Advances in Hybrid Intelligent Systems

Research in Computing Science

Series Editorial Board

Research in Computing Science es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. **Volumen 122,** noviembre 2016. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No. : 04-2005-121611550100-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Queda prohibida la reproducción total o parcial, por cualquier medio, sin el permiso expreso del editor, excepto para uso personal o de estudio haciendo cita explícita en la primera página de cada documento. Impreso en la Ciudad de México, en los Talleres Gráficos del IPN – Dirección de Publicaciones, Tres Guerras 27, Centro Histórico, México, D.F. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

Editor responsable: *Grigori Sidorov, RFC SIGR651028L69*

Research in Computing Science is published by the Center for Computing Research of IPN. **Volume 122**, November 2016. Printing 500. Printed in Mexico City, in the IPN Graphic Workshop – Publication Office.

Volume 122

Advances in Hybrid Intelligent Systems

Julio César Ponce Gallegos,
José Alberto Hernández Aguilar,
Carlos Alberto Ochoa Ortiz (eds.)

Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2016

ISSN: 1870-4069

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

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

Indexed in LATINDEX, DBLP and Periodica

Printing: 500

Printed in Mexico

ISSN 1870-4069

Editorial

This volume of the journal "Research in Computing Science" contains selected papers related to Hybrid Intelligent Systems (HIS) and its applications. The papers were carefully chosen by the editorial board on the basis of the at least two reviews by the members of the reviewing committee or additional reviewers. The reviewers took into account the originality, scientific contribution to the field, soundness and technical quality of the papers. It is worth noting that various papers for this special issue were rejected.

Hybrid Intelligent Systems try to deal with the complexity of real world phenomena, with a multidisciplinary approach and a plurality of techniques. The complex systems where they are applied cover: education, biology, medicine, logistics, management, security, engineering, humanities, among others. All characterized by the difficulty of modeling their usual problems by classical methods. In this framework, this special issue is aimed at discussing research on working progress with social interactions, by using agents or any artificial intelligence technique. Therefore, it will focus on social simulation and dynamic social network systems. It also covers the Hybrid Systems with the capability to hold a negotiation about a specific topic, demonstrate reputation using diverse models, with argumentation procedures as a way for reaching agreements during the negotiation process.

The volume contains 13 papers about various aspects of HIS in the fields of Simulation and Knowledge Discovery in Databases (KDD). Both topics are very important today, because of its potential applications in a wide variety of business.

Knowledge Discovery in Databases is the process of searching for hidden knowledge in the massive amounts of data that is generated by people, public and private institutions. Data are simply a collection of elements, from which have little or null knowledge. With the development and use of data discovery techniques the value of the data is significantly improved, since you get knowledge that may be hidden.

Simulation is the imitation of a real-world process. To simulate something first requires that a model be developed; this model represents an abstract of behaviors or functions of the system or process to be simulated. Simulation can be used in many contexts, such as engineering, testing, training, education, natural behavior, social behavior, games, among others. The more common simulations use mathematical or computer models.

There are a lot of methods available to extracting patterns that when interpreted, provide an add value to the data, possibly previously unknown, this new knowledge allows us to better describe and understand the stored data. Information can be predictive or descriptive in nature. Data mining stage, the pattern extraction phase of KDD, can take many forms; the choice depends on the desired results. KDD is a multi-step process that facilitates the conversion of data to useful information.

We would like to thank Mexican Society for Artificial Intelligence (Sociedad Mexicana de Inteligencia Artificial), and MICAI 2016.

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).

Julio César Ponce Gallegos,
José Alberto Hernández Aguilar,
Carlos Alberto Ochoa Ortiz
Guest Editors
México

November 2016

ISSN 1870-4069

Table of Contents

Page

ISSN 1870-4069

Equations for Describing Behavior Tables in Thermodynamics Using Genetic Programming: Synthesizing the Saturated Water and Steam Table

Martín Montes Rivera[1], Marving Omar Aguilar Justo[1], Alberto Ochoa Zezzatti[2]

[1] Universidad Politécnica de Aguascalientes, Aguascalientes, Mexico

[2] Universidad Autónoma de Ciudad Juárez, Chihuahua, Mexico

martin.montes@upa.edu.mx, marving.aguilar@upa.edu.mx, alberto.ochoa@uacj.mx

Abstract. There are several tables with important data used in the calculus of different processes like machining tables, friction tables and thermodynamics processes tables, or as it is explored in this paper, the description of saturated water and steam table. We propose the generation of equations for describing the entire behavior of numerical values in a table using Genetic Programming (GP), when table data describes the variable behavior of a dependent function. This obtained equations simplify the calculus process without requiring several tables and allowing to work when tables are not available for a desired value of an independent variable, a common situation in thermodynamics. In this case it is tested the proposed algorithm for synthesizing the saturated water and steam table.

