metrics proposed by the state of the art. Where the error functions are defined as:

– Average Relative Error (REL): $\frac{1}{n} \sum_{p}^{n} \frac{|y_p - \hat{y_p}|}{y}$

- Root Mean Squared Error (RMS): $\sqrt{\frac{1}{n}\sum_p^n (y_p - \hat{y_p})^2}$
- Average (Log_{10}) error: $\frac{1}{n}\sum_p^n |log_{10}(y_p) - log_{10}(\hat{y_p})|$
- Threshold accuracy (δ_i): $\delta < thrforthr = 1.25, 1.25^2, 1.25^3$

Where y_p is the value of a pixel of the depth image y, $\hat{y_p}$ is the value of a pixel of the predicted depth image $\hat{y}$ by the trained model, and n is the total number of pixels for each depth image.

Table 1 shows this comparison, where it is necessary to denote that these models with which our proposal is compared are models obtained through various training processes using large RGB image data sets, in the order of millions of data. Our method and model obtained was trained only using 22.5k optical flow images using 50 training times through a training configuration from scratch, the values obtained can be low compared to the other methods with which it is compared, however it is necessary to consider mentioned above.

Table 1. Comparison with other methods of depth estimation.

Method	$\delta_1 \uparrow$	$\delta_2 \uparrow$	$\delta_3 \uparrow$	REL $\downarrow$	RMS $\downarrow$	$Log_{10} \downarrow$
Eigen et al. [5]	0.769	0.950	0.988	0.158	0.641	–
Laina et al. [9]	0.811	0.953	0.988	0.127	0.573	0.055
Alhashim et al. [2]	0.846	0.974	0.994	0.123	0.465	0.053
Ours	0.401	0.600	0.760	0.490	1.080	0.170

In Figure 4, the qualitative results of the system are shown. Where, the first column represents the RGB image only for questions of visual comprehension of the reader, the second column corresponds to the optical flow calculated for the RGB image which is the input to the CNN network, in the third column the Ground Truth is shown, the fourth column corresponds to the output of the CNN network and finally the last column represents the output image of the network by applying a representation in 3D color space for the image for better visual understanding.

Figure 5 shows a fragment of the navigation process of the NAO robot using the proposed system. In it, it is possible to observe the depth estimation obtained by means of optical flow images as input to the CNN network at moments of time during the navigation of the robot.

7 Conclusions and Future Work

In this work we propose the use of optical flow images as input to a CNN network for depth estimation, the idea of using optical flow is due to the fact that by its nature it can represent relative depth according to the values of the vector of optical flow and, in conjunction with a CNN network, we show that it is possible to estimate depth with metric of an arbitrary scene.

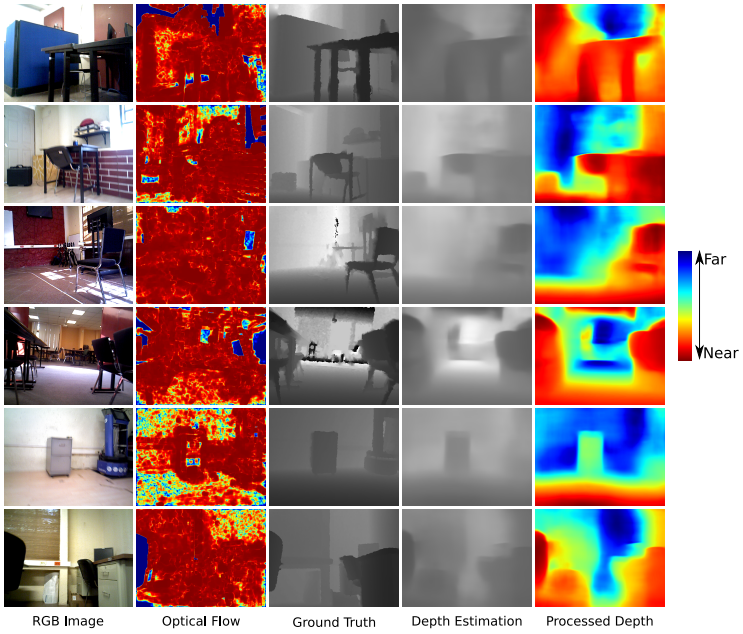


Fig. 4. Qualitative results from proposed method.

According to the experiments carried out, it is concluded that it is possible to use optical flow techniques in conjunction with deep learning techniques, such as the case of Deep Learning, for the estimation of depth maps in autonomous navigation tasks, where it is shown that it is possible to exploit the erratic movement in a humanoid robot to generate depth maps of its environment through the optical flow vectors generated by this movement.

In addition, a new dataset is proposed with images obtained from one of the NAO's frontal camera. In this dataset, we have collected color images obtained from mapping optical flow to the RGB space. The optical flow is generated during the walking motion of the NAO robot. These images are associated to corresponding depth images recorded with a Kinetc sensor. Thus, our main argument is that the erratic motion induced on the NAO's head while walking can be exploited by means of observing the optical flow, and such instantaneous flow, obtained in a frame-to-frame basis, can be exploited to learn depth relative to the robot. For the learning, we used a state of the art CNN architecture used for the problem of depth estimation in a single image, except that instead of using conventional RGB images, we propose to use our coded RGB images mapped from the optical flow. Our results indicate that our approach is feasible and it compares to state of the art methods on depth estimation in a single image.

As future work, we will explore new architectures of CNN networks in order to reduce processing time and to obtain better results in the estimated depth maps. Likewise, we will explore new data augmentation policies and probability

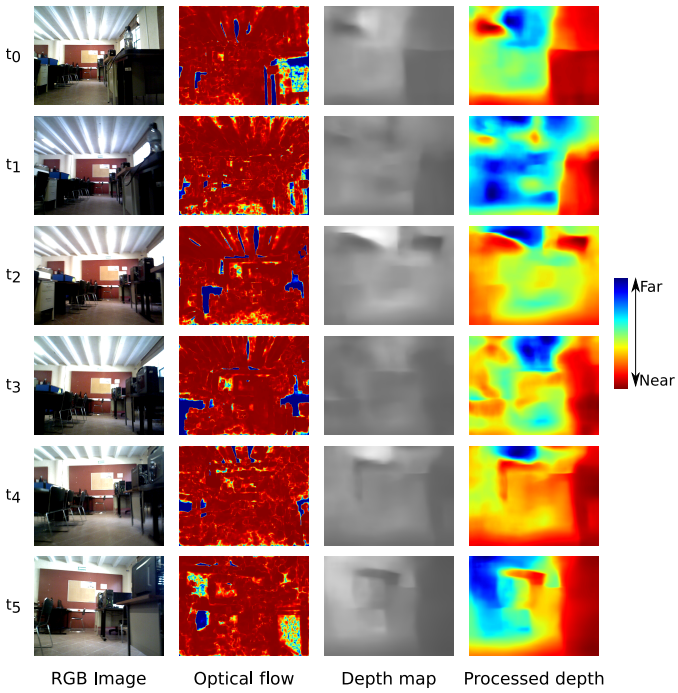


Fig. 5. Results of our proposed method obtained for a sequence of images during navigation of the NAO robot.

values that can help in the training process, generating a faster convergence and requiring a smaller number of iterations in the training process.

References

1. Aldana Murillo, N.G.: Localización de robots humanoides basada en apariencia a partir de una memoria visual. Tesis de Maestría en Optomecatrónica. Centro de Investigaciones en Óptica, A.C. León, Guanajuato p. 75 (2014)
2. Alhashim, I., Wonka, P.: High quality monocular depth estimation via transfer learning. arXiv preprint arXiv:1812.11941 (2018)
3. Changyun, W., Junchao, X., Chang, W., Wiggers, P., Hindriks, K.: An approach to navigation for the humanoid robot nao in domestic environments. In: Towards Autonomous Robotic Systems: 14th Annual Conference, TAROS 2013, Oxford, UK, August 28–30, 2013, Revised Selected Papers. vol. 8069, p. 298. Springer (2014)
4. Delfin, J., Becerra, H.M., Arechavaleta, G.: Humanoid navigation using a visual memory with obstacle avoidance. Robotics and Autonomous Systems (2018)
5. Eigen, D., Puhrsch, C., Fergus, R.: Depth map prediction from a single image using a multi-scale deep network. In: Advances in neural information processing systems. pp. 2366–2374 (2014)

6. George, L., Mazel, A.: Humanoid robot indoor navigation based on 2d bar codes: Application to the nao robot. In: Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on. pp. 329–335. IEEE (2013)
7. Ho, H.W., de Croon, G.C., Chu, Q.: Distance and velocity estimation using optical flow from a monocular camera. International Journal of Micro Air Vehicles 9(3), 198–208 (2017)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
9. Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N.: Deeper depth prediction with fully convolutional residual networks. In: 2016 Fourth international conference on 3D vision (3DV). pp. 239–248. IEEE (2016)
10. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: ACM transactions on graphics (tog). vol. 23, pp. 689–694. ACM (2004)
11. Ponce, H., Brieva, J., Moya-Albor, E.: Distance estimation using a bio-inspired optical flow strategy applied to neuro-robotics. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–7. IEEE (2018)
12. Rioux, A., Suleiman, W.: Autonomous slam based humanoid navigation in a cluttered environment while transporting a heavy load. Robotics and Autonomous Systems 99, 50–62 (2018)
13. Wen, S., Zhang, Z., Ma, C., Wang, Y., Wang, H.: An extended kalman filter-simultaneous localization and mapping method with harris-scale-invariant feature transform feature recognition and laser mapping for humanoid robot navigation in unknown environment. International Journal of Advanced Robotic Systems 14(6), 1729881417744747 (2017)
14. Wirbel, E., Bonnabel, S., de La Fortelle, A., Moutarde, F.: Humanoid robot navigation: getting localization information from vision. Journal of Intelligent Systems 23(2), 113–132 (2014)
15. Xu, X., Hong, B., Guan, Y.: Humanoid robot localization based on hybrid map. In: Security, Pattern Analysis, and Cybernetics (SPAC), 2017 International Conference on. pp. 509–514. IEEE (2017)

ISSN 1870-4069

Semantic Annotation Approach for Information Search

Fernando Pech-May[1], Alicia Martinez-Rebollar[2], Jorge Magaña-Govea[1], Luis A. Lopez-Gomez[1], Edna M. Mil-Chontal[1]

[1] Instituto Tecnológico Superior de los Ríos, Tabasco, Mexico
fpech@tamps.cinvestav.mx, {jgoveaitsr,llopezitsr}@gmail.com, mariled7@hotmail.com

[2] Centro Nacional de Investigación y Desarrollo Tecnológico, CENIDET, Morelos, Mexico
amartinez@cenidet.edu.mx

Abstract. Due to the needs to improve the information search process, new strategies have been created to enhance searches. The semantic search performs the search by means of meaning instead of literals. The semantic search in unstructured documents requires to formalize knowledge through an annotation semantic process. Some annotation proposals use natural language processing tools, ontologies to link document terms; others use the similarity of entities through the weight of the edges, association between pair of concepts or the ontology structure. In this paper we present an alternative for semantic annotation in unstructured documents by semantic context extraction of entities. In the approach we detect the named entities through a data dictionary created from Wikipedia and link the instances in the ontology. The context extraction strategy is based on the concepts similarity; each term is associated with an instance of the ontology and the similarity between relationships explicit is measured by the combination of two types of measures: the association between each pair of concepts and the weight of the relationships. The approach was tested with two ontologies and two datasets in news and business, respectively.

Keywords: semantic annotation, semantic similarity, concept similarity.

1 Introduction

The large amount information stored and shared on the Web in form of unstructured documents [16] has caused difficulties for its search and retrieval. Traditionally two factors are used to classify the results of a search: 1) the relevance that measures the coincidence of the terms, and 2) the documents popularity, which is a complementary factor to documents ranking. Despite this, there are still challenges for searching and information management to reduce effort and search time.

On the other hand, there has been a constant growth in the semantic Web and has opened new opportunities for access and information retrieval and has motivated the development of linked data and knowledge bases for different domains and applications such as DBPedia [2], FreeBase [3], YAGO [32], etc. Also knowledge bases have been developed in specific areas such as Snomed CT [8] and UMLS [27] for medical areas and AGROVOC [4] for the agricultural area. These knowledge bases have become valuable resources for the knowledge extraction. A fundamental component to take advantage of such resources is to formalize knowledge by linking the unstructured text with elements of the knowledge base, called semantic annotation.

Some systems of semantic annotation have been developed in the medical area [20] for the identification of biomedical entities such as proteins, genes, diseases and their relationships. Other approaches have focused on named entities such as people, organizations and places. The first annotation proposals used natural language processing tools for documents analysis; these approaches present problems of: i) ambiguous annotations, when entities have been assigned to more than one concept in the ontology, ii) erroneous annotations, when the meaning of a text is not found in the ontology, and, iii) false annotations, when the annotation does not provide any value for the realization of a semantic search (see Figure 1).

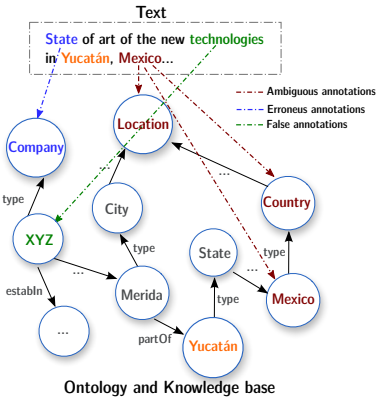


Fig. 1. Problems in semantic annotations: ambiguous, erroneous and false annotations.

This paper presents a semantic annotation approach in unstructured Web documents through its contextual semantic information through the use of ontologies in a limited domain. The proposal of semantic annotation is an improvement presented in [23] to represent unstructured documents by using an ontology, linking the terms or mentions of the document to the entities and to explore the semantic and contextual information by calculating the association of explicit relationships and the weight of the relationships of entities involved.

2 Background

2.1 Ontology

Ontology is a powerful tool for representation and reasoning of formal knowledge [26,20]. It is used by researchers to represent data of different types and areas and is encoded by OWL ontological languages. It consists of a scheme and instances (see Figure 2) to represent the description of knowledge of their concepts and relationships. An S scheme is defined as $< C, D, P >$, where C is the set of classes $C = c_1, c_2, ..., c_n$, D is the set of data types, and P is the set of properties $P = p_1, p_2, ..., p_n$ which are the relations between the classes. Instances represent knowledge and denote an instantiated class and its relationships. Instances can be defined as a graph $G =< V, E >$, where V is the set of instances and E the set of relations or predicates that join the instances. All classes, properties, data types and instances are explicitly identified by their Uniform Resource Identifier (URI) and are entities of the ontology. Each entity in the ontology is characterized by its textual description declared in the property $rdfs : label$ and it is posible to have lexical variations defined as *rdfs:label* = { *"text1", "text2"*}. Figure 2 shows the fragment of an ontology in the research domain. At the schema level, classes (such as *Laboratory* and *Professor*) and properties (such as *interestedIn*) are defined. At the instance level they indicate the instantiated schemas such as ontologies (instance of the *ResearchGroup* class), *Methodology...* and *Alice Perez* belong to the *Publication* and *Author* classes, respectively. The *Acapulco* instance contains two lexical variations *rdfs:label* = { *"acapulco", "acapulco de juarez"*}.

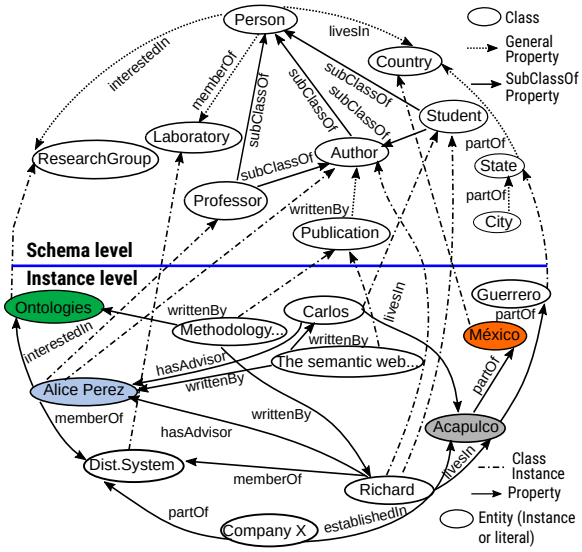


Fig. 2. Ontology example.

2.2 Semantic Annotation

The semantic annotation is the association of some data entity (people, objects, organizations, places, etc. of a text or Web document) to a description or semantic element (ontology) defined in *rdf: label*, in such a way that they are mappings between the fragment of a document term $d = t_1, t_2, \ldots, t_n$ and the concepts that describes the content of the document semantically. The annotation results in metadata that provides information on the classes and instances of entities [34]. It is fundamental in a variety of semantic Web applications such as linked data generation, open information extraction and semantic search. Specifically, semantic search allows users to express their information needs in terms of the knowledge base concepts.

2.3 Related Works

The semantic search involves different processes, which can be divided into: 1) preprocessing, 2) semantic queries translation, 3) semantic annotation, 4) semantic content retrieval, and 5) semantic ranking. In semantic annotation we have classified these works into two categories: 1) general purpose approaches, which help the annotation process and 2) information retrieval based approaches, which use specific domain ontologies and knowledge base.

General-porpuse tools. Let us remark that, AlchemyAPI[3] and OpenCalais [21] use context-based statistical techniques to disambiguate the candidate instances to annotate a term. These tools use proprietary vocabularies and ontologies whose instances are linked to DBpedia through the owl:sameAs relationship. However, OpenCalais provides some limited linkage to DBpedia. Also, OpenCalais is mainly focused on organizations. This approach has two disadvantages. Firstly, it only explores the surface of the graph for each DBpedia instance considering the labels, abstract, links to Wiki pages, and synonyms. Secondly, this approach annotates a term with only one instance of DBpedia. Therefore, this approach does not exploit the semantic information available in DBpedia to disambiguate the instance annotating a given term.

DBpedia Spotlight [17] is a semantic annotation tool for data entities in a document and it is based on DBpedia for the annotation. Also, this tool provides interfaces for disambiguation, including a Web API which supports XML, JSON, and RDF formats. Gate [33] is a tool for text engineering to help users in the process of text annotation manually. This tool provides basic processing functionalities, such as recognition of entity named, sentence dividers, markers, and so on.

Ontea [13] is a tool for semantic metadata extraction from documents. This tool uses regular expressions patterns as text analysis tool, and it detects semantically equivalent elements according to the domain ontology defined in the tool. This tool creates a new individual ontology from a defined class and it assigns

[3] http://www.alchemyapi.com

the detected elements as properties in the ontology class. The patterns of regular expressions are used to annotate the text without format with elements in the ontology. These approaches and tools are based on a dictionary search strategy. This consists of finding occurrences in text by applying a strict match of terms. They also allow for small variations in the matching of words through translation into regular expressions of the words.

Semantic Annotation Approaches Based on Information Retrieval Techniques. Popov and colleagues [24] presented KIM, a platform for information and knowledge management, annotation, and indexed and semantic retrieval. This tool provides a scalar infrastructure for personalized information extraction and also for documents management and its corresponding annotations. The main contribution of KIM is the recognition of the named entities according to ontology. Castells et al. [5] propose an information retrieval model using ontologies for the annotation classification. This model uses an ontology-based schema for semiautomatic semantic annotation of documents. This research was extended by Fernández et al. [28] to provide natural language queries. Berlanga et al. [1] propose a semantic annotation strategy for a corpus using several knowledge bases. This method is based on a statistical framework where the concepts of the knowledge bases and the corpus documents are homogeneously represented through statistical models of language. This enables the effective semantic annotation of the corpus. Nebot and Berlanga [1] explore the use of semantic annotation in the biomedical domain. They present a scalable method to extract domain-independent relationships. They propose a probabilistic approach to measure the synonymy relationship and also a method to discover abstract semantic relationships automatically. Fuentes-Lorenzo et al. [9] propose a tool to improve the quality of results of the Web search engines, performing a better classification of the query results.

3 Approach to Context-Based Semantic Annotation

The paper presents a novel proposal of semantic annotation by unstructured documents representation using an ontology to link the document terms/mentions to the ontology entities and to explore the semantic and contextual information. The annotation approach enriches and describes the documents semantic content using the ontology entities similarity by computation two measures: 1) explicit relationships association and 2) the relationships weight of the entities involved. The semantic annotation approach is shown in Figure 3. Below each step described.

Mentions detection. The documents are analyzed to detect terms or phrases that may be names people, organizations, places, expressions of time, quantities, etc; these terms are known as mentions or named entities. Mentions detected may correspond to entities in the knowledge base [31]. For mentions identification process , the Tagme tool has been used to analyzing the n-grams in documents

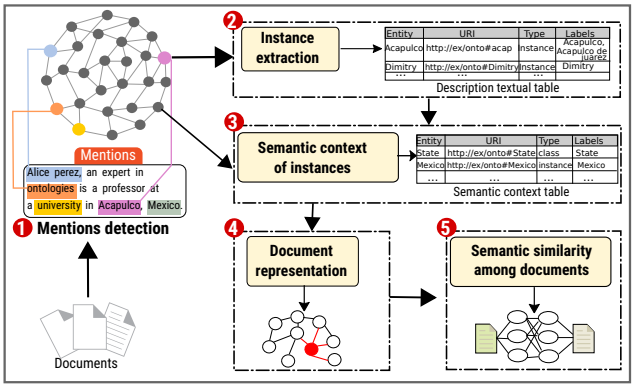


Fig. 3. Approach to context-based semantic annotation.

text by a entities dictionary (see Figure 4). A entities dictionary is built with Wikipedia, taking into account four sources: 1) anchor texts of Wikipedia articles, 2) redirect pages, 3) Wikipedia page titles, and 4) titles variants.

In dictionary construction, the mentions of a single character or with little occurrence are discarded and the further filtering is performed in the words that have low link probability (for example, less than 0.001). Link probability is defined as:

$$Lprobability(m) = P(link|m) = \frac{link(m)}{freq(m)}, \tag{1}$$

where $link(m)$ is the number of times mention m appears as a $link$ and $freq(m)$ denotes the total number of times mention m occurs in Wikipedia.

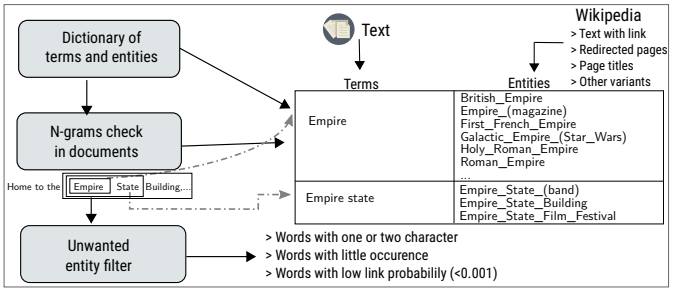


Fig. 4. Process for mentions detection.

The texts of input documents are analyzed to extract terms that may be possible mentions. All n-grams of the input text (up to n = 6), are compared with the entities dictionary. If a n-gram n_1 is contained by another one (that

is to say, that is substring), the shorter n-gram is discarded, if it has lower link probability than the longer one.

Instances Extraction in Knowledge Base. The detected mentions are searched in the ontology by means $rdfs:label$ to find their coincidence in some entity or instance. All values contained in $rdfs:label$ (lexical variations) are considered as labels. Figure 5 shows a code fragment of entity Mexico. In the source code, line 1 shows that the entity is an instance or individual; line 2 entity name; in line 3 the class to which it belongs and 4 its textual description ($rdfs:label$) with two lexical variations: "Mexico"and "Estados Unidos Mexicanos".

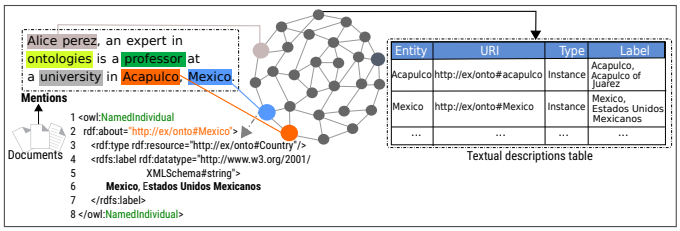


Fig. 5. Instance extraction process.

Semantic Context Extraction of Instances. In this stage, the entities semantic context detected previously is extracted. The explicit relationships in *URI* are also analyzed. The strategy to extract the semantic context is based on measuring the weight of the properties. To measure the association strength among each pair of entities, we have taken into account the entities characteristics and relationships in the knowledge base and it is calculated as a combination of two types of measures: 1) association between each concept pair and 2) the relationships weight.

1. **Concept pairwise Association.** It is used to calculate the relevance degree of a property for entities connected. We compare each pairwise (concepts c_1 and c_2) by calculating similarity. Figure 2 shows the *Acapulco* entity with five explicitly related concepts (*Carlos*, *Guerrero*, *Mexico*, *Richard*, and *CompanyX*). The association strength between each pairwise can be measured taking into account different characteristics, such as the shortest path between concepts pairwise, the depth of their common ancestor, and information content [7]. We have adopted the Resnik approach [25] to measure the similarity between two concepts c_1 and c_2 according to the information content, using the formula:

$$IC = -log_2 \frac{I(D(c))}{I(C)}, \tag{2}$$

where $I(D(c))$ denotes the number instances of the concept c and $I(C)$ represents the number of instances on the ontology.
If we consider that the ontology of Figure 2 contains 1000 resources in *Person*, *Publication* and *ResearchGroup* classes; of which 600 people are interested in a research group (*ResearchGroup*) and 100 people (*Author*) wrote a publication (*Publication*). The information content in *interestedIn* and *writtenBy* is obtained:

$$IC(interestedIn(Person, ResearchGroup)) = -log_2 pr(interestedIn(Person, ResearchGroup)) = -log_2 \frac{600}{1000} = -log_2 0.6 \approx 0.73,$$

$$IC(writtenBy(Publication, Author)) = -log_2 pr(writtenBy(Publication, Author)) = -log_2 \frac{100}{1000} = log_2 0.1 \approx 3.32.$$

Although the information content in a property represents the property discrimination strength, may not be sufficient to determine the entity meaning and extract the semantic context of instances. We propose to measure the weight of each property linked to a concept c.

2. **Relationships Weight.** Based on information theory, the amount of information contained in a random variable over another variable is measured by mutual information (MI). This strategy has been proposed by Cover [6] and we have adapted it to measure the relationship strength of pairwise c_1 and c_2:

$$MI(p(d, r)) = \sum \sum pr(c_1, c_2) \cdot log_2 \frac{pr(c_1, c_2)}{pr(c_1) \cdot pr(c_2)}, \quad (3)$$

where $pr(c_1, c_2)$ is the probability of relationship e belonging to a set of properties of c_1 and c_2. $pr(c_1)$ is the probability of relationship belonging to set of properties of c_1 , whereas $pr(c_2)$ is the probability of relationship e belonging to set of properties c_2. Figure 6 shows the relationships *writtenBy*, *memberOf*, *hasAdvisor*, and *livesIn* belonging to *Richard* entity in the ontology. The instances of these relationships are shown in Figure 2.

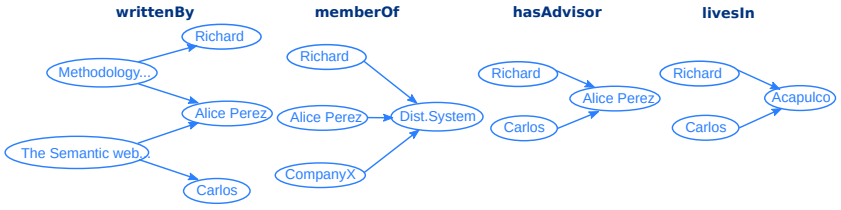


Fig. 6. Examples of *writtenBy*, *memberOf*, *hasAdvisor*, and *livesIn* property.

As an example, and without generality loss, suppose we want to calculate the relationship weight between *Richard* and *Methodology...*, (which is *writ-*

tenBy), is calculated as follows:

$$\begin{aligned}
&\mathbf{MI}(\mathbf{writtenBy}(\mathbf{Publication}, \mathbf{Author})) = pr(Methodology, Richard) \cdot log_2 \\
&\left(\frac{pr(Methodology, Richard)}{pr(Methodology) \cdot pr(Richard)}\right) + pr(Methodology, AlicePerez) \cdot log_2 \\
&\left(\frac{pr(Methodology, AlicePerez)}{pr(Methodology) \cdot pr(AlicePerez)}\right) + pr(TheSemanticWeb, AlicePerez) \cdot log_2 \\
&\left(\frac{pr(TheSemanticWeb, AlicePerez)}{pr(TheSemanticWeb) \cdot pr(AlicePerez)}\right) + pr(TheSemanticWeb, Carlos) \cdot log_2 \\
&\left(\frac{pr(TheSemanticWeb, Carlos)}{pr(TheSemanticWeb) \cdot pr(Carlos)}\right) \\
&= \frac{1}{4} \cdot log_2 \left(\frac{\frac{1}{4}}{\frac{1}{2} \cdot \frac{1}{4}}\right) + \frac{1}{4} \cdot log_2 \left(\frac{\frac{1}{4}}{\frac{1}{2} \cdot \frac{1}{2}}\right) + \frac{1}{4} \cdot log_2 \left(\frac{\frac{1}{4}}{\frac{1}{2} \cdot \frac{1}{2}}\right) + \frac{1}{4} \cdot log_2 \left(\frac{\frac{1}{4}}{\frac{1}{2} \cdot \frac{1}{4}}\right) = 0.5.
\end{aligned} \tag{4}$$

It should be noted that a relationship can have many instances. Consequently, calculating the relationships weight would have a high computational cost. Thus, we calculate the approximate mutual information as stated in:

$$MI(e) \approx log_2 \left(\frac{\frac{1}{[I(e)]}}{\frac{1}{I(c_1)} \cdot \frac{1}{I(c_2)}}\right), \tag{5}$$

where $[I(e)]$ represents all relationships e in the relationships set, $I(c_1)$ represents all relationships in c_1 (subject), and $I(c_2)$ represents all relationships in $c2$ (object).

Combining Association and Relationship Weights. A weighted sum as combination method to adjust the influence of each factor on the total weight was selected. Finally, to combine the association between each pair of concepts (see equation 2) and the weights of the relationships (see equation 3), we calculate the final weight to obtain the entities context, as stated in:

$$W(p(c_i, c_j)) = \alpha \cdot Sim(c_1, c_2) + \beta \cdot MI(p(c_1, c_2)), \tag{6}$$

where $0 \leq \alpha$, $\beta \leq 1$. Sim and MI were normalized to be in the $0, 1$ range by unit-based normalization [13], stated in:

$$\frac{Sim - min_{p \in P} Sim}{max_{p \in P} Sim - min_{p \in P} Sim} \text{ and } \frac{MI - min_{p \in P} MI}{max_{p \in P} MI - min_{p \in P} MI}.$$

3.1 Document Representation

Each document is represented as a contextual graph. The contextual graph is constructed by means of extracted instances in each document and the extraction

of its semantic context by calculating the association between the concepts and the weight of relationship. It can expressed as: Given a document corpus $C = d_1, d_2, ..., d_n$ and a knowledge base, a contextual graph GC_t is constructed to a document d_n; we consider the entities set $E = e_1, e_2, ..., e_m$ that occur in the entire contextual graph.

3.2 Semantic Similarity among Documents

Because the proposal is limited to the semantic annotation process, we have used the strategy of Paul et al. [22], they considers that two documents are similar if many annotations of a document are related to at least one annotation in another document (see Figure 7). The figure shows that the entities of document A are compared with the entities of document B. The edges $e = (v, w)$ with greater similarity are selected to calculate the similarity between both documents by means of the following formula:

$$SimDoc(docA, docB) = \frac{\sum_{a_{1i} \in A_1} (sim_{ent}(a_{1i}, matched(a_{1i})))}{|A_1| + |A_2|}. \quad (7)$$

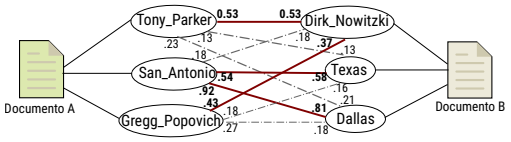


Fig. 7. Similarity approach between pairwise of Paul et al. [22].

4 Evaluation

For the tests, the following resources were used:

- *Ontology and knowledge base.* Two resources were used: DBpedia and KIM platform [24]. The KIM ontology is about politics news, finance and sports. It consists of more than 250 classes, 100 relationships and attributes. The knowledge base consists of 200,000 instances; 50,000 locations 130,000 organizations 6,000 people, etc. DBpedia uses a large multi-domain ontology. It contains 685 classes and 2795 properties and the knowledge base is over 4 million instances.
- *Hierarchy for DBpedia categories.* According to Lam et al. [14] category systems Wikipedia has greater coverage of entities for DBpedia. However, it has two problems: 1) it has no tree structure and 2) it contains cycles.

Kapanipathi et al. [12] created a system of categories[4] using Wikipedia and covering the aforementioned problems. In our approach we have used this classification system for DBpedia and it has been converted into triples for its use.

- *Document corpus.* Two corpus were used, one for each ontology. The corpus used for KIM ontology consists of 100 HTML documents of news in politics and international business and politics in the United Kingdom. For the tests with DBPedia, a corpus compiled by Lee et al. [16] named LP50[5]. It consists of 50 general purpose news documents with lengths between 50 and 126 words.
- *Evaluation metrics.* The Pearson correlation was used to evaluate our similarity results [15,18]. This metric is used to measure the approximation of our context with human judgment. Spearman correlation is a measure between two continuous random variables.

For space issues, Table 1 shows only the results of the first 20 annotated documents of LP50; column 3 shows the mentions detected in each document, in column 4 the entities and their semantic context detected in KIM and column 5 the entities and their semantic context detected by DBpedia. The greater reach of DBpedia ontology and knowledge base with respect to KIM is evident. The tests carried out with ontology and knowledge base KIM, were not very satisfactory; This is due to two factors: 1) the ontology and the instances are limited. If an ontology has a limited scope, there may not be a mention in the ontology and therefore its neighboring entities can not be extracted. On the other hand, an ontology with a larger population is more likely to cover a large part of the mentions obtained in the documents, and 2) the entities must have value in $rdfs : label$, on this depends the link between the mention and the entity of the ontology. Therefore, if an entity lacks the value in $rdfs : label$, it will not be taken into account.The tests performed with DBPedia were more satisfactory, this is because the ontology is greater and the knowledge base contains more than 4 million instances, so its scope is superior.

Table 2 shows the results of the semantic annotation evaluation DBpedia. The measures precision, recall, F measure, and accuracy were used for evaluating the annotations obtained. Precision is the rate between the relevant instances of the ontology and the total number of instances retrieved, and recall is the rate between the number of relevant instances retrieved and the total number of relevant instances existing in the ontology:

$$Presicion = \frac{|TP|}{|TP| + |FP|}, Recall = \frac{|TP|}{|TP| + |FN|}, \tag{8}$$

where TP are the set of retrieved instances that are relevant, FP the set of retrieved instances that are not relevant, and FN are the set of instances that are wrongly retrieved as nonrelevant.

[4] https://github.com/pavan046/higdataset
[5] https://webfiles.uci.edu/mdlee/LeePincombeWelsh.zip

Table 1. Summary of Corpus LP50 annotations with KIM and DBpedia.

# doc.	Words	Mention detection	Linked KIM	Linked DBPedia
1	80	13	8	30
2	98	21	10	35
3	98	17	7	34
4	106	24	4	42
5	80	13	9	47
6	97	15	14	43
7	97	27	8	39
8	82	24	10	35
9	126	12	7	28
10	76	23	11	41
11	83	17	7	31
12	67	15	8	38
13	103	4	10	21
14	105	16	9	24
15	90	17	12	45
16	75	18	11	41
17	73	15	8	29
18	62	16	7	25
19	103	27	13	33
20	122	19	11	34

Comparision to state of art. We compared our approach with different methods in the literature that measure document similarity and use the LP50 data set. Among the methods analyzed are Latent Semantic Analysis (LSA) [19], Explicit Semantic Analysis (ESA) [10], Salient Semantic Analysis (SSA) [29], Graph Edit Distance (GED) [30], and ConceptsLearned [11]. The results are shown in Table 3. The values of Pearson and Spearmen correlation of our approach were 0.745 and 0.65, respectively. This result was best compared to the results of other approaches. Thus, our approach significantly outperforms, to our knowledge, the most competitive related approaches, although ConceptsLearned has better correlation of Pearson and Spearman (0.81 and 0.75). This is because ConceptsLearned uses 17 more features compared to ours, but the computational cost is high.

Table 2. Precision, Recall, F-measure, and accuracy of semantic annotations between context-free and context-based semantic annotation.

Means	Context-free	Context-based
Precision	0.621	0.893
Recall	0.839	0.799
F-measure	0.678	0.815
Accuracy	0.644	0.835

Table 3. Our approach with other methods using LP50 dataset.

Approach	Person correlation	Spearman correlation
LAS	0.59	0.60
ESA	0.68	0.727
GED	0.72	0.63
Our approach	0.745	0.65
ConceptsLearned	0.81	0.75

Comparison with Other Metrics for Information Content (IC) Calculation. We performed tests with different metrics. The information content with the intrinsic approach can be performed using two parameters: (1) the depth of the class and (2) the descendants of a class. Table 4 shows the slight advantage of considering the ontology instances with the extrinsic information content.

Table 4. Information content with others metrics.

Parameters	Pearson correlation
Common ancestor [7]	0.548
Intrinsic IC [30]	0.744
Extrinsic IC (used in our approach) [10]	0.0.745

5 Conclusions

In this paper, we have presented a semantic annotation of unstructured documents approach. By using ontologies of a specific domain ontology. Which considers concepts similarity in ontology through its semantic relations. The unstructured documents are represented as graphs, the nodes represent the mentions, and the edges represent the semantics and relationships. Each semantic relationship has a weighting measure assigned. Thus, the significant relationships have a higher weight.

The context extraction was done through the computation of association between pairwise concepts and the weight of entity relations. The sum of the two values is the one that measures the meaning or context of an entity. We also took advantage of instances in the knowledge base to measure the information content classes and relationships. According to the state of the art the results obtained with our approach give the best results. As future work, we are trying to reduce the knowledge base by selecting the entities whose definition is more likely to be used in the corpus. Additionally, Word2vec tool for semantic extraction of terms and documents can be used. Finally, this approach also has been compared with other proposals available in the literature.

References

1. Berlanga, R., Nebot, V., Jimenez, E.: Semantic annotation of biomedical texts through concept retrieval. Procesamiento del lenguaje natural, 45, 247–250 (2010)
2. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: Dbpedia - a crystallization point for the web of data. Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 7, 154–165 (2009)
3. Bollacker, K., Cook, R., Patrick, T.: Freebase: A shared database of structured general human knowledge. In: Proceedings of the 22 National Conference on Artificial Intelligence, AAAI'7, pp. 1962–1963, AAAI Press (2007)
4. Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jaques, Y., Keizer, J.: The agrovoc linked dataset. Semantic Web, 4, 341–348 (2013)
5. Castells, P., Fernandez, M., Vallet, D.: An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval. Journal of IEEE Transactions on Knowledge and Data Engineering, 19, 261–272 (2009)
6. Cover, T., Joy, T.: Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience (2006)
7. Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Harshman, R.: Indexing by latent semantic analysis. Journal of the american society for information science, 41, 391–407 (1990)
8. Donnelly, K.: SNOMED-CT: The advanced terminology and coding system for eHealth. Journal of Studies in health technology and informatics, 121, 279–90 (2009)
9. Fuentes-Lorenzo, D., Fernandez, N., Fisteus, J., Sanchez, L.: Improving large-scale search engines with semantic annotations. Expert Systems with Application, 40, 2287–2296 (2019)
10. Gabrilovich, E., Markovitch, S.: Computing semantic relatedness using wikipedia-based explicit semantic analysis. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence (IJCAI'07), pp. 1606–1611. Morgan Kaufmann Publishers Inc. (2007)
11. Huang, L., Milne, D., Frank, E., Witten, I.: Learning a Concept-based Document Similarity Measure. Journal of American Society Information science Technology, 63, 1593–1608 (2012)
12. Kapanipathi, P., Jain, P., Venkataramani, C., Sheth, A.: Hierarchical interest graph. http://wiki.knoesis.org/index.php/Hierarchical_Interest_Graph.
13. Laclavik, M., Hluchy, L., Seleng, M., Ciglan, M.: Ontea: Platform for Pattern Based Automated Semantic Annotation. Computing and Informatics, 28, 555–579 (2009)
14. Lam, S., Hayes, C., Galway, N., Dangan, L.: Using the Structure of DBpedia for Exploratory Search. In: KDD 13: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM.(2015)
15. Leal, J., Rodrigues, V., Queiros, R.: Computing Semantic Relatedness using DBPedia. In: Simes, A., Queiros, R., Cruz, da. (eds.). OASIcs-OpenAccess Series in Informatics, volume 21 of OASICS, pp. 133–147. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2014)
16. Lee, J., Kim, K.-S., Kwon, Y., Ogawa, H.: Understanding human perceptual experience in unstructured data on the web. In: Proceedings of the International Conference on Web Intelligence (eds.) WI'17, pp. 491–498. ACM (2017)
17. Mendes, P., Jakob, M., Garcia-Silva, A., Bizer, C.: DBpedia Spotlight: Shedding Light on the Web of Documents. In: Proceedings of the 7th International Conference on Semantic Systems, I-Semantics '11, pp. 1–8. ACM (2011)

18. Lee, M.D., Welsh, M.: An empirical evaluation of models of text document similarity. In: In CogSci2005, pp. 1254–1259. Erlbaum (2006)
19. Nakov, P., Popova, A., Mateev, P.: Weight functions impact on LSA performance. In: Proceedings of the EuroConference Recent Advances in Natural Language Processing (RANLP'01), pp. 187–193 (2001)
20. Nebot, V., Berlanga, R.: Exploiting semantic annotations for open information extraction: an experience in the biomedical domain. Knowledge and Information Systems, 38, 365–389 (2014)
21. OpenCalais http://www.opencalais.com/, 2014, last accessed: 2017-04-02.
22. Paul, C., Rettinger, A., Mogadala, A., Knoblock, C., Szekely, P.: Efficient graph-based document similarity. In: Proceedings of the 13th International Conference on The Semantic Web. Latest Advances and New Domains, Volume 9678, pp. 187–193 (2016)
23. Pech, F., Martinez, A., Estrada, H., Hernandez, Y.: Semantic Annotation of Unstructured Documents Using Concepts Similarity. Scientific Programming, 2017, 1–10 (2017)
24. Popov, B., Kiryakov, A., Kirilov, A., Manov, D., Ognyanoff, D., Goranov M.: KIM: Semantic Annotation Platform. In: Proceedings of the Second International Conference on Semantic Web Conference, pp. 834–849. Springer Verlag (2004)
25. Resnik, P. Using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence-Volume 1, IJCAI'95, pp. 448–453 Morgan Kaufmann Publishers Inc. (2004)
26. Ristoski, P., Paulheim, H.: Semantic Web in data mining and knowledge discovery: A comprehensive survey. Web Semantics: Science, Services and Agents on the World Wide Web, 36, 1–22 (2016)
27. Martinez, J., Valencia R., Fernandez J., Garcia F., Martinez R.: Ontology learning from biomedical natural language documents using umls. Expert Systems with Applications, 38, 12365–12378 (2011)
28. Saha, G.: Web ontology language (owl) and semantic web. Ubiquity 2007, 1:1-1:1 (2007)
29. Samer, H., Rada, M.: Semantic relatedness using salient semantic analysis. In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, pp. 884–889 AAAI Press (2004)
30. Schuhmacher, M., Ponzetto, S.: Knowledge-based graph document modeling. In: Proceedings of the 7th ACM International Conference on Web Search and Data Mining, WSDM'14, pp. 543–552 ACM (2014)
31. Shaalan, K.: A survey of arabic named entity recognition and classification. Computational Linguistics, 40, 469–510 (2014)
32. Suchanek, F., Kasneci, G., Weikum, G.: Yago: A core of semantic knowledge. In: Proceedings of the 16th International Conference on World Wide Web, pp. 697–706 ACM (2007)
33. Department of Computer Science The University of Sheffield: Developing Language Processing Components with GATE. 8 edition, https://gate.ac.uk/userguide (2017)
34. Wei, w., Barnaghi, P., Bargiela A.: Rational research model for ranking semantic entities. Journal of Information Sciences, 181, 2823–2840 (2013)

ISSN 1870-4069

Integrating CBR with Data in Bayesian Networks for Decision Making in an Echelon Supply Chain Distribution Solution

Adrián Francisco Loera Castro[1], Alberto Ochoa Zezzatti[2], Jaime Sánchez[3], Humberto García Castellanos[3]

[1] Tecnológico Nacional de México Campus I.T.C.J., Depto. de Ingeniería Industrial y Logística, Cd. Juárez, Chih. Mexico

aloera@itcj.edu.mx

[2] Universidad Autónoma de Cd. Juárez, Depto. de Ingeniería Industrial y Logística, Cd. Juárez, Chih. Mexico

alberto.ochoa@uacj.mx

[3] Tecnológico Nacional de México Campus I.T.C.J., División de estudios de posgrado e investigación, Cd. Juárez Chih., Mexico

jsanchez@itcj.edu.mx, hgarcia@itcj.edu.mx

Abstract. When developing a causal probabilistic model, that is, a Bayesian network (BN), it is common to incorporate expert knowledge of factors that are important for decision analysis, but there are models where historical data is not available or difficult to obtain, or it is difficult to have a human expert nearby to help. This document explains how data is developed from a discrete/continuous simulated variable through a BN and mixed integer-linear programming (MILP), and the impact of this variable is measured as an important element for the decision-making model. Consider as an additional expert variable. The CBR model and the variable in question is contextualized to support in the decision-making process in a supply chain through two stages, the first is considered multiple factories, with multiple distribution centers (DC) and second, from the multiple distribution centers as it reaches multiple points of sale. As a design of a decision support system for the construction of a supply chain network (SCN) for a range of multiple end products, as well as the determination of factories and distribution centers, it also helps in the design of the distribution network strategy that satisfies all the capacities and requirements of demand of the product imposed through the points of sale. At the end of the work, an evaluation of the performance of two Bayesian networks is carried out, where one of them represents the incorporation of the expert variable using two methods, one of them the receiver operating characteristic (ROC) curve and two a method proposed by Constantinou et al. [2]., Where in both cases the Bayesian network gave a better performance with the expert variable.