Keywords: Genetic Programming, Synthesizing Tables, Saturated Water and Steam Tables.

1 Introduction

Tables are arrays of rows and columns used as tools for presenting information, they are commonly used to show values of responses, relationships, scores, percentages and statistical results, among others [1].

The results of measurements obtained from input-output models and the evaluation of functions are commonly expressed using tables, which describe the behavior of dependent variables in function of independent variables. Examples of function applications are shown in [2] and tables of input-output models with their applications used in thermodynamics are shown in [3].

Tables are common tools in different sciences and processes; thermodynamics is not and exception and students must learn to use them [4].

Thermodynamics like other industrial sciences applied has several mathematical models, which try to describe the behavior of specific variables, sometimes the obtained models based on physical and mathematical principles are complex and tables with the data of the models are used instead [5].

Tables increase accessing information speed, but if the values on tables are used for describing the behavior of a variable, sometimes is better to plot the data or obtain an equation to describe their values for a better comprehension [6].

Nowadays several information is uploaded on internet, this include great amounts of numerical data distributed in several tables which like other information are sometimes difficult to find among other documents [7]. Equations instead tables require less space on internet and are easier to find, in this sense, is better to use equations instead tables when is possible synthesizing them.

Tables in thermodynamics are commonly used to indicate the substitution parameters in process equations, but in this case, would not be better to get an equation that represents the table? And then, substitute it at the process equation which already imply to work with an equation.

There are different ways for obtaining equations from tables related to the kind of equation that represents the numerical data, for example, the processes for obtaining the equation of a table are different if the data chart has straight line shape or parabola shape, because different math considerations are required [2].

Sometimes numerical data in a table could have behaviors that can only be represented with no linear equations, which require complicated mathematical methodologies for obtaining them, in this cases computer optimizing algorithms are an alternative [8].

The numerical data in a table has an equation with unknown structure and an easy form to realize what kind of equation would work for synthesizing it, would imply to plot the dependent variable against the independent and then, human judgment would be required to identify the data chart shape and approximate it to a general equation (supposing that there are mathematical methodologies for approximating the equation).

Optimizing algorithms are tools commonly used for synthesizing equations using known information for improving results. Evolutionary algorithms are optimizing metaheuristic programs that improve equations or parameters by using known information and objective functions combined with natural selection principles [9, 10].

An optimizing evolutionary algorithm which works well with unknown structures is Genetic Programming (GP), even if there are no mathematical techniques to approximate the required equation [8, 11].

In this paper is proposed and algorithm to use GP for obtaining the unknown equations that represent relationships among variables in a table and continue with the process using equations instead of tables. We test the proposed algorithm working with the saturated water and steam table used in thermodynamics for different processes like generation of Pressure, Volume and Temperature (PVT) diagrams.

2 Background

Manufacturing industries are continually improving their models for conceptual design and engineering processes. The work in [5] shows that Thermodynamics demands better models for simulation, which has been a focus of research that will continue looking for nonexistent and more accurate models.

Tables are used in several applications and sciences concentrating important data used in health, economics, research, among others, this has motivated the emerging of new techniques for working with great amount of data in tables and their characteristics extraction like is shown in [12,13].

The work in [4], shows how Thermodynamics has been taught to students and that several tables are required in the generation of PVT diagrams and calculus of different processes, they propose to use tables for the solution of thermodynamic problems and their synthesis, which help students in their understanding.

Teaching thermodynamics has different approaches nowadays, the work in [14], compare the efficiency of use traditional and cooperative methodologies, together with the use of Technologies of Information and Communication (TIC).

The work in [15] exhibit the application of height and weight tables for obtaining equations to predict mid-arm circumference and the correct blood pressure cuff size.

Singapore like other countries concentrates its statistical information using tables, the work in [16] describes Input Output Tables (IOTs), obtained from Supply and Use of Tables (SUTs), used for a variety of proposes like economics, process modeling and impact studies, their paper is focused on the transformation of data from SUTs to IOTs format.

The work in [6] shows the use of tables in education and demonstrates that interpretation of tables by teachers and students is better understood when charts of the columns in tables are generated, improving their comprehension.

Reporting evidence tables in research have better interpretation when they are used for generating charts, like is shown in [17].

Transformation of time series tables to equations, it is a common action when modeling physical processes and GP is commonly used for their synthesis.

The work in [18] shows synthesizing monotonic stochastic processes using time series signals (which could be expressed like tables) and GP.

Similarly, the work in [19] describes the process for synthesizing functions that represent several recollected data of relevant heterogeneous catalysts of different materials, using GP.

Another application using GP for synthesizing equations is shown in [20], where GP is used for equation synthesis with simulated data of main variables and the average of their relations, this work is tested for obtaining equations of Hydraulic processes.

The work in [21], shows a comparison between Artificial Neural Networks (ANN) and GP, applied in the obtaining of new formulations for electrical resistivity based on numerical data, showing the performance of GP in data regression, like the proposed in this work.

3 Water Steam Tables in Thermodynamics

Calculus of different processes in Thermodynamics require tables for describing behavior of gases or liquids, which depend of temperature, pressure or volume [3].

Water and water steam tables describe the behavior of water at different conditions of temperature and pressure, the variables used in water steam table in [3] are: Temperature (T), Pressure (P), Specific volume for saturated liquid (v_f) and saturated gas (v_g), Enthalpy of saturated liquid (h_f) and saturated gas (h_g), Internal energy of saturated liquid(u_f) and saturated gas (u_g) and Entropy of saturated liquid (s_f) and saturated gas (s_g).

The problems that engineers solve with this table are the calculus of the main variables (mentioned before) using T or P like independent variables, they also obtain other variables that require substitution of main variables in equations, for example, determining water intrinsic expansion, behavior of refrigerant cooling with water, among others [3].

Exercise example 3-1 in [3], shows a simple application of what students must learn to do with water and water steam tables, this example asks the pressure and volume of a recipient containing 50 kg of saturated liquid water at 90 $^{\circ}C$. In this case, the value of T is in the table, but if it is not, then, interpolation with equation (1) must be performed:

$$y = y_0 + (y_1 - y_0)\frac{x - x_0}{x_1 - x_0} \tag{1}$$

Where y is the value of the variable when x takes an absent value in the table, and y_0, y_1, x_0 and x_1 are the nearest known conditions in the table for y and x.

Assume that in the previous exercise T = 88 $^{\circ}C$ for understand what happen if there are required absent values in the table and linear interpolation in equation (1) is applied for obtaining the absent values of P, v_f and V (volume of the recipient).

The near section to the required value of T is shown in table (1), obtained from the values in [3].

Table 1. Section of water and water steam table in [3] with near values for T = 88 $^{\circ}C$.

$T(^{\circ}C)$	P(kPa)	$v_f(\frac{m^3}{kg})$
80	47.416	0.001029
85	57.868	0.001032
90	70.183	0.001036
95	84.609	0.001040

Considering table (1), the value of x represents T, which means that x = 88 and the nearest known temperatures are x_0 = 85 and x_1 = 90. For obtaining P when T = 88 $^{\circ}C$, y represents the pressure and its nearest known values in the table are y_0 = 57.868 and y_1 = 70.183. Then, P = 65.257 kPa at 88 $^{\circ}C$ like is shown in equation (2) obtained substituting the nearest known conditions in equation (1):

$$y = 57.868 + (70.183 - 57.868)\frac{88 - 85}{90 - 85} = 65.257 \quad (2)$$

Similarly, v_f (in this case y) is calculated with nearest known conditions $y_0 = 0.001032$, $y_1 = 0.001036$ and temperature conditions represented with x are thesame as before, then, $v_f = 0.0010344 \ \frac{m^3}{kg}$ substituting this values in equation (1).

The calculus of V is obtained with v_f times m (equation (3)), since specific volume is the volume per mass unit, in this case $m = 50$ kg, then, $V = 0.05172$ m^3:

$$V = \left(0.0010344 \ \frac{m^3}{kg}\right)(50 \ kg) = 0.05172 \ m^3 \quad (3)$$

There are other applications with water and water steam tables, but all of them follow the same process in previous exercise, find the required propriety in the table, find the desire value of independent variable (T or P), if this is not in the table, then, interpolate using equation (1) and substitute the obtained values in the specific process.

In this work equation synthesis for testing the proposed algorithm takes T like independent variable and synthesizes dependent variables P, v_f, v_g, h_f and h_g, allowing to get any value without require tables or interpolation, but only T substitution.