Keywords: making decisions, Bayesian networks, case-based reasoning, supply chain networks, mixed integer-linear programming.

1 Introduction

Bayesian networks (BNs) [6] are rapidly becoming a leading technology in applied Artificial Intelligence. By combining a graphical representation of the dependencies between variables with probability theory and efficient inference algorithms, BNs provide a powerful and flexible tool for reasoning under uncertainty.

It has been argued that developing an effective BN requires a combination of expert knowledge and data [2]. Yet, rather than combining both sources of information, in practice, many BN models have been learned purely from data, while others have been built solely on expert knowledge.

Supply chain management is a complex domain where experienced manager practitioners hold much of their knowledge implicitly, making an appealing target for expert systems development, using Case-based reasoning (CBR). The efficiency of case retrieval algorithm is determined and affected directly by the used method for case representation. As a result, it is more logical to introduce case retrieval methods after surveying the representation methods to link them together Accuracy in obtaining the beliefs of experts, it is often unrealistic to expect the expert to provide precise probability values. In this document we present an application of a methodology proposed by [2] to a case of a BN using the learning cause of an Expert System (ES) in combination to model problems of distribution in the Supply Chain Network (SCN).

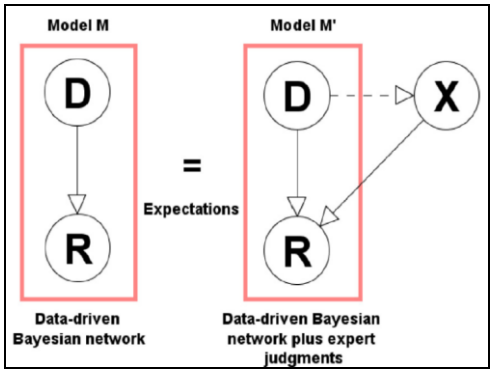


Fig. 1. Illustration where the Model M, with the data variables D and R, extends to the alternative Model M that incorporates the non-human expert variable X. Source: Constantinou et al., Integrating expert knowledge with data in Bayesian networks: Preserving data-driven expectations when the variables remain unobserved, 2006 [2].

Constantinou et al. [2] proposed a method for the evaluation of Bayesian networks, which is described below. The model *M* represents empirically observed data about the influence of D. In the example in figure 1 , the states of D are the investment options {bonds, shares, properties} and R is the *Network objective* , expressed as an observed distribution of values for each different option.

We assume that, from relevant data:

(a) $P(D_i) = d_i$ is known for each $i = 1, \ldots, n$,

(b) $f(R \mid D_i)$ is a known distribution for each $i = 1, \ldots, n$,

Hence, these are the parameters of the model *M*. Let the *expected value* $E(f(R \mid D_i)) = r_i$ for each $i = 1, \ldots, n$. For simplicity, we write this as $E(R \mid D_i) = r_i$. Hence, in model *M* the expected value of *R* is:

$$E_M(r) = \sum_{i=1}^{n} E(r|D_i)P(D_i) = \sum_{i=1}^{n} r_i d_i \tag{1}$$

Now consider the revised BN model *M'* , as shown in Fig. 1. Here *X* is an expert supplied variable with *m* states $X_1, \ldots, X_m$. We assume the expert provides the prior probabilities for *X* , i.e. $P(X_j \mid D_i) = p_{ij}$ for each $i = 1, \ldots, n$ and for each $j = 1, \ldots, m$. When *D* and *X* are not linked, then instead of $n \times m$ priors we only need *m* priors $P(X_j) = p_j$ for each $= 1, \ldots, m$. The challenge for the expert is to complete the conditional probability table (CPT) for R in M' in such a way as to preserve all of the conditional expected values of R given D in the original model M, and also preserve the marginal expectation. Specifically, we require:

$$E_{M'}(R|D_i) = E_M(R|D_i) = r_i \ \ for\ each\ i = 1, \ldots, n \tag{2}$$

Note that, if we can establish Eq. (2), then it follows from Eq. (1) that:

$$E_{M'}(R) = E_M(R)$$

Specifically, Eq. (2) is also sufficient to prove that the unconditional expected value - the expected value of R when D is unobserved- of R is preserved in M'.

Table 1. The CPT for R in M'.

D	D_1					...	D_i					...	D_n				
X	X_1	X_2	…	X_{m-1}	X_m	…	X_1	X_2	…	X_{m-1}	X_m	…	X_1	X_2	…	X_{m-1}	X_m
R	f_{11}	f_{12}	…	f_{1m-1}	f_{1m}	…	f_{i1}	f_{i2}	…	f_{im-1}	f_{im}	…	f_{n1}	f_{n2}	…	f_{nm-1}	f_{nm}

The general form of the CPT for R in M' can be written as a function f_{ij}, whose expected value is r_{ij} for each i = 1 , . . . , n and j = 1 , . . . , m , as shown in Table 1. Specifically,

$$E(f_{ij}) = E_{M'}(R|D_i, X_j) = r_{ij} \ \ for\ each\ i = 1, \ldots, n\ and\ j = 1, \ldots, m$$

Since each X_i is conditioned on D_i we can use marginalization to compute:

$$E_{M'}(R|D_i) = \sum_{j=1}^{m} E(R|D_i, X_j)P(X_j|D_i) = \sum_{j=1}^{m} r_{ij} P_{ij} \tag{3}$$

Since by Eq. (2) we require:

$$E_{M'}(R|D_i) = E_M(R|D_i) = r_i \ for\ each\ i = 1, \ldots, n$$

it, therefore, follows from Eq. (3) that we require:

$$\sum_{j=1}^{m} r_{ij} P_{ij} = r_i \quad for\ each\ 1 = 1, \ldots, n \tag{4}$$

Eq. (4) thus expresses the necessary constraints on the expert elicited values for r_{ij}. We can use Eq. (4) as a consistency check on the expert elicited values if the user wishes to provide them all. However, in practice we would expect the user to provide a subset of the values and so use Eq. (4) to solve for the missing values. There is a unique solution in the case when the expert is able to provide $m - 1$ of the required m values

$$r_{i1}, r_{i2}, \ldots, r_{im-1}, r_{im}.$$

To prove this, without loss of generality suppose that r_{im} is the 'missing value'. Then we can compute the value of r_{im} necessary to satisfy Eq. (4) . We know, by Eq. (4), that:

$$r_i = \sum_{j=1}^{m} r_{ij} P_{ij}$$

So:

$$r_i = \left(\sum_{j=1}^{m} r_{ij} P_{ij} \right) + r_{im} P_{im}$$

Thus:

$$r_{im} = \frac{r_i - (\Sigma_{j=1}^{m} r_{ij} P_{ij})}{P_{im}} \tag{5}$$

For each *i* = 1, . . . , *n* Eq. (5) thus provides the formula for computing the missing CPT values necessary to preserve in the model *M'* all of the conditional expected values of *R* given *D* in the original model *M* .

2 The Role of Distribution in the Chain of Supply

A supply chain is defined as a process with a complete set of activities wherein raw materials are transformed into final products, then delivered to customers by distribution, logistics, and retail. All inter-organizational practices such as planning, purchasing, distribution, delivery process, and reverse logistics are considered as a supply chain management system [21].

3 Distribution Decisions

Development of the new theories and methodologies in logistics and supply chain management can lead to the higher level intelligent and advanced systems. Such kind of systems enable supply chain experts to facilitate information-sharing, highly

qualified decisions and to increase the value to products and services by internal coordination. Over the last decades, the direction of decision support systems has changed drastically. To monitor the materials cost in a garment manufacturer, a decision support model has assisted decision-makers in selecting efficient ways to reduce total manufacturing costs. Decision making is influenced by the characteristics and context of decision situations [37] and it is viewed that understanding the characteristics of different types of organizational decision-making contexts is a prerequisite for understanding the nature of decision-making processes and requirements for decision support within different types of decision-making contexts. There are several ways to characterize different types of decision situations and their associated decision-making contexts within organizations.

4 Case-Based Reasoning (CBR)

A CBR system should be organized with some basic elements: the knowledge representation, to depict the cases, and the similarity measure to define how much a case is similar to another one [16, 17].

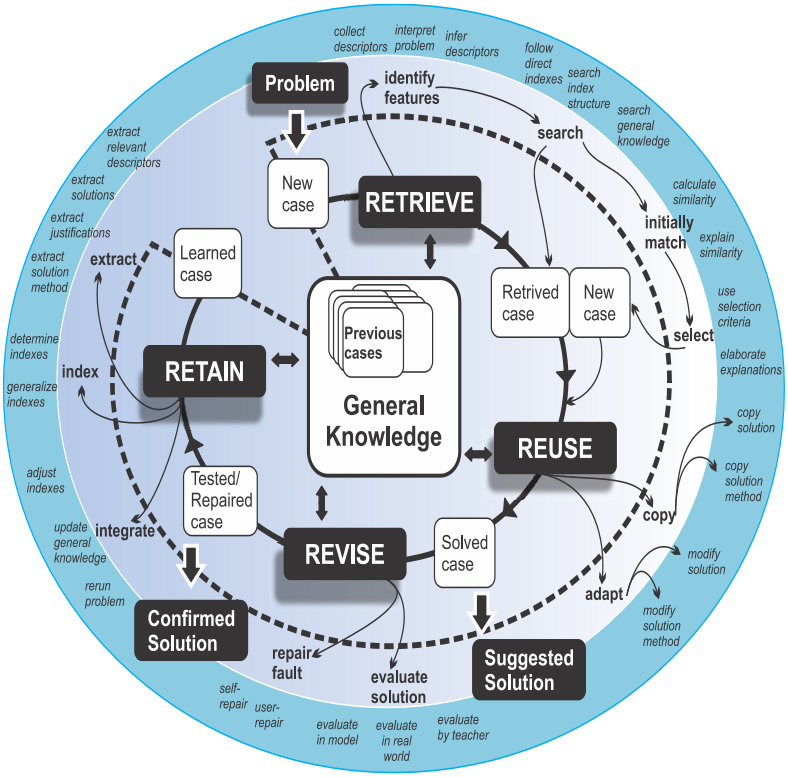


Fig. 2. A process-oriented vision of the CBR adaptation cycle based on Aamodt and Plaza (1994) Source: Loera et al. Implementation of an Intelligent Model for Decision Making Based on CBR for Supply Chain Solution in Retail for a Cluster of Supermarkets.

In figure 2, the tasks are shown with the names of the nodes in bold, while the methods are in italics.

5 R based Framework for Distribution Planning

5.1 Model Formulation

The foundation of a CBR system is the representation and definition of a case. So far, there is no uniformed standard to represent a case [5]. The constituted model represents two echelons, multi-factories, multi-warehouse or distribution centers (DC), and multi-sales points. Decision maker wishes to design of supply chain network (SCN) [35] for the end product, determine the factories and DCs and design the distribution network strategy that will satisfy all capacities and demand requirement for the product imposed via sales points. The problem is a single-product, multi-stage SCN design problem. We formulated the SCN design problem as a Mixed-Integer Linear Programming model (MILP), [18]-[21], as is shown in figure 3.

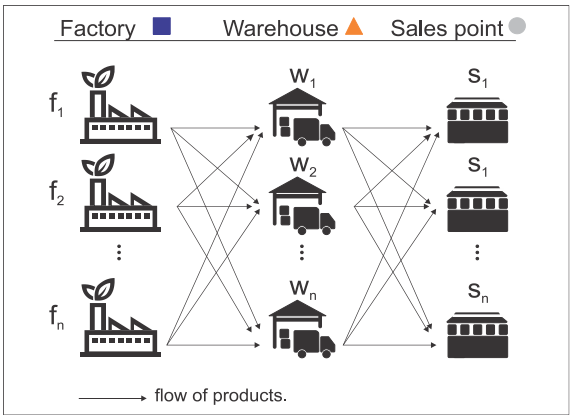


Fig. 3. Simple network of two-stages in supply chain network.

5.2 Model Nomenclature

The indices, parameters and decision variables of the mixed integer linear programming model are listed below:

Index

$f = 1,2,3,\ldots F$	Set of production facilities,
$p = 1,2,3,\ldots P$	Set of product type,
$w = 1,2,3,\ldots W$	Set of warehouse facilities (distribution centers),
$s = 1,2,3,\ldots S$	Set of sales points.

Parameters

m_{fp} Production cost in the factory f for the product p,

D_{fw} Distance between the factory f and the warehouse w,
T_p Cost of transporting the product p from the factory f to the warehouse w,
d_{sw} Distance between the sale point s and the warehouse w,
t_p Cost of transporting the product p from the sales point s to the warehouse w,
r_{sp} The s point of sale demand r for the product p,
C_{fp} The f factory capacity for the product p,
c_w Capacity of the warehouse w,
O_p The turnover rate for each product p,

Decision variables

x_{pfw} The amount of product p that is transported from factory f to warehouse w,
y_{sw} A binary variable taking value 1 when sale point s is associated with warehouse w.

Mathematical model proposed by Adrian Loera et al. in his PhD Thesis (2019):

Minimize

$$Z = \left[\sum_{f=1}^{F}\sum_{p=1}^{P}\sum_{w=1}^{W} x_{pfw}(m_{fp} + T_p D_{fw})\right] + \left[\sum_{s=1}^{S}\sum_{w=1}^{W}\sum_{p=1}^{P} r_{sp} t_p d_{sw} y_{sw}\right] \quad (6)$$

Subject to:

$$\sum_{w=1}^{W} x_{pfw} \leq C_{fp} \quad \forall w \in W, \forall f \in F, \forall p \in P \quad (7)$$

$$\sum_{f=1}^{F} x_{pfw} = \sum_{s=1}^{S} (r_{sp} y_{sw}) \quad \forall s \in S, \forall w \in W, \forall f \in F, \forall p \in P \quad (8)$$

$$\left[\sum_{p=1}^{P}\sum_{s=1}^{S} \frac{r_{sp}}{O_p}\right][y_{sw}] \leq c_w \quad \forall w \in W, \forall s \in S, \forall p \in P \quad (9)$$

$$\sum_{w=1}^{W} y_{sw} = 1 \quad \forall w \in W, \forall s \in S \quad (10)$$

$$m_{fp} \geq 0 \quad \forall f \in f, \forall p \in P \quad (11)$$

$$D_{fw} \geq 0 \quad \forall f \in F, \forall w \in W \quad (12)$$

$$T_p \geq 0 \quad \forall p \in P \quad (13)$$

$$d_{sw} \geq 0 \quad \forall w \in W, \forall s \in S \quad (14)$$

$$t_p \geq 0 \quad \forall p \in P \quad (15)$$

$$r_{sp} \geq 0 \quad \forall p \in P, \forall s \in S \quad (16)$$

$$C_{fp} \geq 0 \quad \forall f \in F, \forall p \in P \quad (17)$$

$$c_w \geq 0 \quad \forall w \in W \tag{18}$$

$$O_p \geq 0 \quad \forall p \in P \tag{19}$$

$$x_{pfw} \geq 0 \quad \forall w \in W, \forall f \in F, \forall p \in P \tag{20}$$

$$y_{sw} \in \{0,1\} \quad \forall w \in W, \forall s \in S \tag{21}$$

5.3 Case Representation

Therefore, taking into consideration of characteristics of SCN, the process case of distribution can be defined as a collection of $three - tuple$:

$$CASE = \{H, D, S\},$$

where H is the case number, $D = \{m_{fp}, T_p, D_{fw}, r_{sp}, t_p, d_{sw}\}$ is the condition feature description of distribution and $S=\{x_{pfw}, y_{sw}\}$ is the corresponding solution of distribution planning. The case representation of distribution planning in terms of condition features is shown in Table 2.

Table 2. Distribution planning case presentation.

Case representation of Distribution planning Case number (H): X
Condition features of distribution problem(D) • Production cost in the factory f for the product p • Distance between the factory f and the warehouse w • Cost of transporting the product p from the factory f to the warehouse w. • Distance between the sale point s and the warehouse w. • Cost of transporting the product p from the sales point s to the warehouse w • The s point of sale demand r for the product p • The f factory capacity for the product p • Capacity of the warehouse w • The turnover rate for each product p
Solution (S) • The amount of product p that is transported from factory f to warehouse w. • A binary variable taking value 1 when sale point $s$is associated with warehouse w.

6 Building a BN

Irrespective of the method used, building a BN involves the following two main steps [2]:

1. Determining the structure of the network: many of the real- world application.
2. Determining the conditional probabilities (CPTs) for each node also, referred to as the parameters of the model.

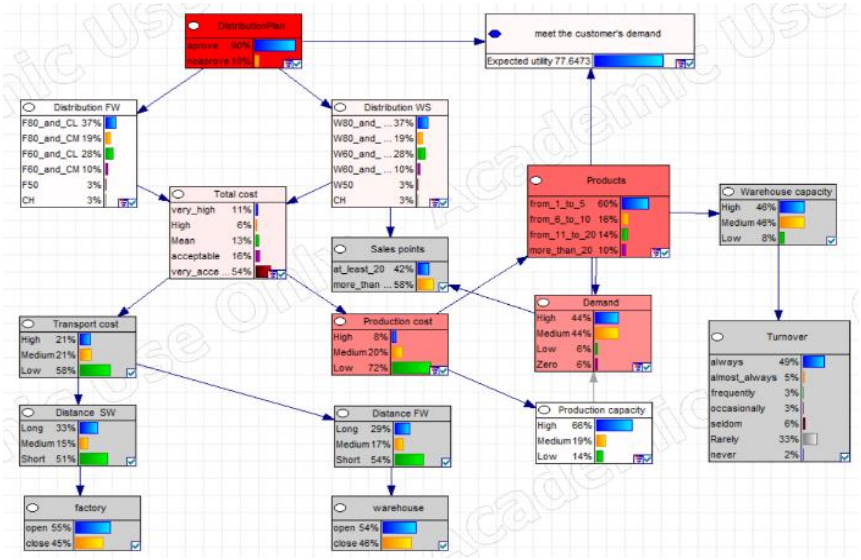


Fig. 4. Bayesian network model to SCN (M model).

6.1 Conditional Probabilities

Once the topology of the BN is specified, the next step is to quantify the relationships between connected nodes – this is done by specifying a conditional probability distribution for each node**,** see table 3.

Table 3. Conditional probability data (M model).

final stage

Turnover		*Warehouse capacity*		*Sales points*		*warehouse*		*Distance FW*		*factory*		*Distance SW*		*Transport cost*	
State	*P(∂|x)**	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*
Always	50%	High	44%	at_least_20	42%	open	55%	Long	37%	open	57%	Long	41%	High	26%
Almost_always	5%	Medium	44%	more_than_21	58%	close	45%	Medium	21%	close	43%	Medium	18%	Medium	31%
Frequently	3%	Low	12%					Short	42%			Short	42%	Low	43%
Occasionally	3%														
Seldom	5%														
Rarely	31%														
Never	2%														
	100%		100%		100%		100%		100%		100%		100%		100%

Distribution WS		*meet the customer's demand*		*Demand*		*Production capacity*		*Products*		*Production cost*		*Distribution FW*		*DistributionPlan*	
State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*	State	*P(∂|x)*
F80_and_CL	37%	Exp. utility	76%	High	41%	High	59%	from_1_to_5	49%	High	15%	F80_and_CL	37%	approved	90%
F80_and_CM	19%			Medium	41%	Medium	21%	from_6_to_10	16%	Medium	27%	F80_and_CM	19%	no_approved	10%
F60_and_CL	28%			Low	9%	Low	20%	from_11_to_20	19%	Low	58%	F60_and_CL	28%		
F60_and_CM	10%			Zero	9%			more_than_20	16%			F60_and_CM	10%		
F50	3%											F50	3%		
CH	3%											CH	3%		
	100%		76%		100%		100%		100%		100%		100%		100%

** Posterior marginal probability distribution P(∂|x)*

6.2 Parameter Learning

The structure of the Bayesian network was imported into GeNie, a general-purpose Bayesian network commercial software [7] for the parameter learning stage.

Likelihood maximization with randomized initial values for the parameters was used so that the process could be repeated from different starting points to avoid local

minima. The process was completed when the expectation maximization algorithm converted; that is when the negative log-likelihood had been minimized.

BNs model the quantitative strength of the connections between variables, allowing probabilistic beliefs about them to be updated automatically as new information becomes available [8]-[11], and this can be observed in tables 5 and 7.

Other outputs of the BN model are represented by the adjacency matrix, which represents a graph with $|V|$vertices, that is, it is a matrix of $|V|\times|V|$ of zeros and ones, where the entry in line i and column j is 1 if and only if the corner (i,j) is in the graph. Case one is represented with an X, as shown in tables 4 and 6.

Table 4. Adjacency Matrix fo M model.

Adjacency Matrix, Lower triangular	Turnover	Warehouse capacity	Sales points	warehouse	Distance FW	factory	Distance SW	Transport cost	Distribution WS	meet the customer's demand	Demand	Production capacity	Products	Production cost	Distribution FW	DistributionPlan
Turnover																
Warehouse capacity	x															
Sales points																
warehouse																
Distance FW				x												
factory																
Distance SW						x										
Transport cost					x		x									
Distribution WS			x					x								
meet the customer's demand																
Demand			x							x						
Production capacity											x					
Products		x									x					
Production cost												x	x			
Distribution FW								x						x		
DistributionPlan									x	x					x	

Table 5. Strength influence of M model.

Strength influence				
Parent	***Child***	***Mean***	***Maximum***	***Weighted***
Demand	meet the customer's demand	0.56	1.00	0.56
Demand	Sales points	0.08	0.20	0.08
Distance SW	factory	0.40	0.60	0.40
Distance FW	warehouse	0.40	0.60	0.40
Distribution FW	Production cost	0.23	0.40	0.23
Distribution FW	Transport cost	0.16	0.45	0.16
Distribution WS	Transport cost	0.15	0.31	0.15
Distribution WS	Sales points	0.09	0.35	0.09
DistributionPlan	Distribution FW	0.40	0.40	0.40
DistributionPlan	meet the customer's demand	0.69	1.00	0.69
DistributionPlan	Distribution WS	0.40	0.40	0.40
Production capacity	Demand	0.00	0.00	0.00
Production cost	Products	0.74	0.84	0.74
Production cost	Production capacity	0.35	0.51	0.35
Products	Demand	0.24	0.35	0.24
Products	Warehouse capacity	0.19	0.29	0.19
Transport cost	Distance SW	0.51	0.73	0.51
Transport cost	Distance FW	0.61	0.84	0.61
Warehouse capacity	Turnover	0.53	0.78	0.53

7 Knowledge Engineering Bayesian Networks (KEBN)

Knowledge Engineering can be viewed as an engineering discipline that involves integrating knowledge into computer systems in order to solve problems normally requiring a high level of human expertise. Similarity assessment techniques (e.g., [17]). The combination is the focus of this article, with the peculiarity that the human expert is replaced by the expert machine, i.e. the CBR [7,9,14,15].

7.1 Integrated BN with CBR

The problem we are interested in solving is the general case where a discrete expert variable -CBR- is inserted into a BN model as a parent of a discrete/continuous data variable, see figure 5. However, when the data variable is discrete some limitations apply proposed per Constantinou et al. [2].

7.2 Data Analysis

Descriptive statistics, Bayesian networks, and Receiver Operating Characteristic (ROC) curve analysis are used in this study for further investigation of the relationships between variables; This is detailed below.

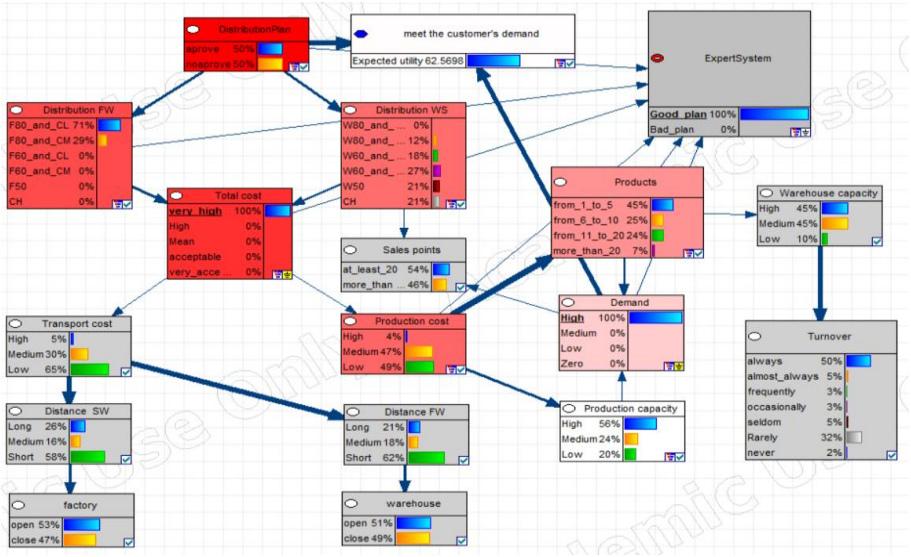


Fig 5. Integrated BN with CBR (M' model).

Table 6. Adjacency Matrix integrated BN and CBR (M' model).

Adjacency Matrix, Lower triangular	ExpertSystem	Turnover	Warehouse capacity	Sales points	warehouse	Distance FW	factory	Distance SW	Transport cost	Meet the customer's demand	Demand	Production capacity	Products	Production cost	Total cost	Distribution WS	Distribution FW	DistributionPlan
ExpertSystem																		
Turnover																		
Warehouse capacity		x																
Sales points																		
warehouse																		
Distance FW					x													
factory																		
Distance SW							x											
Transport cost						x		x										
Meet the customer's demand																		
Demand	x			x						x								
Production capacity											x							
Products	x		x								x							
Production cost	x											x	x					
Total cost	x								x					x				
Distribution WS				x											x			
Distribution FW	x														x			
DistributionPlan	x									x						x	x	

Table 7. Strength influence integrated BN and CBR (M' model).

Strength of influence				
Parent	*Child*	*Average*	*Maximun*	*Weighted*
Demand	meet the customer's demand	0.56	1.00	0.56
Demand	Sales points	0.08	0.20	0.08
Demand	ExpertSystem	0.02	1.00	0.02
Distance SW	factory	0.40	0.60	0.40
Distance FW	warehouse	0.40	0.60	0.40
Distribution FW	Total cost	0.31	0.76	0.31
Distribution FW	ExpertSystem	0.01	1.00	0.01
Distribution WS	Total cost	0.23	0.58	0.23
Distribution WS	Sales points	0.09	0.35	0.09
DistributionPlan	Distribution FW	0.40	0.40	0.40
DistributionPlan	Distribution WS	0.40	0.40	0.40
DistributionPlan	meet the customer's demand	0.67	1.00	0.67
DistributionPlan	ExpertSystem	0.01	1.00	0.01
Production capacity	Demand	0.00	0.00	0.00
Production cost	Products	0.74	0.84	0.74
Production cost	Production capacity	0.35	0.51	0.35
Production cost	ExpertSystem	0.01	1.00	0.01
Products	Demand	0.24	0.35	0.24
Products	Warehouse capacity	0.19	0.29	0.19
Products	ExpertSystem	0.00	1.00	0.00
Total cost	Production cost	0.19	0.41	0.19
Total cost	Transport cost	0.20	0.33	0.20
Total cost	ExpertSystem	0.02	1.00	0.02
Transport cost	Distance SW	0.51	0.73	0.51
Transport cost	Distance FW	0.61	0.84	0.61
Warehouse capacity	Turnover	0.53	0.78	0.53

8 Evaluation Method

The expression "evaluation of a BN" could, in short, be defined as "estimation of the performance of a BN" or "estimation of the quality of recommendations obtained by using a tool based on a BN"[12]. Evaluation constitutes a requisite for the practical application of BNs. Conventional BN evaluation consists of obtaining a set of cases from records or from experts, querying the network for a diagnostic or predictive recommendation for each case, and determining how well the recommendations agree with the actual results known for the cases [1]-[6], [12]-[13]. There are two important issues with regard to the evaluation process of a BN: on the one hand, the selection of the cases and, on the other hand, the method for measuring the performance. The cases can be obtained in two different ways:

— From the BN itself, or
— From a database or with the help of an expert in the domain.

The assessment of performance can be addressed following two distinct strategies:

— By relying on expert opinion to judge the results produced by the BN, or
— By executing a mathematical method whose entries are the cases available and the inferential results.

9 Empirical Results with Constantinou Method

The BN model M and the data-based background are shown for the type of distribution planning (D) and conditional distribution for the solution variables in Table 2 (R). We assume that there is an expert glider. Then R is represented by a set of Condition features of distribution problem described in Table 2. Suppose expert node X now includes states or $u_1, \ldots, u_k$, (where k ≥1) that have been observed. In this case, the problem is that, instead of having to keep the expected value so that:

$$E_M(R|D) = E_{M'}(R|D, X),$$

we only have to ensure that:

$$E_M(R|D) = E_{M'}(R|D, X \text{ not equal to any of } u_1, \ldots, u_k).$$

So, Eq. (5) needs only to preserve the data-driven network in model M' under the states of X for which the expert assumes that they are indirectly captured by data and hence, ignore any $u_1, \ldots, u_k$. This implies that the states $u_1, \ldots, u_k$, which are assumed not to have been captured by data, will now have added impact on R .

Therefore, equation (2) and calculating each of the factors that we obtain the results in terms of a conditional probability:

$E_M(R|D) = 0.35$, $E_{M'}(R|D, X) = 0.27$, thus, $E_M(R|D) \neq E_{M'}(R|D, X)$.

Therefore the final reasoning is that the network based on data in the M' model under the states of X for which the expert assumes that they are captured by a CBR involving the states $u_1, \ldots, u_k$ of X. There is no evidence to say that they are the same, so it is concluded that there is a different impact on R in each M and M' model. In our case the conditional probability for the approval of a distribution plan obtained in the M' model is adjusted, so it is realistic that the conditional probability of the M model.

10 Empirical Results with ROC

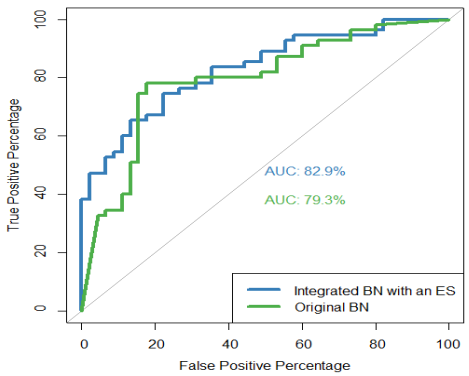


Fig. 6. Comparison of ROC and AUC.

The original Bayesian network obtained an overall model accuracy of 79.3% on validation data and the integrated Bayesian network obtained a slightly higher accuracy of 82.9% (figure 6).

11 Discussion and Conclusions

This research presented an application of a methodology to construct decision support models for BN through the incorporation of a CBR. The main contribution of this

application was the incorporation of an ES to a Bayesian network. Applying the integrated Bayesian network learning method could obtain greater accuracy than the original Bayesian network learning method, due to the fact that there are more interconnections between nodes compared to the more dispersed network of the original method. However, the improvement in overall accuracy was only 3.6%.

12 Future Research

Bayesian networks are now well established as a modeling tool for expert systems in domains with uncertainty. The reasons are its powerful but conceptual transparent representation for probabilistic models in terms of a network, there is no doubt of applicability, the persistent problem is the lack of data, so the recommendation is to apply methods that allow incorporating rare or never seen events and give them a treatment so that in the expectations based on the data of the model, under the assumption that these rare or not observed events known are not established as false within the model.

References

1. Cano, A., Masegosa, A.R., Moral, S.: A Method for Integrating Expert Knowledge When Learning Bayesian Networks from Data. IEEE Trans. Systems, Man, and Cybernetics, Part B 41(5), 1382–1394 (2011)
2. Costa Constantinou, A., Fenton, N.E., Neil, M.: Integrating expert knowledge with data in Bayesian networks: Preserving data-driven expectations when the expert variables remain unobserved. Expert Syst. Appl. 56, 197–208 (2016)
3. Hsu, F.-M., Lin, Y.-T., Ho, T.-K.: Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and Google Maps. Expert Systems with Applications 39(3), 3257–3264 (2012)
4. Guo, Y., Chen, W., Zhu, Y.-X., Guo, Y.-Q.: Research on the integrated system of case-based reasoning and Bayesian network. ISA transactions 90, 213–225 (2019)
5. Jiang, Z., Jiang, Y., Wang, Y., et al.: A hybrid approach of rough set and case-based reasoning to remanufacturing process planning. J Intell Manuf 30, 19 (2019)
6. Flores, M.J., Nicholson, A.E., Brunskill, A., Korb, K.B., Mascaro, S.: Incorporating expert knowledge when learning Bayesian network structure: a medical case study. Artificial intelligence in medicine 53(3), 181–204 (2011)
7. Tang, K., Parsons, D.J., Jude, S.: Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. Reliability Engineering & System Safety, vol. 186, pp. 24–36 (2019)
8. Korb, K.B., Nicholson, A.E.: Bayesian Artificial Intelligence. Second Edition, CRC Press Taylor & Francis Group (2011)
9. Liu, M., Stella, F., Hommersom, A., Lucas, P.J.F., Boer, L., Bischoff, E.: A comparison between discrete and continuous time Bayesian networks in learning from clinical time series data with irregularity. Artificial Intelligence in Medicine 95, 104–117 (2019)
10. Qiu, J., Gu, W., Kong, Q., Zhong, Q., Hu, J.-L.: The emergency response management based on Bayesian decision network. In: Computational Intelligence (SSCI), IEEE Symposium Series on (2016)
11. Neapolitan, R.: Learning Bayesian Networks. Prentice Hall, Series in Artificial Intelligence (2004)

12. Galán, S.F., Arroyo-Figueroa, G., Díez, F.J., Sucar, L.E.: Comparison of Two Types of Event Bayesian Networks: A Case Study. Applied Artificial Intelligence 21(3), 185–209 (2007)
13. Yet, B., Perkins, Z.B., Rasmussen, T.E., Tai, N.R.M., Marsh, D.W.R.: Combining data and meta-analysis to build Bayesian networks for clinical decision support. Journal of biomedical informatics, vol. 52, pp. 373–85 (2014)
14. Na, Y.C., Yang, J.: Distributed Bayesian network structure learning. In: IEEE International Symposium on Industrial Electronics (2010)
15. Yang, Y., Gao, X., Guo, Z., Chen, D.: Learning Bayesian networks using the constrained maximum a posteriori probability method. Pattern Recognition, vol. 91, pp. 123–134 (2019)
16. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations, and system approach. AI Communications 7(1), 39–59 (1994)
17. López De Mántaras, R., Mcsherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Faltings, B., Maher, M.L., Cox, M.T., Forbus, K., Keane, M., Aamodt, A., Watson, I.: Retrieval, reuse, revision, and retention in casebased reasoning. The Knowledge Engineering Review, vol. 00:0, 1–2. Cambridge University Press, United Kingdom (2005)
18. Senoussi, A., Dauzere-Peres, S., Brahimi, N., Penz, B., Kinza Mouss, N.: Heuristics Based on Genetic Algorithms for the Capacitated Multi Vehicle Production Distribution Problem. Computers and Operations Research (2018)
19. Pant, K., Singh,, A.R., Pandey, U., Purohit, R.: A Multi Echelon Mixed Integer Linear Programming Model of a Close Loop Supply Chain Network Design. Materials Today: Proceedings, vol. 5(2), pp. 4838–4846 (2018)
20. Barlett, M., Cussens, J.: Advances in Bayesian Network Learning using Integer Programming (2013)
21. Bartlett, M., Cussens, J.: Integer Linear Programming for the Bayesian network structure learning problem. Artificial Intelligence, vol. 244, pp. 258–271 (2017)

ISSN 1870-4069

Systematic Review of Natural Resource Management using Multiagent Systems and Role-Playing Games

Giovani Farias[1], Bruna Leitzke[1], Míriam Born[2], Marilton Aguiar[2], Diana F. Adamatti[1]

[1] Universidade Federal do Rio Grande (FURG), Programa de Pós-Graduação em Modelagem Computacional (PPGMC), Rio Grande, RS, Brazil
brunaleitzke@hotmail.com, {dianaada,giovanifarias}@gmail.com
[2] Universidade Federal de Pelotas (UFPel), Programa de Pós-Graduação em Computação (PPGC), Pelotas, RS, Brazil
{marilton,mbborn}@inf.ufpel.edu.br

Abstract. A good way of clarifying controversies is to the best quality studies in the area in question. In this way, the systematic literature review is a type of research focused on a well-defined subject, which aims to identify, select, evaluate and synthesize evidence of a specific research topic. The systematic literature review purpose is to build an overview of a specific issue and provide a summary of the literature, guaranteeing standardization and precision. Natural resource management is an area that seeks better ways to manage land, water, plants, and animals, based on the life quality of people now and for future generations. This area gained visibility with the notion of sustainability. Natural resource management focuses specifically on the scientific-technical understanding of resources and ecology and how these resources can support animal life. An approach that can help in the decision-making of this complex area is the integration of multiagent systems and role-playing games. This paper presents a systematic literature review of these three areas: natural resource management, multiagent systems, and role-playing games.

Keywords: natural resource management, multiagent systems, role-playing games.

1 Introduction

Natural resource management aims to think of efficient ways to organize and plan the ecosystem. In addition, it is important using these resources economically and sustainably, thinking about the demand of the population. Such, the management can be carried out in a coherent way and without harming the parties involved, i.e., it is necessary to understand the problems related to the ecosystem and its dynamics, and to deal with the social process involved in this system

[17]. Some models, such as biophysical ones, are often not enough to describe and analyze these processes since they are not based on the decision-making of the organizations and groups involved in the problem. Complex models are determined from the variety of non-linear behaviors within these systems [15]. In this way, it is possible to develop models that describe the dynamics between the parties involved and the management of resources.

Some techniques developed from the advances in the field of Artificial Intelligence (AI) can be used to execute simulations of social phenomena based on their virtual representations [4]. MultiAgent Systems (MAS) and Role-Playing Games (RPG) are some of these techniques, and they have been used to model and simulate the dynamics of complex systems [7].

Multiagent systems can be defined in terms of agents that interact with each other or with the environment. From this, it is possible to observe the decision-making of each agent individually and all of them, collectively. Besides that, the impact of each action on the environment could be analyzed. This concept can be understood as a "bottom-up" model since the development basing in a micro-scale shows results in macro-scale, generating consequences throughout the system. Based on this, a MAS can be designed to simulate the different strategies of resource management and verify their impact on the environment [13]. On the other hand, RPG can be used in different integrated ways to MAS, such as paper or computational models. They can assist in data collection, model validation, or the construction of tools that assist the parties involved during the participatory decision-making process [5].

There are several types of integration between MAS and RPG. These techniques can be used in parallel, i.e., exchanging information throughout the process of model development, or they can be developed in a sequential way, from the same conceptual model. In this integration, players themselves can analyze their decisions and can improve their strategies in each proposed scenario and, consequently, generate different responses to the system [12,9].

Systematic Literature Review (SLR) consists of a specific scientific methodology that goes one step further than the simple overview. A systematic review is a method that allows specialists to obtain relevant and quantified results [18]. The main idea is to identify, select and produce evidence regarding research in a particular topic. It aims to integrate empirical research in order to create generalizations. Each SLR involves specific objectives, which allows the researchers to critically analyze the collected data, to resolve conflicts detected in the literature material and to identify issues for planning a future investigation.

In this paper, we have as main goal to present an SLR about three areas: Natural Resource Management, Multiagent Systems, and Role-Playing Games, because the Natural Resource Management is a very complex area and some studies involving the MAS and RPG areas are being done, but with different approaches, to solve different problems.

In this way, the paper is structured as following: in Section 2 are presented the main concepts in the two computational areas, MAS and RPG, and how to integrate them. Section 3 presents all methodology applied to SLR in this paper,

like databases and keywords. Section 4 shows the papers resulting from the SLR and the approach of each one. Finally, in Section 5 the conclusions and future works of the paper are presented.

2 Theoretical Background

2.1 Multiagent Systems – MAS

Artificial Intelligence emerged in the 1950s on the supported of several areas such as philosophy, mathematics, psychology, neuroscience, economics, linguistics, and computer engineering [27]. In the context of AI, the concept of intelligent agents emerges in 1995, the internet being the ideal environment for the use of these agents in the development of information search mechanisms, recommendation systems, product launches of a company, monitoring of market trends, etc.

According to [27]: "an agent is something able to perceive its environment through sensors and act on this environment through actuators", considering the agent's ability to execute of specific tasks/actions and to reason about a particular domain. MAS is composed of several agents interacting in an environment. Each agent of a system has individual behavior, but they must be able to interact with others in an organized way, in this way characteristics such as cooperation, coordination, competition, and negotiation are relevant [3,8], since most problems to be solved aim a way distributed of the solution.

In the literature, there are several benefits of using MAS: (i) speed in solving problems, due to the inherent complexity of concurrent processing; (ii) increased flexibility and scalability by connecting multiple systems; (iii) increasing the capacity to respond to a given problem because all resources are located in the same environment, among others. MAS is currently used in the natural resources management domain. With this technique, it is possible to reproduce the knowledge and reasoning of several heterogeneous agents, which together need to solve common planning problems [10]. This SLR presents that MAS is widely used in this domain and, together with RPG, have satisfactory solutions.

2.2 Role-Playing Games – RPG

Role-playing game is a technique very used in training since it can put the players in situations of decision-making similar to the real ones but without effective consequences. In particular, large companies have used RPG in training courses, because of the playfulness involved in games, facilitating training and/or learning of specific subject [24]. Role-playing is a type of game where players "play/interpret" a character, created within a certain scenario (also called environment). The characters respect a system of rules, which serves to organize their actions, determining the limits of what can and can not be done [23].

Role-playing game is situated between games and theater and consist of a technique where players' rules and behaviors are determined, as well as an

imaginary (environment) context [2]. In this way, RPG could reveal some aspects of social relationships, allowing direct observation of interactions between players [6]. In an RPG, there are no winners and losers, given that it has an aspect of collaboration rather than competition. In the end, players must complete a story built from the rules of the game, in the achievement of individual and/or collective goals [2].

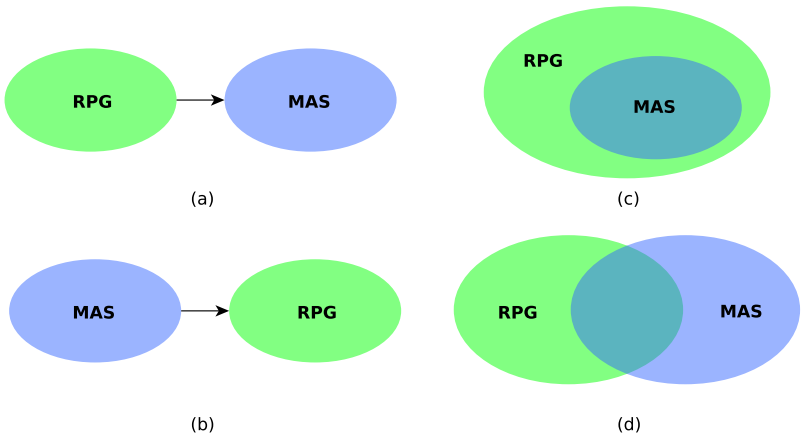


Fig. 1. Integration between RPG and MAS: (a) RPG → MAS; (b) MAS → RPG; (c) RPG + MAS; (d) RPG ++ MAS.

2.3 Integration between RPG and MAS

There are in the literature some ways to integrate RPG and MAS. Figure 1 presents four ways for this integration[3]:

(a) (RPG → MAS): RPG is playing for the stakeholders (using cards and chips in a table) and, in the final of the game, all cards are collected and information about the game is transformed in a multiagent simulation. In this type of integration, normally, the RPG is used to collect information about the problem and how the stakeholder take their decision-making;
(b) (MAS → RPG): the MAS is defined based on the knowledge of the developers. After, an RPG (with the same ideas of the MAS) is paying for the stakeholders, to validate the simulation;

[3] The components in green in (a), (b) and (c) represent activities in "paper", i.e., activities that are not computational. In the component (d) all components are computational.

(c) (RPG + MAS): an RPG is developed, but it needs computational support to execute because the rules and the calculations of the game are complex. In this way of integration, the RPG and MAS working together. The stakeholders play the game using cards, but all actions in the game process in a multiagent simulation;
(d) (RPG ++ MAS): this way of integration is unique that the two techniques are computational. The RPG is developed as a software and all calculation is done by the multiagent simulation. In this type of integration, if the RPG is developed to execute in the Web, the stakeholders could play in remote places.

In next Sections, when we address each type of integration, we have used the nicknames: (RPG → MAS), (MAS → RPG), (RPG + MAS) and (RPG ++ MAS).

3 Methodology

In this section, we describe the protocol used to perform the SLR [22] about natural resource management using multiagent systems and role-playing games. Four reviewers conducted and defined the steps of this SLR. In this work, the goal was to search in the literature for works addressing MAS and RPG used to enhance natural resource management. The main question defined for the SLR was: "What are the main works addressing natural resource management using a multiagent systems and role-playing games?" (Table 1).

Table 1. Protocol of the systematic literature review.

Main question	"What are the main works addressing natural resource management using multiagent systems and role-playing games?"
Objective	The purpose of this SLR was to search in the literature for studies, especially in the last 10 years (2009 – 2019), addressing natural resource management using multiagent systems and role-playing games.
Inclusion criteria	Studies that contain the keywords and their constraints (Table 3): – Multiagent System; – Role-Playing Game; – Natural Resource Management.
Exclusion criteria	The study is not in English; The study is not in the searched databases; The study has not the searched keywords.