4 Genetic Programming

GP is an evolutionary algorithm proposed by John R. Koza in 1992, this algorithm like other evolutionary algorithms follows natural selection principles in [23], i.e. its main operations are population initializing, fitness value calculation, fitness-based selection, crossover and mutation (Fig. 1) [11, 22].

Most used code representations of elements in GP are tree base-individuals (Fig. 2), which have disadvantages like conversions from tree representation to equations, and uses several pointers that sometimes make very slow its execution. Nevertheless, there are other many representation with faster manipulation, in this case we need to generate equations so a variation of linear GP with native equation structure is used like described in [24, 25].

Equation structure is a linear GP representation that stores individuals in arrays with operators and numbers and parenthesis that separate operations, similarly to depth in tree-based individuals [24, 25].

Fitness evaluation comes after population initializing using fitness function, which is different depending of the desire objective.

After fitness evaluation fitness based selection must define which chromosomes will be selected for crossover operation. Tournament selection is a commonly used selection technique, because allows to control selection pressure with tournament size (t_s) and population size (p_s). Tournament is made by selecting t_s random individuals for match and winners according to its fitness value are taken into mating pool.

Crossover operation takes individuals from mating pool for offspring generation. Equation structure crossover is similar to haploid crossover in Genetic Algorithms (GA), which exchange genes between two parents in a random crossover point for generating offspring, like is shown in Fig. 3 with crossover point at 5th bit [24, 25].

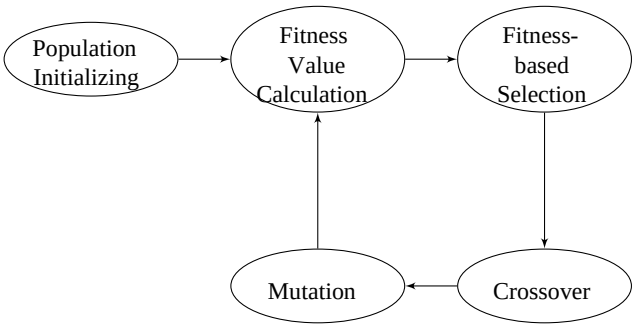


Fig. 1. Evolutionary algorithm diagram.

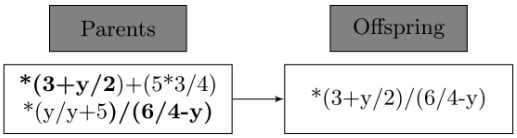


Fig. 2. GP tree-based individual that represents $(4 - 2) \times (8 + y)$.

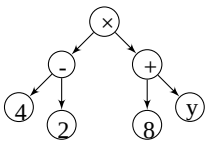


Fig. 3. GP structure equation individual and its crossover operation.

Mutation in linear GP with equation structure is similar to mutation in GA, which could be Single Gene Mutation (SGM) or Multiple Gene Mutation (MGM), where SGM changes aleatory single selected bits and MGM selects two points for changing all alleles or bits between them [8,24]. MGM has shown to be better in GP since there are more possible values than 1 or 0 for its alleles (like occurs in GA), therefore diversity plays a crucial place for obtaining good solutions in GP and MGM supply more genetic material than SGM for a single individual . SGM and MGM are shown in Fig. 4 [26].

Fig. 4. Single (left figure) and multiple gene mutation (right figure) [8].

5 Applying Equation Structure for Synthesizing Tables

Our proposed algorithm starts by generating two arrays, first one (X_i) contains all values of T in water and water steam table and second one (Y_i) with all values of required dependent variable (P, v_f, v_g, h_f or h_g).

Since T is the independent variable, the individuals in population are expected to be functions of T, therefore, operands are T and integer numbers [0,9]; operators are plus (+), minus(-), times, (×), divide (/) and power (^).

The control values generating population are number of operators (n_o), depth or number of operations separated with parenthesis (d), a dividing factor (div) for increase or decrease values in equations (required when Y_i has small values) and population size (p_s).

After generating population, fitness function ($F(p_j)$) evaluates fitness in the j_{th} individual of the population by calculating accumulated absolute error produced between the dependent variable Y_i and the evaluation of substituting T in equation of p_j with all values of X_i, for every i of the n available values in the table, like is shown in equation (4), therefore, best adapted individuals has $F(p_j) = 0$ and it worst $F(p_j) = \infty$:

$$F(p_j) = \sum_{i=1}^{n} abs[Y_i - p_j(X_i)] \tag{4}$$

Selection operation used is tournament with selected tournament size (t_s) control and crossover operation is performed like described in section 4, mutation is MGM with control value in number of mutations per generation (m_{gen}). Overall control in the algorithm is the number of generations (n_{gen}).