The keywords used in the query were defined in the reviewers meeting, and they were changed based on the first searches. We used five databases shown in Table 2. The searches were performed in April 2019 and the string to search in all databases was the same presented in Table 3.

Table 2. Databases used in this work.

Name	URL
ACM	https://dl.acm.org/
IEEE	https://ieeexplore.ieee.org/Xplore/home.jsp
ScienceDirect	https://www.sciencedirect.com/
Scopus (Elsevier)	https://www.scopus.com/
SpringerLink	https://link.springer.com/

Table 3. Keywords used to define the string to SLR.

("multiagent system"	OR	"multiagent systems"	OR
"multiagents system"	OR	"multiagents systems"	OR
"multi-agent system"	OR	"multi-agent systems"	OR
"multi-agents system"	OR	"multi-agents systems")	
AND			
("role-playing game"	OR	"role-playing games"	OR
"role playing game"	OR	"role playing games"	OR
"role game"	OR	"role games")	
AND			
("natural resource management"	OR	"natural resources management")	

A total of *352 works* were collected from the databases (0 from ACM, 1 from IEEE, 27 from ScienceDirect, 234 from Scopus (Elsevier) and 90 from SpringerLink). The records were downloaded in bibtex and comma-separated values (csv) formats. The collected references file was edited/organized by software, Mendeley[4] and JabRef[5]. The JabRef was used to remove redundant references and the Mendeley was used to share the reading list and collaboratively tag and annotate research papers. A total of *281 references* were qualified for the data evaluation step.

The data evaluation consists of the phase in which the reviewers analyzed each reference collected and they define the papers that were included in the SLR list. This data evaluation was performed according to the following steps: *title and abstract evaluation* was the step where each reviewer evaluated if the paper were related to the main question based on its title and abstract. In this step, it was evaluated if the papers complied with the main question, keywords related to the main question and objectives were searched in the title and abstract. This step was important to eliminate papers that had no relationship with the searched theme following the exclusion criteria. In the title and abstract evaluation, the reviewers approved *22 papers*.

In the *introduction reading evaluation*, the reviewers read the introduction of each paper classified for this step. Thus, the reviewers verified which papers described better the main question and attend to the objective of this SLR. Also,

[4] (https://www.mendeley.com/)
[5] (http://www.jabref.org/)

it was possible to verify if the keywords used in the database search were present in these papers. In the introduction reading, the reviewers approved *12 papers*.

Finally, in the *complete reading evaluation*, the final step of data evaluation, the reviewers performed full-text reading and evaluate the studies that better complies with the requirements of the main question of the SLR. In this step, the reviewers approved *10 papers*, as presented in Table 4.

4 Results

In this section is presented the results obtained in SLR. Table 4 presents the list of authors, as well as how the MAS and RPG was integrated and where they were applied.

Table 4. Result of the systematic review.

Paper	Integration	Application
Adamatti et al., 2009 [1]	RPG ++ MAS	quality of water resources
Campo et al., 2009 [11]	RPG + MAS	community forest management
Farolfi et al., 2010 [14]	RPG → MAS	management of water resources
Ruankaew et al., 2010 [26]	RPG → MAS	management of forest resources
Souchère et al., 2010 [28]	RPG → MAS	management of erosive runoff
Gourmelon et al., 2013 [16]	RPG + MAS	land use management
Le Page et al., 2014 [19]	RPG → MAS	land/water and labor migration
Rebaudo et al., 2014 [25]	MAS → RPG	integrated pest management
Le Page et al., 2016 [20]	RPG ++ MAS	conservation and management of natural resources
Le Page and Perrotton, 2018 [21]	RPG + MAS	coexistence between human populations and wildlife

In the work [1], the authors have used the methodology called GMABS, that was created from the integration of RPG and MAS simulator. Using this methodology, the authors developed two prototypes in the context of periurban catchment basin. The first prototype, called JogoMan, is a role-based game: all players need to be physically present at the same place and time, and there is a minimum number of participants required to play the game. In the second prototype, called ViP-JogoMan, it was possible to insert virtual players who could imitate real behaviors and capture autonomy, social skills, reaction and adaptation of real players. For the modeling of virtual players, the BDI (belief, desire, and intention) architecture was used. In this work, the first version is the (RPG + MAS) integration and the second version is the (RPG ++ MAS) integration. As a conclusion of the paper, the authors present some test results obtained with the two prototypes as well as a preliminary discussion on how the insertion of virtual players affected the results of the game.

The work [11] reports on the experiences and lessons learned from applying a multiagent system model to study the dynamics and complex interactions among

stakeholders in forest management, managed by three villages on Palawan Island (Philippines). This model was developed using the ComMod (The Companion Modeling) approach, which consists of the application of RPG and computational simulation, in order to develop a collaborative resource management plan. Based on this approach, the authors concluded that the use of RPG and the multiagent simulation allowed researchers to understand how stakeholders act in relation to decision-making about a controlled environment and also how they devise resource management strategies. In this work, the integration was (RPG+MAS).

In the paper [14], the authors present a multiagent model created from the ComMod approach of the Kat River Valley, Eastern Cape, South Africa. This model was called KatAWARE and, in this work, the goal was to propose a detailed methodology to formalize and systematize the modeling phases of this approach. In the context, the case study was the Kat River basin and the specification of the system structure and its dynamics was represented in diagrams based on Unified Modeling Language (UML). In this way, the work presented three iterations/models with the purpose of aggregating the innumerable sources of knowledge and data with the objective of developing a collective management plan. RPG proved to be a facilitating tool in these discussions, implementation and simulations were important to explore scenarios and discuss results. It is a (RPG $\rightarrow$ MAS) integration because the information obtained from the RPG sessions served to support the scenarios of the simulations.

The paper [26] presents theoretical and experimental aspects of the ComMod approach, using as experiment the conflict between two ethnic communities into a national park in northern Thailand. In the RPG sessions, fundamental issues such as deforestation, biodiversity conservation, and community livelihoods were discussed, and from this discussion, the information was represented in a multiagent simulator. The integration in this work is (RPG $\rightarrow$ MAS) since the data collected in the RPG workshops supported the multiagent system. The conclusion of this research showed that collaborative interactions between researchers and stakeholders increased communication and collective learning, improving the integrated management of sustainable development of renewable resources.

In the [28], the authors propose the creation of a RPG, based on the ComMod approach, in order to facilitate negotiation on erosive runoff management. The game was organized in two sessions with two different river basin management committees in the Pays de Caux (France), in order to discuss and share knowledge about the environment and the stakeholders. This works also is a (RPG $\rightarrow$ MAS) integration because the information obtained from the RPG sessions provided support of the simulations. In the results reported by the authors, despite the inherent complexity of the management problem, the group of players managed to reduce flow by 20% to 50% through a dialogue on grassland, storage tanks, and management of the off-season.

The work of [16] presents the changes in land use in the island of Ushan, which is part of the Armorique Regional Nature Park in Brittany – France, has suffered

from landslides, resulting in consequences in the landscape, traditional activities and biodiversity. From the ComMod approach, researchers, together with the biosphere reserve manager and the Center d'Etude du Milieu d'Ouessant, conducted a study on the interactions between social and environmental dynamics on this protected island. The objectives of the work were: the conception of a model, the implementation of a geographic MAS and the design of a RPG game, to assist in the decision making of the interested parties and in the sustainable management of land use.

The MAS was developed based on the ARDI method and was designed and implemented on the CORMAS platform, where it was refined and validated at each stage of the discussion. The integration used among the techniques was (MAS + RPG) because, at the end of each turn of the sessions of RPG, the MAS provided the changes of land induced by the actions of the players. RPG sessions were held with individuals involved in the project where they had a critical eye and compared the results of the game to reality.

The work [19] models a MAS with a group of rain-fed rice farmers in Thailand, to investigate interactions between the availability of water and the migration of labor in rice production. Firstly, the stakeholders played a RPG to discovery the dynamic of the envirnoment and improve new game sessions and after, to develop the MAS. This work is a (RPG $\rightarrow$ MAS) integration. As results, the authors concluded that this model is a communication tool used by local scientists and farmers, and it has been developed to exchange and integrate knowledge on land-use-water interactions and labor migration. The model has three social levels: individual, family and village.

In the work [25], the authors analyze the dissemination of information on Integrated Pest Management (IPM) in communities of small Andean farmers, Peru. The researchers designed a MAS of the Andean agricultural system and from that, developed the RPG as a paper model, in this case, the integration used between the techniques was (MAS $\rightarrow$ RPG). They tested the game with 90 potato farmers from 6 communities in three countries. After RPG sessions, they conducted interviews with farmers and concluded that the RPG based on MAS is a tool that can be used for the teaching and dissemination of IPM information, which may reduce vulnerability to pest risks in the region.

The work [20] presents the ReHab game. It is an RPG-based game used to introduce the main ideas of natural management to students in the first years of a graduate course in the area. The main goal of the game is to find the equilibrium of the environment. The game has two phases: the first one, where the players can not communicate with the colleagues and they must choose their actions individually; the second one, where they communicate with the colleagues and take their decisions about the game actions. It is an (RPG ++ MAS) integration because all system is computational and each player plays in a computer. However, each simulation allows to one player (the game is not a collective simulation). The authors concluded that a well-designed RPG allows players (individually and collectively) to shape, learn about, and reflect upon responses from the socio-ecological system to various management regimes.

The idea of the work in [21] is to propose a new approach to model participatory multiagent simulation, called KILT. This approach differs to KISS and KIDS approaches, well-known in literature because the simulations could be stylized to socio-ecosystems and stimulate the social learning of the stakeholders. It is an (RPG + MAS) integration because the stakeholders play in cards and the actions are computed in a multiagent simulation. The authors conclude that with this approach, the focus of interaction can be set on computer-participant interactions (participants observe the simulation run in the manner of a cinema audience), or on participant-participant interactions (participants can intervene while the simulation runs or at intervals provided during the run).

5 Conclusions

This work proposed a SLR about three areas, Natural Resource Management, MAS and RPG, in order to identify relevant research developed in last years. The main objective of this type of review is to obtain consistent subsidies, in relation to a specific research, bringing relevant works with significant scientific contributions.

In addition to the identification of research in these areas, we have defined the ways of integration between RPG and MAS. This definition is essential for the understanding of related works and possible researches involved in this context. The works mentioned here present four ways of integration between RPG and MAS, and mostly, were developed from the ComMod approach, created by the CIRAD group, France.

Other important finding of this SLR is that until 2010 were published several researches and, in the last years, this number was reduced. However, the results of all works used in this SLR were very satisfactory, which proves that this area has huge potential to bring contributions.

Acknowledgments. This study was financed in part by the Coordenação de Aperfeiçoameto de Pessoal de Nível Superior (CAPES/Brasil) and Agência Nacional de Águas (ANA/Brasil) – Edital N° 16/2017.

References

1. Adamatti, D.F., Sichman, J.S., Coelho, H.: An analysis of the insertion of virtual players in GMABS methodology using the ViP-JogoMan prototype. Journal of Artificial Societies and Social Simulation 12(3) (2009)
2. Adamatti, D.F., Sichman, J.S., Coelho, H.: Utilização de RPG e MABS no desenvolvimento de sistemas de apoio a decisão em grupos. Anais do IV Simpósio Brasileiro de Sistemas Colaborativos SBSC 2007 p. 15 (2007)
3. Alvares, L.O., Sichman, J.S.: Introdução aos sistemas multiagentes. In: XVII Congresso da SBC-Anais JAI'97 (1997)

4. Barnaud, C., Bousquet, F., Trebuil, G.: Multi-agent simulations to explore rules for rural credit in a highland farming community of northern thailand. Ecological Economics 66(4), 615–627 (jul 2008), https://linkinghub.elsevier.com/retrieve/pii/S0921800907005228
5. Barreteau, O., Abrami, G.: Variable time scales, agent-based models, and role-playing games: The PIEPLUE river basin management game. Simulation & Gaming 38(3), 364–381 (sep 2007), http://journals.sagepub.com/doi/10.1177/1046878107300668
6. Barreteau, O., Le Page, C., D'aquino, P.: Role-playing games, models and negotiation processes. Journal of Artificial Societies and Social Simulation 6(2) (2003)
7. Boissau, S., Castella, J.C.: Constructing a Common Representation of Local Institutions and Land Use Systems through Simulation-Gaming and Multiagent Modeling in Rural Areas of Northern Vietnam: The SAMBA-Week Methodology. Simulation & Gaming 34(3), 342–357 (sep 2003), http://journals.sagepub.com/doi/10.1177/1046878103255789
8. Bordini, R.H., Vieira, R., Moreira, Á.F.: Fundamentos de sistemas multiagentes. XX Jornada de atualização em informática (JAI) 2 (2001)
9. Bousquet, F., Castella, J.C., Trébuil, G., Barnaud, C., Boissau, S., Kam, S.P.: Using Multi-Agent Systems in a Companion Modelling Approach for Agroecosystem Management in South-East Asia. Outlook on Agriculture 36(1), 57–62 (mar 2007), http://journals.sagepub.com/doi/10.5367/000000007780223650
10. Bousquet, F., Le Page, C.: Multi-agent simulations and ecosystem management: a review. Ecological modelling 176(3-4), 313–332 (2004)
11. Campo, P.C., Mendoza, G.A., Guizol, P., Villanueva, T.R., Bousquet, F.: Exploring management strategies for community-based forests using multi-agent systems: A case study in Palawan, Philippines. Journal of Environmental Management 90(11), 3607–3615 (aug 2009), https://linkinghub.elsevier.com/retrieve/pii/S0301479709002321
12. Dray, A., Perez, P., LePage, C., D'Aquino, P., White, I.: Companion modelling approach: The AtollGame experience in Tarawa atoll (Republic of Kiribati). In: MODSIM05 - International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making, Proceedings. pp. 1601–1609. Melbourne, VIC (2005)
13. Etienne, M.: SYLVOPAST: A multiple target role-playing game to assess negotiation processes in sylvopastoral management planning. Journal of Artificial Societies and Social Simulation 6(2), 1–26 (2003)
14. Farolfi, S., Müller, J.P., Bonté, B.: An iterative construction of multi-agent models to represent water supply and demand dynamics at the catchment level. Environmental Modelling & Software 25(10), 1130–1148 (oct 2010), https://linkinghub.elsevier.com/retrieve/pii/S1364815210000770
15. García-Barrios, L., Speelman, E., Pimm, M.: An educational simulation tool for negotiating sustainable natural resource management strategies among stakeholders with conflicting interests. Ecological Modelling 210(1-2), 115–126 (jan 2008), https://linkinghub.elsevier.com/retrieve/pii/S0304380007003663
16. Gourmelon, F., Chlous-Ducharme, F., Kerbiriou, C., Rouan, M., Bioret, F.: Role-playing game developed from a modelling process: A relevant participatory tool for sustainable development? A co-construction experiment in an insular biosphere reserve. Land Use Policy 32, 96–107 (may 2013), https://linkinghub.elsevier.com/retrieve/pii/S0264837712002037

17. Gurung, T.R., Bousquet, F., Trébuil, G.: Companion Modeling, Conflict Resolution, and Institution Building: Sharing Irrigation Water in the Lingmuteychu Watershed, Bhutan. Ecology and Society 11(2) (2006)
18. Kitchenham, B.: Procedures for performing systematic reviews. Tech. rep., Joint Technical Report Software Engineering GroupKeele University, United Kingdom and Empirical Software Engineering, National ICT Australia Ltd, Australia (2004)
19. Le Page, C., Naivinit, W., Trébuil, G., Gajaseni, N.: Companion modelling with rice farmers to characterise and parameterise an agent-based model on the land/water use and labour migration in northeast Thailand. In: Empirical Agent-Based Modelling - Challenges and Solutions, pp. 207–221. Springer, New York, NY (2014), http://link.springer.com/10.1007/978-1-4614-6134-0
20. Le Page, C., Dray, A., Perez, P., Garcia, C.: Exploring how knowledge and communication influence natural resources management with ReHab. Simulation & Gaming 47(2), 257–284 (apr 2016), http://journals.sagepub.com/doi/10.1177/1046878116632900
21. Le Page, C., Perrotton, A.: KILT: A modelling approach based on participatory agent-based simulation of stylized socio-ecosystems to stimulate social learning with local stakeholders. In: Dimuro G.P., A.L. (ed.) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10798 LNAI, pp. 156–169. Springer Verlag (2018), http://link.springer.com/10.1007/978-3-319-91587-6_11
22. Mariano, D.C.B., Leite, C., Santos, L.H.S., Rocha, R.E.O., de Melo-Minardi, R.C.: A guide to performing systematic literature reviews in bioinformatics. arXiv e-prints arXiv:1707.05813 (Jul 2017), https://ui.adsabs.harvard.edu/abs/2017arXiv170705813M
23. Pereira, C.E.K.: Construção de personagem & aquisição de linguagem: O desafio do rpg no ines. In: vol. 10,(jul/dez) Rio de Janeiro INES, 2004 Semestral ISSN 1518-2509 1–Forum–Instituto Nacional de Educação de Surdos. p. 7 (2003)
24. Perrotton, A., Garine-Wichatitsky, D., Valls Fox, H., Le Page, C., et al.: My cattle and your park: codesigning a role-playing game with rural communities to promote multistakeholder dialogue at the edge of protected areas. Ecology and Society 22(1) (2017)
25. Rebaudo, F., Carpio, C., Crespo-Pérez, V., Herrera, M., de Scurrah, M.M., Canto, R.C., Montañez, A.G., Bonifacio, A., Mamani, M., Saravia, R., Dangles, O.: Agent-based models and integrated pest management diffusion in small scale farmer communities. In: Integrated Pest Management, pp. 367–383. Springer Netherlands, Dordrecht (2014), http://link.springer.com/10.1007/978-94-007-7802-3_15
26. Ruankaew, N., Le Page, C., Dumrongrojwattana, P., Barnaud, C., Gajaseni, N., van Paassen, A., Trébuil, G.: Companion modelling for integrated renewable resource management: a new collaborative approach to create common values for sustainable development. International Journal of Sustainable Development & World Ecology 17(1), 15–23 (feb 2010), https://www.tandfonline.com/doi/full/10.1080/13504500903481474
27. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall (2003)
28. Souchère, V., Millair, L., Echeverria, J., Bousquet, F., Le Page, C., Etienne, M.: Co-constructing with stakeholders a role-playing game to initiate collective management of erosive runoff risks at the watershed scale. Environmental Modelling & Software 25(11), 1359–1370 (nov 2010), https://linkinghub.elsevier.com/retrieve/pii/S1364815209000656

ISSN 1870-4069

Collective Behaviors in Swarms of Builder Robots

Erick Ordaz-Rivas, Angel Rodriguez-Liñan, Luis Torres-Treviño

Universidad Autónoma de Nuevo León,
Facultad de Ingeniería Mecánica y Eléctrica,
San Nicolás de los Garza, N.L., Mexico
luis.torres.ciidit@gmail.com

Abstract. Swarm robotics is inspired by the behavior of social animals for the coordination of a large number of low cost and insufficient robots that in performing a task requires collaboration. The behavior in a swarm of robots can be manipulated by changing the parameters of repulsion, attraction, orientation and influence (RAOI). In the case of repulsion, attraction and orientation modify the basic behavior of the swarm creating functional groups of robots keeping them close or dispersed, even forming chains. While the influence parameter is associated with specific stimuli to guide the swarm to perform simple tasks. To demonstrate this, a simulation platform presents the impact of these parameters in a swarm of builder robots considering a task of transporting materials.

Keywords: collective behavior, collective construction, builder robots, swarm robotics, robotics in construction.

1 Introduction

There is a widespread trend in the use of small low-cost robots rather than a single robot for certain tasks of exploration, localization, formation generation, etc [1,3,5,7,14]. Robot coordination is inspired by the behavior of social animals such as insects and mammals [10]. When multiple individual organisms meet and move as a coordinated entity is called a swarm.

Swarm behavior also allows groups of animals to accomplish tasks they could not solve individually. This leads to enormous advantages over a single individual because it allows him to solve problems in parallel and the exclusion of some members does not imply a deterioration in the elaboration of task [4].

A swarm of robots is constituted by simple robots with sensory limitations of perception at the local level that follows very simple rules. However, when interacting with the nearest neighbors or through indirect signals, very complex behaviors emerge that can be governed to perform complex tasks such as those required under construction [12,13].

Inspired by the behavior that exists in swarms, herds, hordes, etc. and considering builder animals such as bees, termites, beavers to name a few, rules

of behavior can be proposed to govern a swarm of building robots to perform construction tasks using the information they collect in their environment [8].

In literature, swarms of robots are structured to work with a single type of material, in other cases, the materials are sensors or marks inserted to help robots identify objects. On the other hand, both robots and materials are structured so that they can perform the construction [11]. In our case, we try not to structure the robots or materials. The objects that can be gripped by each robot depends on the maximum opening of the gripper and the size of each object, therefore we only deal with small objects. The challenge of the project lies not so much in the hardware but the software because it is necessary to modify the equations inspired in the computation of swarms to create local behavior policies without relying on the position of each robot but to use only the local information that each member of the swarm.

The motivation is to develop behavior rules that are based on the repulsion, attraction, orientation and influence parameters and local information to perform construction tasks such as transporting materials, their location and placement. The proposed method is not based on centralized control and does not require to know the exact (Cartesian) position of swarm members. We selected metrics of search time and delivery, and the distance reached by the robots to evaluate the performance of the swarm.

2 Kinematics and Dynamics Model

To represent each member of the swarm through simulations, it is considered a mobile robot with differential configuration. Based on the work of A. Bara [2], the kinematics and dynamics of swarm members are described by the expressions (1) and (2), respectively:

$$\begin{bmatrix} \dot{x}_g \\ \dot{y}_g \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \frac{r}{2}\cos(\theta) - \frac{rd}{2R}\sin(\theta) & \frac{r}{2}\cos(\theta) + \frac{rd}{2R}\sin(\theta) \\ \frac{r}{2}\sin(\theta) + \frac{rd}{2R}\cos(\theta) & \frac{r}{2}\sin(\theta) - \frac{rd}{2R}\cos(\theta) \\ \frac{r}{2R} & -\frac{r}{2R} \end{bmatrix} \begin{bmatrix} \omega_r \\ \omega_l \end{bmatrix}, \tag{1}$$

where θ is the orientation the center of mass G of the mobile with respect to the inertial frame $\{I\}$, ω_r and ω_l are the angular velocities applied to the right and left wheels, d is the length between the center of mass G and the origin of mobile frame $\{M\}$, r is the wheel radius, R is the length between wheel and origin C of mobile frame and, finally, x_c and y_c are the cartesian coordinates that determine the position of the origin C of mobile frame.

$$\begin{bmatrix} m & 0 \\ 0 & I + md^2 \end{bmatrix} \dot{\mathbf{v}} + \begin{bmatrix} -md\dot{\theta}^2 \\ mdv_c\dot{\theta} \end{bmatrix} (\mathbf{v}) = \begin{bmatrix} \frac{1}{r} & \frac{1}{r} \\ \frac{R}{r} & -\frac{R}{r} \end{bmatrix} \tau, \tag{2}$$

where m is the mass of robot, I is the inertia moment, $\mathbf{v} = [v_c \ \dot{\theta}]^T$ is the vector of velocities, and $\tau = [\tau_r \ \tau_l]^T$ is the vector of wheel torques.

3 Behavior Rules

Behavior rules are based on mathematical equations that have four parameters that are associated with specific behaviors. In repulsion the robot seeks to move away from its neighbors to avoid collisions, in attraction the robot seeks to approach other members of the swarm, in orientation the robot aligns in the direction of its neighbors and with influence the robot associates a stimulus with a specific task that in our case is the execution of simple construction tasks [9]. To design new rules of behavior, different influence signals are associated with specific stimuli that are detected by different sensors.

The repulsion, attraction and orientation radius are represented by r_r, r_o and r_a, and limit the local zones shown in figure 1, where the robot is in the center of the zones [6].

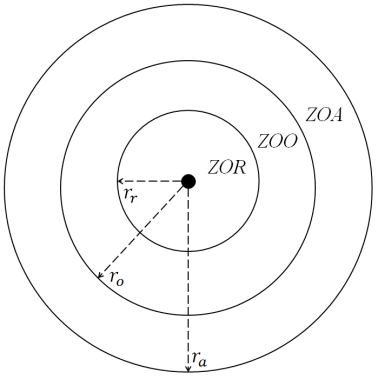
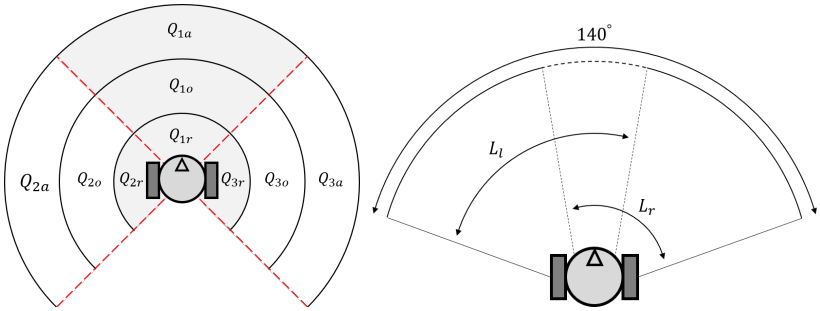


Fig. 1. Zones of repulsion (ZOR), orientation (ZOO) and attraction (ZOA).

Each member is provided with proximity and light sensors to detect and measure the distance with neighbors or objects and get information from the light perceived in the environment. These sensors are located at the front, left and right of the robots. Also, they are equipped with a gripping mechanism for holding and transporting objects

The location of the sensors allows zones in figure 1 to be divided into new sections shown in figure 2. In this work only the Q_1a, Q_1o, and Q_kr, with $k = 1, 2, 3$, zones are considered. The Q_1a zone is the only one that allows the robot to be attracted with its neighbors, the Q_1o zone keeps the robot in its current direction and the Q_kr zones evade their neighbors to avoid collisions. Besides, L_l and L_r are the light detection zones perceived by the left and right sensors respectively.

Because of sensory limitations, robots only have information about an approximate distance from their neighbors or objects in direction of the corresponding proximity sensor. There is no information about orientation or the number



(a) Repulsion and attraction zones detected by distance sensors

(b) Light detection zones by photoresists

Fig. 2. Perception zones by sensors for each robot.

of neighbors, and the allowed movements in the robots are turns in their axis and forward.

When a robot detects neighbors in ZOR, it decreases its speed and changes its direction relative to the equation 3. If it does not detect other robots in ZOR, the orientation is governed by the influence detected in the environment.

If the gripping mechanism is open, the robot searches for objects placed in the test area guided by the influence, otherwise it deposits them in the collecting area. If it detects neighbors in ZOA, it increases its speed in its current direction (equation 4). These rules return according to the status of each robot.

Where d_r and d_a represent the direction vector of robot in ZOR or ZOA, respectively.

$$d_r = -(q_{1r}[1,0] + q_{2r}[0,1] + q_{3r}[0,-1]), \tag{3}$$

where

$$q_{kr} = \begin{cases} 1, & \text{if neighbors are detected in } Q_{kr} \text{ zone,} \\ 0, & \text{if neighbors are not detected in } Q_{kr} \text{ zone,} \end{cases}$$

$$d_a = q_{1a}[1,0], \tag{4}$$

where

$$q_{1a} = \begin{cases} 1, & \text{if neighbors are detected in } Q_{1a} \text{ zone,} \\ 0, & \text{if neighbors are not detected in } Q_{1a} \text{ zone.} \end{cases}$$

4 Simulation Environment

4.1 Experimental Setup

Experiments based on simulations were carried out to understand the effects of repulsion, attraction, orientation and influence on the swarm. Simulations were performed with 5, 10 and 20 robots with differential configuration and 20

Table 1. Robot parameters.

Parameter	Value
m	$0.325kg.$
I_p	$0.005kg \cdot m.^2$
r	$0.03m.$
R	$0.05m.$
d	$0.02m.$
$c_r = c_l$	$0.434\ s$
$k_{L,r} = k_{L,l}$	$2.745\ \frac{rad}{s \cdot N - m}$
$k_{s,r} = k_{s,l}$	$1460.2705\ \frac{rad}{s \cdot V}$

randomly placed objects in an area of 10 x 10 m, the physical parameters of the robots are shown in table 1. The object location zone is at coordinates [7.5, 7.5] and has a radius of 4 m, while the delivery zone is at coordinates [2, 2] and has a radius of 4 m. These experiments were performed on Scilab 6.0.1 software on Intel(R) Core(TM) i7-7500U CPU 2.9 GHz, RAM 8GB and 64-bit operating system.

4.2 Object Transport Task

The experiment consists of a task of transporting objects, considering a limited area and randomly placed objects, the goal is to collect these objects and group them in the desired zone. Dynamics of simulated robots are governed by RAOI parameters and environmental conditions. The influence factor helps robots to find objects by giving an approximation of their location, as they cannot be detected unless they are a minimum distance away. When a robot collects an object, a new stimulus moves it to the desired zone. This allows the main task to be divided into search and delivery subtasks. The stimuli are generated by a simulated light source in the environment. The simulation ends when the robots deposit all objects in the desired zone.

In these experiments, all members of the swarm are given the same parametric settings. Tables 2 and 3 show an experimental design for changes to parameters. In the parametric settings of the table 2, the values of r_o and r_a are set as constants and only r_r varies, while in the table 3 r_e and r_o are set as constants and only r_a varies. This is to explore the behavioral changes that arise when values of repulsion and attraction change. Three replicas were performed for each experiment to complete 54 tests. Table 4 shows the speeds of the robots when in a specific zone.

5 Results and Discussions

To illustrate the swarm's performance, graphics simulations are carried out. The figures 3 and 4 show some snapshots of a simulation with 20 robots and 20 objects. The robots are represented with black arrows when they look for objects

Table 2. Parametric settings changing repulsion values with $r_o = 0.15$ and $r_a = 0.2$ as constants.

Objects	Robots	Repulsion radius (m)
20	5	0.01
		0.05
		0.1
	10	0.01
		0.05
		0.1
	20	0.01
		0.05
		0.1

Table 3. Parametric settings changing attraction values with $r_r = 0.05$ and $r_o = 0.15$ as constants.

Objects	Robots	Attraction radius (m)
20	5	0.2
		0.6
		1
	10	0.2
		0.6
		1
	20	0.2
		0.6
		1

Table 4. Speeds in perception zones.

Zone	Speed (cm/s)
Repulsion	5
Orientation	10
Attraction	$(v_i + 20)/2$
Influence	10 - 20
Out of range	10

and blue when they deliver them, while objects are represented with yellow circles. These figures give us a visual perspective about the behavior of the swarm through perform task time.

The swarm of figure 3 has low repulsion values and high attraction. In this configuration the object collection is constant, that is to say, in the snapshots times shown there are always robots searching and delivering objects. Robot chains of different lengths are formed, this helps the robots to reach the object search zone more quickly, however, sometimes these chains are headed by robots that are in their delivery task and divert attention from those that are in a

search task. On the other hand, the swarm of figure 4 has high repulsion and low attraction values. Unlike the previous configuration, the swarm's behavior is more closely joined and causes most objects to collect at first. However, the behavior is cycled when the swarm revolves around the last objects. When robots search objects, they converge to them through influence. In some cases, one or two robots are separated from the swarm because they aren't detecting objects but finally converge towards them.

Tables 5 and 6 show results of experiments performed with parameter settings from tables 2 and 3, respectively in a average scenario. Best-performing results are marked in bold for each swarm population size. A better perspective to the results is shown in figures 5 and 6.

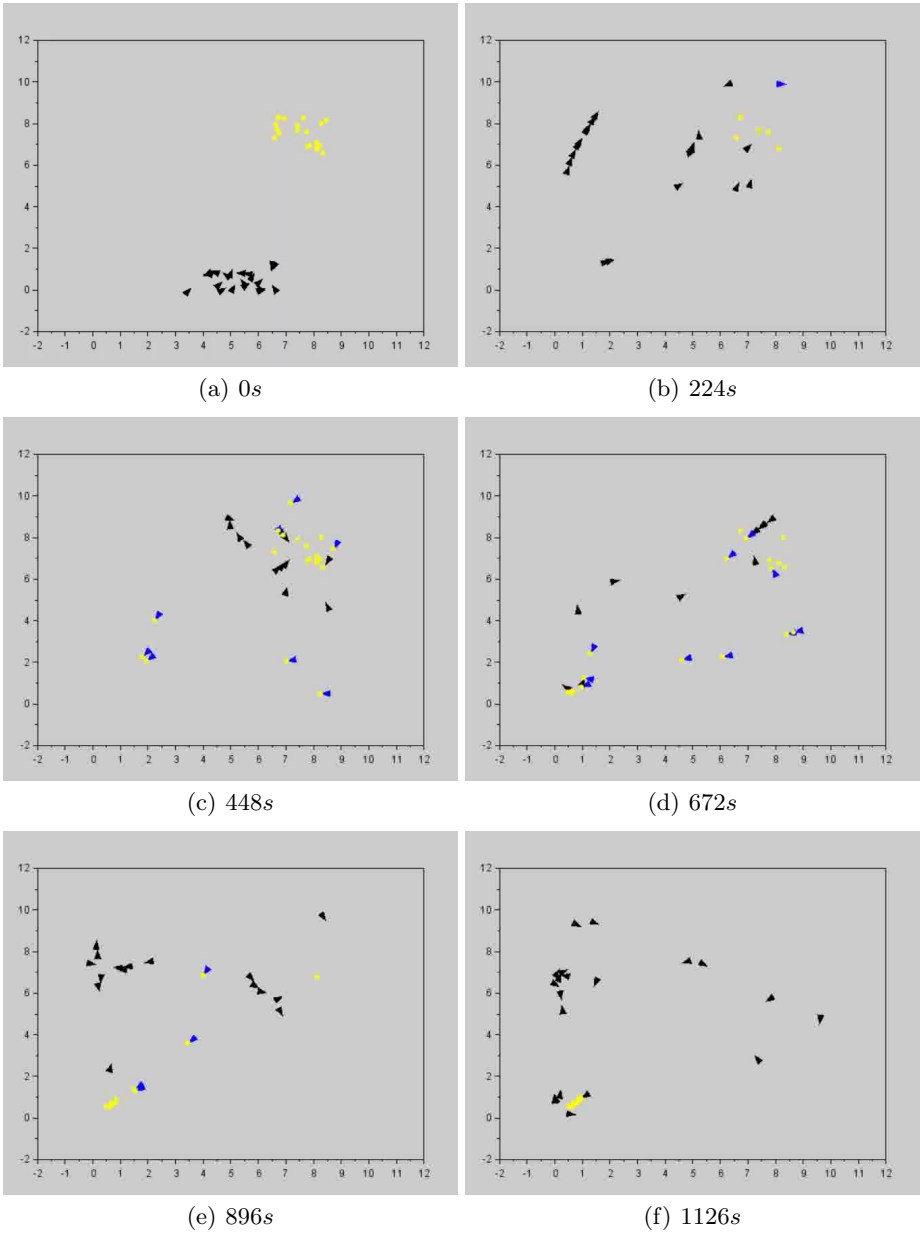
Table 5. Swarm behavior by changing repulsion values.

Robots	Repulsion radius (m)	Delivery time (s)	Search time (s)	Time (s)	Distance reached (m)
5	0.01	**797**	6966	7763	913.38
	0.05	801	9380	10181	1099
	0.1	865	**4364**	**5229**	**590**
10	0.01	413	**2117**	**2530**	**262.47**
	0.05	**382.47**	4542	4925	548.15
	0.1	390	3149	3539	377.94
20	0.01	**199**	1877	2076	221.44
	0.05	201	2012	2213	233.13
	0.1	381	**680**	**1061**	**79.68**

Table 6. Swarm behavior by changing attraction values.

Robots	Attraction radius (m)	Delivery time (s)	Search time (s)	Time (s)	Distance reached (m)
5	0.2	801	9380	10181	1099.12
	0.6	793	**3070**	**3863**	**392.15**
	1	**781**	5450	6232	588.16
10	0.2	**382**	4542	4925	548.15
	0.6	393.80	**3129**	**3522**	**379.44**
	1	395	4209	4604	510.32
20	0.2	201	2012	2213	233.13
	0.6	207	**914**	**1121**	**111.93**
	1	**195**	1517	1711	178.70

The results in figure 5 show that with average repulsion values the swarm performs worse, while at the extremes (low and high repulsion) the swarm performs its tasks faster and with better performance. The results in figure 6 show the opposite behavior but with attraction values. With average values,



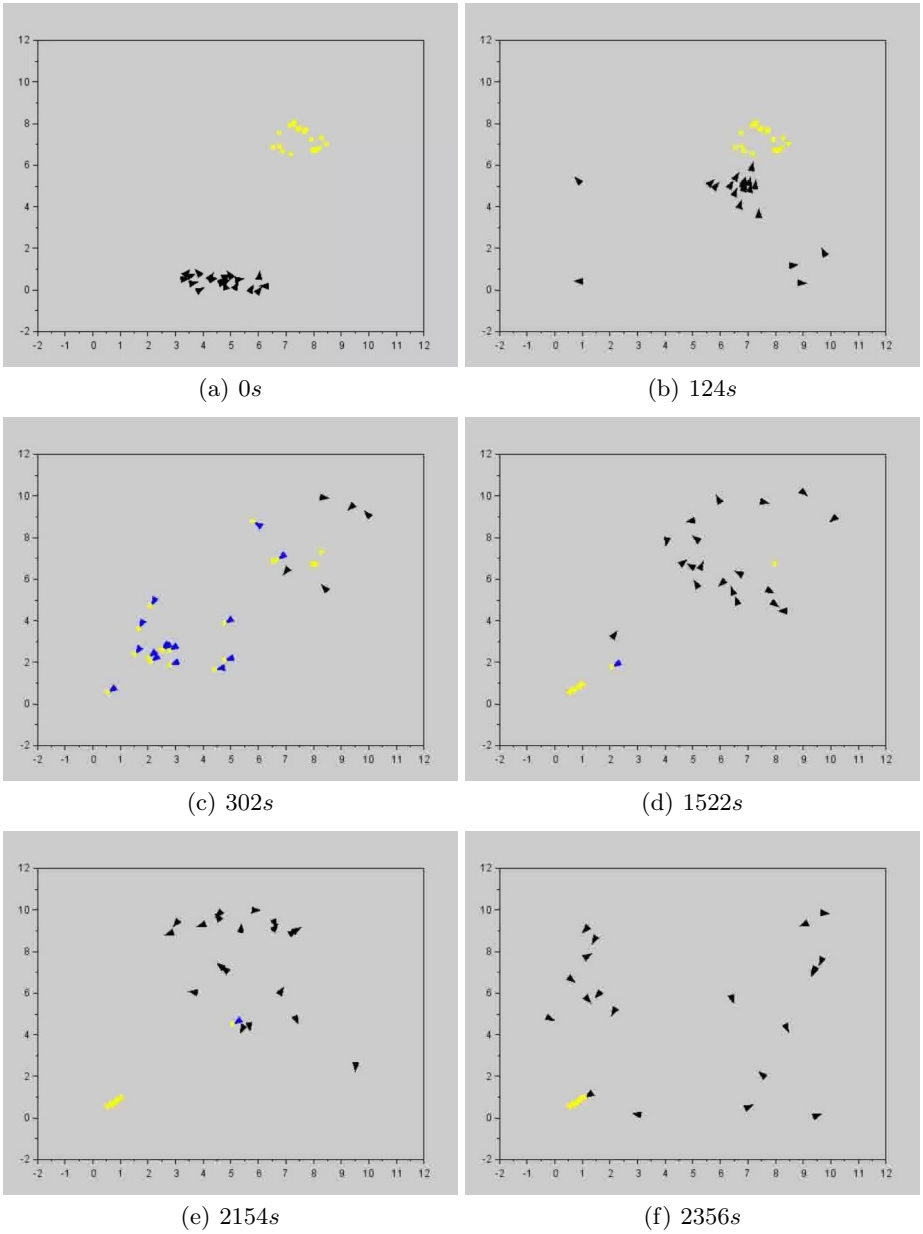
(a) 0s (b) 224s

(c) 448s (d) 672s

(e) 896s (f) 1126s

Fig. 3. Swarm simulation with 20 robots performing an object transportation task with $r_r = 0.01$ and $r_a = 1$.

better performance is obtained concerning low and high values. In both cases, as the population size increases, these properties remain but their intensity



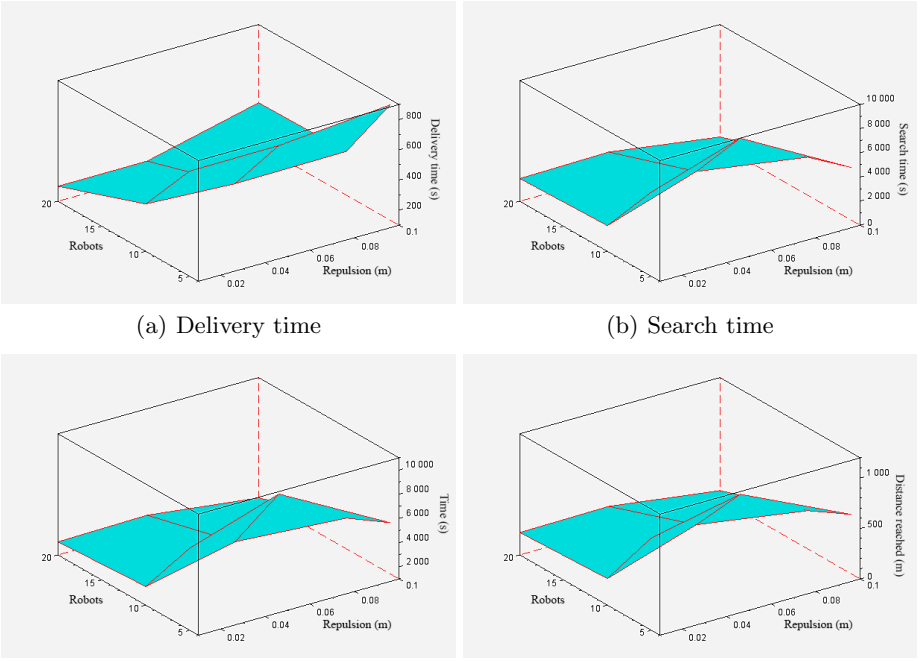
(a) $0s$ (b) $124s$

(c) $302s$ (d) $1522s$

(e) $2154s$ (f) $2356s$

Fig. 4. Swarm simulation with 20 robots performing an object transportation task with $r_r = 0.1$ and $r_a = 0.2$.

decreases. However, the measurements showed only reflect a change in behavior through parametric variations but formation properties or how robots achieve



(a) Delivery time (b) Search time

(c) Time (d) Distance reached

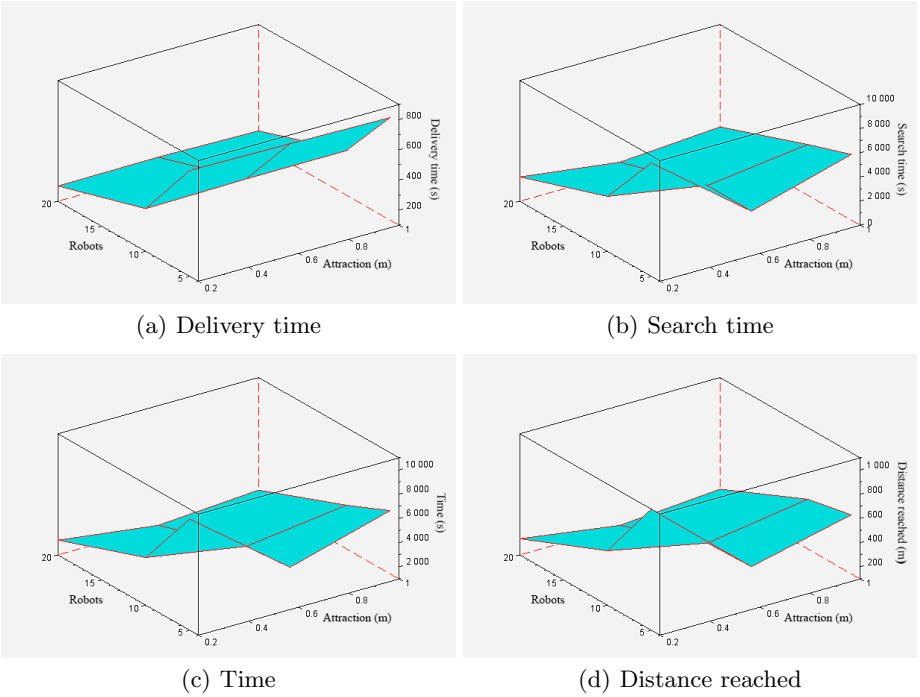
Fig. 5. Swarm behavior by changing repulsion values.

their objectives are explained below based on the snapshots of figures 3 and 4. When repulsion is low, robots tend to hold together to avoid collisions. As repulsion increases, they disperse and have priority to remain away from their neighbors rather than collect objects. A high attraction causes chains of robots to form, this causes collisions between them to be avoided and reach their goal more easily

6 Conclusions and Future Work

Different parametric settings were explored through simulations by changing the repulsion and attraction parameters, through these changes it is demonstrated that it is possible to govern the swarm behavior to perform search and delivery tasks more effectively. However, these parameters have not been fully explored because while one changed the others remained constant. Due to the stimuli received by the robots in the environment it is possible to switch the parameters concerning each sub-task to reduce the total time of the main task.