6 Results

The resulting equations for dependent variables P, v_f, v_g, h_f and h_g, were obtained with seed value of 1 for MATLAB™ random numbers (platform used for implementing the proposed algorithm) and the control parameters listed in table 2, obtained heuristically since they

depend of the specific application, therefore, they are increased until find acceptable equations for the processed model.

Table 2. GP parameters for obtaining synthesized equations.

Parameter: \ Model:	P	v_{f}	v_{g}	h_{f}	h_{g}
n_{gen}	2500	2500	3000	2500	3000
p_{s}	200	200	200	200	200
n_{o}	5	5	6	6	6
d	5	5	6	6	6
div	1	1000	1	1	1
t_{s}	3	3	3	3	3
m_{gen}	20	20	20	20	20

Charts of fitness behavior (in this case, accumulated absolute error) against generations and its final accumulated error in best individuals are shown per model from Fig. 5 to 9.

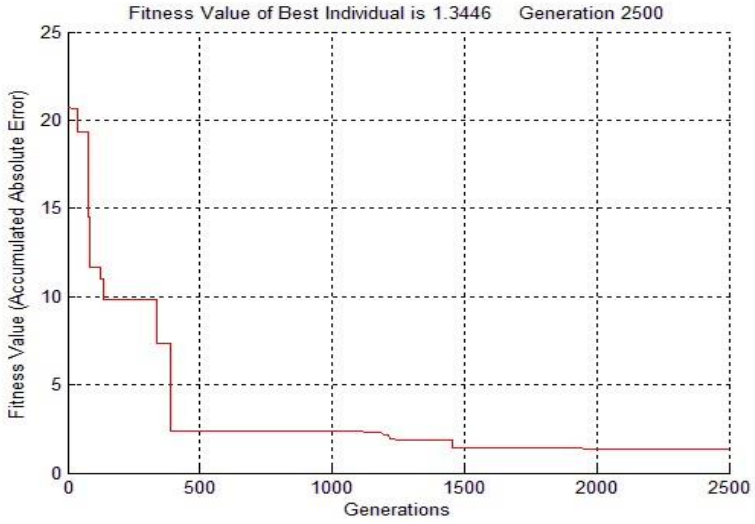


Fig. 5. Fitness against generations synthesizing *P* equation.

In every chart must be consider that the fitness value or accumulated error in the best chromosome is the total absolute error obtained between the generated equation and the data in the table, substituting the 100 independent values available in the saturated water and steam table.

The relative error is calculated as the quotient of the average error in the obtained equation and the average value of data in the table.

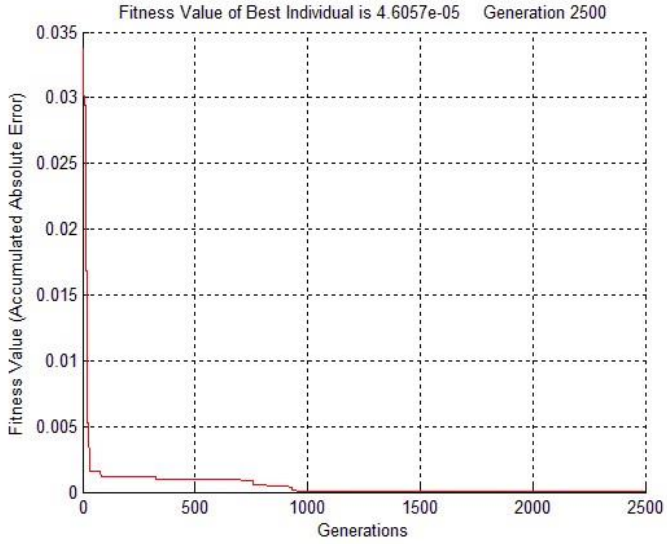


Fig. 6. Fitness against generations synthesizing v_f equation.