Although these rules have already been explored by simulation for construction tasks, they have not yet been implemented in real robots. Even though there are swarms of builder robots in the literature, they are still very structured, so as a future work we propose the development of a more open platform generating



(a) Delivery time (b) Search time

(c) Time (d) Distance reached

Fig. 6. Swarm behavior by changing attraction values.

technology and own designs implementing these rules to a swarm of builder robots. In addition to this, it is desired to test more parametric values to generate new behaviors that have not yet been explored and to change these values according to subtasks they perform when perceiving factors of influence in environment.

References

1. Ardiny, H., Witwicki, S.J., Mondada, F.: Are autonomous mobile robots able to take over construction? a review. International Journal of Robotics 4(3), 10–21 (2015)
2. Bara, A., Dale, S.: Dynamic modeling and stabilization of wheeled mobile robot. In: Proceedings of the 5th WSEAS International Conference on Dynamical Systems and Control. pp. 87–92. CONTROL'09, World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA (2009), http://dl.acm.org/citation.cfm?id=1628055.1628077
3. Bayindir, L.: A review of swarm robotics tasks. Neurocomputing 172, 292–321 (2016), http://www.sciencedirect.com/science/article/pii/S0925231215010486
4. Calovi, D., Bardunias, P., Carey, N., Turner, S., Nagpal, R., Werfel, J.: Surface curvature guides early construction activity in mound-building termites. Philosophical

transactions of the Royal Society of London. Series B, Biological sciences 374 (06 2019)
5. Conti, R., Meli, E., Ridolfi, A., Allotta, B.: An innovative decentralized strategy for i-auvs cooperative manipulation tasks. Robotics and Autonomous Systems 72, 261–276 (2015), http://www.sciencedirect.com/science/article/pii/S0921889015001347
6. Couzin, I.D., Krause, J., James, R., Ruxton, G.D., Franks, N.R.: Collective memory and spatial sorting in animal groups. Journal of Theoretical Biology 218(1), 1–11 (2002), http://www.sciencedirect.com/science/article/pii/S0022519302930651
7. Gerling, V., Mammen, S.V.: Robotics for self-organised construction. In: 2016 IEEE 1st International Workshops on Foundations and Applications of Self* Systems (FAS*W). pp. 162–167 (Sept 2016)
8. Hamann, H.: Swarm Robotics: A Formal Approach (01 2018)
9. Ordaz-Rivas, E., Rodriguez-Liñan, A., Aguilera-Ruíz, M., Torres-Treviño, L.: Collective tasks for a flock of robots using influence factor. Journal of Intelligent & Robotic Systems 94(2), 439–453 (May 2019), https://doi.org/10.1007/s10846-018-0941-2
10. Parker, C.A.C., Zhang, H.: Biologically inspired collective comparisons by robotic swarms. The International Journal of Robotics Research 30(5), 524–535 (2011), https://doi.org/10.1177/0278364910397621
11. Petersen, K.H., Napp, N., Stuart-Smith, R., Rus, D., Kovac, M.: A review of collective robotic construction. Science Robotics 4(28) (2019), https://robotics.sciencemag.org/content/4/28/eaau8479
12. Tuci, E., Alkilabi, M.H.M., Akanyeti, O.: Cooperative object transport in multi-robot systems: A review of the state-of-the-art. Frontiers in Robotics and AI 5, 59 (2018), https://www.frontiersin.org/article/10.3389/frobt.2018.00059
13. Vardy, A.: Orbital construction: Swarms of simple robots building enclosures. pp. 147–153 (09 2018)
14. Wareham, T., Vardy, A.: Putting it together: the computational complexity of designing robot controllers and environments for distributed construction. Swarm Intelligence 12(2), 111–128 (Jun 2018), https://doi.org/10.1007/s11721-017-0152-7

ISSN 1870-4069

How to Learn Picture Languages

David Kuboň, František Mráz

Charles University, Prague, Czech Republic
{dkubon,mraz}@ksvi.mff.cuni.cz

Abstract. Analysis of sentences in a natural language is often based on similar methods as analysis of formal languages. Analogically, analysis of pictures could be based on analysis of formal picture languages. However, the field of formal picture languages is not developed enough for this purpose. This paper presents several models of automata accepting two-dimensional languages and outlines their learning capabilities. Further, it examines the possibility of transforming a two-dimensional language into a one-dimensional language and applying machine learning techniques in a single dimension. In this paper, we propose a new representation for formal picture languages consisting of two components – a picture-to-string function and a string language. The function rewrites any two-dimensional picture into a string. A picture language is then the set of all pictures that the function maps into the given string language. Using this representation, picture languages can be learned by applying methods of grammatical inference for string languages.

Keywords: learning, grammatical inference, automata, formal languages, picture languages.

1 Introduction

In comparison to one-dimensional (string) languages, our knowledge about two-dimensional (picture) languages is limited [14], even though their theoretical and practical significance is comparable – they can be both used as formal models of practical problems. In the case of picture languages it can be for instance automatic detection of different shapes (e.g. road signs) or more generally any problem on two-dimensional data which has some pattern regularity [30].

In this work we focus on formal two-dimensional languages, i.e. sets of two-dimensional pictures that have formally exact description. They are referred to as *picture languages*, but they are not sets of pictures in the common sense, as such sets, like the set of photos containing cars, cannot be defined mathematically. While deep neural networks are known as the best tools for recognizing objects in images [19], their application for picture languages is limited. On the other hand, powerful models of automata working on two-dimensional inputs are usually non-deterministic and inefficient for applications.

Here we propose a method for learning picture languages from positive and negative samples. Learning a model (a grammar) for a target language based on some information about the words of the language is called *grammatical*

inference [15]. In case of one-dimensional (string) languages, there are several known algorithms of grammatical inference for a number of classes of languages. Much less is known about grammatical inference for two-dimensional languages.

Two types of representations of pictures can be found in the literature. The first one is generative – such representation describes how a picture can be generated from a string. Freeman in [12] introduced an 8-letter alphabet interpreted as movements north, south, east, west, northeast, southeast, northwest, and southwest. By interpreting a word over the alphabet called a "chain code" we obtain a drawing. Hence, a picture is a set of unit length lines on a plane. This was later simplified to a 4-letter alphabet with the first four movements from the Freeman's alphabet [23]. Later, Costagliola [11] extended the alphabet in order to generate colored pictures or pictures with labels.

The second representation of a picture is a rectangular array of symbols that can be interpreted as colors of pixels in the image.

Using the first representation of pictures, a picture language can be represented as a set of strings describing all pictures in the picture language. Using the second representation of pictures, a picture language is the set of pictures accepted by a device (automaton) working on two-dimensional inputs. For example, the class of recognizable picture languages is accepted by non-deterministic online tessellation automata [14], even more powerful are sgraffito automata [33] and two-dimensional limited context restarting automata [20]. These automata models are quite powerful but they share high complexity – the problem whether given input picture is accepted by such automaton is NP-complete.

Here we propose a new representation for picture languages, which uses rectangular arrays of symbols for representing pictures and string languages for representing sets of pictures. The new representation consists of a function R that rewrites any two-dimensional picture p into a string $R(p)$ and a one-dimensional (string) language L. The picture language is then the set of all pictures p for which $R(p)$ is in L. Further we propose one such function R and combine it with a classical algorithm for inferring regular languages from positive and negative samples. In this way we obtain a method for learning picture languages.

The paper is structured as follows. After introducing basic definitions for pictures and picture languages in Section 2, the next section introduces the used learning paradigm. Then, in Section 4 we discuss the generative representation of picture languages and some known results on learning picture languages represented in this way. Next, in Section 5 we informally present several models of two-dimensional automata which work on pictures as rectangular arrays of symbols. Also some results on learning such automata are included. The crucial Section 6 introduces the new representation of picture languages and lists results of experiments with inferring representations for several simple picture languages using grammatical inference for regular languages. Concluding section contains outline for further research.

2 Pictures and Picture Languages

In the past, several authors recognized that describing multidimensional objects by strings enables them to apply the means developed in the field of formal languages for studying sets of objects [23]. The first approach to describe two-dimensional pictures by one-dimensional strings were chain codes. Here we use its simplified version from [23]. According to it, a picture is a set of unit length lines in the Cartesian plane. A word in the alphabet $\Pi = \{l, r, u, d\}$ is a *picture description*. A set of picture descriptions is a *picture description language*. For a picture description $q \in \Pi^*$, its interpretation denoted as $pic(q)$ is the drawing obtained in the following way: The interpretation starts by placing a pen at any point with integer coordinates in the Cartesian plane. Next each letter x of q, from left to right, is interpreted as moving the pen by unit distance in the direction left, for $x = l$, right, for $x = r$, up, for $x = u$, and down, for $x = d$. E.g., the letter 'L' can be drawn as $uuddr$. Obviously, the resulting picture is connected. A picture consisting of several non-connected parts would be possible to represent by extending the alphabet Π by symbols $\uparrow$ and $\downarrow$, for lifting and lowering the pen above and down to the drawing plane, respectively.

Of course, such interpretation can draw a picture at any position in the plane. We usually do not distinguish between pictures which differ by their position in the Cartesian plane – the detailed definitions can be found in [23].

For a nonempty fixed picture q consisting of a connected set of unit length horizontal and vertical lines there exists infinite number of its picture descriptions. Consider for example the set of picture descriptions given by the regular expression $(rl)^+$ that all describe a horizontal line of unit length. The set of all picture descriptions of q is denoted as $des(q)$. It is known that $des(q)$ is a regular language [23].

The second approach defines a *picture* P as a two-dimensional rectangular array of elements from a finite alphabet Σ (see [13]). We say that P has dimensions (m, n), if it has m rows and n columns. Then $P_{i,j}$ from Σ denotes the symbol at position j in row i. The set of all rectangular pictures over Σ of dimensions (m, n) will be denoted as $\Sigma^{m,n}$ and the set of all rectangular pictures over Σ of any dimension will be denoted as $\Sigma^{*,*}$. A *picture language* is then any subset of $\Sigma^{*,*}$.

Any automaton working on an picture P of dimensions (m, n) needs to know where is the border of the picture, therefore the picture is usually surrounded by sentinels $\#$, where $\# \notin \Sigma$. Delimited picture P is called *boundary picture* $\widehat{P}$ over $\Sigma \cup \{\#\}$ of dimensions $(m + 2) \times (n + 2)$ – see Fig. 1.

3 The Learning Paradigm of Exact Identification

The learning paradigm of exact identification [3] has already been applied in a number of domains including regular [2] and context-free languages [34].

A *concept* in a universe of objects is any subset of the universe. A *class of concepts* is a set of concepts. Concepts can be usually represented by words

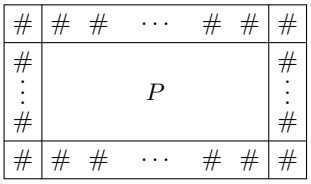


Fig. 1. The boundary picture $\widehat{P}$.

over some fixed finite alphabet. Then, the learning lies in exactly identifying an unknown (target) concept chosen from a specified class of concepts $\mathcal{C}$, i.e. computing a representation of the unknown concept. To learn, the learner may use some given information about the target concept (for example, a set of positive and negative samples, i.e. words from the concept and words not belonging to the concept) or it can even pose certain types of queries to a teacher returning the correct answers. Usually, the following two types of queries are considered (see [2]):

- a membership query, where the learner proposes an object x: the reply is *yes* if x belongs to the target concept and *no* otherwise, and
- an equivalence query, where the learner proposes a representation p (in the specified system) of a concept in $\mathcal{C}$: the reply is either *yes* if p represents the unknown concept or *no* if p is wrong and in this case the teacher also returns an arbitrary object x that p and the target concept classify differently.

Any class of concepts is trivial to learn with equivalence queries given a recursively enumerable set of representations: we just ask until the *yes* answer is received; and thus the main interest lies in the efficiency of learning algorithms.

Several known methods (protocols) can be used for learning (string) languages (see [15]). A combination of membership and equivalence queries in a polynomial-time algorithm for learning regular languages was presented by Angluin [2]. In her approach, regular languages are represented by state-minimal deterministic finite automata and the algorithm is polynomial in the number of states of a minimal automaton for the target language and the length of the longest counterexample received as response from an equivalence query. It was shown in [4,24] that there exists no polynomial-time learning algorithm that would identify the regular languages either from equivalence queries or from membership queries alone.

Here we will concentrate on learning from positive and negative samples. In particular, we will apply one of the well-known algorithms for learning regular languages from positive and negative examples called RPNI —- *Regular Positive and Negative Inference* by Oncina and Garcia [27].

Informally, the RPNI algorithm starts by separating the set S of all sample words (also called a training set) into a set of positive samples S^+, which are words belonging to the target language, and a set of negative samples S^-, which

are words not belonging to the target language. At first, the algorithm builds a prefix tree automaton from all positive samples. The automaton has states which correspond to all prefixes for all positive samples in S^+ and it accepts exactly all words from S^+ (i.e. a finite language). Then the algorithm traverses the states of the current automaton and tries to merge pairs of states. If, after merging a pair, no negative sample from S^- is accepted by the resulting quotient automaton, the merged automaton becomes the current automaton, otherwise the automaton is not changed and the algorithm proceeds trying to merge following pairs in a predetermined order.

The running time of the algorithm is $O(\|S^+\| + \|S^-\| \cdot \|S^+\|^2)$, where $\|X\|$, for a set of words X, denotes the number of states of a prefix tree automaton accepting all strings from X. The algorithm is guaranteed to produce a finite state automaton consistent with the set of samples S. This means that the resulting automaton accepts all words from S^+ and rejects all words from S^-.

4 Learning Picture Description Languages

What follows is an informal and intuitive insight into formal definitions and results from [8].

Using the definitions from Section 2, any picture consisting of a connected set of unit-length lines in the square grid of the Cartesian plane together with its starting and ending point can be represented by a string over the alphabet $\Pi = \{l, r, u, d\}$. Then, the picture $pic(w)$, for $w \in \Pi^*$, can be viewed as an equivalence class as many pictures can be translated into one another and because a pen may traverse the same line multiple times. Thus, there are infinitely many words that represent a given nonempty picture. For instance, both pictures $pic(uuddr)$ and $pic(rluduuddr)$ are identical as they define the same set of unit lines and they have the same starting and end points.

We say that a regular language L is a *description language* of a picture q if $\forall w \in L : pic(w) = q$. Let $\mathcal{B}_q$ consist of all description languages of q. The regular language $des(q)$ of all descriptions of a picture q belongs to $\mathcal{B}_q$.

For any pictures q_1, q_2, their concatenation $q_1 q_2$ is the picture obtained by joining the end point of q_1 and the starting point of q_2. Now we can define the class of *regular picture sets* $\mathcal{P}$. It is the minimal class of picture sets satisfying the following:

1. The empty set is in $\mathcal{P}$.
2. For each picture q, $\{q\}$ is in $\mathcal{P}$.
3. If $\mathcal{B}_1$ and $\mathcal{B}_2$ are in $\mathcal{P}$, then also
   (a) $\mathcal{B}_1 \cup \mathcal{B}_2$ is in $\mathcal{P}$,
   (b) $\mathcal{B}_1 \cdot \mathcal{B}_2 = \{q_1 q_2 \mid q_1 \in \mathcal{B}_1,\ q_2 \in \mathcal{B}_2\}$ is in $\mathcal{P}$, and
   (c) $\mathcal{B}^* = \bigcup_{i \geq 0} \mathcal{B}^i$ is in $\mathcal{P}$, where $\mathcal{B}^{i+1} = \mathcal{B}^i \cdot \mathcal{B}$.

Brüggemann-Klein et al. [8] considered several grammatical inference problems for regular picture sets. They showed that inferring a description language for a fixed picture q can be as hard as inferring arbitrary regular language. Let

$\mathcal{B}_n$ denote the set of all complete descriptions of pictures of size n, i.e. having n unit line segments. Each reasonable learning algorithm for $\mathcal{B}_n$ requires $\Omega(2^n)$ membership queries in the worst case.

The inductive structure of such $\mathcal{P}$ allows [8] to resent them as regular expressions over the alphabet $\Pi = l, r, u, d$. However, some regular expressions over Π do denote different regular languages of words but the same picture set. When inferring regular pictures sets, membership queries are not sufficient for exact identification of regular picture sets.

Due to the ambiguity of representation of regular picture sets by regular expressions over Π, the authors in [8] conclude that the methods they proposed cannot be applied to them except for a very special case of *staircase pictures*, whose regular sets can be denoted by regular expressions over the alphabet $\{r, u\}$.

The results on inferring regular picture sets carry over to drawn symbolic languages of [11]. A drawn picture is another term for describing unit lines drawn on the Cartesian plane. By adding an alphabet symbol to each point of the picture we get an intuitive and simple extension called *drawn symbolic picture*. It can be easily seen that these picture languages can also describe picture languages as per the definition in Section 2. Similarly to picture sets, they can be represented by strings. Several important results are proven for regular drawn picture languages: The membership problem is NP-complete [35] and the equivalence, containment and ambiguity problems are undecidable [17,18].

With respect to grammatical inference of picture languages, the above regular picture sets and drawn symbolic picture languages share one problem: on given input picture (without its starting and ending points), how to obtain its string description. This problem was not studied yet.

5 Two-Dimensional Automata for Picture Languages

While for string languages we have hierarchies such as the Chomsky hierarchy [10], picture languages lack a similar concept. Moreover, it seems that the taxonomy of languages accepted by various two-dimensional automata is more complicated and so far there has been no complete description of such.

There are known several models of 2D automata [14,6,16,31,32] that generalize the class of regular languages into two dimensions and whose restriction to a single dimension leads to a regular language.

Naturally, for learning of automata it is necessary for us to be able to do all computations effectively. This is hardly achievable for nondeterministic models like online tessellation automaton [14] or sgraffito automaton [31,33] or two-dimensional limited context restarting automaton (2LCRA, for short) [21]. For all these two-dimensional automata models, the membership problem is NP-complete as it is NP-complete already for the REC class accepted by online tessellation automata. The classes of languages accepted by sgrafitto automata and by 2LCRA coincide and they are proper supersets of REC. In contrast to these automata, for the nondeterministic versions of 2LCRA the membership problem is solvable in polynomial time, if the automaton has a

so-called correctness preserving property, i.e. if the input word is from the accepted language, then each rewriting operation leads again to a word of the same language [21]. Unfortunately, to find out whether a given automaton has the correctness preserving property is algorithmically undecidable problem [21].

5.1 Learning Two-Dimensional Automata

There have been many experiments with learning of automata. Typically those were supervised learning methods. For one-dimensional automata there exist interesting algorithms such as L* [2], RPNI [26], EDSM [22], Biermann-Feldman algorithm [7] and more, whose analogies or generalizations into two dimensions are not known.

Each learning algorithm for two-dimensional automata requires to test acceptance by such automata. Hence, deterministic automata with polynomial complexity of computations are necessary. Nondeterminism of computations of two-dimensional automata can be reduced by limiting the freedom of movements of a head within the picture. E.g., a scanning strategy prescribes the sequence of movements of a head within the picture that visits each field of the picture in a tiling automaton [5] and restarting tiling automaton [32].

Determinism can be achieved by other means as well. It is possible to limit the freedom of application of rules – for example, the first plausible rule is used [29] or the automaton is allowed to (deterministically) choose the movement through the picture by itself based on what it finds in the picture [25,28,29].

The author of [20] proposed an alternative approach to simulation via sgraffito automaton with employing a solver for Constraint Satisfaction Problem: A list of all feasible reductions is generated by the algorithm for every position of the picture. Then sets of locally compatible reductions are chosen until either the generated reductions are exhausted, in which case the input picture is rejected, or a set is found that forms a valid computation of 2LCRA, resulting in acceptance of the input picture. This method is not very efficient and in order to obtain results in a reasonable time, several enhancements and dedicated heuristics and data structures need to be used.

6 A New Representation for Picture Languages

Below we propose a new method for learning picture languages based on a new type of representations of picture languages. The new representation of a picture language consists of two parts: a function $R : \Sigma^{*,*} \to (\Sigma \cup \{\#\})^*$ which rewrites any two-dimensional picture over Σ into a string over $\Sigma \cup \{\#\}$ and a (string) language $L \subseteq (\Sigma \cup \{\#\})^*$. Then (R, L)-picture language is the set of all pictures $P \in \Sigma^{*,*}$ such that $R(P)$ is in L. We say that a string language L^{S} R-represents a picture language L^{P}, if for each picture P it holds that $P \in L^{\mathrm{P}}$ if and only if $R(P) \in L^{\mathrm{S}}$.

There are possible many different methods for rewriting a picture into a string. Let us discuss some of them.

Any rectangular picture $P \in \Sigma^{*,*}$ can be represented by storing all symbols (colors of pixels) into a string in the row-by-row fashion. Such simple representation does not contain the information about dimensions of the picture. Therefore, we can add delimiter # for marking ends of rows. E.g., the picture of dimension 3-by-4 containing only letters a will be represented by the string $aaaa\#aaaa\#aaaa$ and the language L_1^{P} of all nonempty pictures filled with b's will be represented by the (string) language $L_1^{\mathrm{S}} = \{(b^n\#)^{m-1}b^n \mid m, n > 0\}$. Evidently, the language L_1^{S} is context-sensitive but not context-free. Not all words over $\{b, \#\}$ represent pictures. We would prefer a representation which allows to represent simple picture languages (like L_1^{P}) by simple (e.g. low in the Chomsky hierarchy) string languages. Fortunately, the sample picture language L_1^{P} can be also represented by the string language defined by the regular expression $(b^+\#)^*b^+$.

The above representation does not allow any simple representation for many elementary picture languages, like the picture language $L_{\mathrm{vl}}^{\mathrm{P}}$ of all pictures containing a column of b's and a's elsewhere. Such language could be represented using a function R which rewrites the picture into a string column-by-column. However instead of allowing different paths during rewriting pictures into strings, we propose to use a window of size 3-by-3. For a picture P, the function R_{w} will scan the extended picture $\widehat{P}$ with the window row-by-row and store the contents of the window (also row-by-row) into a string. E.g., the picture of dimensions 2-by-3 containing a's in the first and last column and b's in the middle column will rewritten into the string

$$\#\#\#\#ab\#ab|\#\#\#abaaba|\#\#\#ba\#ba\#|\#ab\#ab\#\#\#|abaaba\#\#\#|ba\#ba\#\#\#\#$$

(the vertical lines are not present in the string, we have added them in order to separate contents of the window from different positions). Intuitively, such representation can preserve the information about neighboring rows and columns of the picture.

6.1 Experimental Results

Our ultimate goal is to learn target picture language L from given sets S^+ and S^- of positive and negative sample pictures. That is, we know that $S^+ \subseteq L$ and $S^- \subseteq \Sigma^{*,*} \setminus L$. First, we apply the picture-to-string function R_{rw} to all known sample pictures. Then we use $R_{\mathrm{rw}}(S^+)$ and $R_{\mathrm{rw}}(S^-)$ as positive and negative samples to infer an R_{rw}-representation of L using RPNI algorithm.

To experimentally verify the proposed learning protocol, we have generated random positive and negative examples of four picture languages. All our sample picture languages are over the binary alphabet $\{a, b\}$:

- L_1 is the set of all rectangular pictures (not necessarily square) filled with b's and containing a diagonal of a's either from the top-left corner or from the top-right corner. The diagonal leads till the border of the picture.
- L_2 is the set of all pictures which contain several rows filled with a's followed by rows of b' till the bottom of the picture.
- L_3 is the set of all pictures of dimensions at least $(3, 3)$ filled with b's except border filled with a's.

L_4 is the set of all pictures with regular chessboard pattern of a's and b's. The top-left corner of such picture can contain a or b, but the whole picture must have the chessboard pattern.

Sample pictures from each of the sample languages are shown in Table 1.

Table 1. Sample pictures from the picture languages L_1, L_2, L_3 and L_4.

#	#	#	#	#	#
#	a	b	b	b	#
#	b	a	b	b	#
#	b	b	a	b	#
#	b	b	b	a	#
#	#	#	#	#	#

#	#	#	#	#	#
#	a	a	a	a	#
#	a	a	a	a	#
#	a	a	a	a	#
#	b	b	b	b	#
#	#	#	#	#	#

#	#	#	#	#	#
#	a	a	a	a	#
#	a	b	b	a	#
#	a	b	b	a	#
#	a	a	a	a	#
#	#	#	#	#	#

#	#	#	#	#	#
#	a	b	a	b	#
#	b	a	b	a	#
#	a	b	a	b	#
#	b	a	b	a	#
#	#	#	#	#	#

In our experiments we have generated different training and testing sets of positive and negative sample pictures for each of the sample languages. Our sets of sample pictures contained pictures of dimensions between 3 and 8 such that each training set contained approximately the same number of positive and negative samples.

By applying the function R_{rw} with window size 3-by-3 we have rewritten the sample pictures into sample strings. Then for each set S of sample strings, we have learned a deterministic finite state automaton consistent with S. For that we have used an RPNI implementation in MATLAB from [1]. Afterwards, we tested the resulting automaton on an independent test set of pictures (rewritten into strings by the function R_{rw}). The learned finite automata correctly recognize the training set of pictures, however they do not accept/reject correctly all pictures from the test sets.

Below in Table 2 we report accuracy – the ratio of correctly accepted/rejected sample pictures within a test set of the same size as the training set used to infer the corresponding automaton. We have noted also running times for the algorithm RPNI for different sizes of training sets. All experiments were performed on a 64 bit Windows PC with Intel Core i5-4460 processor running at 3.2 GHz.

The absolute value for the running time is not so important as we believe that the running times can be substantially reduced by re-implementing the RPNI algorithm in C or C++ and by using a more recent model of CPU. The interesting information is that the running time of the RPNI algorithm varies for different picture languages considerably. In particular, the language L_2 seems to be "easy" for RPNI compared to other sample languages.

6.2 Possible Extensions

The above representation does not allow any simple representation for many elementary picture languages, like the picture language $L^{\mathrm{P}}_{\mathrm{sq}}$ of all square pictures containing b's. Let us extend the above representation. Instead of collecting all symbols (colors of pixels) of a picture in a fixed order, we will allow "to draw" the picture by traversing in an arbitrary order encoded in the string representation together with symbols of the fields of the picture. For that let $\Pi_0 = \{l, r, u, d\}$ be a fixed alphabet encoding directions of movements to the left, right, up and down. Then a word $w = w_1 w_2 \dots w_n$, where

Table 2. Accuracy and time (in seconds) for learning sample languages for different sizes of training sets.

Lang.	100 examples		200 examples		400 examples		800 examples		1600 examples	
	Acc	Time	Acc	Time	Acc	Time	Acc	Time	Acc	Time
L_1	0.84	35.67	0.84	244.22	0.91	996.12	0.90	1316.04	0.97	964.99
L_2	1.00	1.68	1.00	3.72	1.00	7.67	1.00	11.63	0.99	317.12
L_3	0.86	31.62	0.88	134.52	0.89	396.32	0.95	531.85	0.95	2883.87
L_4	0.83	62.31	0.83	185.98	0.91	413.62	0.91	1532.75	0.91	2175.82

$w_i \in \Pi^*$, where $\Pi = \Pi_0 \cup \Sigma \cup \{\#\}$ for $i = 1, \ldots n$, can be interpreted as follows. The interpretation starts with $i = 1$ and the row-column position $(x_1, y_1) = (1, 1)$ at the top left corner of the picture. If $w_1 = \#$, then $n = 1$ and the word encodes the empty picture. Otherwise, if $w_i \in \Sigma \cup \{\#\}$, then the field at position (x_i, y_i) is "paint" by the symbol w_i and the position for the next step does not change: $(x_{i+1}, y_{i+1}) = (x_i, y_i)$. Let $\eta : \Pi \to \mathbb{Z} \times \mathbb{Z}$ be the mapping $\eta(l) = (0, -1)$, $\eta(r) = (0, 1)$, $\eta(u) = (-1, 0)$, $\eta(d) = (0, 1)$, and $\eta(x) = (0, 0)$ for all $x \in \Sigma \cup \{\#\}$. If $w_i \in \Pi$, then the position changes to $(x_{i+1}, y_{i+1}) = (x_i, y_i) + \eta(w_i)$. Interpreting a word w in this way does not ensure, that each position of the drawn picture is assigned a symbol from Σ. Therefore, we will assume that the fields of the picture which were not visited or visited without assigning a symbol from Σ have a fixed "background" color $\alpha \in \Sigma$. We say that a picture $P \in \Sigma^{*,*}$ of dimension (m, n) was painted by the word $w \in \Pi^*$ if

- starting in the top left field of the rectangular picture $P_0 \in \{\alpha\}^{m,n}$, where α is the fixed background symbol, during interpretation of w on the bordered picture $\widehat{P_0}$, we do not leave the tape where $\widehat{P_0}$ is stored and
- the resulting picture is P.

For example, if a is the background symbol, by interpreting $(brd)^{n-1}br\#dl\#$ we paint the square picture of dimension (n, n), for $n \geq 1$, with diagonal marked by the symbols b and with symbols a elsewhere. Then the string language $(rd)^*u\#dl\#$ represents all "empty" square pictures. On the other hand, single word $brbdblb$ represents all pictures of size (m, n) where $m, n \geq 2$ consisting of symbols a except of small 2-by-2 square containing four symbols b located at the top left corner of the picture.

7 Conclusions

We have introduced a new representation of picture languages which differs from known ones. It is based on a function that rewrites given picture into a string and a string language that is used to decide whether the picture belongs to the picture language or not. Evidently, the complexity of recognizing whether the picture is from the picture language depends mainly on the complexity of the membership problem for the string language.

We experimented with the rewriting function R_{rw} which rewrites the picture by storing the contents of a scanning window of size 3-by-3 when scanning given picture

row by row. The obtained results show that using this function and a classical grammatical inference algorithm RPNI, which is capable of inferring regular languages, we can infer some simple picture languages.

Nevertheless, this paper is only the first step in applying the new representation of picture languages for inferring picture languages. Further theoretical study is needed to establish the power of such representations with respect to known classes of picture languages. By combining different functions for rewriting pictures into strings and different classes of string languages (or automata to recognize them) we can obtain a rich set of picture language classes.

The results of our experiments encourage more thorough study employing other known methods for inferring string languages, like evidence driven state merging algorithms [9].

Acknowledgement. This research was supported by the Charles University Grant Agency (GAUK) project no. 1198519.

References

1. Akram, H.I., De La Higuera, C., Xiao, H., Eckert, C.: Grammatical inference algorithms in matlab. In: International Colloquium on Grammatical Inference. pp. 262–266. Springer (2010)
2. Angluin, D.: Learning regular sets from queries and counterexamples. Information and computation 75(2), 87–106 (1987)
3. Angluin, D.: Queries and concept learning. Machine learning 2(4), 319–342 (1988)
4. Angluin, D.: Negative results for equivalence queries. Machine Learning 5(2), 121–150 (1990)
5. Anselmo, M., Giammarresi, D., Madonia, M.: Tiling automaton: A computational model for recognizable two-dimensional languages. In: International Conference on Implementation and Application of Automata. pp. 290–302. Springer (2007)
6. Anselmo, M., Giammarresi, D., Madonia, M.: A computational model for tiling recognizable two-dimensional languages. Theoretical Computer Science 410(37), 3520–3529 (2009)
7. Biermann, A.W., Feldman, J.A.: On the synthesis of finite-state machines from samples of their behavior. IEEE Trans. Comput. 21(6), 592–597 (Jun 1972)
8. Brüggemann-Klein, A., Fischer, P., Ottmann, T.: Learning picture sets from examples. In: Results and Trends in Theoretical Computer Science, pp. 34–43. Springer (1994)
9. Bugalho, M., Oliveira, A.L.: Inference of regular languages using state merging algorithms with search. Pattern Recognition 38(9), 1457–1467 (2005)
10. Chomsky, N.: Three models for the description of language. IRE Transactions on information theory 2(3), 113–124 (1956)
11. Costagliola, G., Deufemia, V., Ferrucci, F., Gravino, C.: On regular drawn symbolic picture languages. Information and Computation 187(2), 209–245 (2003)
12. Freeman, H.: On the encoding of arbitrary geometric configurations. IRE Transactions on Electronic Computers EC-10(2), 260–268 (June 1961)
13. Giammarresi, D., Restivo, A.: Recognizable picture languages. International Journal of Pattern Recognition and Artificial Intelligence 6(02n03), 241–256 (1992)
14. Giammarresi, D., Restivo, A.: Two-Dimensional Languages, pp. 215–267. Springer Berlin Heidelberg, Berlin, Heidelberg (1997)

15. De la Higuera, C.: Grammatical inference: learning automata and grammars. Cambridge University Press (2010)
16. Inoue, K., Nakamura, A.: Two-dimensional multipass on-line tessellation acceptors. Information and Control 41(3), 305–323 (1979)
17. Kim, C.: Complexity and decidability for restricted classes of picture languages. Theoretical Computer Science 73(3), 295–311 (1990)
18. Kim, C., Sudborough, I.H.: The membership and equivalence problems for picture languages. Theoretical Computer Science 52(3), 177–191 (1987)
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
20. Krtek, L.: Learning picture languages using restarting automata. master thesis, Charles University, Faculty of Mathematics and Physics (2014)
21. Krtek, L., Mráz, F.: Two-dimensional limited context restarting automata. Fundamenta Informaticae 148(3-4), 309–340 (2016)
22. Lang, K.J., Pearlmutter, B.A., Price, R.A.: Results of the abbadingo one dfa learning competition and a new evidence-driven state merging algorithm. In: International Colloquium on Grammatical Inference. pp. 1–12. Springer (1998)
23. Maurer, H.A., Rozenberg, G., Welzl, E.: Using string languages to describe picture languages. Information and Control 54(3), 155–185 (1982)
24. Moore, E.F.: Gedanken-experiments on sequential machines. Automata studies 34, 129–153 (1956)
25. Mráz, F., Otto, F.: Ordered restarting automata for picture languages. In: International Conference on Current Trends in Theory and Practice of Informatics. pp. 431–442. Springer (2014)
26. Oncina, J., Garcia, P.: Identifying regular languages in polynomial time. In: Advances in structural and syntactic pattern recognition, pp. 99–108. World Scientific (1992)
27. Oncina, J., García, P.: Inferring regular languages in polynomial updated time. In: Pattern recognition and image analysis: selected papers from the IVth Spanish Symposium. pp. 49–61. World Scientific (1992)
28. Otto, F., Mráz, F.: Extended two-way ordered restarting automata for picture languages. In: International Conference on Language and Automata Theory and Applications. pp. 541–552. Springer (2014)
29. Otto, F., Mráz, F.: Deterministic ordered restarting automata for picture languages. Acta Informatica 52(7-8), 593–623 (2015)
30. Pradella, M., Crespi Reghizzi, S.: A sat-based parser and completer for pictures specified by tiling. Pattern Recogn. 41(2), 555–566 (Feb 2008)
31. Průša, D., Mráz, F.: Two-dimensional sgraffito automata. In: International Conference on Developments in Language Theory. pp. 251–262. Springer (2012)
32. Průša, D., Mráz, F.: Restarting tiling automata. International Journal of Foundations of Computer Science 24(06), 863–878 (2013)
33. Průša, D., Mráz, F., Otto, F.: Two-dimensional sgraffito automata. RAIRO-Theoretical Informatics and Applications 48(5), 505–539 (2014)
34. Sakakibara, Y.: Learning context-free grammars from structural data in polynomial time. Theoretical Computer Science 76(2-3), 223–242 (1990)
35. Sudborough, I.H., Welzl, E.: Complexity and decidability for chain code picture languages. Theoretical Computer Science 36, 173–202 (1985)

ISSN 1870-4069

On the Use of CSP Semantic Information in SAT Models

Claudia Vasconcellos-Gaete, Vincent Barichard, Frédéric Lardeux

Université d'Angers, Université Bretagne Loire,
Laboratoire d'Étude et de Recherche en Informatique d'Angers (LERIA),
Angers, France
{claudia.vasconcellos, vincent.barichard,
frederic.lardeux}@univ-angers.fr

Abstract. Constraint Satisfaction Problems (CSP) and Propositional Satisfiability Problems (SAT) are two paradigms intended to deal with constraint-based problems. In CSP modeling, it results natural to differentiate between *decision* and *auxiliary* variables. In SAT, instances do not contain any information about the nature of variables; solvers use the Variable Selection heuristic to determine the next decision to make. This article studies the effect of transfer semantic information from a CSP model to its corresponding SAT instance, in order to guide the branching only to variables directly related to the CSP model. The results obtained suggest that this modification can speed up the resolution for some instances.

Keywords: CSP, SAT, decision variables.

1 Introduction

Constraint Satisfaction Problems (CSP) and Propositional Satisfiability Problems (SAT) are two paradigms intended to deal with constraint-based problems.

A CSP problem ($\mathcal{P}$) is defined as a triple $\mathcal{P} = \langle X, D, C\rangle$. This triple contains a set of variables $X = \{x_1, x_2 \ldots, x_n\}$, a set of finite domains $D = \{d_1, d_2 \ldots, d_n\}$ and a set of constraints $C = \{C_1, C_2 \ldots, C_m\}$. A constraint C_j is a relation between the domains $c \subseteq D_1 \times \ldots \times D_n$.

Usually, a CSP model has decision and auxiliary variables. Decision variables represent any value that decision maker needs to determine and they are used by the solver as decision points. Auxiliary variables could be introduced to support the modeling; they are not decision points, but they are invoked during propagation.

A SAT problem ($\mathcal{S}$) is defined as a tuple $\langle X, L, \Phi\rangle$. This triple contains a set of Boolean variables $X = \{x_1, x_2, \ldots, x_n\}$, a set of literals $L = \{l_{1,1}, l_{1,2}, \ldots, l_{2n}\}$ and a Boolean formula in Conjunctive Normal Form (CNF) $\Phi : \{0,1\}^n \to \{0,1\}$. A literal represents a variable (x_i) or its negation ($\neg x_i$). The problem is *satisfiable* only if there exists an assignment of truth values for X in which the formula Φ is true; otherwise, the problem is *unsatisfiable*.

Modeling a problem directly in SAT is a complex task, so there are almost no problems modeled originally as a SAT formula. Also, formats to describe SAT CNF instances (i.e.: DIMACS CNF) are limited describing models; for example, they do not differentiate between decision or auxiliary variables. Typically in a SAT (CNF) instance all variables are equally considered (in their semantics) to be picked as the next variable to branch.

In this article, we study if by adding CSP semantic information to a SAT instance the resolution can improve, basing the analysis in restricting the set of variables with the potential to be picked by the SAT variable selection heuristic. Also, we propose a graph-based tool able to receive a CSP model, produce its corresponding SAT model and trace (in a single graph) all the transformations occurred in the path from CSP to SAT. The results show that focusing branching only in variables directly related to the CSP model, it is possible to reduce the number of decisions made by solver.

This article is structured as follows: Section 2 introduces the framework proposed to address CSP and SAT modeling languages, Section 3 presents a discussion about the branching heuristics in CSP and SAT solvers and how information could be kept when going from CSP to SAT. Section 4 presents the Magic Square problem. Finally, the results obtained are given in Section 5 and the upcoming work is discussed at Section 6.

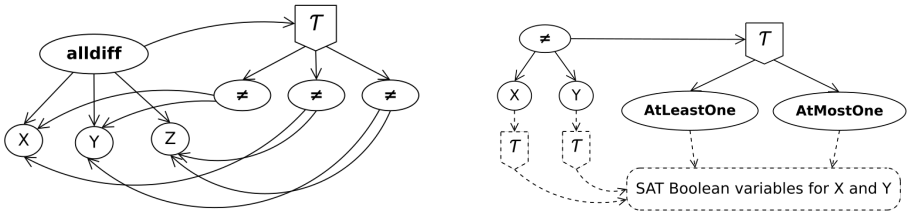
2 Transforming CSP Models into SAT Models

Based on the main elements from CSP (variables, domains and constraints) and SAT (Boolean variables, literals and clauses) we propose a graph able to support both specifications by generalizing them in terms of: *variables*, *relations* and *transformations*. Variables are any symbol representing a value in a finite domain specified by the problem. Relations represents any relationship between one or more variables in the same space (CSP constraint, SAT clause, etc.). Transformations are any function able to produce a set of variables and relations that conserves the original solution (like CSP to SAT encodings).

2.1 The Graph-based Model

We propose an *acyclic labeled graph* $G = (\mathcal{N}, \mathcal{E})$ where the set $\mathcal{N}$ of nodes is composed of three non overlapped sets $\mathcal{N} = \mathcal{V} \bigcup \mathcal{R} \bigcup \mathcal{T}$ with $\mathcal{V} \bigcap \mathcal{R} = \emptyset$, $\mathcal{V} \bigcap \mathcal{T} = \emptyset$ and $\mathcal{R} \bigcap \mathcal{T} = \emptyset$. $\mathcal{V}$ is a subset of nodes labeled with variables names, $\mathcal{R}$ is a subset of nodes labeled with CSP constraints or SAT clauses names, and $\mathcal{T}$ is a subset of nodes labeled with transformation names. The set of edges $\mathcal{E} = \{e_1, \ldots, e_m\}, e_i : n_x \rightarrow n_y \mid n_x, n_y \in \mathcal{N}$ expresses link between two elements. For example, Figure 1 shows the graph representations for transformations `alldiff_to_diff()` and `diff_to_cnf()`.

Using a single graph to maintain the CSP and SAT models allow us to profit from the classical graph operations. For example, a search operation can trace the path followed by a variable or constraint from CSP to SAT and viceversa,



(a) `alldiff_to_diff()` (b) `diff_to_cnf()`

Fig. 1. Graph representation of some CSP/SAT relations and transformations.

determine when a SAT variable was created, identify the relations working with some variables, among others.

Finally, the CSP to SAT encoding chosen produce the SAT model in CNF, facilitating the generation of the corresponding DIMACS CNF output.

2.2 SAT Encoding for Arithmetic Constraints

We extended the idea described in [5] to encode linear arithmetic constraints with digital adders. To be compatible with the machine representation required by them, we use *Log Encoding* [7]. In this encoding, each integer variable v_i is represented with n Boolean variables x_i^k, where $x_i^k = 1$ only if the k-th bit of the domain value assigned to v_i is 1.

To encode a linear constraint $\sum a_i x_i = c$ we consider the schema of a full adder. As this schema only permits to sum two values at the same time, we decompose the linear constraint into a bunch of sums in the form $x + y = z$. Each logical gate correspond to one or more CNF clauses, and the intermediate results (between gates) were modeled with Boolean auxiliary variables.

3 Decision versus Auxiliary Variables

In CSP modeling, a *decision variable* represents values that a decision maker needs to determine. Also, some *auxiliary variables* could be introduced to support the modeling when constraints are difficult to express or, to help the model to propagate better. This difference between variables is used later by the *variable ordering heuristic* to extend nodes in the search tree until all decision variables have been valuated.

Unlike CSP, SAT does not differentiate between decision or auxiliary variables directly in the model. Instead, what some SAT solvers call decision variable is actually any Boolean variable chosen by the branching heuristic during the solving phase. Then, if the branched variables are related (or not) to the decision variables in the problem or if they are only to store intermediate results, it is not concerning to the SAT solver.

In the graph proposed, the semantic information is transferred from CSP to SAT during the encoding of the CSP model; then, the origins of SAT Boolean

variables can be traced to the variable encoding (Log encoding in this case) or as part of the constraint encoding. Moreover, the variable encoding can apply over CSP decision or auxiliary variables (Figure 2).

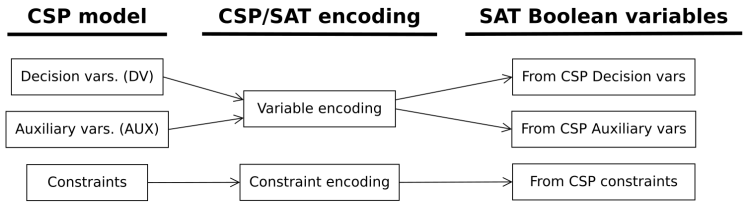


Fig. 2. Origins of SAT Boolean variables.

As our interest is to investigate if by adding CSP semantic information to a SAT instance the resolution can improve, we based the analysis in restricting the set of variables with the potential to be picked by the SAT Variable Selection heuristic. Technically, the graph tags all SAT Boolean variables in the instance as *decision* or *auxiliary* (based on the CSP model given) and then, the DIMACS CNF output generator will sort the variables and add a new parameter to the CNF to indicate the first ID corresponding to the auxiliary variables. The new description in the DIMACS CNF file will be:

```
p cnf nvars nclauses first_auxiliary_var
```

4 The Magic Square

The *Magic Square* (MS) is a mathematical puzzle that consists in finding an assignment of different natural values for a $N \times N$ matrix so that the sum across the rows, columns and diagonals always results in the same number, called the *magic number* [4]. The number N in this puzzle is known as *order*.

In the classical version of the problem, variables are in the domain $[1, N^2]$ and the magic number is calculated as $M = N(N^2 + 1)/2$ (the *open* version of MS uses non-consecutive values in the square, but it is not our interest to consider this case).

The CSP model for Magic Square is defined by the following constraints (Equations 1, 2, 3 and 4):

$$\forall\, i, j \in \{1, N\} \quad \text{alldifferent}(x_{ij}), \tag{1}$$

$$\forall\, i \in \{1, N\} \quad \sum_{j=1}^{N} x_{ij} = M, \tag{2}$$

$$\forall\, j \in \{1, N\} \quad \sum_{i=1}^{N} x_{ij} = M, \tag{3}$$

$$\sum_{i=1}^{N} x_{ii} = \sum_{i=1}^{N} x_{i(N-i+1)} = M, \tag{4}$$

$$x_{ij} \in \{1, N^2\}. \tag{5}$$

5 Experimental Results

All experiments reported in this section use instances of the Magic Square of orders between 4 and 9. They were performed on a machine Intel® Core™ i7-2620M CPU @ 2.70GHz (quad-core, 64 bits) with 8 GB RAM, running Ubuntu 16.04 LTS. Runtimes reported are in seconds, with an upper limit of 2 hours (beyond that time, we will report ∞ values). The solvers used are Gecode (version 5.0.0) [6], and Glucose (version 4.1), a CDCL SAT solver [1].

All instances have been produced with the graph proposed. Variables were encoded in Log encoding and linear constraints are based in the full adder described in subsection 2.2.

5.1 CSP Solving

First, we measure the effort required by a CSP solver to deal with the MS selected instances. For each instance we developed a full CSP model, using Global Constraints and symmetry breaking ("Standard CSP model") and a second model, with no Global constraints and no symmetry breaking ("Decomposed CSP").

The values reported are: runtime (in seconds), the number of propagations and the number of failures occurred during the solving.

Table 1. CSP Results for the Magic Square.

	Instance	Solutions Found	Runtime	Propagations	Failures
Standard CSP	MS4	1	0.014	27152	892
	MS5	1	0.443	2292251	72227
	MS6	1	0.001	1382	27
	MS7	1	2.698	9841603	481301
Decomposed CSP	MS4	1	0.001	2075	14
	MS5	1	0.011	82083	498
	MS6	1	0.847	7866816	47162
	MS7	1	∞	∞	∞

The results in Table 1 show how the difficulty increases when symmetry is allowed in the model (MS7 case), even when the instances have known magic numbers (M) and the domain sizes are limited to the range $[1, N^2]$.

5.2 Structure of CNF Instances Obtained

We describe the structure of the SAT CNF instances produced (in terms of variables and clauses) and observe the changes that a typical SAT minimization step (preprocessing) can do over them.

Structure of Clauses pre/post Minimization Based on Glucose preprocessor (option `-dimacs`), we compare how the structure of the SAT CNF instance changes before/after the minimization (*Raw* versus *Minimized* CNF). Values reported are number of variables, number of clauses and distribution of clauses by arity. The "raw" instances are the CNF obtained using the proposed graph.

Table 2. Structure of SAT CNF instances produced.

	Instance	vars	clauses	CNF clause arity									
				10	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
Raw CNF	MS4	1773	6829	-	-	-	20	376	-	120	5010	1260	43
	MS5	3674	14008	-	-	24	12	475	-	192	11004	2232	69
	MS6	7484	29528	-	28	28	1638	-	-	280	23842	3612	100
	MS7	12728	49825	16	48	16	1911	-	-	384	42032	5280	138
Minimized CNF	MS4	675	3688	-	-	603	10	178	16	63	2520	278	-
	MS5	1378	7984	-	-	2200	29	158	25	-	5154	398	-
	MS6	3089	17794	-	4744	13	277	36	-	240	12235	249	-
	MS7	5390	31214	-	9096	71	292	49	-	60	21120	526	-

Results in Table 2 show that minimized instances reduced -in average- a 60.2% of variables and a 41.5% of clauses. Regarding binary clauses, they do not represent more than a 4% in the minimized instances, which could affect directly in the searching step.

Another effect of minimization techniques, like *Strengthening* [1] [3], is the generation of clauses longer than the longest clause in the raw instance (i.e.: the raw MS4 instance has clauses of length 7, while the minimized instance has clauses of length 8 and 9).

Distribution of variables pre/post Minimization In Section 3, we explained that SAT variables obtained after encoding a CSP instance can be linked to three sources: to CSP decision variables, to CSP auxiliary variables and, as result of the encoding of constraints. For the Magic Square instances, the origins of variables are:

- Decision variables, from the CSP decision variables (CSP_{DV}).
- Auxiliary variables, from the CSP auxiliary variables (CSP_{AUX}).
- Auxiliary variables, from the encoding of the `alldiff()` constraints (*DIFF*).

[1] If there exist two clauses $C_1 = \{l, D\}$ and $C_2 = \{\neg l, E, D\}$, then the clause C_2 can be replaced by the resolvent $C_1 \otimes C_2 = \{D, E\}$.

– Auxiliary variables, from the encoding of full adders (*ADDER*).

Table 3 reports the total number of variables and the distribution (in percentage) of the variables by their origins. The results obtained show that preprocessing does not affect the distribution of variables but, as long as the instances grow, the percentage of auxiliary variables coming from constraint encodings increases.

Table 3. Distribution of variables pre/post minimization.

	Instance	vars.	Decision(%)	Auxiliary(%)		
			CSP_{DV}	CSP_{AUX}	DIFF	ADDER
Raw CNF	MS4	1773	12.9	0.8	45.7	40.6
	MS5	3674	11.2	0.6	39.2	49.0
	MS6	7484	9.4	0.5	31.2	59.0
	MS7	12728	8.1	0.4	26.9	64.7
Minimized CNF	MS4	675	11.1	0.7	45.7	40.6
	MS5	1372	9.7	0.3	37.1	53.0
	MS6	3092	8.2	0.4	31.2	60.1
	MS7	5381	7.1	0.4	25.7	67.6

5.3 Analysis of the Branching Behavior of a SAT Solver

This analysis focuses in the decisions made by the SAT solver, based on the origin of variables previously stated. Out interest is to determine if a smaller set of variables used available for branching can improve the solving process in SAT.

Origin of SAT Branched Variables By combining the tracing capabilities of our graph-based tool and some modifications in the Glucose solver to visualize each new branch, we determine the origins for each variable in the MS instances, allowing us to quantify which type of variable was branched the most.

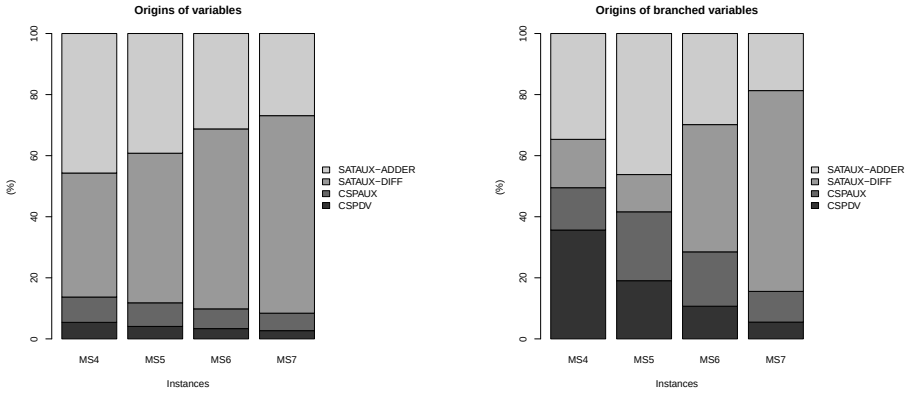
Table 4 shows the distribution by origin for all variables and for the subset of branched variables. To support the comparison, we present the same results expressed as percentages in Figures 3(a) and 3(b).

From the results obtained, we observe that even if the proportion of decision variables (CSP_{DV}) is low compared to auxiliary variables (CSP_{AUX}+ DIFF + ADDER), there is a big number of variables coming from CSP_{DV} which are branched. In average, a 82.5% of variables related to CSP_{DV} are considered during branching, making them the most branched category. For the other categories, results are 54.8% (CSP_{AUX}), 17.7% (DIFF) and 21.2% (ADDER).

Results show that variables directly linked to CSP decision variables, as they are (usually) part of multiple constraints, appear in several clauses and provide more information. In contrast, variables generated due to constraint encodings tend to appear less, their number is determined by the modeling thus, their branching utilization depends only of the CSP model provided.

Table 4. Origins of SAT variables.

	Instance	Total	Decision	Auxiliary		
			CSP_{DV}	CSP_{AUX}	DIFF	ADDER
All Variables	MS4	1773	96	147	720	810
	MS5	3674	150	284	1800	1440
	MS6	7484	252	484	4410	2338
	MS7	12728	343	729	8232	3424
Branched Vars.	MS4	202	72	28	32	70
	MS5	656	125	148	80	303
	MS6	2028	217	361	845	605
	MS7	5349	294	538	3517	1000



(a) Variables (b) Branched variables

Fig. 3. Distribution (%) of variables in CNF instances by origin.

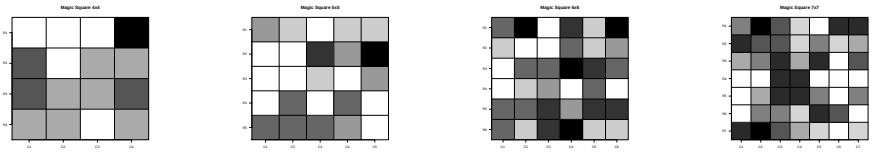
The "Solution Branch" (the set of assignments that lead to the resolution of the instances), has the same behavior. The most frequently picked variables are the ones related to CSP_{DV}, while all others are much less considered. Table 5 describes them in terms of depth (how long the branch is), and variable distribution per category. The parameter *max branches* indicates the maximum number of times that a CSP_{DV} was indirectly branched (a CSP variable is encoded in multiple SAT variables so, technically, the same CSP variable can be branched several times).

Figure 4 shows the projection of the Solution Branches in their corresponding MS grids. The colors represent how many times a CSPDV was indirectly branched; the darker the color, the higher the frequency.

These results reinforces the idea that CSP_{DV} is the most branched category so, it worths to prioritize them. Particularly for the MS, branching tends to pick variables related to the sum involving the magic number M. For example, in the grid for MS4 (Figure 4(a)), the most branched variables are linked to the linear constraint for the first row; the same for MS5 (Figure 4(b)) where the linear

Table 5. Description of Solution Branches.

Instance depth		Distribution			Max branches
	CSP_{DV}	CSP_{AUX}	DIFF	ADDER	
MS4	22	20	2	-	-	3
MS5	49	36	4	-	9	5
MS6	91	84	3	-	4	5
MS7	154	128	7	2	17	6



(a) MS4 (b) MS5 (c) MS6 (d) MS7

Fig. 4. Projection of branched variables.

constraint for the second row is the most branched. For MS6 and MS7 (Figures 4(c) and 4(d)), the most branched variables are linked to the last additions of some rows and columns. This is an interesting fact, as M is a constant and, consequently, the only which makes possible the Unit Propagation.

Sorted versus Shuffled Instances Our graph encodes to SAT following the order initially given to CSP constraints, making that CSP_{DV} variables will appear first in the CNF file. We analyze this, because studies like [2] suggests that SAT solvers are sensitive to the CNF file organization.

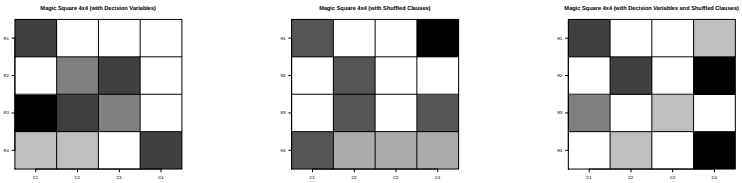
We analyze 4 versions derived from the MS4 instance: the classical one (MS4), with clauses shuffled (MS4s), branching only on CSP_{DV} variables (MS4dv) and, with clauses shuffled and branching only on CSP_{DV} variables (MS4dvs). Table 6 and Figure 5 show the results for the MS4 Solution Branches.

Table 6. Solution Branches for MS4 versions.

Instance	depth	Distribution			
		CSP_{DV}	CSP_{AUX}	DIFF	ADDER
MS4	22	20	2	-	-
MS4dv	22	22	-	-	-
MS4s	20	16	2	-	2
MS4dvs	19	19	-	-	-

The results show again the predominance of SAT_{DV} variables in the branch. The size of branches remains almost the same in the four cases, so nothing can be concluded from that.

But, observing the cases limited to branch over decision variables (MS4dv and MS4dvs), the grids projected looks more sparse than in the cases where all variables are candidate for branching. The cause for this could be the presence of the *CDCL learning processes*. The work by adding new clauses derived from the knowledge obtained after branch and propagation. It may seems, that by reducing the variables set, we are also reducing the learning speed of the solver.



(a) MS4dv (b) MS4s (c) MS4dvs

Fig. 5. Projection of branched variables for MS4.

5.4 SAT Solving for Different Branching Sets

Finally, we present the SAT solving results for different branching sets. We added two new instances (MS8 and MS9) to observe better the changes produced by the different sets.

The values reported are the runtime (in seconds), the number of propagations and the number of decision points; all of them are provided directly by the solver. The runtime has an upper limit of 2 hours (if any instance goes beyond that time, we will report ∞ values). In Figure 6, runtimes are log-scaled in order to handle the huge differences between instances.

Table 7 shows that number of decision is considerably less for the first three instances (MS 4,5,6) when branching uses all variables coming from CSP (Branching on SAT_{CSP}) and, it is the only group where MS8 and MS9 instances were solved in the time frame given. The same is observed in Figure 6.

Regarding the second group of values (Branching on SAT_{DV}), we observe a fast increment in the number of Decisions. This seems to agree with the discussion about branching and learning presented in Subsection 5.3. Moreover, if we consider that CSP_{AUX} variables belong to the magic number, then, we conclude that avoid SAT variables linked to CSP auxiliary variables reduce the chances for the Unit Propagation.

A comparison between Tables 1 and 7 show that SAT solver outperforms the resolution of the decomposed CSP model from MS7 instance onwards. In the "Standard" cases, both solvers perform similar on the small instances (MS4, MS5 and MS6), but for MS7, the CSP outperforms the SAT solver.

Table 7. SAT Results for the Magic Square.

	Instance	Runtime	Propagations	Decisions
Standard model	MS4	0.03	92944	1927
	MS5	0.11	384221	8272
	MS6	0.54	2034502	37941
	MS7	78.90	155917224	2551515
	MS8	156.83	537569994	7602086
	MS9	∞	∞	∞
Branching on SAT$_{DV}$	MS4	0.08	142316	2394
	MS5	0.12	277330	5110
	MS6	1.27	7745848	124110
	MS7	76.14	245845662	3904006
	MS8	∞	∞	∞
	MS9	∞	∞	∞
Branching on SAT$_{CSP}$	MS4	0.04	29814	536
	MS5	0.12	218608	3638
	MS6	1.22	6489299	111836
	MS7	74.99	224437398	3653155
	MS8	466.68	1176044383	16278155
	MS9	4828.31	8677420297	125249474

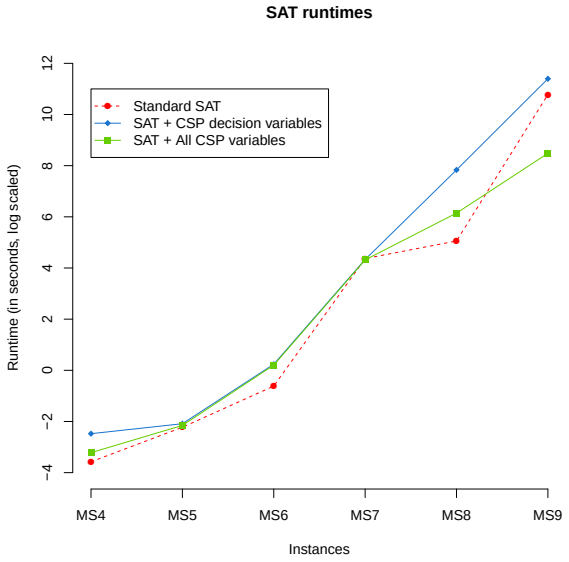


Fig. 6. SAT runtime (log scaled).

6 Conclusions and Future Work

This article analyzes the behavior of the branching heuristics of a SAT CDCL solver when CSP semantic information about their variables is added to SAT models. From the results we conclude that SAT solver intuitively branches on

variables directly linked to CSP model, despite the fact that these variables are a minority in the universe of SAT variables for a single instance. To achieve this analysis, we also propose a graph-based tool able to keep tracking of the transformation between CSP and SAT models. With it, we traced successfully the CSP origin of each SAT variable in the problem instances, giving us a CSP point of view over the SAT resolution.

Is clear that studying the transformations between CSP and SAT models opens a lot of possibilities to understand better their interactions and to improve the current techniques involved into modeling and solving. The upcoming work goes in the line of investigate if more, and which kind, of CSP semantic information could be added to a SAT model and, if it is possible to replicate a CSP-like propagation in SAT. We expect that our findings can contribute to expand the knowledge in this area.

References

1. Audemard, G., Simon, L.: Glucose, www.labri.fr/perso/lsimon/glucose/
2. Audemard, G., Simon, L.: Experimenting with small changes in conflict-driven clause learning algorithms. In: Principles and Practice of Constraint Programming. pp. 630–634 (2008)
3. Balint, A., Manthey, N.: Boosting the performance of SLS and CDCL solvers by preprocessor tuning. In: POS@ SAT. pp. 1–14 (2013)
4. Derksen, H., Eggermont, C., Van Den Essen, A.: Multimagic squares. American Mathematical Monthly 114(8), 703–713 (2007)
5. Eén, N., Sorensson, N.: Translating pseudo-boolean constraints into SAT. Journal on Satisfiability, Boolean Modeling and Computation 2, 1–26 (2006)
6. Schulte, C., Lagerkvist, M., Tack, G.: Gecode, www.gecode.org
7. Walsh, T.: SAT vs CSP. In: Principles and Practice of Constraint Programming – CP 2000: 6th International Conference. pp. 441–456 (2000)

ISSN 1870-4069

Design of a Brain-Computer System to Measure Brain Activity during a Dolphin-Assisted Therapy using the TGAM1 EEG Sensor

Jaime Moreno[1], Oswaldo Morales[1], Liliana Chanona[1], Ricardo Tejeida[2], Pedro Flores[3], Víctor Calderón[4]

[1] Instituto Politécnico Nacional, ESIME, Zacatenco, Mexico
[2] Instituto Politécnico Nacional, EST, Mexico
[3] Secretaría de Salud, Mexico
[4] Delfiniti, S.A. de C.V. Ixtapa, Mexico
jemoreno@esimez.mx

Abstract. The recent proliferation of sensors technology applications in therapies to children disabilities to promote positive behavior among such children has produced optimistic results in developing a variety of skills and abilities in them. Dolphin-Assisted Therapy (DAT) has also become a topic of public and research interest for these disorders intervention and treatment. This work exposes the development of a system that controls brain-computer interaction when a patient with different abilities takes a DAT. The study was carried out at Definiti Ixtapa facilities and shows that brain activity increases by 376% during a DAT. A TAGM1 sensor was used to develop the system, which is connected to the Bluetooth 4.0 communication protocol, which is isolated from environmental conditions, which is brackish and humid. In this way, we explore the behavior of Obsessive Compulsive Disorder and neurotypic children using Fast Fourier Transform (FFT) from Electroencephalogram (EEG).

Keywords: artificial intelligence, EEG, FFT, TGAM1, dolphin-assisted therapy, BCI.

1 Introduction

Dolphins have had a therapeutic effect in the treatment of a number of human race conditions, both psychological and physical. Underwater communication sounds made by the dolphins play a part in this therapeutic effect. Dolphins appear to sense electrical fields from humans and attempt to communicate using the same frequencies. This has generated several aspects: (a) dolphin acoustic emissions, recorded in sea water, may bring about modifications in human brain-wave activity, (b) dolphin interaction may give patients pain relief due to the increased release of hormones into the blood, (c) dolphin interaction may produce complex neurological, (d) stimulation which helps relaxation, helps reduce stress levels and thereby strengthens the immune system, and (e) dolphin's ultrasonic

energy may cause significant cellular changes within the living tissue of the central nervous system [3].

Based on the above aspects, there has been developed an alternate treatment for people with diverse psychological and physical disabilities or disorders like Autism, Attention Deficit, Down Syndrome, Infantile Cerebral Palsy, and Obsessive Compulsive Disorder. This treatment is called Dolphin-Assisted Therapy (DAT). DAT is a growing area of interest to the general public, with reports of pain relief, extinction of depression and improved learning in children with different disabilities. [2]. DAT aims to improve functioning through complementing and reinforcing existing therapies, rather than replacing them [3].

In case of Obsessive Compulsive Disorder (OCD), this is characterized by intrusive, troubling thoughts (obsessions), and repetitive, ritualistic behaviors (compulsions) which are time consuming, significantly impair functioning and/or cause distress. When an obsession occurs, it almost always corresponds with a massive increase in anxiety and distress. Common obsessions include contamination fears, worries about harm to self or others, the need for symmetry, exactness and order, religious/moralistic concerns, forbidden thoughts (e.g., sexual or aggressive), or a need to seek reassurance or confess. Common compulsions include: cleaning/washing, checking, counting, repeating, straightening, routinized behaviors, confessing, praying, seeking reassurance, touching, tapping or rubbing, and avoidance. Unlike in adults, children need not view their symptoms as nonsensical to meet diagnostic criteria [6].

Younger children will not be able to recognize that their obsessions and compulsions are both unnecessary (e.g., you don't really need to wash your hands) and extreme (e.g., washing hands for 15-20 times is fine, but 5 min in scalding water is too much) in nature. In young children, compulsions often occur without the patient being able to report their obsessions, while adolescents are often able to report multiple obsessions and compulsions. Children and adolescents are also more likely to include family members in their rituals and can be highly demanding of adherence to rituals and rules, leading to disruptive and oppositional behavior and even episodes of rage [6].

A way the estimate the efficiency of DAT in children with different disabilities such as OCD is to observe changes in the children' brainwave activity from Electroencephalogram signals recorded before and during a session of DAT.

The electroencephalography (EEG) recorded from scalp give rise to the study of human cognitive activity through the measure of neurons' electrical impulses. At the beginning, the EEG data were often collected with the use of multiple electrodes and/or wired EEG headset to retain signal accuracy and spatial resolution, which greatly diminishes the mobility and actual application pragmatism. To facilitate a mobile driver safety program, [8] demonstrated the effectiveness of wireless single-channel (electrode) EEG headset NeuroSky's MindWave in distinguish user's eyes conditions – open or closed.

2 Related Work

For a System Artificial Intelligence is defined, it must include a model of the physical environment in which it is to operate [12]. In our case, we develop a model that measures the brain activity (physical feature) and the meaning of its behavioural changes. Thus, this proposal can be considered as an Artificial Intelligence model since it incorporates the knowledge of the domain of the application (in this case, in Neuroscience) to be able to measure the efficiency of the DAT from the study of biosignals.

For the study of these biosignals, it is necessary to use a Brain Computer Interface (BCI). In this way, BCI research programs have arisen due to more understanding of brain functions and powerful low cost computer equipment. Under BCI, the users can communicate with environments and the external devices through their brain activities. In particular, people suffering from disabilities like Obsessive Compulsive Disorder (OCD) have growing number of demands and needs of such system. The goal of these BCI systems are to translate the neuro-physiological signals for controlling the external devices. For acquiring the brain dynamics, electroencephalography (EEG) is relatively convenient, comfortable, and inexpensive [10].

To get the rid out of economical loss, but still taking in account the improvement in the reliability, accuracy, latency and false alarm rate, [4] have used the TAGM1 sensor as a probe to detect the brain wave pattern of the user and designed filter and threshold algorithm to detect driver's fatigues and drowsiness index with low cost and high reliability.

The aims of this work is trying to create a brain-computer system based on EEG. In this paper is developed a system to analyze and measure the both Obsessive Compulsive Disorder and neurotypic children brain activity during a DAT using the TGAM1 EEG signal sensor. Moreover, we measure, in a novel way, the Power Spectrum Density from EEG by means of BCI system, in order to estimate really changes in the children' brainwave activity recorded before, during and after a session of DAT. This paper is divided as follows: Section 3 introduces the system and architecture of BCI communication system; Section 4 describes the way we measure the brain activity before and during DAT; Section 5 shows the results and discussion and Section 6 is the conclusion.

3 Theoretical Framework

3.1 Electroencephalography

The brain is made up of hundreds of thousands of cells called neurons, which interact with each other in a bioelectrical phenomenon called synapses; By means of impulses that are transmitted around the membrane, information is sent that helps the brain to coordinate sensory, motor and cognitive functions. The activity of the present electric field can be manifested by a lapse of tens to hundreds of milliseconds, which is sufficient for equipment such as electroencephalographs,

which use non-invasive electrodes on the cranial surface, to detect and record such activity.

Electroencephalography (EEG) studies, compared with other imaging techniques or behavioral observations, have advantages because of their excellent time resolution since several records can be taken in a second through multiple sensors. Other studies, such as MagnetoEncephaloGraphy (MEG), record the electric field activity generated by the neurons, have excellent resolution in time, but require large, stationary and expensive detectors, in addition to their maintenance and special training to operate such equipment. Functional Magnetic Resonance Imaging (fMRI) measures the change in blood flow associated with neuronal activity, since oxygenated blood, in comparison to the rest, causes magnetic distortions that are generated by the protons present in it, therefore fMRI has an excellent spatial resolution. Since human neuronal activity has a high degree of complexity in its nature, neuronal oscillations can be measured as a mixture of several underlying base frequencies, which reflect certain cognitive, attentional and affective states. These oscillations were defined in specific frequency ranges: delta, theta, alpha, beta and gamma.

- Delta (0.5-4 Hz): This type of wave has been examined during periods of deep sleep, locating its point of greater power in the right hemisphere of the brain. Since sleep is associated with the consolidation of memory, it plays a central role in learning functions. Its amplitude varies from 20 - 200 μV.
- Theta (4-8 Hz): Some studies report that the activity of this wave is related to cognitive activities such as: selective attention, assimilation of information, processing, learning and working memory. You can have a better record from the prefrontal, central, parietal and temporal zones. Its amplitude varies from 20 - 100 μV.
- This oscillatory rhythmic activity is involved when there is a response to sensory stimuli, motor and memory functions. High levels of alpha can be noted when there is relaxation with closed eyes, the suppression of alpha implies a characteristic signature of mental states of interest, for example, when there is sustained attention to any type of stimulus at a particular time. Its amplitude varies from 20 - 60 μV and the source of activation of these signals is in the occipital region of the brain
- Beta (12-25 Hz: This frequency is generated both in the occipital and frontal regions, it is associated with the states of active thinking, motor functions, visual and spatial coordination, anxiety and concentration. Its amplitude varies from 2 - 20 μV.
- Gamma (more than 25 Hz): They tend to have the highest frequency and the lowest amplitude. So far there is no strong association of what this type of waves reflect in the brain, there are some researchers who suggest being the reflection of the interconnection of several sensory responses to an object in a coherent way, therefore, represent a process of attention due to intense brain activities. On the other hand, others think that it is the effect of neural processes to achieve vision control, and, therefore, it would not represent a cognitive process after all.

There is discussion to date about a cerebral rhythm called mu (μ), its frequency band covers practically the same as alpha, however, it is only observed in the sensory-motor cortex.

3.2 Electroencephalography Sensor: TGAM1

MindWave Mobile by NeuroSky (Figure 1 (b)) is Brain Computer Interface (BCI) in a headband that captures the EGG waves (Figure 1 (a)) and the eye blink of an user[7]. This device is an interesting for developers because they can program powerfull algorithms with an interface to connect to mobile devices easily and be able to use it in research applications [11].

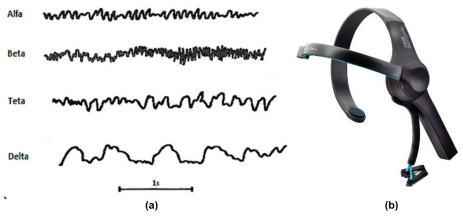


Fig. 1. (a) Types of EEG waves and (b) NeuroSky Mindwave Mobile.

From Figure 2, NeuroSky Mindwave Mobile can be divided in four parts:

1. Dry Electrode (Figure 4(b)),
2. Reference and Ground Electrodes (Figure 4(a)),
3. EEG biosensor TGAM1 (Figure 2(c)), and
4. Communication Module (Figure 2(c)).

NeuroSky's EEG biosensor digitizes and amplifies raw analog brain signals to deliver concise inputs the most important features of this biosensor are the following:

- Direct connect to dry electrode since it does not need a special solution or gel,
- One EEG channel, Reference, and Ground,
- Extremely low-level signal detection,
- Advanced filter with high noise immunity, and
- RAW EEG at 512Hz.

The code that may appear in the ThinkGear packets are listed in the Table 1.

TGAM1 has configuration pads that can be used to change two default settings that are applied at chip power up. The configuration pads are located on

(a) Dry Electrode. (b) Reference and Ground Electrodes. (c) EEG biosensor TGAM1 with Communication Module.

Fig. 2. NeuroSky Mindwave Mobile system main parts.

Table 1. ThinkGear CODE.

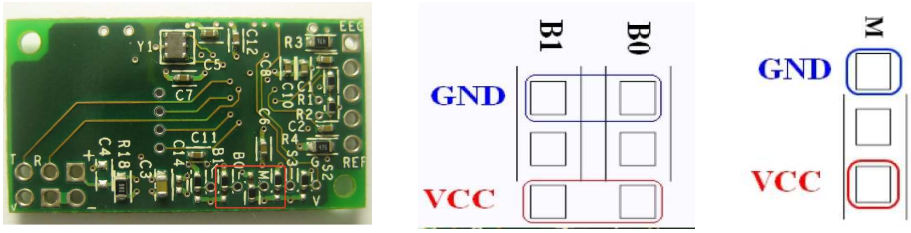
Code	Length	Value	Default Setting
0x02	N/A	Poor Quality (0-200)	On
0x04	N/A	eSense Attention (0-100)	On
0x05	N/A	eSense Meditation (0-100)	On
0x80	2	10-bit Raw EEG	Off
0x83	24	EEG Powers (integer)	On

the backside of the TGAM1, as indicated by the red square in Figure 3(a). From Table 2, the BR0 and BR1 pads configure the output baud rate and data content, after the TGAM1 powers up. The M pad configure the notch filter frequency. Normal Output mode includes the following output: poor quality value, EEG value, Attention value and Meditation value.

A magnified picture of the B1 and B0 pads are shown in Figure 3(b). The first row of pads are GND and third row of pads are VCC. The TGAM1 output baud rate and data content after power up behavior depends on the pad setting as described in table above. For example, the stuff option in the module in Figure 3(a) has both BR1 and BR0 tie to GND pads for a 9600 baud with Normal Output Mode.

The baud rate can also be configured after the module is powered up by sending commands through the UART interface. The commands are listed in the Table 3. When the module is reset, the baud rate setting will revert back to the default set by BR0 and BR1.

TGAM1's notch filter frequency can be configured with the M configuration



(a) Main Pads (b) B0 and B1. (c) M.

Fig. 3. Configuration pads in TGAM1 chip.

Table 2. Configuration pads.

BR1	BR0	Function
GND	GND	9600 Baud with Normal Output Mode
GND	VCC	1200 Baud with Normal Output Mode
VCC	GND	57.6k Baud with Normal + Raw Output Mode
VCC	VCC	N/A

Table 3. Baud rate configuration.

Command	Function
0x00	9600 Baud with Normal Output Mode
0x01	1200 Baud with Normal Output Mode
0x02	57.6k Baud with Normal + Raw Output Mode

pads. It is used to select either 50Hz or 60Hz to reduce the AC noise specific to a targeted market. As indicated in Figure 3(c), the top pad is GND and bottom pad is VCC. Tie the M pad to VCC pad to select 60Hz, and to GND pad to select 50Hz notch filtering frequency.

Unlike the BR0, BR1 configuration, there is no equivalent software configuration for the M configuration. The most common stuff option for these configuration pads are illustrated in Figure 3(a), configuring the TGAM1 for 9600 Baud, normal output and 60Hz notch filtering frequency.

3.3 Filtering Biosignals using Fast Fourier Transform

The Fast Fourier Transform (FFT) algorithm is based on the method known as successive bending. It is defined by Equation 1 as:

$$F_u = \frac{1}{2 \cdot M} \left[\sum_{x=0}^{N-1} \left(a_x \cdot w\left(u, x\right) \right) \right], \tag{1}$$

where $w\,(u,x)$ can be defined by Equation 2 as following:

$$w(u,x) = e^{(-j\cdot 2\cdot \frac{u\cdot \pi\cdot x}{2})}. \tag{2}$$

The result of the transformation is represented in a vector called F of subscript u. The value N is the amount of element of the sample, M is half of N. The values of the sample are loaded in the vector called a_x and x represents the subscript. Equations 3 and 4 are obtained from Equation 1:

$$Feven_u = \frac{1}{M}\cdot\left[\sum_{x=0}^{M-1}\left(a_{2\cdot x}\cdot w_1\,(u,x)\right)\right], \tag{3}$$

$$Fodd_u = \frac{1}{M}\cdot\left[\sum_{x=0}^{M-1}\left(a_{2\cdot(x+1)}\cdot w_1\,(u,x)\right)\right]. \tag{4}$$

All the even subscripts of x belonging to a_x are grouped in the vector $Feven$ and the odd ones in the vector $Fodd$. In this case $w_1\,(u,x)$ is defined by Equation 5 as:

$$w_1\,(u,x) = e^{\left(-j\cdot 2\cdot \frac{u\cdot u\cdot x}{\frac{N}{2}}\right)}. \tag{5}$$

Frequently, the relevant information of a signal has a characteristic wave that is known in a general way, for this it is possible to obtain time series containing this information, later study measures of relations between signals in both the time domain and the frequency (correlation or crossed spectrum) and more complex transformations for the extraction of characteristics.

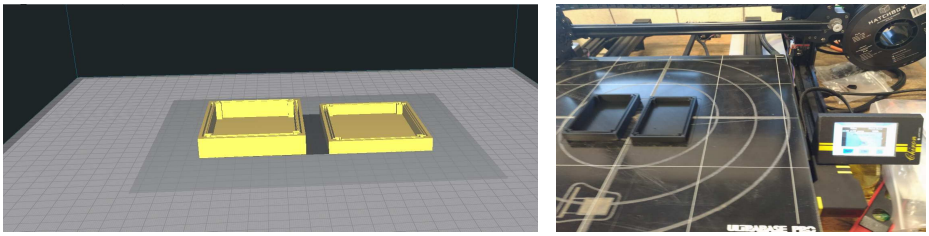
In the case of EEG signal analysis, considering the time plotted on the X axis and the voltage on the Y axis; The Fast Fourier Transform transforms the signal from the time domain to the frequency. Basically what is done during this transformation is to examine the raw data of the signal how much can be approximated to sine waves consisting of pure frequencies, that is, free of redundancy; the more they adjust, the greater their correlation. Since the frequencies of the human brain associated with affective and cognitive activities are in the range of 0.5-60 Hz, an analysis by the FFT gives a lot of relevant information, for example, if a person is in a general state of concentration (theta band) or the response of their neuronal activity describes a dream state (delta band).

To implement the FFT in the microcontroller it is necessary to implement $Feven$ and $Fodd$. We used MatLab 2018b to implement both the programming of the embedded devices and the analysis of these signals.

4 Measuring Brain Activity during a Dolphin-assisted Therapy: DAT-TGAM1 System

4.1 Designing of the Waterproof Case

The original BCI adaptation consists of keeping the headband but introducing the TGAM1 sensor in a waterproof case, Figure 4 shows the Modeling and Printing of this case.



(a) 3D Modeling and Designing. (b) 3D Printing.

Fig. 4. Designing of the waterproof case where TGAM1 Microcontroller is located.

An important part of the choice of material for 3D printing are the parameters of the water quality of confinements in closed facilities must meet the following characteristics:

- Salinity: 18 to 36 parts per thousand.
- Hydrogen Potential (pH): between 6 and 8 units.
- Temperature: from 5 to 27°C.
- Pressure: 1 Atmosphere (ATM).

It is important to mention, even this case supports at least one atmosphere of pressure, that is to say, that it does not leak at 10 meters of depth, it was designed for floating.

4.2 Processing EEG Signals

For this work, we explore the behavior of two children (control and intervention patients) using the quantitative mathematical tool such as Fast Fourier Transform (FFT) from ElectroEncephaloGram (EEG) signals[1]. The EEG RAW data are time series that showed the cerebral brain activity, voltage versus time, at rest or before DAT (Figure 5(a)), during a DAT (Figure 5(b)), and after DAT (Figure 5(c)[9]). As recorded by the first frontopolar electrode (F_{P1}) by means of a EEG biosensor TGAM1 Module.

For analyzing EEG data we decomposed the signal into functionally distinct frequency bands: δ (0.5 - 4 Hz), θ (4 - 8 Hz), α (8 - 12 Hz), β (12 - 30 Hz), γ

(a) Before DAT (b) During DAT (c) After DAT.

Fig. 5. Method of obtaining EEG RAW samples.

(30 - 60 Hz), and All Bands (0.5 - 60 Hz) through FFT, in order to obtain an estimate of the Power Spectral Density (PSD), expressed in $\mu Volts/Hz$[5].

5 Experimental Results

5.1 Initial Conditions

The main goal of this proposal is to test our BCI when a patient who has a problem in neurodevelopment, we have chosen a patient with Obsessive Compulsive Disorder (OCD), who is taking a Dolphin Assisted Therapy (DAT). The whole system is subdivided into three fundamental parts:

1. Female dolphin of the bottle nose species
2. Patient with OCD.
3. DAT-TGAM1 system.

All three subsystems interact in order to determine if DAT is effective for OCD patients

5.2 Results

Figure 6 shows the results of our experiments for measuring the efficiency of DAT for RAW EEG Brain activity (first row) of the control and intervention child patients, first and second columns respectively, using FFT - Power Spectrum Density (second and third rows) Before DAT, During DAT and After DAT. In this image we can realize that there are meaningful changes in behavior experienced by these two-treated children. Figures 6(a) and 6(b) show that there is a higher brain activity during DAT than Before or After, namely, FFT analysis points out an average great increment of 376% in the average power spectral density of data in both patients during the DAT respect to before it, which yields the increment on every spectral band, Figures 6(c) and 6(d). While Figures 6(e) and 6(f) show the entire power spectral density from 0 to 256 Hz (half of δ_t) and it can be noticed that the overall power is higher in all cases when a DAT is developed.

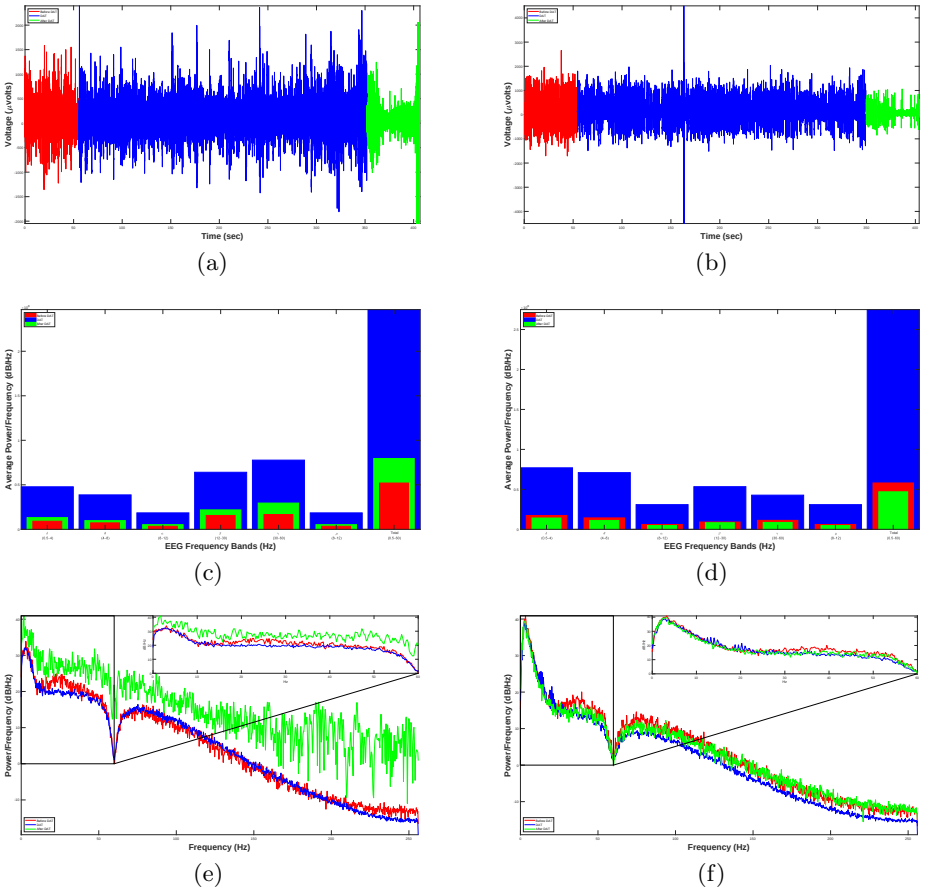


Fig. 6. Results of efficiency of DAT for RAW EEG Brain activity (first row) of the control and intervention child patients, first and second columns respectively, using FFT - Power Spectrum Density (second and third rows) Before DAT, During DAT and After DAT.

5.3 Discussion

Table 4 is the summary of the experimental results. When we compare Average Power Spectrum Density (PSD) During DAT regarding before it, we can realize that the brain of the patients increment its activity in approximately 376% in both cases. After the DAT, the brain activity of the control patient was reduced in 67.92%, while the intervention patient reduces its activity in 83.02%, namely, the patient with OCD gets an apparent relaxation since its PSD decreases 20% regarding its initial rest state meanwhile the control patient increases its PSD 53% regarding its initial rest state.

Table 4. Summary of the experimental results.

	Power Spectrum Density		
	BEFORE	**DAT**	AFTER
Control Patient	15345000	73158000	23470000
Intervention Patient	17433000	82959000	14088000

6 Conclusions

Brain Computer Interface (BCI) can have variety of application areas such as Neuroscience research, where it can help Obsessive Compulsive Disorder and neurotypic children where its sources mainly include EEG signals obtained from brain using head sensors such as the TAGM1, used to develop the proposed system in this paper.

Based on the outcomes from the PSD analysis, we obtain the results of the Table 4 which show that with this mathematical method, there is a huge increment of brain activity during DAT regarding the other two states, before and after therapy.

Our results show the efficiency of DAT is around of the 376% of increment of neuronal PSD. The novelty of the PSD analyses in a TGAM1 EEG sensor is the ability to identify and give a mathematical meaning of the beneficial changes in behavior of the neuronal system of children with or without Obsessive Compulsive Disorder when their treatment is assisted by a bottle-nose dolphin.

Finally, our findings point out an increment of neural activity (more PSD) during DAT, displaying collective behaviors, i.e., positive increments of neuronal activity could be followed by much more neural activity. Thus, the brain could react to DAT gradually over a period of time, which indicates the existence of external features that modify the brain behaviour of children with diverse psychological and physical disabilities or disorders.

Acknowledgment. The research described in this work was carried out at Delfiniti México, Ixtapa in collaboration with the Instituto Politécnico Nacional and the Secretaría de Salud of Mexico by means of projects No. 20190046 and 20195208. Also, Hugo Quintana, Ixchel Lina, Teresa Ivonne Contreras, Ana Gabriela Ramírez, Alfredo Durand, and Pedro Arrechea are thanked for the methodological and technical support.

References

1. Abtahi, F., Ro, T., Li, W., Zhu, Z.: Emotion analysis using audio/video, emg and eeg: A dataset and comparison study. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 10–19 (March 2018)

2. Birch, S.: Dolphin sonar pulse intervals and human resonance characteristics. In: Proceedings of the 2nd International Conference on Bioelectromagnetism (Cat. No.98TH8269). pp. 141–142 (Feb 1998)
3. Chengwei, L., Xiaoming, H., Limei, Z.: The study on brain paralysis ultrasonic therapy instrument simulating dolphin. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. pp. 6056–6059 (Jan 2005)
4. Joshi, D.H., Jaliya, U.K., Thakore, D.G.: A.r.g.o.s: Alertness rating gamma brainwave observation system. In: 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE). pp. 374–377 (March 2016)
5. Kantelhardt, J.W., Tismer, S., Gans, F., Schumann, A.Y., Penzel, T.: Scaling behavior of eeg amplitude and frequency time series across sleep stages. EPL (Europhysics Letters) 112(1), 18001 (October 2015)
6. Lack, C.: Obsessive-compulsive disorder: Evidence-based treatments and future directions for research. World journal of psychiatry 2, 86–90 (12 2012)
7. Li, K.G., Shapiai, M.I., Adam, A., Ibrahim, Z.: Feature scaling for eeg human concentration using particle swarm optimization. In: 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE). pp. 1–6 (Oct 2016)
8. Lim, C.K.A., Chia, W.C., Chin, S.W.: A mobile driver safety system: Analysis of single-channel eeg on drowsiness detection. In: 2014 International Conference on Computational Science and Technology (ICCST). pp. 1–5 (Aug 2014)
9. Senevirathna, B., Abshire, P.: Spatio-temporal compressed sensing for real-time wireless eeg monitoring. In: 2018 IEEE International Symposium on Circuits and Systems (ISCAS). pp. 1–5 (May 2018)
10. Singh, A.K., Wang, Y.K., King, J.T., Lin, C.T., Ko, L.W.: A simple communication system based on brain computer interface. In: 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI). pp. 363–366 (Nov 2015)
11. Sosa-Jimenez, C.O., Mesa, H.G.A., Rebolledo-Mendez, G., de Freitas, S.: Classification of cognitive states of attention and relaxation using supervised learning algorithms. In: 2011 IEEE International Games Innovation Conference (IGIC). pp. 31–34 (Nov 2011)
12. Wichmann, T.: Detail zooming in artificial intelligence world modeling. In: SoutheastCon 2015. pp. 1–5 (April 2015)

ISSN 1870-4069

A Word Embedding Analysis towards Ontology Enrichment

Mikael Poetsch[1], Ulisses Brisolara Correa[1,2], Larissa Astrogildo de Freitas[1]

[1] Federal University of Pelotas (UFPel), Pelotas, RS, Brazil
{mpoetsch, ub.correa, larissa}@inf.ufpel.edu.br
[2] Sul-rio-grandense Federal Institute of Education, Science, and Technology (IFSul), Charqueadas, RS, Brazil
ulissescorrea@charqueadas.ifsul.edu.br

Abstract. Word Embedding is a set of language modeling and feature learning techniques in Natural Language Processing where words or phrases are mapped to vectors of real numbers. This approach could be used in many tasks of Natural Language Processing, such as Text Classification, Part-Of-Speech Tagging, Named Entity Recognition, Sentiment Analysis, and others. In this paper we created different Word Embedding models, using TripAdvisor's hotel reviews. The corpus was pre-processed, in order to reduce noise, and then submitted to four Word Embedding algorithms: Word2Vec, FastText, Wang2Vec, and GloVe. Finally, HOntology concepts and relations are compared with the outputs of models created aiming to improve it, enriching this domain ontology.

Keywords: word embedding, domain ontology, natural language processing.

1 Introduction

Word Embedding (WE) is a set of language modeling and feature learning techniques in Natural Language Processing (NLP) where words or phrases are mapped to vectors of real numbers [16]. These models create vector spaces where words that are semantically similar are mapped to nearby points.

Nowadays, one trend in NLP is to use vectors of words whose syntactic similarities correlate with semantic similarities. These vectors are used to calculate similarities between terms [10,19,23].

Bengio et al. (2003) [2] were among the first to introduce the term Word Embedding, where words or phrases are mapped to vectors of real numbers. The first work that shows the usefulness of WE pre-trained models in NLP tasks was proposed by [6].

The dissemination of WE can be attributed to [16], when the author created Word2Vec, a toolkit that allows the training of WE models, as well as make it easy to use pre-trained models.

Likewise other NLP tasks, research about WE applied to Portuguese texts is still in its first steps [23]. [10] made syntactic, semantic, intrinsic and extrinsic

analysis of different trained approaches of WE. Authors publicly share trained models in the repository http://nilc.icmc.usp.br/embeddings.

Another research in Portuguese is proposed by [19]. In [19] the intrinsic analysis using different trained models of WE with some tweets dataset were made.

Vargas and Pardo [23] used WE to aspect (explicit and implicit) identification and clustering that are critical tasks for sentiment analysis.

In this work, we created different trained models of WE, available in https://github.com/mikaellllll/Word-Embedding-Setor-Hoteleiro, using hotels reviews posted on the TripAdvisor with the intention of expanding the area of WE applied in the Portuguese. Finally, the models created are compared with a hotel ontology, with the aim of enriching it.

The remaining of this text is organized as follow. Section 2 discusses relevant related works. Next, Section 3 presents an overview of concepts involved in this work. In Section 4 we present the methodology used to assess WE performance, as well as the experimental results obtained. Finally, Section 5 summarizes our contributions and future work.

2 Related Works

Technical literature presents few related work, we found the following works about WE applied to Portuguese texts: [10] , [19], and [23].

Hartmann et al. (2007) [10] trained 31 WE models using the Word2Vec, FastText, Wang2Vec and GloVe algorithms. Their experiments were made varying the dimension of the model (using 50, 100, 300, 600 and 1000 inner dimensions) and using different corpora (mixed, encyclopedic genres, informative, didactic). The corpora used in his work are large. To compare the results, an intrinsic evaluation was performed on syntactic and semantic analogies.

The experiments of [10] shows that GloVe produced the best results for syntactic and semantic analysis, with 46.70% accuracy using 300 dimensions. Wang2Vec Skip-Gram produced the best results for POS-tagging, with 95.94% accuracy using 1000 dimensions.

Saleiro et al. (2017) [19] created WE models from tweets in Portuguese. First, authors begin with a relatively small sample and focus on three challenges, (i) the volume of training dataset, (ii) the size of the vocabulary, and (iii) the intrinsic metrics. The intrinsic metrics aims to establish a good combination between the number of dimensions and the size of the vocabulary. Through this work, the authors realized that producing WE from tweets is challenging due to the specificity of the vocabulary in the social media. Results show that using less than 50% of the available training examples for each vocabulary size might result in overfitting.

Automatic aspect identification and clustering are critical tasks for sentiment analysis. [23] presented a new approach to group explicit and implicit aspects from online reviews (about Books, Cameras, and Smartphones). The authors

achieved the 95.7% accuracy using a Skip-Gram based Word2Vec model, with 300 dimensions.

3 Theoretical Reference

This section describes a few key concepts to understand our proposed approach.

3.1 Corpus

Corpus is a set of collected text to be subject of linguistic research. This resource should be composed of texts, described in natural language by native speakers [20]. Corpora have been used as a key validation resource of solutions of NLP Tasks [8].

Developing a corpus demand to deal with several key factors, such as: definition of texts domain, definition of texts origin, definition of the amount of data to be collected, and how the corpus will be used (practical applications) [8]. Besides that, [8] highlights the importance of planning in corpus construction. Lack of planning could turn invalid experimental data produced using a malformed corpus.

As to size, according to [21], corpus can be classified as small, small-medium, medium, medium-large and large size. [21] created these classes based on research reported in Linguistics conferences and meetings. In this paper we use a medium-size corpus, with 656 thousand words.

3.2 Word Embedding

WE is a set of language modeling and feature learning techniques in Natural Language Processing (NLP) where words or phrases are mapped to vectors of real numbers [16]. These models create vector spaces where words that are semantically similar are mapped to nearby points and where some operations can extract logical results.

Mikolov et al. (2013) [16] present Word2Vec, that allow us to apply vector arithmetics to work with concepts. For instance, if we subtract the vector representation of 'man' from the vector representation of 'king', then we add the vector representation of 'women' the result will be near of the vector representation of 'queen' ($king - man + woman = queen$).

In this work we trained WE models using several different techniques (Word2Vec, FastText, Wang2Vec, and Glove), in order to evaluate which one can present better results in the task of Domain Ontology Enrichment.

Word2Vec. Word2Vec is a widely used method and have two model architectures for computing continuous word vectors using simple model. In the first model, called Continuous Bag-of-Words (CBOW), the non-linear hidden layer is removed and the projection layer is shared for all words; thus, all words get projected into the same position.

Whereas, in the Skip-Gram model the prediction of the current word is based on the context. The author use each current word as an input to a log-linear classifier with continuous projection layer and predict words within a certain range before and after the current word.

FastText. FastText is an approach based on the Skip-Gram model, this algorithm include representing sentences with bag-of-words and bag-of-n-grams, using sub-word information. The method attempts to capture morphological information to insert in WE model.

Wang2Vec. Wang2Vec is a modification of Word2Vec that intends to improve the WE obtained for syntactic tasks. The modification of CBOW is named Continuous Window (CWindow). The modification of Skip-Gram is named Structured Skip-Gram. CWindow and Structured Skip-Gram were made in order to make the network aware of the relative positioning of context words.

GloVe. The Global Vectors (GloVe) approach was proposed by [18]. It combines the advantages of the global matrix factorization and the local context window methods. This method model efficiently leverages statistical information by training only on the non-zero elements in a word to word co-occurrence matrix.

3.3 Ontology

According to [9], an ontology is a formal representation of knowledge based on conceptualization. This kind of resource encompasses a representation of the concepts, relations between concepts and instances.

There are many methodologies that helped in creating an ontology, such as: Enterprise [7] and OnToKnowledge [22]. In general, ontologies previously constructed could be adapted and expanded based in its first version. Ontologies are created and not discovered, so one domain can have different ontologies created by different methodologies.

In this work we evaluate WE models to enrich a domain ontology created with some of the methodologies aforementioned.

4 Experiments

In order to evaluate the applicability of WE models to domain ontologies enrichment we used a corpus to train a set of WE models and then manually analyzed models searching for relations between ontology's concepts absent in ontology.

The corpus used in our experiments is a set of hotel reviews from TripAdvisor [12]. Before the WE training all hotel reviews are pre-processed as follow.

First step was mapping emails to a token "EMAIL", mapping numbers to a token "0", mapping URLs to a token "URL", remove texts within square

brackets, and remove sentences with less than 5 tokens. This pre-processing approach was first proposed by [10].

Second step was removing punctuation of corpus and third step was correcting the spelling of the corpus through the LibreOffice's VERO library[3].

Fourth step was identifying n-grams using N-Gram Statistics Package [1]. N-gram is a sequence of words in a sentence [24]. For instance, the bi-grams of the sentence "O hotel é ótimo" ["The hotel is excellent"] are "# O" ["# The"], "O hotel" ["The hotel"], "hotel é" ["hotel is"], "é ótimo" ["is excellent"], "ótimo #" ["excellent #"], were # is empty. [26] states that use 4 or more grams can cause noise. For this reason, in this paper, we use n-grams with $n < 4$. Besides that, n-grams with less than 10 occurrences were discarded.

Fifth step was removing corpus stopwords. This was made using the Python library Natural Language Toolkit (NLTK)[4].

Sixth step was lemmatizing and identifying the grammatical category of words. We used SpaCy[5] to do so. For example, the lemma of "tem" ["has"] and "tinha" ["had"] is "ter" ["have"].

Sixth step was creating WE models. We used just the grammatical categories of verbs and nouns and its lemmas in the four methods (Word2Vec, FastText, Wang2Vec, and GloVe).

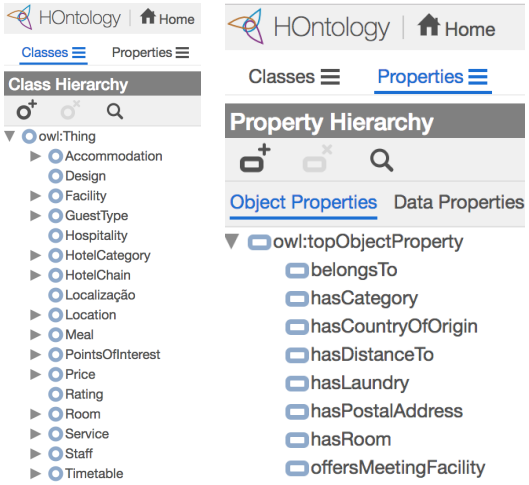
We used 300 dimensions, because according to [10] the increase in performance not reward the increase in memory. Besides that, as configuration we used 10 windows, because according to [17] a bigger number can lead to loss of performance.

To create CBOW and Skip-Gram of Word2Vec and FastText we used Gensim [25]. To create CWindow and Structured Skip-Gram of Wang2Vec we used the codes created by [11] and Glove we used the codes created by [18].

To visualize the WE models we used Bokeh[6] (Figures 2, 3, 4, and 5). More specifically the technique t-distributed Stochastic Neighbor Embedding (t-SNE). t-SNE was created by [15] to reduce the dimensions of the WE models to two dimensions facilitating the visualization of high-dimensional data. It is important to note that there are other techniques to decrease the dimensions, as "Principal Component Analysis"or "TruncatedSVD". We chose to use t-SNE because many works in literature has used this technique, such as: "Learning a Parametric Embedding by Preserving Local Structure [13]" and "Visualizing non-metric similarities in multiple maps [14]".

Next, we searched and visualized domain ontologies. We obtained the hotel domain ontology, HOntology, from the repository OntoLP[7]. To visualize the domain ontology we used Web Protégé[8]. As can be seen in Figure 1. Figure 1(a) presents HOntolgy classes and Figure 1(b) presents HOntolgy properties.

[3] https://pt-br.libreoffice.org/projetos/vero/
[4] https://www.nltk.org
[5] https://spacy.io/
[6] https://bokeh.pydata.org/en/latest
[7] http://ontolp.inf.pucrs.br/Recursos/downloads-Hontology.php
[8] https://webprotege.stanford.edu



(a) Classes (b) Properties

Fig. 1. HOntology.

HOntology is a multilingual ontology of the accommodation sector described in English, Portuguese, Spanish and French [5]. This ontology reuses concepts of another vocabulary as Dbpedia.org[9] and Schema.org[10]. At first, this resource was created to support managerial decision making and end-user applications. Some experts updated HOntology based on existing concepts in others ontologies, as can be seen in the work of [5].

HOntology was developed in seven phases: (1) identify existing ontologies on related domains, (2) select the main concepts and properties, (3) organize concepts and properties hierarchically, (4) manually translate concepts and properties, (5) expand concepts and properties based on online reviews manually evaluated, (6) manually translate the new concepts and properties,(7) export the ontology [5]. Our work could be applied to complement phase 5.

Currently, HOntology contains 282 concepts categorized into 16 top-level concepts, with maximum depth of concept hierarchy equals to 5 [5].

After visualize the WE models we compare them. Figure 2 shows the two variations of Word2Vec, CBOW and Skip-Gram. Glove model is depicted in Figure 3. Figure 4 presents the FastText models based in CBOW and Skip-Gram. In Figure 5, them Wang2Vec models (CWindow and Structured-Skip-Gram) can be seen. Each figure presents the whole structure of the corpus, following different WE models.

We perform two types of analysis. First, we analyze classes and subclasses with each WE model. The classes and subclasses were plotted to facilitate the identification in the clusters.

[9] https://wiki.dbpedia.org
[10] https://schema.org

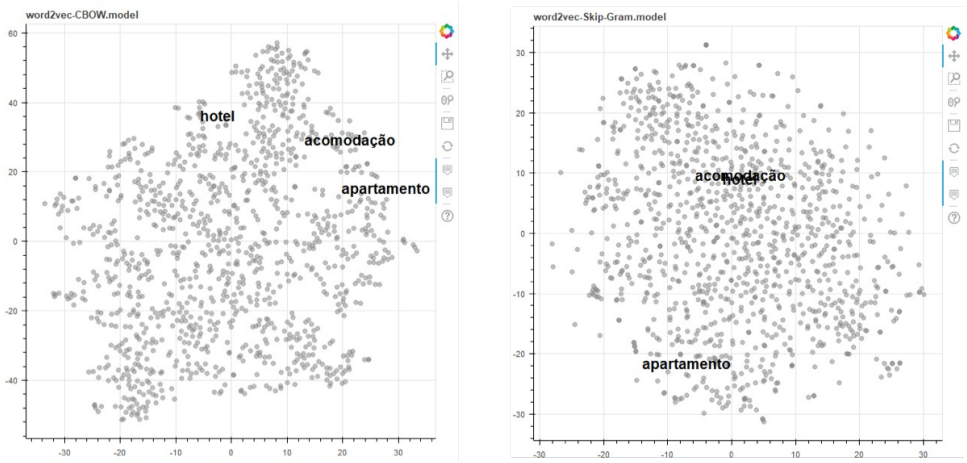


Fig. 2. Word2Vec models, "Acomodação"["Accommodation"] class and its subclasses "Apartamento" ["Apartment"] and "Hotel" ["Hotel"].

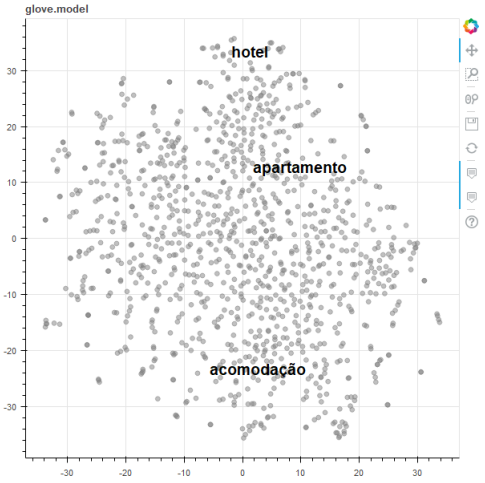


Fig. 3. Glove model, "Acomodação" ["Accommodation"] class and its subclasses "Apartamento" ["Apartment"] and "Hotel" ["Hotel"].

For instance, to the "Acomodação" ["Accommodation"] class some of its subclasses were not found in the models created. This may occur because the ontology is a formal representation of information while the reviews collected are informal texts, where formal terms are not common. Still, we observe that many subclasses were not found. For example: "Instalação do Banheiro" ["Bathroom Facility"], "Preço do Café da Manhã" ["Breakfast Price"] and "Serviço de Câmbio" ["Exchange Service"]. In general, these subclasses are related to

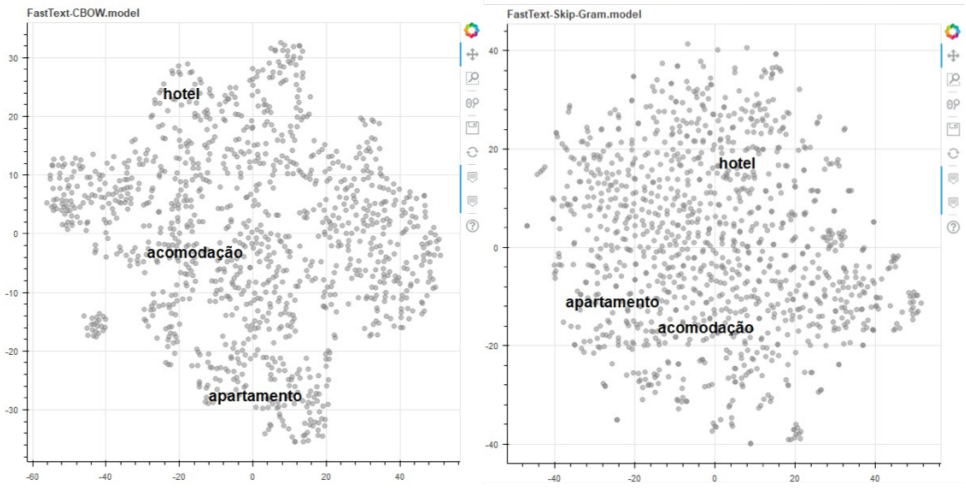


Fig. 4. FastText models, "Acomodação" ["Accommodation"] class and its subclasses "Apartamento" ["Apartment"] and "Hotel" ["Hotel"].

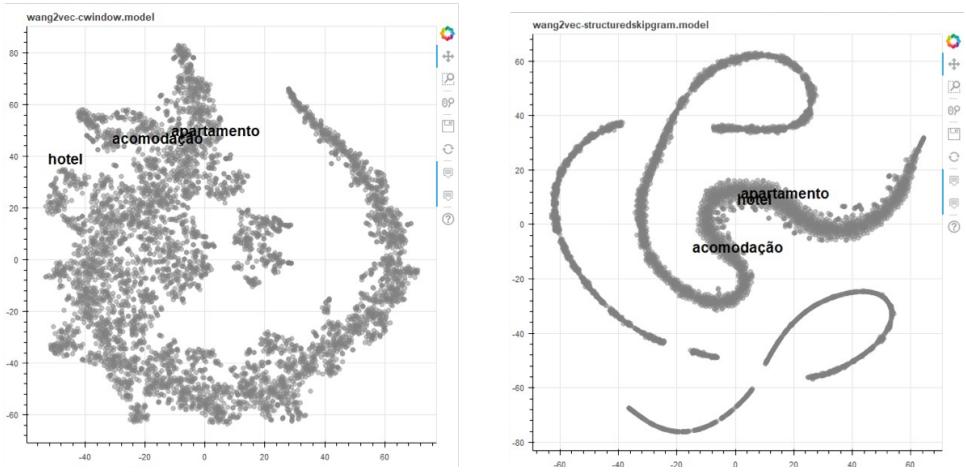


Fig. 5. Wang2Vec models, "Acomodação" ["Accommodation"] class and its subclasses "Apartamento" ["Apartment"] and "Hotel" ["Hotel"].

location or services offered by an accommodation. It is interesting to notice that the subclasses "Café da Manhã" ["Breakfest"] and "Jantar" ["Dinner"] of class "Refeição" ["Meal"] were found, but "Almoço" ["Lunch"] was not.

According to our results, "Acomodação" ["Accommodation"] class has no similarity to "Apartamento" ["Apartment"] subclass in Word2Vec, FastText and Glove due to low percentage of similarity and long distance between the words.

Wang2Vec and Word2Vec CBOW presents similarity bigger than 90% between "Acomodação" ["Accommodation"] class and "Hotel" ["Hotel"] subclass, as can be seen in TABLE 1,.

Table 1. Similarity between "Acomodação" ["Accommodation"] class and its subclasses.

	Subclasses	
Model	**Apartment**	**Hotel**
Word2Vec CBOW	36,7%	92,3%
Word2Vec Skip Gram	38,5%	61,9%
FastText CBOW	36,0%	73,6%
FastText Skip Gram	41,9%	54,5%
Wang2Vec CWindow	98,9%	90,5%
Wang2Vec Structured Skip Gram	99,9%	99,9%
Glove	30,8%	24,8%

Wang2Vec Structured-Skip-Gram presented the best similarity for both subclasses, "Apartamento" ["Apartment"] and "Hotel" ["Hotel"]. Observing the similarity between all classes and subclasses (depth 1) of HOntology, it was possible to notice that in the most of the models (Word2Vec, FastText, and Glove) there is no consensus on the degree of similarity between classes and subclasses, with except for the Wang2Vec. "Acomodação" ["Accommodation"] and "Refeição" ["Meal"] classes exist on all models. "Pontos de Interesse" ["Points of Interest"] exist only in FastText e Wang2Vec models. "Categoria de Hotel" ["Hotel Category"], "Endereço" ["Address"], "Serviço" ["Service"] and "Tipo de Hóspede" ["Guest Type"'] exist only on FastText model.

We also analyzed HOntology classes and words with each WE model, considering or not considering grammatical categories (verbs and nouns). In TABLE 2 we show words most similar to "Acomodação" ["Accommodation"] class.

In Word2Vec CBOW the word most similar to "Acomodação" ["Accommodation"] class was "Instalação" ["'Facility"], this word also is a class of HOntology. Thus, we could insert a new relation between these two classes.

In Wang2Vec CWindow one of the words most similar was the trigram "É limpo E" ["is clear and"]. The third column of the TABLE 2 should include verbs, which indicates an error in the POS-tagging, where "É limpo E" ["is clear and"] was classified as verb.

Finally, we analyze HOntology relations with each WE model. Among the HOntology relations we can mention: "pertenceA" ["belongsTo"], "temCategoria" ["hasCategory"], "temPaísDeOrigem" ["hasCountryOfOrigin"], "temDistânciaA" ["hasDistanceTo"], "temLavanderia" ["hasLaundry"], "temEndereçoPostal" ["hasPostalAddress"], "temQuarto" ["hasRoom"], and "ofereceLocalDeReunião" ["offersMeetingFacility"] (TABLE 3). For example, the relation "pertenceA" ["be-

longsTo"] represent that "Hotel" ["Hotel"] belongs to "Rede de Hotéis" ["Hotel Chain"].

Table 2. Words Most Similar to "Acomodação" ["Accommodation"] Class.

Model	Word Most Similar	Word Most Similar Considering Grammatical Category
Word2Vec CBOW	Instalação	Adequar, Gostar
Word2Vec Skip Gram	Boutique	Reduzir, Surpreender
FastText CBOW	Fomos Muito Bem	Fomos Muito Bem, Muito A Desejar
FastText Skip Gram	Acomodar	Acomodar, Surpreender
Wang2Vec CWindow	Hotel É Muito	Praticar, **É Limpo E**
Wang2Vec Structured Skip Gram	É Uma Cidade	Cado, Significar
Glove	Instalação	Adequar, Treinar

Table 3. Words that correspond to the relations of HOntology.

Relation	Domain	Extension	Word
pertenceA	Hotel	Rede de Hotéis	pertencer, hotel, rede, hotel de rede, hotel da rede, hotéis da rede, da rede acor, da rede ibis
'belongsTo'	'Hotel'	'Hotel Chain'	
temCategoria	Hotel	Categoria de Hotel	ter, categoria, hotel, padrão, turista
'hasCategory'	'Hotel'	'Hotel Category'	
temPaísDeOrigem	Tipo de Hóspede	-	ter, país, origem, casal, família
'hasCountryOfOrigin'	'Gest Type'		
temDistânciaA	Hospitalidade	Pontos de Interesse	ter, distância, distante do centro, hospitaleiro, hospitalidade, pouco distante do
'hasDistanceTo'	'Hospitality'	'Points Of Interest'	
temLavanderia	Hospitalidade	Lavanderia	ter, lavanderia, hospitalidade, hospitaleiro
'hasLaundry'	'Hospitality'	'Laundry'	
temEndereçoPostal	Hospitalidade	Localização	ter, endereçar, hospitalidade, hospitaleiro
'hasPostalAddress'	'Hospitality'	'Location'	
temQuarto	Hotel, Hospitalidade	Quarto de Hotel, Hotel	ter, quarto, quarto de hotel, hotel, hospitalidade, hospitaleiro
'hasRoom'	'Hotel', 'Hospitality'	"Hotel Room', 'Hotel'	
ofereceLocalDeReunião	Hotel	Reunião	oferecer, hotel, reunião, reunir
'offersMeetingFacility'	'Hotel'	'Meeting'	

After defining the words for each relation defined in HOntology, we find the similarity between all words searched for each model. Wang2Vec presented a big similarity between words searched and FastText did not obtain any conclusive results to "pertenceA" [" belongsTo"].

Negative similarity means that the relation between words has opposite meanings. If two words have similarity equals to zero, they are unrelated. And,

if two words have similarity equals to -100%, they have a perfectly opposite relation.

5 Final Remarks and Future Work

In this work we created different models of WE using hotel reviews published on TripAdvisor in order to assess its applicability to the task of domain ontology enrichment.

After analyzing the HOntology and the different WE models created (Word2Vec CBOW and Skip-Gram, Glove, FastText CBOW and Skip-Gram, Wang2Vec CWindow and Structured-Skip-Gram), we realize that some concepts that are found in the reviews are not found in the domain ontology. Then, we conclude that it is possible to use WE models to enrich domain ontologies by incorporating new properties, such as a relation between "Accommodation " and "Installation".

Wang2Vec algorithm presented the best results for relations identification, accusing a great similarity between words related to an already existing relation in the HOntology. However, according to [4] the relevance of these relations must be evaluated before being inserted into the domain ontology to avoid an excessive increase of complexity in the representation.

As future works we intend to use other WE algorithms. Also, we intend to explore other levels of HOntology, such as: subclasses (depth 2, 3 and 4) and instances. And analyze how their results can be used to enrich domain ontologies. In addition, we could analyze the use of hyponymy, hyperonymy, and coreference.

References

1. Banerjee, S., Pedersen, T.: The design, implementation, and use of the ngram statistics package. In: 4th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing), pp. 370–381 (2003)
2. Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A neural probabilistic language model. Journal of Machine Learning Research 3, 1137–1155 (2003)
3. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5(1), 135–146 (2016)
4. Carvalho, M.G.P., Braganholo, V., Machado Campos, M. L., de Almeida Campos, M.L.: Enriquecimento de ontologias: uma abordagem para extração de conhecimento do campo definição. In: 3th Seminário de Pesquisa em Ontologias no Brasil (Ontobras) (2010)
5. Chaves, M., Freitas, L., Vieira, R.: Hontology: a multilingual ontology for the accommodation sector in the tourism industry. In: 4th International Conference on Knowledge Engineering and Ontology Development, pp. 1–6 (2012)
6. Collobert, R., Weston, J.: A unified architecture for natural language processing: deep neural networks with multitask learning. In: 25th International Conference on Machine Learning (ICML), pp. 160–167 (2008)
7. Dietz, J.: Enterprise Ontology: Theory and Methodology. Springer-Verlag, Berlin, Heidelberg (2006)

8. Fromm, G.: O uso de corpora na análise linguística. Factus 1(1), 69–76 (2003)
9. Gruber, T. R.: Toward principles for the design of ontologies used for knowledge sharing? International Journal of Human-Computer Studies, 43(5-6), 907–928 (1995)
10. Hartmann, N., Fonseca, E. R., Shulby, C., Treviso, M. V., Silva, J., Aluísio, S.: Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In: 11th Brazilian Symposium in Information and Human Language Technology (STIL), pp. 122–131 (2017)
11. Ling, W., Dyer, C., Black, A. W., Trancoso, I.: Two/too simple adaptations of word2vec for syntax problems. In: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1299–1304 (2015)
12. López Barbosa, R. R.: Aplicación del análisis de sentimientos a la evaluación de datos generados en medios sociales. Ph.D. thesis, Universidad de Alcalá (2015)
13. Maaten, L.: Learning a parametric embedding by preserving local structure. In: Artificial Intelligence and Statistics, pp. 384–391 (2009)
14. Van der Maaten, L., Hinton, G.: Visualizing non-metric similarities in multiple maps. Machine learning 87(1), 33–55 (2012)
15. van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research 9(Nov), 2579–2605 (2008)
16. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
17. Minarro-Giménez, J. A., Marín-Alonso, O., Samwald, M.: Applying deep learning techniques on medical corpora from the world wide web: a prototypical system and evaluation. CoRR, pp. 1–14 (2015)
18. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543 (2014)
19. Saleiro, P., Sarmento, L., Mendes Rodrigues, E., Soares, C., Oliveira, E.: Learning word embeddings from the portuguese twitter stream: A study of some practical aspects. In: Portuguese Conference on Artificial Intelligence, pp. 880–891 (2017)
20. Sardinha, T. B.: Linguística de corpus: histórico e problemática. Delta 16(2), 323–367 (2000)
21. Sardinha, T. B.: Tamanho de corpus. The ESPecialist 23(2), 103–122 (2002)
22. Sure, Y., Staab, S., Studer, R.: On-to-knowledge methodology (otkm). In: Handbook on Ontologies, International Handbooks on Information Systems, pp. 117–132 (2003)
23. Vargas, F., Pardo, T.: Aspect clustering methods for sentiment analysis. In: 13th International Conference on the Computational Processing of Portuguese (PROPOR), pp. 365–374 (2018)
24. de Castro Sonnenfeld Vilela, P.: Classificação de Sentimento para Notícias sobre a Petrobras no Mercado Financeiro. Ph.D. thesis, PUC-Rio (2011)
25. Řehůřek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: Workshop on New Challenges for NLP Frameworks, pp. 45–50 (2010)
26. Wiebe, J., Wilson, T., Bruce, R., Bell, M., Martin, M.: Learning subjective language. Computational linguistics 30(3), 277–308 (2004)

ISSN 1870-4069

Intelligent Agent based System for Crop Monitoring

Ahad Hanif[1], Aslam Muhammad[2], Ana María Martinez-Enriquez[3], Adrees Muhammad[4]

[1] Department of Mechatronics Engineering, UET, Lahore, Pakistan
ahad.hanif@umt.edu.pk
[2] Department of CS, UET, Lahore, Pakistan
maslam@uet.edu.pk
[3] Department of CS, CINVESTAV, CDMX, Mexico
ammartin@cinvestav.mx
[4] Department of CS, Superior University, Lahore, Pakistan
adreesgujer@gmail.com

Abstract. This paper describes the utility of environmental sensors for the monitoring of crops remotely, grown in a green house or a tunnel. The work combines the instrumentation techniques with knowledge based system. Agriculturalists frequently visit their farms to monitor the condition of plants and crop. This monitoring is time consuming. Conventional farms monitoring techniques and inspections give low crop production, attack of severe diseases parasites. Adequate amount of water, fertilizers, and nutrients can improve the yield and at the same time reduces the cost. In order to assist farmers and landowner, we designed and developed an agent based crop monitoring system. The objective of this research is to develop an agent based system to help the agriculturists to monitor the crops remotely by using recent technology and hence advise the farmers to take appropriate measures. The system is divided into main two main parts. First one is data acquisition and the other one is recommendation part. The readings of crucial environmental parameters are taken and sent to a central computer service. Color processing techniques are used to find the nitrogen deficiency in the green plants. Manual leaf color chart has been replaced by electronic one which is capable of providing sufficient information about future scenario of nitrogen content in the plants. The system uses real time information and recommends an appropriate solution of the problem caused by environmental stresses. In case of any critical situation, the system generates warnings in the form of an email, to be sent to the landowners or agriculturist.

Keywords: remote sensing, knowledge base, color processing, leaf color charts, intelligent agent.

1 Introduction

As the requirements of the researchers in the area of crop sciences change, a need is felt for developing the tools that assists them with their day to day activities. The system required for such a purpose needs to be intelligent in nature regarding the acquisition of data from the sensors, interfacing with the system and manipulating it

in certain cases. Remote Sensing is the science and art of obtaining information about an object, area or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area or phenomenon under investigation. Many systems of monitoring like RADARs, Satellite imaging and aerial photography exist but these systems are quite expensive, complex and require technical expertise to operate them. The environmental parameters are monitored and relayed to the monitoring system. The monitored parameters include temperature, humidity and light [1]. By maintaining optimum range of parameters like light, temperature and humidity one can protect the plants from pests and diseases [2]. Excessive water and low soil oxygen damages the roots of the plants [3].

Manual monitoring of crop depends on the knowledge of the farmers. The problem arises when farmer is not well educated and he needs the expert's opinion. Furthermore, manual monitoring is time consuming and ends up with low yield. It is not possible for the farmer to make sure his presence in the field round the clock. We have to monitor the crops/plants in real time. There are two main factors, which are crucial for the development of plants. They are crop water status and availability of fertilizers to meet the requirements of essential nutrients. Unsuitable ranges of environmental factors (temperature, humidity) disturb the mobility of nutrients in the plants. Plants may suffer from dehydration, chilling and freezing effects. They become vulnerable against the attack of pests and fungus.

Leaf color charts are being used by the farmers to check the nitrogen deficiency in the plants. Manual leaf color charts require personal technical expertise. If plants are supplied with insufficient amount of nutrients, then yield will be low. In other case production cost will be increased. Small scale farming requires a system which is not only cost effective but also provides good yield.

This paper proposes intelligent agent based system for monitoring the crops, which are located in the remote areas. It was designed for the acquisition of necessary parameters which are crucial for the development of the plants like humidity and temperature of the environment. The intelligent agent manipulate the data acquired by the sensors and also provide us with a suitable solution of the problem based on the expert's opinion. Green houses and tunnel farming techniques would be more productive if they are equipped with such a remote monitoring system. It gives suggestions to the farmers so that they take appropriate measures before the damage is done to the crop. In this work, all the knowledge regarding the normal growth of the plants is gathered within the knowledge base of the system. System utilizes this knowledge base while analysing the information and taking decisions.

The rest of the paper has four main parts. It includes the literature review, which provides insight into the existing woke done in Phytomonitoring techniques. Then comes the Materials and Methodology section, which elaborates our contribution in crop monitoring. Next section is about experimental results and discussions. The last section of the paper is conclusion, which summarizes the whole work.

2 Literature Review

Phytomonitoring techniques are very helpful when it comes to detect the physiological disorders in the plants. Environmental parameters including

temperature, air humidity and soil moisture are recorded through sensors. For remote data transmission GSM modem is used [4] [5]. Growth rate of trunk, flower and fruits are most sensitive parameters in Phytomonitoring which may change with a slight modification in water balance. It is important to monitor and control the environmental parameters in real time to help the farmers to bring innovation in the irrigation strategy [6]. Image processing is now being used extensively in agriculture sector. Image processing may be used to find the age of wheat crop. The information was quite helpful in providing the suitable amount of fertilizers and nutrients to the plants. Overall cost reduces by the application of this technique [7]. In image processing, three prime colors combine to form an image. They are red, green and blue [9]. Rodríguez (2002) proposed hierarchical control of green house. Fungus damages the fruit. Its growth can be controlled by maintaining suitable ranges of temperature and humidity. So fruits storage properties were improved [8].

A trapezoidal two dimensional index was empirically derived by Thomas, which could detect water stress even with a low percentage of canopy cover [11]. Continuous monitoring and scheduling the irrigation process by using Infrared images have provided early warning of water stress [10]. The best way to find the plant water status is to detect the soil water content by evapotranspiration estimation. Infrared images were produced to find the vegetation index [12]. Solar energy based green house (GH) monitoring system was designed for the monitoring of green house. All the nodes and sensors were powered by solar energy to solve the low battery issue. Data was transferred through ZigBee and GSM technology [13].

With the increase in world population we need more food to meet the requirements of the people. Many people die each year due to starvation. We need techno-economic solution of this problem. Conventional monitoring systems are not techno-economic. So much space is available to develop such a system which is cheaper, reliable and give quick response.

The objective of this research is to develop such a system which suggests the farmers and agriculturists, a solution of the problem caused by environmental stresses so that they take appropriate measures before the damage is done to the crop. The knowledge base of the system possesses the expertise of the experts. We would be able to maintain large amount of information which is quite helpful in making reasoning and rules. We have tried to bring flexibility in Leaf Color Chart (LCC) so that ever person can use it conveniently.

3 Materials and Methodology

The most crucial atmospheric element,which effects the growth and yield of plants is temperature. Optimum temperature is required for photosynthesis respiration. Second important parameter is humidity. Relative humidity is the ratio of two water vapour contents i.e. the real water vapour contents and saturated water vapour contents in the air. Both values are taken at same pressure & temperature. It is expressed in percentage. Plants use air carbon dioxide for their photosynthesis process. Plant's leaves have pores to take in CO2. During respiration, some moisture goes out in the air through these pores. If air has high humidity, then plants will transpire moisture more slowly. The reverse happens when air has low moisture contents. Therefore,

when air is dry then more water is evaporated from the plants. In this case plants become deficient in water and leaves close their pores. Although water outgoing is reduced but at the same time it reduces the Carbon dioxide intake. Plants absorb water through roots from soil. If evapotranspiration rate is high than absorption rate, then plants under go severe atmospheric and chemical stresses. It may cause the plant cells to die.

A sensors based crop monitoring system was developed which utilizes the field sensors to acquire information about the environmental parameters. It possesses the knowledge of the experts and provides farmers with a solution of the problem. Secondly, manual leaf color charts (LCC) have been replaced by electronic leaf color charts, which are not only flexible but are also easy to use. It assists the farmers to use the fertilizers in appropriate amount to minimize the production cost. Temperature sensor (DS18B20) and humidity sensor (HSU-04) were used to monitor the environment of the enclosed area in which plants were grown. IP camera was installed in the field to get the images of the plants to be compared with the standard leaf color chart.

The agent is basically a software which is designed to utilize the intelligence to carry out an assigned task automatically like retrieving and delivering information. Intelligent agent is also closely related to software agent (an autonomous computer program that carries out tasks on behalf of users). In computer science, the term intelligent agent may be used to refer to a software agent that has some intelligence. For example, an autonomous program, set for operator assistance or data mining, is also called an intelligent agent. An intelligent system incrementally accommodates new problem solving rules. It should have the capacity to analyze itself, behavior, faults and corrections. An agent maps every possible percept to an action. An agent perceives from its environment and surroundings through sensors and reciprocates through actuators.

The agent collects real time data and manipulates it by using its knowledge base and rules. Information about environmental parameters was transferred to a computer through RS-232 serial port. All the data is manipulated by the intelligent agent based system. The system monitors the temperature, humidity and the quantity of nutrients (Nitrogen) required for normal growth of the plants. Real time images of plants are taken by camera to find the nitrogen deficiency in the plants. The results are transferable through an email. There are two basic parts in this system. One is remote monitoring unit (RMU) and the other one is data manipulation unit (DMU). The color of the leaf changes with the change in the amount of nutrients. There are three main nutrients which are called prime nutrients. They are Nitrogen, phosphorous and potassium. Green color of leaf is due to nitrogen. If plant is deficient in nitrogen, then the color of the leaf becomes pale green. In addition, if plant has sufficient amount of nitrogen then leaves appear in dark green color. Real time images of plants are compared with a standard color scheme to find out the percentage of nitrogen in the plants.

Information from field sensors is transferred to the computer through a microcontroller (PIC18F452) and RS-232 serial port. Computer has an intelligent agent based system which utilizes this information. For better visualization, values which are coming from field sensors are also displayed on a 16x2 LCD. Agent, Matlab program, uses this information and on the bases of rules it generates an email

which is sent to the user inbox to intimate him about the current situation. RS-232 Data Logger software has been incorporated. It generates a text file of the value of temperature and humidity. This file is automatically updated after regular intervals. This file is used by the Matlab code to apply the rules. System architecture is shown in Figure 1.

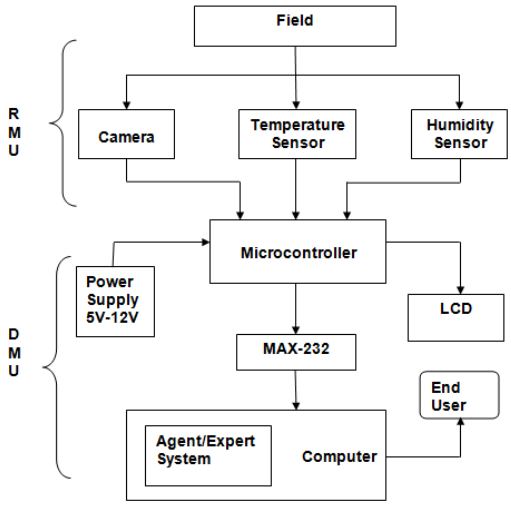


Fig. 1. System architecture.

A graphical user interface (GUI) has been developed to set the ranges for both temperature and humidity. Till the real time values remain between the ranges, no intimation or warning email is generated. But as soon as the value goes out of the range an email is generated, intimating the farmer about the unwanted conditions. Screen shot of GUI window is shown in Figure 2.

While doing color image processing we have to separate the desired color, which is under observation, from the original image. In our case we are working on the green color.

So firstly we have to identify the green color in the original image. There are many disturbing factors like sun shine, shades, background and presence of non-green colors in the original image which create hurdles in detecting the pure green color.

A technique has been developed to identify the green color in the image. Mask of the required color is created by using image processing tools and apply it to the original image. The program does the simple color detection in RGB color space. We first find out the individual color bands from the image. Histogram of the image is computed first and color threshold ranges are selected. This mask is now applied to the original image. After applying we can identify the green portion in the original image. The processed image is then compared with the standard leaf color chart.

Leaf color chart provides a very simple technique to check the nitrogen deficiency in the plants. It provides farmers with an opportunity to find out the plant's nitrogen demand so that they can apply the nitrogen fertilizer in appropriate amount to minimize the production cost. There are usually 6 shades of green colors on LCC

which vary from light green to dark green. Each shade represents a specific amount of nitrogen deficiency. Different shades of green color are shown in Figure 3.

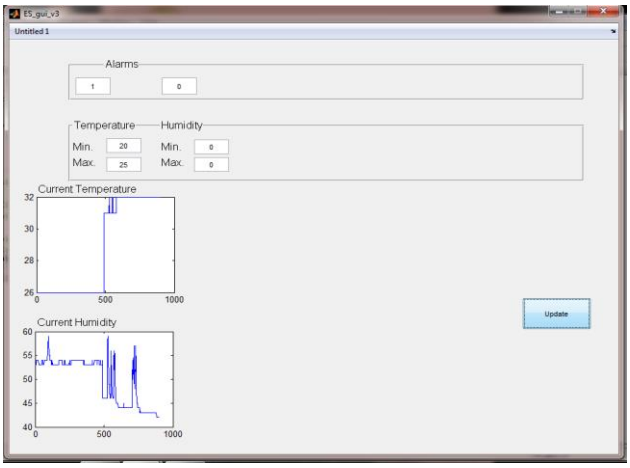


Fig. 2. GUI window.

Fig. 3. Color scheme of LCC.

Light green color means that plants have high deficiency of nitrogen and dark green color means that plants have sufficient amount of nitrogen. Sometimes it happens that the color of leaf lies between the two shades and does not completely match with any shade. This issue has been addressed in this work. Electronic LCC gives actual comparison and tells if the color lies between the two shades.

Fig. 4. Image from field.

Figure 4 shows the field image. It is compared with the standard color scheme. But before comparing we have to use some image processing tools to extract the green

color from the image as described above. In this way, we can minimize the effects of non-green parts on the results.

4 Experimental Results and Discussions

There are total 256 pixels in the image. The range is from 0 to 255. Each color has its own histogram. Histograms of all bands 'red, green and blue' are shown in Figure 5 (a) and 5 (b) respectively.

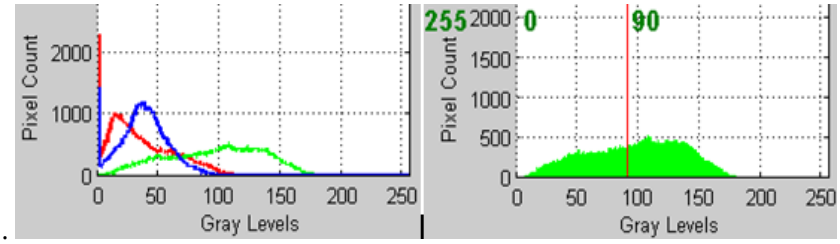


Fig. 5. (a) Histogram of all bands (b) Histogram of green band.

After the identification of green color we compared it with the standard ELCC. Shades of ELCC may vary depending upon the type of crop and the area in which it is grown. Therefore, for a specific region and crop, reference images for ELCC are collected first. These images make a color scheme. Now user can monitor the nitrogen contents in the plants by comparing the real time image with the standard color chart. Each pixel of field image is divided by the average value of reference images. But before finding the average value we have to use image processing tools as described above to find out the true green color from the image. The results are shown in Figure 6.

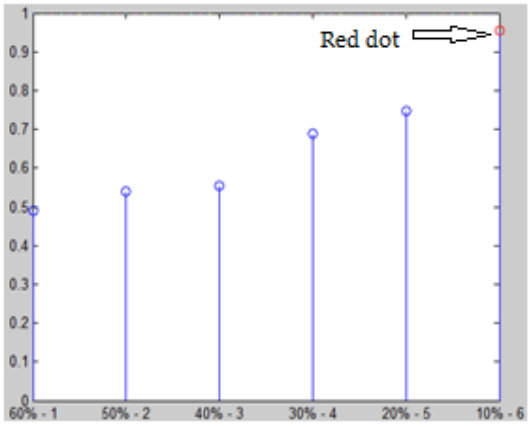


Fig. 6. Results of color processing.

X-axis is labeled as the percentage of nitrogen deficiency in the plants. There are six shades of green color in ELCC so there are 6 vertical lines, representing different shades of green color. The line, with red circle at top, is closest to the dotted horizontal line (passing through y-axis 1). This vertical line shows that the field image matches with that corresponding shade. If the circle is on the dotted line, it means that the corresponding shade from the scheme exactly matches the field image. If it is below or above the dotted line it means that the color of field image lies between the two shades. On x-axis we can find that how much nitrogen deficiency is there in the plants.

On the basis of that information the agent suggests the amount of nitrogen fertilizer (like urea) to be added in the soil. Electronic leaf color chart (ELCC) provides us cost effective, efficient and quick measurement of nitrogen deficiency in the green plants. Intensity of Green color is associated with nitrogen contents. Sun light is a disturbing factor when camera takes the images of the plants. It may cause slight difference in color. We need to check whether camera is facing or not facing the sun light. If a picture is taken in sunlight, then it is possible that it shows different shade as compared to one taken in overcast weather.

Through this technique we have tried to minimize the effects of sun light and other factors. For better results, it is essential that camera should focus on green portion of the plants and should not include the non green parts like the color of trunk or soil etc.

5 Conclusion

In addition to remotely monitoring and controlling the plant environment, the system minimizes the human efforts and helps the farmers to produce techno-economical crops by utilizing nutrients in appropriate amount and minimizing the energy consumption.

The growth of plants becomes more sophisticated. By applying modern instrumentation and artificial Intelligence techniques in agriculture sector we can make this sector more productive and efficient. This system may also be used as testing tool by deliberately adjusting the water supply, temperature and humidity. We can analyze the impacts of environmental stresses on the growth of plants. In addition to that it is quite helpful to see the past trends of nitrogen deficiency is the plants. Results have been improved by the application of image processing tools.

To avoid any abrupt fluctuation in temperature and humidity value, agent takes the average of the previous values. So it avoids temporary alarming conditions. Real time monitoring of the plants through color image processing has made the system quite useful. Farmers are provided with the best possible solution through email.Plants can be grown in those areas which do not have suitable environmental conditions for a specific crop.

The results of this system could be used for future research especially by those departments which are concerned with data acquisition. It would be a great contribution if the system enhances its knowledge base automatically to compensate those conditions and rules which are not incorporate in its knowledge base.

References

1. Zuo, X. et al.: Design of Environmental Parameters Monitoring System for Watermelon-Seedlings Based on Wireless Sensor Networks. App Mathe. & Infor. Sci., 243S-250S (2011)
2. Albright L. D. et al.: Environmental Control for Plants on Earth and Space. IEEE Control Systems Magazine 21(5), 28–47 (2011)
3. Schaffer, B.: Effects of Soil Oxygen Deficiency on Avocado (persea americana mill.) Trees, Seminario International. En: Manejo del Riego y Suelo en el Cultivo del Palto La Cruz, Chile (2006)
4. Avidan, A., Hazan, A.: Application of the Phytomonitoring Technique for Table Grapes. Submitted for the Proceedings of the International Workshop on Advances in Grape vine and Wine Research. Venosa (Italy), September 15-17 (2005)
5. Puig, V. et al.: Optimal predictive control of water transport systems: Arrêt-Darré/Arros, a case study. Water Sci Technol 60(8), 2125–33 (2009)
6. Ton, Y., Kopyt, M.: Phytomonitoring in Realization of Irrigation Strategies for Win grapes. Acta Hort. (ISHS) 652, pp. 167–173 (2004)
7. Kakran, A., Mahajan, R.: Monitoring Growth of Wheat Crop using Digital Image Processing. International Journal of Computer Applications 50(10) (2012)
8. Ibrahim, M., Rabah, A.B.: Effect of Temperature and Relative Humidity on the Growth of Helminthosporium fulvum. Nigerian Journal of Basic and Applied Science 19(1), 127–129 (2011)
9. Plataniotis, K.N., Venetsanopoulos, A.N.: Color Image Processing and Applications. (2000)
10. Rodríguez, F.: Modeling and Hierarchical Control of Greenhouse Crop Production (in Spanish). PhD thesis, University of Almería, Spain (2000)
11. Kopyt, M., Ton, Y.: Phytomonitoring Technique for Table Grapes Application Guide. 2ND Edition, PhyTech Ltd. (2005)
12. Clarke, T. R.: An Empirical Approach for Detecting Crop Water Stress Using Multispectral Airborne Sensors. HortTechnology 7(1), January-March (1997)
13. Dupin, S., Gobrecht, A., Tisseyre, B.: Airborne thermography of vines canopy: effect of the atmosphere and mixed pixels on observed canopy temperature. UMR ITAP, Montpellier SupAgro, bat. 21, 2 place Viala, 34060 Montpellier, France (2011)
14. Gao, L., Cheng, M., Tang, J.: A Wireless Greenhouse Monitoring System based on Solar Energy. Telkomnika 11(9), 5448–5454 (2013)

ISSN 1870-4069

Análisis automático de estilo de escritura en textos de longitud variable

Germán Ríos-Toledo

Tecnológico Nacional de México, Campus Tuxtla Gutiérrez, Chiapas, México
german_rios@ittg.edu.mx

Resumen. El análisis automático del estilo de escritura es una tarea que consiste en la creación de un modelo que represente el estilo de un autor. La representación del estilo de un autor, generalmente, se obtiene mediante la conteo de la frecuencia de uso de marcadores de estilo presentes en un texto. Para un análisis confiable, es importante considerar la cantidad de información de los textos en términos de palabras o en oraciones. En esta investigación se evaluaron novelas de diez diferentes autores utilizando un marcador de estilo conocido como n-grama de cuatro diferentes categorías. EL conjunto de novelas de cada autor se dividió en dos etapas. Cada novela se dividió en fragmentos de distintos tamaños en los que disminuye gradualmente el número de oraciones. Se entrenó un algoritmo de aprendizaje automático supervisado para predecir a cuál de las dos etapas de un autor pertenece un fragmento de texto. Los resultados de los experimentos mostraron por un lado, que la categoría de n-gramas basados en información sintáctica mostraron un porcentaje de clasificación superior a las categorías restantes y por el otro, que determinar la cantidad apropiada de texto depende en gran medida del estilo de escritura de cada autor.

Palabras clave: marcador de estilo, n-gramas, estilometría.

Automatic Analysis of Writing Style in Texts of Variable Length

Abstract. Automatic analysis of the writing style is a task that consists in creating a model that represents the style of an author. The representation of an author's style is generally obtained by counting the frequency of use of style markers present in a text. For a reliable analysis, it is important to consider the amount of information in the texts in terms of words or sentences. In this research, novels by ten different authors were evaluated using a style marker known as n-grams of four different categories. The set of novels of each author was divided into two stages. Each novel was divided into fragments of different sizes in which the number of sentences gradually decreases. A supervised machine learning algorithm was trained to predict to which of the two stages

of an author a piece of text belongs. The results of the experiments showed, on the one hand, that the category of n-grams based on syntactic information showed a higher classification percentage than the remaining categories, and on the other, that determining the appropriate amount of text depends largely on the style of writing of each author.

Keywords: style markers, n-grams, stylometry.

1. Introducción

La estilometría es una disciplina que se basa en la presunción de que cada persona tiene un estilo de escritura. El análisis automático del estilo de escritura tiene utilidad en áreas como el derecho penal o civil, pues coadyuva a la detección de plagio, la creación de perfiles de autor y a la protección del anonimato.

En el análisis de estilo comúnmente se utilizan las palabras como marcadores de estilo: palabras con contenido semántico (sustantivos, verbos, adjetivos, adverbios); palabras funcionales (preposiciones, adverbios, artículos, pronombres, adjetivos); longitud de palabras; categorías gramaticales; errores de escritura; lemas; entre otras.

El análisis de estilo se enfoca en la forma del texto y no en su contenido. Actualmente, el reto principal en este tipo de estilo es identificar marcadores de estilo robustos al contenido temático y al tipo de documento. En este sentido, el uso de la información sintáctica es una alternativa potencial. Los analizadores sintácticos modernos obtienen la información sintáctica de cada oración y la representan de forma estructurada en un árbol de dependencia. Dichos árboles muestran el orden no lineal entre las palabras, las categorías gramaticales y los nombres de las relaciones sintácticas existentes. Las características estilométricas basadas en información sintáctica posibilitan el desarrollo de análisis de estilo más completos.

El análisis automático de estilo puede abordarse con un enfoque de aprendizaje supervisado. Tanto los textos de entrenamiento como los de prueba se representan como vectores de características bidimensionales (matrices).Las celdas de la matriz contienen la frecuencia un marcador en un texto particular. Los algoritmos de aprendizaje reconocen patrones en las frecuencias y generan un modelo, dicho modelo recibe textos no vistos en la etapa de entrenamiento y genera predicciones. Un modelo ideal sería aquel que clasifica correctamente todos los textos.

Las secciones restantes que conforman el artículo son: Trabajos relacionados, este apartado describe los estudios previos que abordan el problema de la cantidad de texto requerida para tareas del Procesamiento de Lenguaje Natural como Atribución de Autoría. Metodología, aquí se detalla el proceso típico que se sigue para el análisis automático de textos. La sección de Organización de los experimentos describe la forma en que se realizaron las pruebas de clasificación con el algoritmo de aprendizaje automático supervisado y por último, las Conclusiones derivadas de los resultados obtenidos.

2. Trabajos relacionados

En la tarea del Atribución de Autoría, Eder [4] utilizó pruebas controladas con textos de diferente longitud, idiomas y géneros. utilizó la métrica Delta propuesta por John F. Burrows [2], la cual mide la diferencia entre dos textos basándose en las palabras frecuentes. Extrajo muestras de textos formadas por 200, 400, 600, 800, ..., 20000 palabras. Las palabras se obtuvieron en forma aleatoria y en secuencia. Eder concluyó que el uso de ejemplos de 2500 palabras difícilmente proporcionará un resultado confiable.

En [14] aplicaron la técnica de los núcleos de secuencia de palabras en tarea de Atribución de Autoría. Evaluaron un conjunto de textos relativamente cortos de 50 periodistas que cubrían más de un tema aplicando el enfoque de cadenas de Markov. Crearon textos de 312, 625, 1250, 2500 y 5000 palabras y de ellos obtuvieron 1750, 3500, 7000, 14000 y 28000 caracteres. Los investigadores indicaron que la cantidad de texto para el entrenamiento de los algoritmos tiene más influencia que la cantidad de textos de prueba. Además concluyeron que se requieren aproximadamente entre 1250 y 5000 palabras en los textos de entrenamiento para obtener un rendimiento relativamente bueno.

Corney et al. [3] realizaron experimentos para identificar la autoría de correos electrónicos utilizando marcadores apropiados para este tipo de mensajes. Dichos mensajes contenían hasta 964 palabras, con una longitud promedio de 92 palabras. Utilizaron el algoritmo de aprendizaje automático supervisado SVM para discriminar entre las clases de autoría. Descubrieron que aproximadamente 20 mensajes con aproximadamente 100 palabras cada uno, deberían ser suficientes para discriminar la autoría en la mayoría de los casos. Mencionaron que el rendimiento del clasificador mejoró cuando agregaron un conjunto de características específicas de correo electrónico.

Luyckx y Daelemans [9] analizaron ensayos de un mismo tópico. Los ensayos contenían aproximadamente 1400 y provenían de 145 estudiantes. Utilizaron palabras y n-gramas de etiquetas POS con el algoritmo de aprendizaje automático SVM. En sus conclusiones, argumentaron que su propuesta mostró solidez al tratar con datos limitados, ya que de los 145 autores, casi el 50 % de los textos fueron clasificados correctamente.

En el ámbito de la Atribución de Autoría, Luyckx y Daelemans [10] evaluaron el efecto que tienen el número de autores y la cantidad de textos de entrenamiento en correos electrónicos, validaron su propuesta por medio de Aprendizaje Automático. Algunos de marcadores que utilizaron fueron longitud de sentencia, longitud de palabra, palabras función y de contenido y n-gramas de caracteres. En sus conclusiones indicaron que, como se esperaba, la precisión del clasificador se deterioró al aumentar el número de autores y disminuir el tamaño de los textos de entrenamiento. Además, indicaron que los n-gramas de caracteres son robustos a los cambios en el tamaño del conjunto de autores y el tamaño de los textos.

Feiguina y Hirst [6] evaluaron bigramas sintácticos como marcadores de estilo para la Atribución de Autoría en textos cortos. Utilizaron un analizador sintáctico y el algoritmo de aprendizaje automático SVM. Evaluaron novelas del

corpus Brontë Sister, mismas que dividieron en fragmentos de 200, 500 y 1000 palabras. Encontraron que los bigramas sintácticos fueron útiles al discriminar textos cortos del Corpus Brontë pero no para datos forenses simulados donde las distinciones sintácticas parecían menos necesarias. Argumentan que esto podría atribuirse a un desbalance de datos y al pequeño tamaño del conjunto de datos forense.

3. Metodología

La Figura 1 muestra el esquema general propuesto para el análisis de cambio de estilo de escritura con enfoque de aprendizaje automático supervisado. Las novelas evaluadas se descargaron del proyecto *Gutenberg*[1], éstas incluyen información de la editorial, datos biográficos del autor, semblanzas entre otras cosas. En el preprocesamiento se eliminó toda esa información así como las oraciones de uno y dos palabras, ya que los 3-gramas de palabras, etiquetas POS y relaciones sintácticas requieren al menos tres palabras.

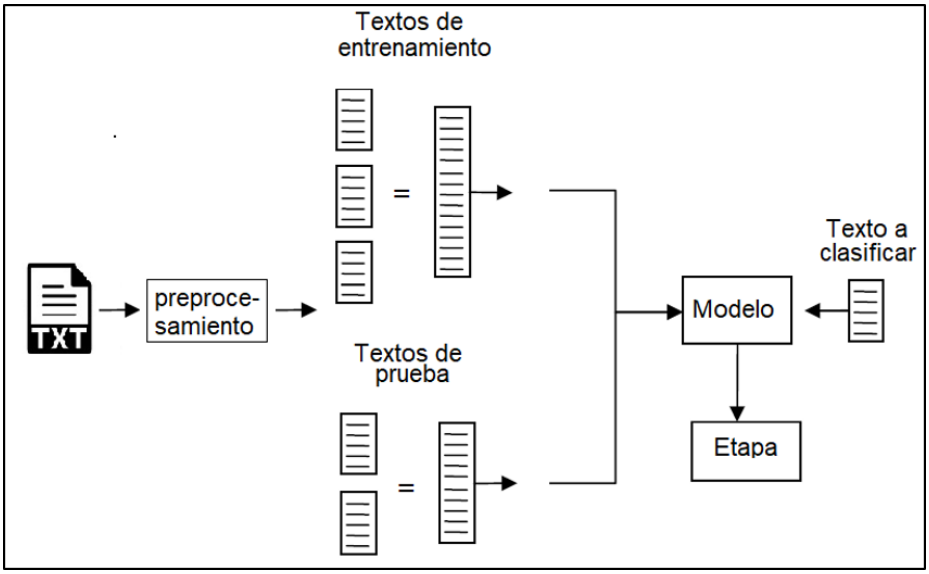


Fig. 1. Esquema de detección de cambio de estilo de escritura

La parte medular de esta investigación fue dividir cada novela en textos de diferente tamaños. Esta acción tiene dos consecuencias: primero, aumenta el número de textos disponibles y segundo, disminuye la cantidad de información en cada uno de ellos. Cada novela se dividió en 10 fragmentos con el mismo

[1] `https://www.gutenberg.org/`

número de sentencias. La notación para referir a cada tamaño de texto fue la siguiente: **1** fragmento que contiene la novela completa, **2** fragmento que contiene la mitad de una novela, **3** fragmento que contiene un tercio de una novela, así sucesivamente. La Tabla 1 muestra el número de sentencias del autor Booth Tarkington. Para el resto de los autores se aplicó el mismo procedimiento.

Tabla 1. Distribución de sentencias en novelas de Booth Tarkington.

Novela	**Tamaño del texto**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Canaan	4,598	2,299	1,532	1,149	919	766	656	574	510	459
Gentleman	5,350	2,675	1,783	1,337	1,070	891	764	668	594	535
Penrod	3,841	1,740	1,160	870	768	640	548	480	426	384
Seventeen	3,917	1,958	1,305	979	783	652	559	489	435	391
Turmoil	5,892	2,946	1,964	1,473	1,178	982	841	736	654	589
Vanrevels	2,802	1,401	934	700	560	467	400	350	311	280

Después de preprocesamiento, se realizó el etiquetado y el análisis sintáctico de cada texto. Las etiquetas POS se generaron con el etiquetador de la herramienta NLTK[2] y el análisis sintáctico se llevó a cabo con Stanford Parser[3]. Cuando se utilizan n-gramas se debe definir el valor de n, dicho valor indica el número de términos del n-grama. Valores comunes de n son 1, 2, 3, 4 y 5. Estudios previos reportaron que el más apropiado para n es 3: detección de plagio [1], atribución de autoría [5,11,15], categorización de textos [13] e identificación de autores[8].

Por otro lado, debe establecerse la frecuencia mínima con la que debe aparecer un n-grama. Las frecuencias 1 y 2 generan una gran cantidad de marcadores y contribuyen muy poco o nada al estilo de escritura del autor. En esta investigación se trabajó con n-gramas de frecuencia mayor o igual a 3. Para obtener los n-gramas de caracteres, palabras y etiquetas POS se utilizó el programa *text2ngram*[4], un software libre bajo licencia GPL. Dicho programa requiere como parámetros el tipo de n-grama, la longitud, la frecuencia y el texto de entrada. Los 3-gramas de relaciones sintácticas se generaron con un programa desarrollado en Python [12], este programa genera tres tipos de n-gramas sintácticos: de palabras, etiquetas POS y relaciones de dependencia; requiere como parámetros el valor de n y el archivo que contiene la información sintáctica de cada las oraciones. Además, se programó una rutina en Python para realizar el conteo de 3-gramas de cada sentencia, la frecuencia se establece dentro del propio código. La Tabla 2 muestra la información generada en esta etapa. Información similar se generó para los 3-gramas de caracteres, etiquetas POS y relaciones de dependencia.

[2] https://www.nltk.org/

[3] `https://nlp.stanford.edu/software/lex-parser.shtml`

[4] `https://homepages.inf.ed.ac.uk/lzhang10/ngram.html`

Tabla 2. Total de 3-gramas de palabras por autor.

Autor	Tamaño de texto									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
BT	2,762	1,679	1,215	1,020	878	771	714	660	593	572
CD	19,896	13,068	10,138	8,503	7,420	6,703	6,176	5,750	5,400	5,100
ER	3,693	2,423	1,847	1,511	1,284	1,138	1,042	965	881	824
FM	7,488	4,742	3,609	2,998	2,565	2,327	2,082	1,911	1,805	1,646
GM	5,285	3,232	2,416	1,951	1,690	1,495	1,321	1,191	1,108	1,060
GV	2,841	1,617	1,131	914	725	637	559	478	429	402
IM	8,761	5,554	4,176	3,534	2,965	2,631	2,401	2,248	2,052	1,979
JB	3,982	2,393	1,794	1,444	1,235	1,077	952	871	804	740
LT	2,601	1,483	1,056	780	669	566	471	396	358	339
MT	5,785	3,821	2,961	2,656	2,326	2,072	1,918	1,864	1,706	1,636

Posteriormente se crearon matrices *término-documento.* para grupo de 3-gramas. En estas matrices, las filas representan documentos, las columnas representan 3-gramas y las celdas contienen la frecuencia. La Tabla 3 muestra una vista parcial de la matriz de de caracteres en las novelas del autor Booth Tarkington.

Tabla 3. Matriz término-documento de Booth Tarkington.

Novela	3-gramas de caracteres									
	the	**and**	**ing**	**her**	**you**	**his**	**ere**	**tha**	**was**	**she**
Gentleman	8,977	4,307	3,018	2,465	1,525	1,652	1,820	1,421	1,393	1,095
Vanrevels	5,270	2,358	1,905	1,630	1,024	1,040	924	1,036	934	755
Canaan	6,370	2,913	2,420	1,873	1,479	1,487	1,287	1,250	1,117	933
Ramsey	2,929	1,577	1,210	1,027	789	646	642	698	606	447
AliceAdams	5,147	2,511	2,803	2,761	2,412	1,176	1,068	1,232	880	1,551
Julia	4,729	2,438	2,081	2,316	1,320	1,155	996	1,073	1,009	1,156

Algunos 3-gramas pueden tener frecuencias muy altas y en consecuencia predominan sobre las frecuencias bajas. Previendo este fenómeno, se realizó una normalización de las frecuencias. Existen diversos métodos de estandarización. Aquí se aplicó el método que consiste en extraer la media del grupo del valor de cada variable y dividir el valor resultante por la desviación estándar. Existen otros métodos para normalizar como el valor *tf-idf* entre otros.

4. Entorno experimental

La colección de novelas de cada autor se ordenó de forma cronológica con base en el año de publicación, de la más antigua a la mas reciente. Tomando una novela como punto medio en el tiempo, las novelas que la preceden pertenecen a la etapa 1 y las que la suceden pertenecen a la etapa 2. El conjunto de autores evaluados se muestran en la Tabla 4. De aquí en adelante, los autores se refieren a través de la leyenda que se muestra en el Cuadro.

Tabla 4. Conjunto de autores evaluados

Autor	Leyenda
Booth Tarkington	BT
Charles Dickens	CD
Edgar Rice	ER
Frederick McDonald	FM
George McDonald	GM
George Vaizey	GV
Iris Murdoch	IM
John Buchan	JB
Louis Tracy	LT
Mark Twain	MT

En Aprendizaje Automático los datos se dividen en conjuntos de entrenamiento y prueba, en estos experimentos la proporción de cada conjunto fue 66 % y 33 % respectivamente. La regresión logística es un algoritmo de clasificación supervisada que se utiliza para predecir una variable dependiente categórica. En los experimentos se utilizó la implementación de *scikit-learn*[5]. En cada una de las dos etapas el número textos es el mismo, por lo que la métrica *exactitud* resulta apropiada para medir la eficiencia del clasificador [7]. La exactitud indica la proporción de textos clasificados correctamente. Dado que la clasificación es binaria, la probabilidad de asignar al azar un texto a la etapa correcta es de 50 %, este valor se consideró la línea de base.

5. Resultados

5.1. Promedios generales de exactitud por autor

En la categoría de 3-gramas de caracteres, la Figura 2 muestra que los autores mejor clasificados fueron BT y ER al superar el 90 % de exactitud. Por otro lado, el autor GV muestra el promedio más bajo al no superar la línea base. El resto de los autores muestran valores que oscilan entre el 70 % y el 80 %. Aparentemente para algunos autores, disminuir la cantidad de información en los textos no fue un factor relevante.

La Figura 3 muestra la exactitud de los autores utilizando 3-gramas de palabras. La mayor parte de los autores superan el 70 % de exactitud. Al comparar con 3-gramas de caracteres, se observa que autores como GB obtuvieron mejor resultado mientras que otros como ER y BT disminuyeron sus promedios de exactitud.

La Figura 4 muestra los resultados en la categoría de etiquetas POS. La tendencia en los porcentajes de clasificación correcta se mantiene. Sin embargo, contrario a lo obtenido en 3-gramas de caracter y palabras, el autor GM presenta una disminución significativa ya que no supera el 60 %. En este punto es posible

[5] `https://scikit-learn.org/stable/`

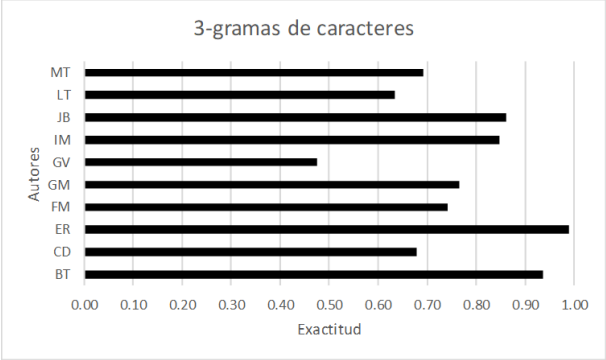


Fig. 2. Exactitud promedio en 3-gramas de caracter

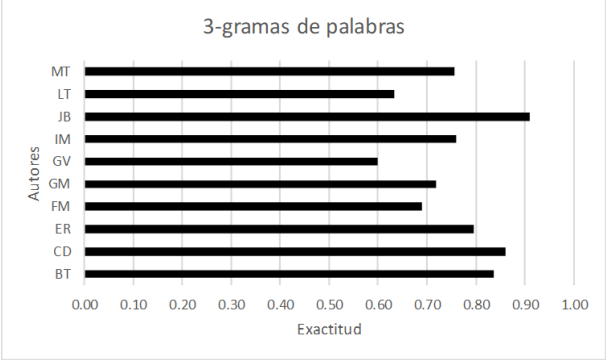


Fig. 3. Exactitud promedio en en 3-gramas de palabras

observar que los autores GV y LT son los que mantienen resultados muy cercanos a la línea base.

La Figura 5 muestra los resultados que se obtuvieron en los 3-gramas de relaciones de dependencia. Destaca que con excepción de LT, todos los autores superan holgadamente el 70 % de exactitud. Contrario a la tendencia mostrada en los otros 3-gramas, GV mostró un incremento significativo con exactitud cercana al 80 %. Algo similar ocurrió con IM, logrando casi el 100 % de exactitud. En general, en esta categoría de 3-gramas todos los autores mejores sus porcentajes independientemente del tamaño del texto.

5.2. Exactitud por tamaño de texto

La Tabla 5 muestra los resultados en 3-gramas de caracteres. Los autores mejor clasificados son BT y ER, BT obtiene 96 % en bloques de tamaño 4 y en el resto supera el 90 %, mientras que ER logra 100 % en los primeros cuatro bloques. Los casos atípicos son CD, GM, GV y LT, quienes logran sus mejores

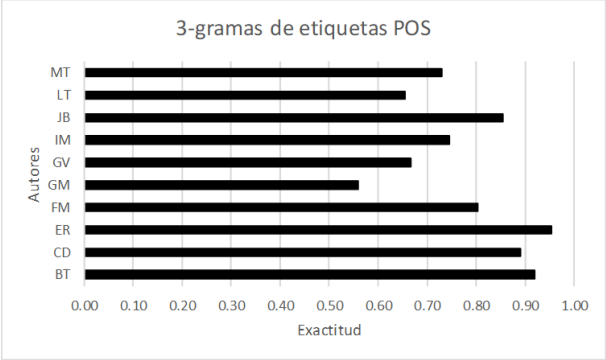


Fig. 4. Exactitud promedio en 3-gramas de etiquetas POS

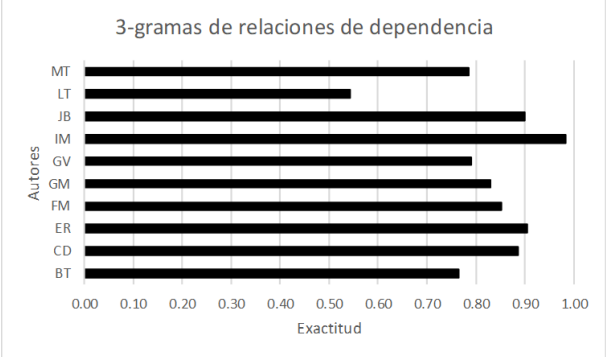


Fig. 5. Exactitud promedio en 3-gramas de relaciones de dependencia

resultados a partir de bloques de tamaño 7. También se observa que GV no superó el 50 % de exactitud en ninguno de los distintos bloques. El resto de los autores muestran mejores resultados en los bloques de tamaño 1 y 2, siendo éstos los que poseen mayor cantidad de información.

En la Tabla 6 se observa que en los 3-gramas de palabras todos los autores logran porcentajes más altos en textos de tamaño 1 y 2. Por otro lado, CD, IM y JB tienen resultados constantes a traves de los distintos tamaños. En el resto de los autores la exactitud disminuyó de forma gradual al reducir la cantidad de información.

La Tabla 7 muestran los resultados en 3-gramas de etiquetas POS. Los autores BT, ER y JB son los que muestran los porcentajes de clasificación más altos. El autor FM logra máxima exactitud en bloque de tamaño 4, mientras que el resto de los autores lo hace en bloques de texto de tamaño 1 y 2. Un caso peculiar se observó con CD. En el bloque 1 logra 72 % de exactitud y en los bloques 2, 5 y 8 un 92 %. Además, en el resto de los bloques supera el 90 %.

Tabla 5. Exactitud por tamaño de texto en 3-gramas de caracteres.

Autor	Tamaño de texto									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
BT	94	92	94	**96**	94	94	93	94	91	94
CD	56	67	65	65	67	69	72	**74**	72	72
ER	**100**	**100**	**100**	**100**	97	**100**	99	99	96	98
FM	61	69	76	76	74	76	**79**	75	78	77
GM	72	78	76	78	77	77	**79**	76	77	76
GV	**50**	**50**	48	46	46	49	47	44	48	47
IM	83	**92**	85	86	83	83	86	84	85	81
JB	**89**	86	87	82	84	85	87	87	88	87
LT	**72**	69	63	64	63	57	65	64	61	57
MT	61	58	72	69	72	71	69	72	73	**76**

Tabla 6. Exactitud por tamaño de texto en 3-gramas de palabras.

Autor	Tamaño de texto									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
BT	**100**	97	91	86	82	79	76	81	72	73
CD	**89**	**89**	87	86	87	85	88	85	84	80
ER	**89**	83	**89**	82	81	81	78	78	67	68
FM	**78**	72	65	65	69	69	72	64	67	69
GM	**78**	72	61	76	77	73	**78**	67	70	67
GV	61	**67**	61	61	46	59	60	63	56	66
IM	72	**81**	78	75	73	75	78	74	77	76
JB	89	**92**	91	90	93	90	92	92	90	91
LT	72	**78**	65	68	61	68	58	59	51	53
MT	83	83	**85**	76	78	74	73	69	69	67

Tabla 7. Exactitud por tamaño de texto en 3-gramas de etiquetas POS.

Autor	Tamaño de texto									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
BT	**100**	**100**	96	97	97	88	87	85	85	84
CD	72	92	91	89	**92**	91	90	**92**	90	91
ER	**100**	**100**	94	96	97	97	95	92	92	91
FM	78	81	81	**83**	82	79	82	79	78	80
GM	**61**	47	56	53	56	56	60	57	56	57
GV	**72**	67	61	**72**	67	67	63	65	66	67
IM	**78**	69	76	71	76	74	75	74	75	77
JB	**100**	83	89	85	83	82	83	84	83	83
LT	67	**72**	63	68	69	63	62	61	64	65
MT	**78**	75	74	71	72	74	73	69	72	71

La Tabla 8 muestra que en la categoría de 3-gramas de relaciones de dependencia. La máxima exactitud de cada autor ocurrió en textos de tamaño 1 y 2. IM mostró una clasificación casi perfecto sin importar el tamaño de texto. LT

apenas superó la línea base, sin embargo logra su mejor resultado en bloques de texto de tamaño 9.

Tabla 8. Exactitud por tamaño de texto en 3-gramas de relaciones de dependencia.

Autor	**Tamaño de texto**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
BT	**83**	**83**	80	78	78	70	73	71	76	73
CD	89	**92**	91	86	90	86	87	90	88	87
ER	**100**	94	93	92	88	89	87	88	88	85
FM	89	**92**	85	89	83	81	85	85	81	82
GM	**89**	86	85	85	84	84	81	81	78	77
GV	**83**	78	80	79	78	79	79	80	78	76
IM	**100**	**100**	**100**	96	97	**100**	98	98	96	98
JB	**100**	92	94	92	87	90	83	87	89	86
LT	50	50	56	53	53	56	55	56	**61**	52
MT	78	**83**	78	81	78	82	79	76	77	74

6. Conclusiones

El análisis automático de estilo de escritura se refiere a la forma de un texto y no a su contenido. Por ello, se deben tomar en cuenta dos aspectos. Primero, el marcador estilo utilizado. Es recomendable recurrir a marcadores que en la medida de lo posible sean inmunes al tipo de documento y al tópico que estos tratan. En cierta forma, la frecuencia de uso de las palabras dan cuenta del tópico de documento y de alguna forma influirán en la frecuencia de los n-gramas de caracteres. Para superar esta barrera, se ha explorado el uso n-gramas de etiquetas POS y relaciones de dependencia. Las etiquetas POS o categorías gramaticales indican el tipo de palabra que se esta usando en el texto, como adjetivos, verbos o sustantivos. Así, el análisis de estilo mostrará con que frecuencia una persona utiliza categorías de palabras. Estos marcadores revelan patrones muy distintos a los que ocurren cuando el texto se analiza en su forma lineal. La información sintáctica de una oración se muestra en forma de árboles, que muestran como incluso palabras que aparecen distantes entre sí están relacionadas, lo que permite descubrir nuevos patrones. Los resultados de estos experimentos mostraron que el uso de la información sintáctica es una alternativa viable para crear marcadores de estilo.

El segundo aspecto a considerar es la cantidad de información que hay en los textos en términos de palabras. Los trabajos relacionados dan cifras exactas para la tarea de Atribución de Autoría, las cuales varían en función del tipo de marcador y el tipo de documento (correos, ensayos, novelas). Esta investigación se enfoca en el cambio de estilo de escritura a través del tiempo por medio de aprendizaje automático supervisado. El objetivo principal fue observar el efecto que tiene sobre el clasificador disminuir la cantidad de oraciones en cada uno de los textos. Cada autor tiene un estilo de escritura propio por lo que

una cifra que para un autor resultó favorable en algún tipo particular de n-grama puede no serlo para otro. Por otro lado, en algunos autores disminuir la cantidad de información no influyó de forma significativa en sus resultados. Es necesario realizar investigaciones más exhaustivas para garantizar que ciertos valores resultan apropiados para un análisis de cambio de estilo de escritura independientemente del autor.

Referencias

1. Barrón-Cedeño, A., Rosso, P.: On automatic plagiarism detection based on n-grams comparison. In: European conference on information retrieval. pp. 696–700. Springer (2009)
2. Burrows, J.: 'delta': a measure of stylistic difference and a guide to likely authorship. Literary and linguistic computing 17(3), 267–287 (2002)
3. Corney, M.W., Anderson, A.M., Mohay, G.M., de Vel, O.: Identifying the authors of suspect email (2001)
4. Eder, M.: Does size matter? authorship attribution, small samples, big problem. Digital Scholarship in the Humanities 30(2), 167–182 (2015)
5. Escalante, H.J., Solorio, T., Montes-y Gómez, M.: Local histograms of character n-grams for authorship attribution. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. pp. 288–298. Association for Computational Linguistics (2011)
6. Feiguina, O., Hirst, G.: Authorship attribution for small texts: Literary and forensic experiments (2007)
7. García, V., Mollineda, R.A., Sánchez, J.S.: Index of balanced accuracy: A performance measure for skewed class distributions. In: Iberian conference on pattern recognition and image analysis. pp. 441–448. Springer (2009)
8. Houvardas, J., Stamatatos, E.: N-gram feature selection for authorship identification. In: International conference on artificial intelligence: Methodology, systems, and applications. pp. 77–86. Springer (2006)
9. Luyckx, K., Daelemans, W.: Authorship attribution and verification with many authors and limited data. In: Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008). pp. 513–520 (2008)
10. Luyckx, K., Daelemans, W.: The effect of author set size and data size in authorship attribution. Literary and linguistic Computing 26(1), 35–55 (2011)
11. Posadas-Duran, J.P., Sidorov, G., Batyrshin, I.: Complete syntactic n-grams as style markers for authorship attribution. In: Mexican International Conference on Artificial Intelligence. pp. 9–17. Springer (2014)
12. Posadas-Durán, J.P., Sidorov, G., Batyrshin, I., Mirasol-Meléndez, E.: Author verification using syntactic n-grams. Working notes papers of the CLEF (2015)
13. Rahmoun, A., Elberrichi, Z.: Experimenting n-grams in text categorization. Int. Arab J. Inf. Technol. 4(4), 377–385 (2007)
14. Sanderson, C., Guenter, S.: Short text authorship attribution via sequence kernels, markov chains and author unmasking: An investigation. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. pp. 482–491 (2006)
15. Sapkota, U., Solorio, T., Montes, M., Bethard, S., Rosso, P.: Cross-topic authorship attribution: Will out-of-topic data help? In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. pp. 1228–1237 (2014)

ISSN 1870-4069

Desarrollo de ontologías agrícolas mediante el reúso de recursos semánticos

Fernando Pech-May

Instituto Tecnológico Superior de los Ríos, Tabasco, México
{fernando.pech@cinvestav.mx }

Resumen. En México, la agricultura juega un papel muy importante debido a la dependencia económica y alimentaria. Este sector genera grandes volúmenes de datos que no son procesados y están representados en forma de texto, tablas, etc, lo que carece de significado o valor para ser utilizados en la toma de decisiones por parte de empresarios agrícolas, organismos de gobierno o investigadores del área. Esto ha ocasionado que algunas organizaciones lideradas por la Organización de la Naciones Unidas para la Agricultura y Alimentación implementen tecnologías para estructurar y formalizar el conocimiento a través de los recursos semánticos; lo que ha llevado al desarrollo de una gran cantidad de recursos semánticos agrícolas. Sin embargo, dichos recursos han sido desarrollado con las características de un país en particular, lo que hace imposible utilizarlos en el nuestro. En este documento se analizan los recursos existentes con el objetivo de brindar una información relevante que permita ser útil para la adaptación y desarrollo de nuevos recursos semánticos de los cultivos más comunes en México. Asimismo se desarrollaron tres recursos semánticos adaptados a la región de Tabasco mediante el reúso de recursos de la Organización para la Agricultura y Alimentos.

Palabras clave: ontologías agrícolas, grafo de conocimiento, OWL FAO.

Development of Agricultural Ontologies through the Reúse of Semantic Resources

Abstract. In Mexico, agriculture plays a very important role due to economic and food dependence. This sector generates large volumes of data that are not processed and are represented in the form of text, tables, etc., which lacks meaning or value to be used in decision-making by agricultural entrepreneurs, government agencies or researchers of the area. This has caused some organizations led by the Food and Agriculture Organization of the United Nations to implement technologies to structure and formalize knowledge through semantic resources; which has led to the development of a large amount of semantic agricultural resources. However, these resources have been developed with the characteristics of

a particular country, which makes it impossible to use them in ours. This document analyzes the existing resources with the objective of providing relevant information that will be useful for the adaptation and development of new semantic resources of the most common crops in Mexico. Likewise, three semantic resources adapted to the Tabasco region were developed through the reuse of resources from the Food and Agriculture Organization.

Keywords: agricultural ontologies, knowledge graph, OWL FAO.

1. Introducción

La agricultura es el sector productivo más importante para distintos países debido a la dependencia alimentaria de millones de personas. Además, constituye un estímulo para potenciar el progreso y crecimiento productivo. Sin embargo, el sector afronta múltiples cambios y desafíos, como el incremento de la población, falta de inversión y la disponibilidad de algunos recursos naturales que son esenciales para el éxito o fracaso de los cultivos [13]. Esto hace necesario de la ayuda de tecnologías que permitan el desarrollo de agricultura sostenible y precisa. Bajo este contexto, distintas organizaciones, gobiernos y empresas, lideradas por la Organización de las Naciones Unidas para la Agricultura y Alimentación (FAO) han creado estrategias para la formalización de la información de los grandes volúmenes de datos agrícolas generadas por distintas fuentes como sensores de suelo, drones, estaciones meteorológicas, etc.

Todos estos datos no son procesados, y se encuentran representados en forma de texto, tablas, etc.; lo que carece de significado o valor para ser utilizados en la toma de decisiones por parte de empresas, gobierno o agricultores. Para que los datos sean de utilidad, es necesario formalizarlos para proporcionarle contexto o significado. Para la formalización, es necesario de la ingeniería del conocimiento, las tecnologías y herramientas de la web semántica, principalmente de las ontologías, para la representación del conocimiento en forma de grafos. Uno de los objetivos de la FAO es la formalización del conocimiento y la conjunción e integración de recursos agrícolas mediante la web semántica. Dichos recursos involucran ontologías, vocabularios, taxonomías, etc; estos elementos pueden ser de gran utilidad para el desarrollo de aplicaciones para cultivos sostenibles y precisos [4].

Actualmente se están realizando distintas investigaciones para el desarrollo de recursos semánticos que incluyen vocabularios, taxonomías y ontologías que comprenden distintos dominios de la agricultura y áreas relacionadas en distintos idiomas, tales como: Agrovoc[1], Tesauro agrícola de la biblioteca nacional de agricultura (NALT)[2], etc; éstos son vocabularios generalizados que contienen términos específicos de distintos subdominios agrícolas tales como fertilizantes, clima, pesticidas, etc.

[1] http://aims.fao.org/vest-registry/vocabularies/agrovoc
[2] https://agclass.nal.usda.gov/

También se han desarrollado grafos de conocimiento de cultivos específicos tales como Plant Ontology [3], AgriOnt [10], OntoAGroHidro [2], etc. Todos estos recursos semánticos han sido desarrollados y adaptados a zonas geográficas específicas europeas, lo que lo hace inadecuado para México debido a que contiene distintas características tales como el idioma, conceptos específicos de la región, entre otros.

Esto hace necesario el desarrollo de ontologías que brinden una información organizada respecto a los principales cultivos en México y que favorezca la toma de decisiones para usuarios interesados en las actividades agrícolas. En este artículo se presenta el desarrollo de 3 recursos semánticos mediante el reúso de ontologías; esto, con el propósito que sea utilizado para sistemas de soporte para la toma de decisiones agrícolas para mejorar su sostenibilidad y precisión en nuestro país.

2. Trabajos relacionados

La Web Semántica es una extensión de la web actual que tiene como propósito organizar y estructurar el conocimiento, de manera que sea entendida o comprendida por las computadoras [9]. Además, cuenta con una serie de herramientas, entre las que destacan las ontologías [12], que representan el conocimiento en forma de grafos mediante una jerarquía de conceptos con atributos y relaciones; para su diseño y/o desarrollo, existen lenguajes entre los que destacan RDF, RDFS y OWL [1].

El conocimiento estructurado permite nuevos descubrimientos científicos mediante el enlace de distintos conjuntos de datos, de los cuales la agronomía, agricultura, alimentación, ciencia de las plantas y biotecnología son beneficiadas. En el campo de la agricultura, las ontologías han sido utilizadas para la representación del conocimiento y para el desarrollo de distintas aplicaciones y que pueden clasificarse como: 1) sistemas basados en conocimiento (aplicaciones que realizan razonamiento y sugieren soluciones mediante bases de conocimiento), 2) sensores remotos (que integran datos a través de dispositivos remotos), 3) sistema de soporte de decisiones (que representan el conocimiento y sugieren recomendaciones) y 4) sistemas expertos (aplicaciones que toman decisiones a través del razonamiento de la información).

Actualmente existe una variedad de ontologías desde las más generales hasta las más específicas; asimismo, se han creado una serie de portales que permiten visualizar y administrar recursos semánticos agrícolas. Los trabajos relacionados han sido clasificados en dos categorías: 1) recursos semánticos agrícolas y 2) portales web agrícolas.

2.1. Recursos semánticos agrícolas

La adopción de tecnologías web semánticas depende de la disponibilidad de los recursos semánticos existentes. Los recursos semánticos para la agricultura

son recursos que utilizan tecnologías semánticas para describir el conocimiento recopilado por una organización o individuo. Actualmente existe una gran variedad de recursos semánticos del dominio agrícola, como son vocabularios, taxonomías, tesauros y grafos de conocimiento; a continuación, se explican las más relevantes.

- Agrovoc[3] es uno de los recursos de mayor tamaño. Contiene términos y conceptos sobre la agricultura, alimentos, nutrición pesca, silvicultura y medio ambiente; está disponible en 27 idiomas y cuenta con 35,000 conceptos y 40,000 términos; es compatible con datos enlazados abiertos (LOD, del inglés Link Open Data) y se encuentra enlazado con 16 recursos. Agrovoc contiene distintas irregularidades en las relaciones entre conceptos; sin embargo, es utilizado para crear nuevos recursos ontológicos y adaptarlas a una región en particular.
- Chinese Agricultural Thesaurus (CAT) [9] contiene 40 categorías principales, más de 63 mil conceptos e incluye más de 130 mil relaciones semánticas.
- Tesauro Agrícola de la Biblioteca Nacional de Agricultura (NALT)[4] contiene 128,25 términos agrícolas en inglés y español, 40 categorías, como la clasificación de cultivos y 63,000 conceptos de cultivos de leguminosa, frijol etc. También es compatible con los esquemas de datos abiertos vinculados y se encuentra integrado a otros recursos semánticos.
- Plant Ontology (PO) [3]. Relacionada con anatomía multiespecífica y desarrollada para la anotación de genes y fenotipos. Se encuentra en distintos idiomas. Contiene más de 2. 2 millones de anotaciones que unen los términos de PO a 110,000 objetos de datos que representan genes o modelos de genes, proteínas, ARN, germoplasma y loci de rasgos cuantitativos (QTL) de 22 especies de plantas.
- OntoAgroHidro [2]. Ontología de dominio creada por expertos investigadores con interés en los recursos hídricos, cambio climático y uso de la tierra. Cuenta con 6 principales clases de las que se desglosan subclases de otros, relaciones, propiedades y casos.
- Agricultural Ontology (AgriOnt) [10]. Proporciona una visión general del dominio de la agricultura, conceptos agrícolas y ciclos de vida entre semillas, plantas, cosecha, transporte y consumo; también proporciona relaciones entre los conceptos agrícolas y conceptos relacionados con el clima, condiciones del suelo y fertilizantes. Además, se compone por 447 clases y más de 700 axiomas.

2.2. Portales web agrícolas

Existen portales web que alojan un conjunto de recursos semánticos del dominio agrícola tales como:

- Agroportal [8]. Contiene 98 ontologías y tesauros; permite la anotación semántica, almacena y explora los datos enlazados.

[3] http://aims.fao.org/vest-registry/vocabularies/agrovoc
[4] https://agclass.nal.usda.gov/

- Crop Ontology [6]. Aloja un conjunto de ontologías enfocadas en el dominio agrícola, permite cargar diccionarios de rasgos para el mejoramiento del cultivo y la creación directa de ontologías; asimismo, permite consultar, crear, actualizar o eliminar ontologías.
- CIARD Ring [11]. Es un portal con servicios web semánticos que contiene 3201 conjuntos de datos y 5327 servicios de datos. Además, permite indexar una serie de herramientas de software para analizarlos.
- Vest[5]. Es un repositorio perteneciente a la FAO y Godan. Cuenta con 398 recursos y descripción de grafos con LOD. Además, permite hacer consultas en SPARQL [5].

3. Desarrollo de ontologías agrícolas

El desarrollo de ontologías requiere de un gran esfuerzo, expertos en el dominio del conocimiento y una gran cantidad considerable de tiempo. Sin embargo, existen distintos recursos semánticos que pueden reusarse para la creación de nuevos recursos agrícolas adaptados a una región en particular. En este trabajo se desarrollaron las ontologías del maíz y frijol mediante el reúso de recursos semánticos obtenidos en Crop Ontology. Asimismo, fue necesario el análisis y creación de distintos subdominios tales como cultivo, fertilizantes, pesticidas, recursos geográficos, los cuales se relacionan con los recursos antes mencionados. Parte de estos recursos fueron desarrollados por organizaciones antes mencionadas.

Para la reutilización de las ontologías, se analizaron su estructura e identificación de métricas (ver Cuadro 1) así como el código fuente para la identificación de etiquetas del lenguaje utilizado (*owl:class*, *rdfs:label*, etc.).

Tabla 1. Descripción de las métricas ontológicas.

Métricas Ontológicas	
Nombre	**Descripción**
Axiomas	Propiedades adicional que describen el comportamiento de una clase
Clases	Individuos/objetos con características en común
Propiedades de objeto	Descripción de características y atributos de un objeto
Individuos	Objetos particulares del dominio
Propiedades de anotación	Descripción más detallada de las clases
Subclases	Clase hija de otra clase

[5] http://aims.fao.org/vest-registry

3.1. Ontología del maíz

La ontología del maíz fue desarrollada con la finalidad de medir los rasgos del maíz a través de diferentes variables según el método o escala utilizada. La ontología proporciona nombres de rasgos de mejoradores armonizados, métodos de medición, escalas y variables estándar. Además, hace uso del recurso SKOS[6] para documentación de la transitividad, definición, etiquetado, asignación de acrónimos y asignación de prefijos. Asimismo, adopta términos de ontologías biomédicas abiertas. La Figura 1 muestra las clases principales de la ontología y en el Cuadro 2 se aprecia las métricas totales de la ontología. A continuación, se describe cada clase.

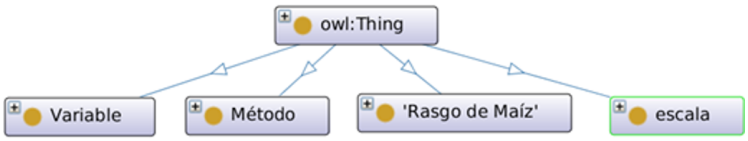


Fig. 1. Clases principales de la ontología del maíz; vista desde Protege [7].

- **Rasgos del maíz**. Esta clase (ver Figura 2) contiene características físicas y químicas de la planta tales como: 1) *rasgos agronómicos*, que se conforma por una lista de características físicas (altura de la oreja de la hoja, aspectos de la planta, número de orejas, número de granos, entre otros); 2) *rasgos de calidad*, que presenta una lista de características respecto a la calidad del grano del maíz y se establece mediante la constitución física para determinar la textura, dureza y la composición química (contenido del hierro, zinc, textura del grano, etc.); 3) *rasgos de estrés abiótico*, que se conforma por dos subclases que determinan los rasgos negativos de la adaptación de la planta ante distintas condiciones de estrés (gravedad de la hoja y de voladura); 4) *rasgos de estrés biótico*, subclases que determinan el impacto negativo de factores como hongos, virus y herbívoros que dañan la planta (gravedad del óxido común, incidencia del haz negro, incidencia de la pudrición del tallo, incidencia del virus de la raya del maíz, presencia de enfermedades fúngicas, severidad de la raya tropical, etc.); 5) *rasgos fenológicos*, proporciona una lista de características que determinan el ciclo de vida de la planta (intervalo de sedación de la antesis, tiempo de antesis, madurez, seda y senescencia de la hoja del oído); 6)*rasgos fisiológicos*, presenta una lista de características respecto a procesos químicos y físicos asociados a la vida de la planta (contenido de la clorofila, glucosa y prolina en el grano, ajuste osmótico, índice de agua, etc.); 7) *rasgos morfológicos*, se conforma por una serie de características respecto a los órganos que componen el cuerpo de la planta (tallo, hoja, raíces, etc.).

[6] https://skos.um.es/

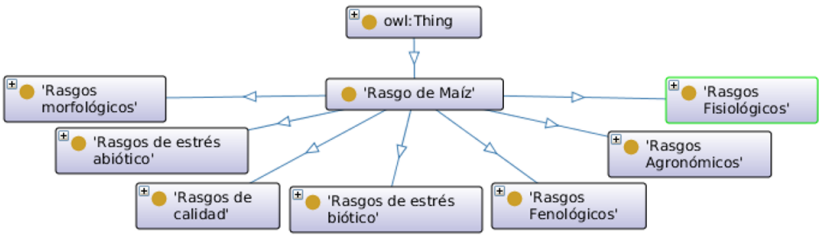


Fig. 2. Subclases de la clase "Rasgo de Maíz".

- **Método**. Clase que se conforma por algunos de los métodos de medición que permiten evaluar el cultivo e identificar problemas tales como: 1) conteo, subclase que se conforma por una serie de métodos que permiten el conteo de algunos rasgos de la planta (número de plantas dañadas por busseola y parcela, número de plantas afectada por el complejo de acrobacia de maíz por parcela, gorgojos del maíz en una muestra, etc.); 2) cálculo, se compone por una serie de métodos que permiten el cálculo de algunos de los rasgos de la planta (días para la antesis, senescencia, de la oreja, días hasta la madurez, pesos secos de la base, fecha de la senescencia de la hoja del oído, etc.); 3) estimación, subclases que determinan una estimación de algunos rasgos del maíz (fecha de sedación, vencimiento, etc.); 4) medición, se integra por una serie de métodos que permiten medir rasgos químicos y físicos de la planta (azúcar, sacarosa, glucosa, zinc, peso seco de la mazorca, etc.).
- **Escala**. Integrada por 7 subclases: 1) *tiempo*, permite establecer el día, mes y años de algunos de los métodos preestablecidos (estimación o cálculo de ciertos rasgos de la planta); 2)*nominal*, permiten la medición de rasgos físicos específicos de la planta (escala de la forma de la oreja, arreglo de hilera de granos de la hoja, color del tallo, color del grano, etc.); 3) *numérico*, contiene una serie de escalas numéricas (porcentajes, relación de la posición del oído, escala de proporción, número de rango, etc.); 4) *ordinal*, se conforma por una serie de intervalos que permiten medir los rasgos físicos y químicos de la planta (escala de senescencia, colores de la mazorca, posición del oído, textura del grano, germinación estándar, apertura de la borla, etc.); 5) *texto*, permite relacionar la estimación de la presencia de enfermedades fúngicas del maíz; 6) *código* y 7) *duración*, fungen como soporte en cuestión a métodos extras.
- **Variable**. Cuenta con más de 100 subclases que son datos susceptibles de ser modificados, de acuerdo con la relación entre cada una de sus relaciones.

3.2. Ontología del frijol

La ontología fue desarrollada para medir los rasgos del frijol, métodos de medición y escalas. Se utiliza SKOS para la transitividad, etiquetados, asignación acrónimos, etc. Consta de 3 clases principales (ver Figura 3) denominadas: 1)

escala, 2) rasgos del frijol común y 3) método. En el Cuadro 3 se aprecia las métricas totales de la ontología. A continuación, se describe cada uno.

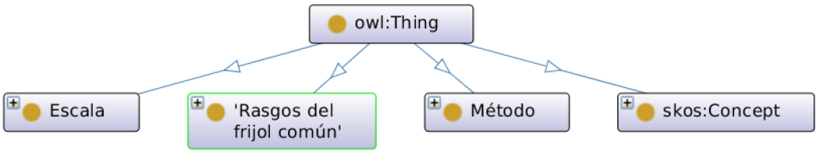


Fig. 3. Clases principales de la ontología del Frijol.

- **Rasgos del frijol común**. Esta clase se conforma por subclases (ver Figura 4) que definen características físicas y químicas de la planta del frijol, tales como: 1) *rasgos agronómicos*, define una lista de características físicas (altura de la planta, aspectos de la planta, vainas por planta, tamaño de semilla, etc.); 2) *rasgos de calidad*, describe características o componente respecto a la calidad de la semilla del frijol (hierro, fósforo, proteína, Zinc, etc); 3) *rasgos bioquímicos*; determinan la composición química de la planta (presencia del marcador de ADN unido al gen bc-3, bgm-1, BGYMV, etc); 4) *rasgos de estrés biótico*, determinan el impacto negativo de factores como hongos, virus y herbívoros que dañan la planta (gusano de tallo, virus de la necrosis del mosaico común, mancha de hoja harinosa, etc.); 5) *rasgos fenológicos*, proporciona características que determinan el ciclo de vida de la planta (días para florecer, días hasta la madurez fisiológica y etapa de crecimiento); 6) *rasgos fisiológicos*, subclase que presenta una lista de características de procesos químicos y físicos asociados a la vida de la planta (discriminación de isótopos de carbono del grano, pérdida de hojas, cantidad de nódulos efectivos en frijol arbustivo, contenido de calcio de semilla en campo, etc.); 7) *rasgos morfológicos*, conforma una serie de características que componen el cuerpo de la planta (vaina, hoja, raíces, etc.).

- **Método**. La clase se conforma con más de 100 métodos de medición que permiten evaluar el cultivo y a su vez identificar problemas (método agronómico de eficiencia del agua, longitud basal de la raíz, orientación del pico de la vaina, etc.).

- **Escala**. Clase derivada de la clase método. Se encuentra integrada por más de 100 tipos de escalas tales como porcentajes, intervalos del 1 al 9, evaluación de escalas en categorías, gramos, entre otras.

3.3. Ontología de subdominios relacionados

Esta ontología fue realizada para albergar características locales en relación con el clima, la cultura, idiomas y las variedades de plantas locales relacionadas

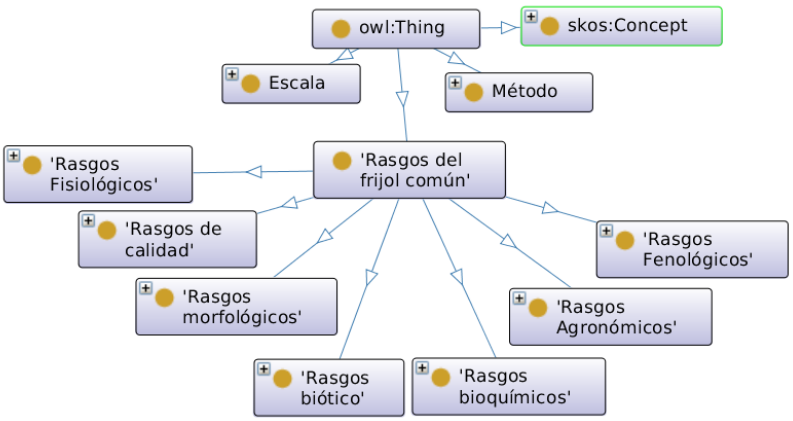


Fig. 4. Subclase de la clase "Rasgos del frijol común".

con actividades agrícolas. Contiene subdominios para distintos cultivos. Entre los subdominios de mayor relevancia creados son fertilizantes, pesticida y ambiente (ver Figura 5). En el Cuadro 4 se aprecia las métricas totales de la ontología. A continuación, se describen las clases principales.

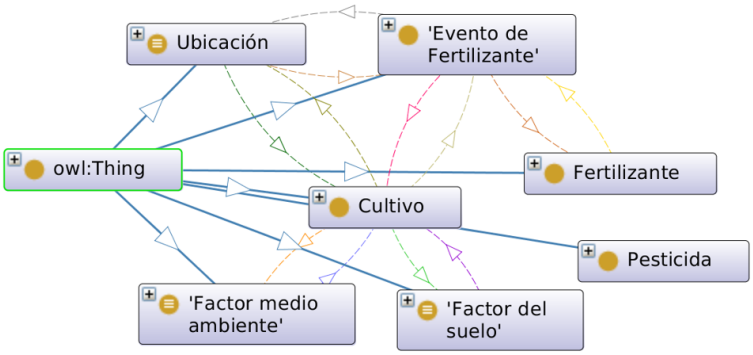


Fig. 5. Clases principales de la ontología de subdominios.

- **Fertilizantes**. Los fertilizantes se definen como un material orgánico o inorgánico que se agrega a un suelo para suministrar uno o más nutrientes vegetales esenciales para la planta. La ontología contiene instancias que establecen información relevante respecto a fertilizantes existentes tales como: composta, abono encalado, abono orgánico, abono verde, etc. También se define una clase denominada "evento de fertilizantes", que contiene concep-

tos tales como síntoma, cantidad, tiempo, problema creciente, método de cosecha, etc (ver Figura 6)

- **Ambiente**. Los factores y recursos ambientales garantizan la calidad de los cultivos; en esta ontología aborda estos conceptos clasificándolos en dos clases: 1) factores de medio ambiente, proporciona información sobre factores externos basados en cultivos o granjas, como humedad, luz solar, viento, CO2 y fuente de agua; 2) factores de suelo, incluye todas las condiciones necesarias del suelo relacionadas con los cultivos (tipo de suelo, valor del ph, textura y características del suelo, etc.).
- **Pesticida**. Contiene una lista de pesticidas, orgánicos e inorgánicos, que pueden aplicarse a los cultivos.

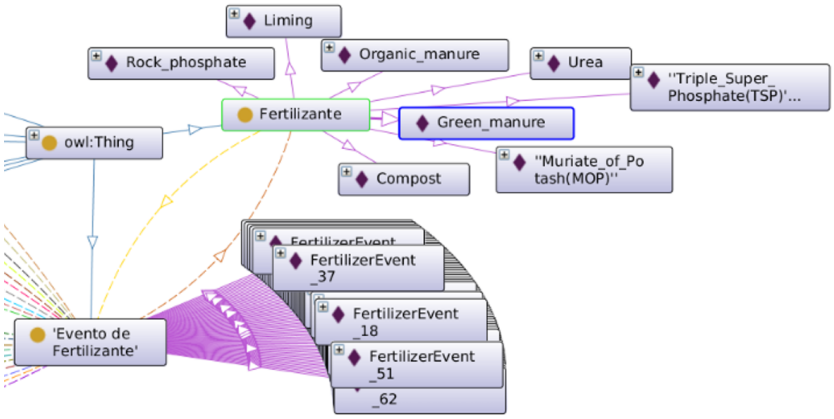


Fig. 6. Instancias de la subclase "Fertilizante".

Para el desarrollo de los recursos semánticos (ontologías) antes mencionadas, se reusaron una variedad existente. En la ontología del maíz se obtuvo más de 11500 axiomas, 1103 clases y más de 2700 subclases (ver Cuadro 2). Respecto a la ontología del frijol, 1070 clases, 1434 subclases y 1056 individuos (ver Cuadro 3). Cabe destacar que la ontología de subdominios aún está en desarrollo, hasta el momento se han obtenido 1070 clases, 1435 subclases y 1056 individuos (ver Cuadro 4).

4. Conclusiones

La web semántica juega un papel muy importante para la estructuración del conocimiento; proporciona tecnologías y herramientas para la formalización del conocimiento. Esto ha ayudado a formalizar los datos agrícolas obtenidos desde distintas fuentes como sensores de suelo, drones, estaciones meteorologías, etc.

Tabla 2. Métricas de la ontología del maíz.

Métricas Ontológicas del Maíz	
Métrica	**Cantidad**
Axiomas	11542
Axiomas lógicos	3810
Axiomas de declaración	2196
Clases	1103
Propiedades de objeto	3
Individuos	1088
Propiedades de anotación	9
Subclases	2722

Tabla 3. Métricas de la ontología del frijol.

Métricas Ontológicas del Frijol	
Nombre	**Cantidad**
Axiomas	8315
Axiomas lógicos	2490
Axiomas de declaración	2131
Clases	1070
Propiedades de objeto	2
Individuos	1056
Propiedades de anotación	8
Subclases	1434

Tabla 4. Métricas de la ontología de subdominios.

Métricas Ontológicas de Subdominios relacionados	
Nombre	**Cantidad**
Axiomas	5591
Axiomas lógicos	4041
Axiomas de declaración	977
Clases	83
Propiedades de objeto	191
Propiedades de datos	45
Individuos	657
Propiedades de anotación	4
Subclases	33

Por otra parte, a pesar de la gran cantidad de recursos semánticos disponibles, en México aún existe pocos trabajos en la formalización del conocimiento agrícola, lo que podría ayudar en la desarrollo de aplicaciones para el cultivo sostenible y preciso. La presencia de la web semántica en el sector agrícola abona a esta nueva etapa del campo en México y se manifiesta en diversos aspectos tales como: 1) la evolución en la aceptación de estas tecnologías por parte de los empresarios y técnicos es consecuencia del cambio generacional, la profesionalización del sector y la penetración social de las ciencias compu-

tacionales y 2) la madurez de las aplicaciones avanzadas de gestión empresarial (gestión de clientes, de insumos, control de producción, etc.) que han pasado de ser productos solo disponibles para grandes corporaciones a ser accesibles por cooperativas y empresas agroalimentarias de cualquier tamaño.

Referencias

1. Arenas, M., Ugarte, M.: Designing a query language for rdf: Marrying open and closed worlds. ACM Trans. Database System 42(4), 21:1–21:46 (oct 2017), http://doi.acm.org/10.1145/3129247
2. Bonacin, R., Nabuco, O.F., Ivo, P.J.: Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery. Future Generation Computer Systems 54, 423–434 (2016)
3. Cooper, L., Jaiswal, P.: The Plant Ontology: A Tool for Plant Genomics, vol. 1374. Humana Press, New York, NY (2016)
4. Drury, B., Fernandes, R., Moura, M.F., de Andrade, A.: A survey of semantic web technology for agriculture. Information Processing in Agriculture 6, 487–501 (2019)
5. DuCharme, B.: Learning SPARQL. O'Reilly Media, Inc. (2013)
6. Durufle, H., Laporte, M.A., Matteis, L., Valette, L., Agbona, A., Agrama, H., Agrawal, S.K.: The crop ontology: Improving the quality of 18 crop trait dictionaries for the breeding management system and adding new crops. In: General Research Meeting of the Generation Challenge Programme (2014)
7. Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C.: A Practical Guide To Building OWL Ontologies With The Protege-OWL Plugin. University of Manchester, 1 edn. (2004), http://home.skku.edu/~samoh/class/sw/ProtegeOWLTutorial.pdf
8. Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Yeumo, E.D., Emonet, V., Graybeal, J., Larmande, P.: Agroportal: A vocabulary and ontology repository for agronomy. Computers and Electronics in Agriculture 144, 126–143 (2018)
9. Liang, A., Sini, M.: Mapping agrovoc and the chinese agricultural thesaurus: definitions, tolos, procederus. New Review of Hypermedia and Multimedia pp. 51–62 (2007)
10. Ngo, Q.H., Le-Khac, N., Kechadi, T.: Ontology Based Approach for Precision Agriculture, vol. 11248. Springer, Cham (2018)
11. Pesce, V., Maru, A., Archer, P., Malapela, T., Keizer, J.: Setting up a global linked data catalog of datasets for agriculture. In: Garoufallou, E., Hartley, R.J., Gaitanou, P. (eds.) Metadata and Semantics Research. pp. 357–368. Springer International Publishing (2015)
12. Saha, G.: Web ontology language (owl) and semantic web. Ubiquity 2008(September), 1:1–1:1 (Sep 2007), http://doi.acm.org/10.1145/1295289.1295290
13. Trendov, N., Varas, S., Zeng, M.: Tecnologías digitales en la agricultura y las zonas rurales (2019), http://www.fao.org/3/ca4887es/ca4887es.pdf

ISSN 1870-4069

Identification of Static and Dynamic Signs of the Mexican Sign Language Alphabet for Smartphones using Deep Learning and Image Processing

Bella Martinez-Seis, Obdulia Pichardo-Lagunas, Edgar Rodriguez-Aguilar, Enrique-Ruben Saucedo-Diaz

Instituto Politecnico Nacional, Interdisciplinary Professional Unit for Engineering and Advanced Technologies (UPIITA-IPN), Mexico
{bcmartinez,opichardola}@ipn.mx

Abstract. The Mexican Sign Language (MSL) is a language with its own syntax and lexicon. It is used by the deaf people, who use it to express thoughts, ideas and emotions. However, most of hearing people are unable to understand this language. The alphabet of any Sign Language (SL) is composed of signs where each sign corresponds to a letter of the alphabet of the dominant language in the region, for example, Spanish or English. Most signs of a signed alphabet are static, that means, they are only composed by the configuration of the hands. However, there are letters that are represented by signs that include movement. The present work proposes a system that, using artificial vision techniques and image processing, identify the 27 letters -including dynamic and static signs- of the Spanish alphabet in a mobile application. To solve the problem of sign identification it was used a combination of image processing techniques and deep learning. Canny and Camshift algortihms was implemented for the recognition of edges and trajectories in signs with movement. Once the characteristics were identified, the K-means and Tensorflow algorithms were used to classify the signs. The system achieves a 92% accuracy in the alphabet sign detection.

Keywords: Mexican sign language, automatic sign detection, trajectory tracking.

1 Introduction

The translation of any oral language implies a major challenge, since not all the structures of the original language can be correctly expressed under the rules of the second language. The challenge acquires a new level of difficulty when it comes to a sign language. The automatic translation of sign language can be done based on the oral language -normally are used written texts- or starting from a set of images that can be represent words or letters.

The alphabets of the different sign languages in the world are usually based on the dominant oral language used in their geographical region.

These alphabets use a different symbol for each letter, the symbols used can be static or dynamic. The static symbols only involve the configuration of the hand, the dynamic signs also include movement. The translation of sign language is usually approached as image processing. The identification of static signs is a common task but is not the case of dynamic signs, the identification of movement implies the use of different algorithms with greater degree of difficulty.

The present work proposes the identification of the signs that compose the Mexican Sign Language alphabet including the dynamic signs. For this job it is proposed use techniques of image processing and deep learning . Unlike other works the system reduces the tasks and processing time in order to perform this process on a mobile device.

2 Computational Tools for Automatic Sign Detection

To implement the automatic identification of the alphabet of a sign language it is essential to use different types of images or videos if the identification of dynamic signs is included.

The images or videos to be used must be pre-processed to identify their characteristics and then be classified in a category. The algorithms used in this pre-processing are described below.

2.1 Image Processing

Image processing is a technique that allows to improve and identify the characteristics of the images obtained for various applications. There are different techniques in image processing that have been developed during the last decades, however, the rapid advance in hardware and software technologies has allowed a greater development in recent years.

Digital image processing generally refers to the processing of a two-dimensional image by a digital computer [1]. A digital image is represented as a matrix of real numbers represented by a finite number of bits. Image processing techniques range from image pre-processing, image enhancement, feature extraction or image classification.

The pre-processing of images is responsible for correcting errors that may be related to the geometry or the brightness values of the pixels. These errors are corrected using mathematical models in order to improve the visual characteristics of the image or to represent it in a model that suits the possible applications[2]. The process of improving the image seeks to accentuate the characteristics of the image that allow a subsequent analysis. Some of these tasks are contrast and edge identification, noise filtering or scaling. The improvement process emphasizes the specific characteristics of the image. Image segmentation is the process that subdivides an image into its constituent parts or objects. The subdivision of the image depends largely on the problem that is being solved.

Image processing is a basic tool in the identification of symbols in different sign languages. Konwar et al. identified American Sign Language (ASL) using

the HSV color model to detect the shape of the hand using skin color and edge detection[3]. Another work is the one made by Adithya et all in 2013 employed ANN forward backward algorithm to automatically recognize alphabets and numbers of Indian Sign Language with 91.1% of accuracy [4].

Image Segmentation The goal of segmentation is to identify the objects of interest in an image. Image threshold techniques are used for image segmentation [5]. The best threshold is the one that selects all the pixels of the object and assigns them to "black". The threshold can be defined as the gray scale assignment in the binary set $\{0, 1\}$:

$$S(x,y) = \left\{ \begin{array}{l} 0 \text{ if } g(x,y) < T(x,y) \\ 1 \text{ if } g(x,y) \geq T(x,y) \end{array} \right\}, \tag{1}$$

where $S(x, y)$ is the value of the segmented image, $g(x, y)$ is the gray level of the pixel (x, y) and $T(x, y)$ is the threshold value at the coordinates (x, y). In the simplest case $T(x, y)$ is coordinate independent and a constant for the whole image. It can be selected, for instance, on the basis of the gray level histogram. When the histogram has two pronounced maxim, which reflect gray levels of object(s) and background, it is possible to select a single threshold for the entire image[6].

Segmentation of images involves sometimes not only the discrimination between objects and the background, but also separation between different regions. The segmentation of images is of vital importance in the recognition of signs. Identifying the components in an image will later allow to evaluate characteristics such as manual configuration.

Feature Extraction Feature extraction techniques are developed to extract characteristics that allow the classification of objects. The characteristics are the specific elements that describe an object can be: color, size, shape, location, etc. Segmentation techniques are used to isolate the objects of a plane to be later grouped according to their characteristics[7]. Once the pre-processing and segmentation has been carried out, some feature extraction technique is applied to the segments to subsequently apply classification techniques. The phase of extraction of characteristics is one of the most relevant processes in the recognition systems and has an important impact on the efficiency of the systems. The selection of techniques for extracting characteristics must be made taking into account various factors[8].

Canny Edge Detection To be able to detect the edges with the Canny method, the image P is first blurred and then it is convolved with a pair of orthogonal filters, like the Prewitt, to create two images H and V that contain the horizontal and vertical derived directions. For a pixel (i,j) with orientation θ_{ij} and magnitude i_j the gradient is calculated as follows[9]:

$$\theta_{ij} = \arctan\left[\frac{v_{ij}}{h_{ij}}\right], \tag{2}$$

$$a_{ij} = \sqrt{h_{i\,j}^2 + v_{i\,j}^2}. \tag{3}$$

Track Tracking There are different ways to calculate or follow the trajectory of an object or several. There are several algorithms that rely exclusively on artificial vision, using the camera of a device as the only source of information. This set of algorithms is known as video tracking. To carry out video tracking, an algorithm analyzes sequential video frames and determines the movement of the targets between the frames. Each of the algorithms has its strengths and weaknesses. The use or application sought must be considered before choosing the appropriate algorithm.

One of this options is the Meanshift Algorithm, the objective of this algorithm is to select a region of interest and evaluate it in a window where we look for the average value and move the center of our region of interest to this point. The algorithm continues this process until it finds a convergence of the area of interest. It must be assumed that every pixel that directs us to the same local maximum of convergence is part of the same region. That is why we do the same process in parallel, sweeping the image and finding how many local maxim exist. From this we label the image according to the maximum to which each pixel belongs.

Other option is the Camshift Algorithm that corrects the errors that may occur with the previous method. In the Meanshift algorithm the window always keeps the same size, even when the object moves away from the camera, which sometimes causes an inconsistent result. The size of the window needs to adapt to the size and rotation of the lens, CAMshift (Continuously Adaptive Meanshift) is responsible for addressing the problem. The CAMshift algorithm first applies Meanshift. Once Meanshift converges, update the size of the window and calculate the orientation of the best ellipse that fits in it. Again apply Meanshift with the new window and the location of the previous window. The process continues until the required accuracy is achieved[10].

Image Classification Image classification refers to the process of computer vision that allows you to group images according to their content. An image classification algorithm could indicate if an image contains the shape of a particular object. The robust classification of images is a challenge in the applications of artificial vision. The techniques used in this field will be described more precisely in the next section of this title.

2.2 Machine Learning

The techniques of Machine Learning are responsible for studying and computationally modeling learning processes in their various manifestations. It seeks to build a program that automatically improves with experience.

The Artificial Neural Networks (ANN), are inspired by the biological neural networks of the human brain. The minimum unit of RNA is an element that behaves, or tries to behave, similar to the biological neuron.

The ANN learn from experience and abstract the characteristics of a series of data [11]. Neural networks are composed of nodes connected through directed connections. A connection from unit j to unit i serves to propagate the activation a_j from j to i. In addition, each connection has a numeric weight W_j, i associated, which determines the strength and the sign of the connection. Each unit i first calculates a weighted sum of its inputs:

$$in_i = \sum_{j=0}^{n} W_{j,i} a_i, \tag{4}$$

finally applies an activation function g to this sum to produce the output:

$$a_i = g(in_i) = \left(\sum_{j=0}^{n} W_{j,i} a_i \right). \tag{5}$$

Much of the image classification tasks are performed using automatic learning techniques. The most commonly used are conventional neural networks, deep learning networks and convolutional neuronal networks. Traditional machine learning is based on networks composed of one input and one output layer with a hidden layer at most.

Deep learning networks are distinguished from common neural networks by their depth; that is, the number of node layers through which the data passes in a pattern recognition process. More than three layers (including entry and exit) qualify as learning.

Tensorflow. $TensorFlow^{TM}$is an open source software library for numerical calculation that uses data flow graphs. The nodes in the graph represent mathematical operations, while the edges of the graph represent arrays of multidimensional data (tensors) communicated with each other. Its flexible architecture allows to implement calculations in one or more CPUs or GPUs in a desktop, server or mobile device with a single API. TensorFlow was developed by Google Brain Team researchers and engineers within the Google Artificial Intelligence research organization to perform machine learning and deep neural network research, but the system is general enough to apply to a wide variety of other domains.

3 Sign Detection Model in Smartphones

We propose two parallels solutions to alphabet sign detection that involves hand gesture: the first one is focused on the signs that are static and the second one in the ones that are dynamic. In both of them we have some common stages as we can see in Fig. 1.

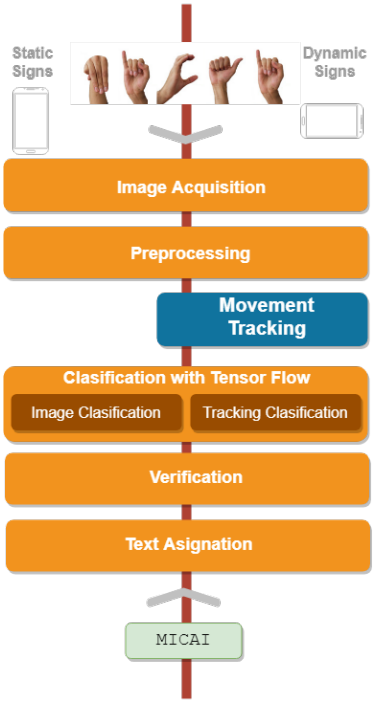


Fig. 1. Process for sign detection in smartphones.

First we do the image acquisition throw the smartphone and the image is prepossessed. Then, it follows the classification stage that depends on the type of sign (static or dynamic), for the static signs the image classification using tensor flow is the most important part, while in the dynamic signs the classification is simpler but it requires previous stages. Finally, there is a verification and text assignation stage to show the corresponding letter of the alphabet in an mobile application.

In the following sections, we will explain the stages that depend on the type of sign, such as classification and the ones for dynamic signs. Then we will show the integration of those stages to the model and to the smartphone.

3.1 Static Signs Detection with Deep Learning

In static sign detection, even when the signs are statics, there is a continuous image acquisition. The systems has to select when do we have a new sign to translate. To do so, we process continuously the images, and we selected them through the verification phase.

For each image, the system has to classify it into one of the 21 possible letters that correspond to static signs. We use TensorFlow because it offers first level solutions for computational learning problems, it helps us to recognize the signs

of the MSL alphabet. Two of the main models are Mobilenet and Inception V3, the first one uses depthwise separable convolution while Inception V3 uses standard convolution. Then, Mobilnet requires less parameters than Inception V3 but also there is a slight decrease in the performance. Nevertheless, Inception V3 is optimized to be precise, while the MobileNet is optimized to be small and efficient (as expected in a classifier that operates on a mobile device) at a minimum cost of accuracy, so we selected MobilNet [12]. Considering the decrease of performance, with Mobilnet, a previous stage was used in the images as we can see in Fig. 2.

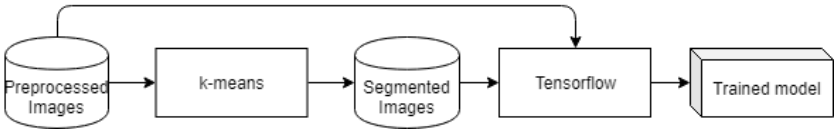


Fig. 2. Training data process for static sign classification.

In the training phase, with Tensorflow algorithm, the algorithm uses original images and segmented images. Segmentation is performed by k-means algorithm, it allows us to get two segments: one for the background and clothes and the other one for the hand. The Figure 3 shows an example of the two detected segments. For training phase, we use the ones that contain our Region of Interest (RoI) that involves the hand and the other one was discarded.

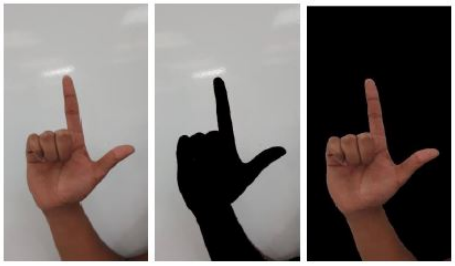


Fig. 3. Original image (left); Discarded image (in the middle); Used segment (right).

3.2 Dynamic Signs Detection with Deep Learning

There are six dynamic signs corresponding to letters; "J", "K", "N", "Q", "X" and "Z". The dynamic sign detection requires an image processing of a series of images in a time window of three seconds... For dynamic signs, we need algorithms capable of identifying an object and tracking it, then the process

we follow is shown in Figure 4. First, a RGBA Selection is done in order to identify the hand through its colors, and then we convert the image to gray-scale. Second, the hand tracking is performed using the CamShift algorithm. Third, the tracking generates a trajectory that is drown in a canvas, its pixels are used for the classification process. Finally, the Bitmap of the canvas pixels is generated because it is used for the process that recognize the image getting the translation of the signs.

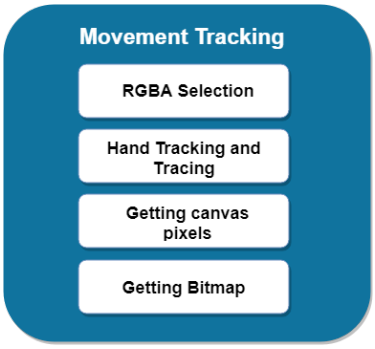


Fig. 4. Movement tracking process.

All tracking algorithms, basically, have the same principle of operation: segmentation, edge detection, create binary masks and enclose the object in a tracking window. Some steps can be omitted or used in a different order, but the method is generally the same.

By comparing the different trajectory tracking algorithms, with CamShift its window adapts to the distance and rotation of the object with respect to the camera, while in Meanshift, the window remains the same regardless of the distance from the objective. Camshift first applies Meanshift, but it adjusts the window according to the total movement of the objective despite the processing consumption.

Tracking by color-based methods it is possible to find the coordinates of specific points in the hand, which are stored as soon as a dynamic signal is executed (limited by a time window). After performing a series of operations on these coordinates, it is possible to construct an image like the one in Figure 5 with an artificial equivalent.

With hundreds of trajectories used as training information, the model generated is able to identify patterns and associate them with a trace obtained by the smartphone application of the system. The inference in the mobile is then made with a second model, in a process totally independent of the recognition of the static signs.

The algorithm of Tensorflow receives the image as a Bitmap, it obtains the pixels in an integer array, copies the input information to the algorithm,

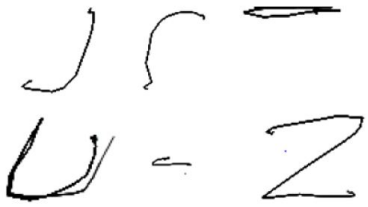


Fig. 5. Track of letters "J", "K", "N", "Q", "X" and "Z".

and proceeds to read the results thrown by the model, finally, by means of ordering methods, create a list of probabilities finding the better classification. The classifier is independent of the static signs and sends the Bitmap generated by the description of the trajectory of the hand.

3.3 Integration of the Classification to Smartphone

In order to classify we need two previous stages: Image Acquisistion and Verification. And once we have the classification we have two final stages. All of them are explained in the following section. Once we have define those stages the complete model integrated to the smartphone is presented.

Common Stages for both Types of Signs. The commons stages are image acquisition, pre-processing, verification and text assignation, those work as follows.

Image Acquisition. This stage has the purpose of capturing continuous images of the sign that is going to be translated using the camera of the device. Because the aim of the system is the translation of the entire MSL alphabet, it was required to implement two *views* in different *layouts*. It mainly depends on the smartphone orientation. For the statics signs the system uses the natural device orientation (vertical) and it gets continuous images. For the dynamic sign, the images are capture, the tracking process gives a trajectory that was process to get the corresponding bitmap that correspond to the pixels of the path that the hand follows, this bitmap is used to classify the letter. For the static signs, the smartphone captures the images from the camera of the device and uses a *TextureView* to show it inside the application. For the dynamic signs, the application uses an OpenCV view integrated with a canvas to the program, so that the user see the camera, the trajectory of the hand and the app interface.

Preprocessing. The images for the training phase and the pictures that are taken from the smartphone have a preprocessing phase. This consist of a transformation of the image to grayscale and then an edge detection phase using the Canny Algorithm with double thresholding as we can see in Figure 6.

Verification. Because of the continuous reception of the images to be process, the system does previous verification:

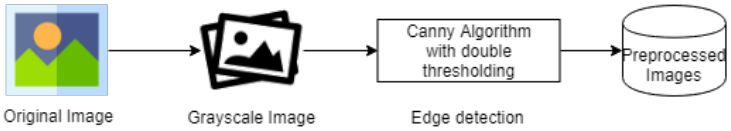


Fig. 6. Pre-processing the image to get the training set.

- That the previous letter is different, except in the case of the first written letter. This rules out the existence of results such as "aa", "bb", and so on that are not used in spanish.
- The probability thrown by the classifier model must be greater than 0.9.
- There must be 3 readings (of the same symbol) with a probability greater than 0.9 before displaying the translation.

Text Assignation. Once the verification is done, the letters are display to form words.

Integration to Smartphone. The training phases of the proposed model were processed in a computer, the results of this phase is used by the application of the smarthphone. The system uses three main elements of the device: the camera, the display and the processor. The camera is used to continuously take pictures. Then the device preprocesses the images and used them on the classifiers using resources of the smartphone. During the use of the application the shown images correspond to the ones taken by the camera, and for the dynamic signs, over the images, the application displays the trajectory of the sign.

The application detects the orientation of the smartphone, with the vertical position the application waits for a static sign and in the horizontal position it waits for a dynamic sign.

4 Experiments and Results

4.1 Tensorflow Validation for Sign Classification

The resolution of the input images were 128, 160, 192, or 224 pixels. Which is enough considering that images are finally taken from smartphones. We compare two of the main algorithms of Tensorflow: Inception 3 and MobileNet. In the tests carried out, the Inception V3 model uses approximately 85 MB of space, even after performing compression operations; MobileNet models, on the other hand, uses 5 423, 10 343 and 16 797 KB of space, depending on the selected architecture. In short, the Inception V3 model is five times heavier than the more complex MobileNet, considerably slower, and it did not offer a significant advantage in terms of accuracy. The configuration of all the tests remained fixed in terms of the resolution of the input images: 224x224 pixels. The relative size

of the network was varied throughout the tests, finding that 1.0 and 0.75 of the largest MobileNet obtained the best results.

We worked with transfer learning, which means that we started with a model that had been previously trained to solve another problem. A model trained in the ImageNet Large Visual Recognition Challenge data set was used and re-trained to differentiate between a small number of classes (21 signs).

4.2 Training Model for Static Signs

The used Mobilnet has a learning rate of 0.015 with 4000 training epochs with 795 images per each category, so that the size is up to 16797 KB. For the recognition of the static signs, more than 30 models were trained with different parameters and using different sets of photographs. Table 1 summarizes the parameters and results of some of the most relevant models, that is, those that presented significant improvements with respect to the previous batch of tests. The % accuracy depends on 15 attempts for each static sign.

Table 1. Classifiers for static signs.

	Learning loop	Epochs	MobileNet Section	Images by category	k-means Images	Accuracy (%)	Failure
1	0.5	500	0.5	100	0	47.60	d,e,f,i,r,s,t,v,w,y
5	0.0001	4 000	0.75	400	0	69.04	d,i,l,p,t,y
9	0.001	4 000	0.75	500	0	77.38	d,i,t
14	0.0001	4 000	1	550	0	52.80	d,e,i,m,n,r
15	0.001	4 000	1	550	0	78.57	i,s,t
17	0.001	4 000	1	600	70	75	d,p,t
K	0.001	4 000	1	70	70	52.38	a,b,d,i,l,m,n,s,t,u
27	0.01	4 000	1	630	100	83.33	p,s,t
29	0.01	4 000	1	630	150	85.71	i,t
30	0.04	4 000	1	795	150	88.09	s,t
31	0.02	4 000	1	795	150	90.47	t
32	0.015	4 000	1	795	150	95.23	-

We performed experiments with different backgrounds and similar light conditions, Table 2 shows the results of those experiments where the background not only changes on color but also on texture. When we have a white background the contrast increases so the accuracy is high, up to 93%. The wort case is with not smooth surface because the shadows of the texture of the background make noise to the algorithm of border detection and the classification fails. In that case, the signs that are difficult to classify are the ones that look like the fist, or that only have one finger up.

Table 2. Test with different backgrounds.

Background	RGB	HSV	Accuracy (%)	Failure
Random background, not smooth	-	-	52	c, i, l, n, p, s, t
Dark red background	#B62906	(12°, 97%, 71%)	60	c, i, l, n, p, s
Black background	#191114	(330°, 25%, 7%)	74	c, i, l, p, t
Light gray background, rough surface	#BDB6B3	(18°, 5%, 74%)	76.1	c, i, p
Light gray background	#BDB6B3	(18°, 5%, 74%)	85.9	p, t, r
Light gray / light blue background	#BAB6BB	(288°, 3%, 73%)	89.71	p, t
White background	#F0FBFF	(196%, 6%, 100%)	93.1	s

4.3 Training Model for Dynamic Signs

The used Mobilnet has a learning rate of 0.02 with 4000 training epochs with 110 images per each category (a seventh amount of static images). For the recognition of the dynamic signs, we tested it with the best results if static signs. Table 3 summarizes the parameters and results of some of the most relevant models, that is, those that presented significant improvements with respect to the previous batch of tests. The % accuracy depends on 15 attempts for each static sign.

Table 3. Classifiers for dynamic signs.

	Learning loop	Epochs	MobileNet Section	Images by category	Accuracy (%)	Failure
1	0.02	4 000	1	15	≈ 66.66	q, k
3	0.02	4 000	1	50	≈ 83.33	q
4	0.02	4 000	1	110	≈ 92	

Once the trajectory of the sign is read properly, the classification process is simpler, since there are only 6 dynamic signs, which facilitates the work of the classifier model and improves its accuracy. In addition, each of the trajectories has particular characteristics and is easily distinguishable. Even in the second of the tests, which had a training corpus of only 50 images per category, an acceptable level of accuracy was achieved.

5 Conclusions

We proposed a two phases training model, the first one for image segmentation using k-means, and the second one for image recognition with tensor flow. This process increases the accurancy of sign detection because the k-means method allowed to improve the extraction of particular traits in some signs, identifying false positives.

The implementation of trajectory tracking algorithms allowed to identify signs with movement grouping them according to their trajectory with a success of at least 90%.

The need of using a glove of contrasting color was discarded because the precision achieved with the natural color already exceeded 90%.

Separating the identification stages by performing the training in an external device allowed the process to be efficient, leaving the smartphone only with the task of classification.

References

1. Young, I.T., Gerbrands, J.J., Van Vliet, L.J.: Fundamentals of Image Processing, 2nd ed., pp. 2–32 (2007)
2. Chitradevi, B., Srimathi, P.: An Overview on Image Processing Techniques. International Journal of Innovative Research in Computer and Communication Engineering IJIRCCE 2(11) (2014)
3. Pansare, J., Ingle, M.: Vision-based approach for American Sign Language recognition using Edge Orientation Histogram. In: International Conference in Image, Vision and Computing. IEEE (2016)
4. Adithya, V., Vinod, P.R., Gopalakrishnan, U.: Artificial Neural Network Based Method for Indian Sign Language Recognition. IEEE (2013)
5. Nixon, M., Aguado, A.: Feature extraction and image processing for computer vision. 1st ed. Elsevier, Academic Press, Amsterdam (2013)
6. Monga, P., Ghogare, S.: Scrutiny on Image Processing. International Journal of Advanced Research in Computer Science and Software Engineering 5(1) (2015)
7. Rao, R.M., Arora, M.K.: Overview of Image Processing. In: Varshney, P.K., Arora, M.K. (eds.) Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data, pp- 51–85. Springer, Berlin, Heidelberg (2004)
8. Kumar, G., Bhatia, P.: A Detailed Review of Feature Extraction in Image Processing Systems. In: 2014 Fourth International Conference on Advanced Computing and Communication Technologies, IEEE (2014)
9. Canny, J.: A Computational Approach to Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligence 8(6), 679-–698 (1986)
10. OpenCV: Meanshift and Camshift. Docs.opencv.org. (2015) `http://docs.opencv.org/3.1.0/db/df8/tutorial_py_meanshift.html`. Último acceso: 01/04/2018
11. Dreiseitl, S., Ohno-Machado, L.: Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 35(5–6), 352–359 (2002)
12. Codelabs.developers.google.com. (2017). TensorFlow For Poets. [En línea] https://codelabs.developers.google.com/codelabs/tensorflow-for-poets/. Último acceso: 16/11/2017.

Electronic edition
Available online: http://www.rcs.cic.ipn.mx