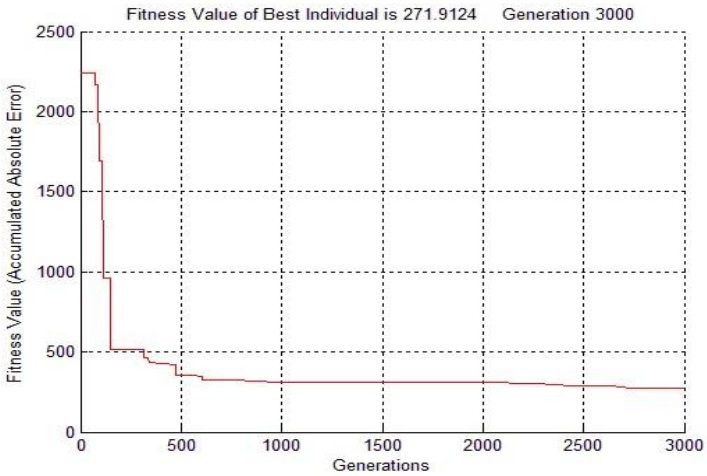


Fig. 7. Fitness against generations synthesizing v_g equation.

P synthesized equation obtains and average error of 0.013446 units, which is 5.38% relative error.

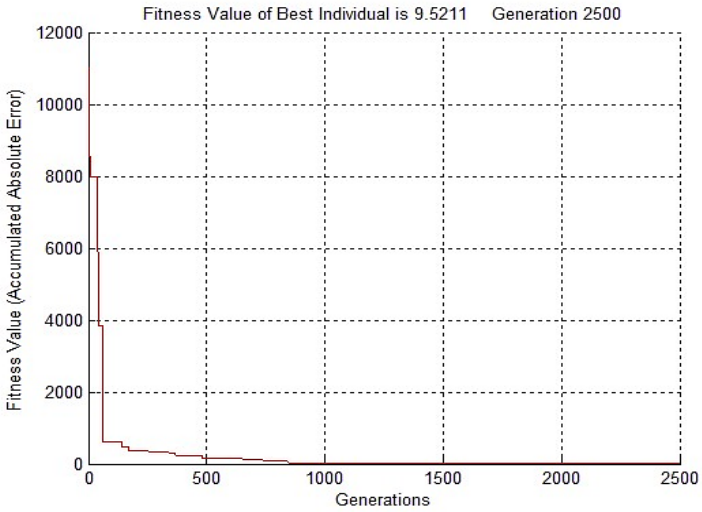


Fig. 8. Fitness against generations synthesizing h_f equation.

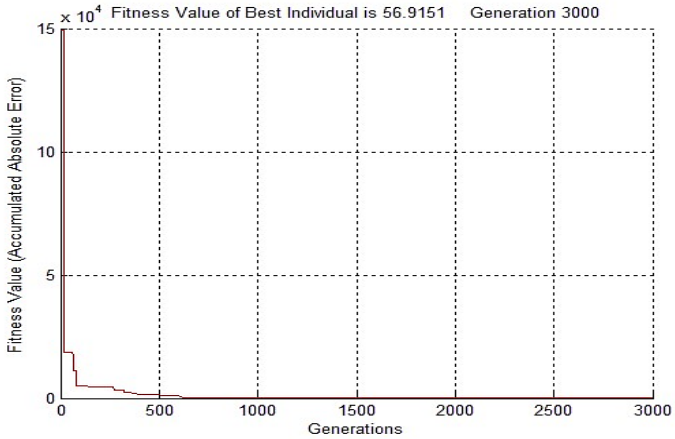


Fig. 9. Fitness against generations synthesizing h_g equation.

v_f synthesized in equation obtains and average error of 4.6057×10^{-7} units, which is 0.0461% relative error.

v_g synthesized equation obtains and average error of 2.719124 units, which is 7.77% relative error.

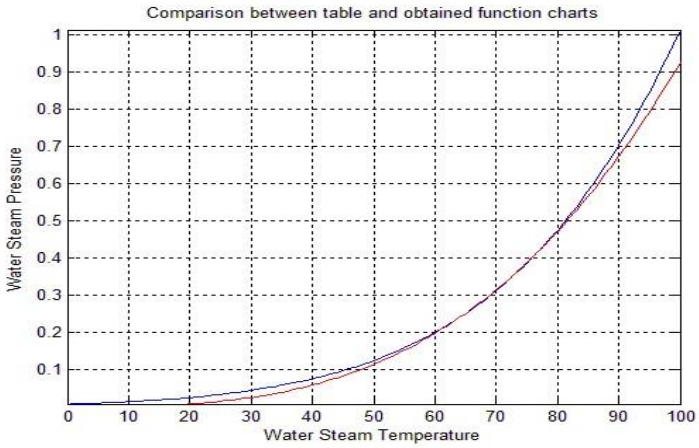


Fig. 10. Comparison between table data and obtained equation for *P* model.

.

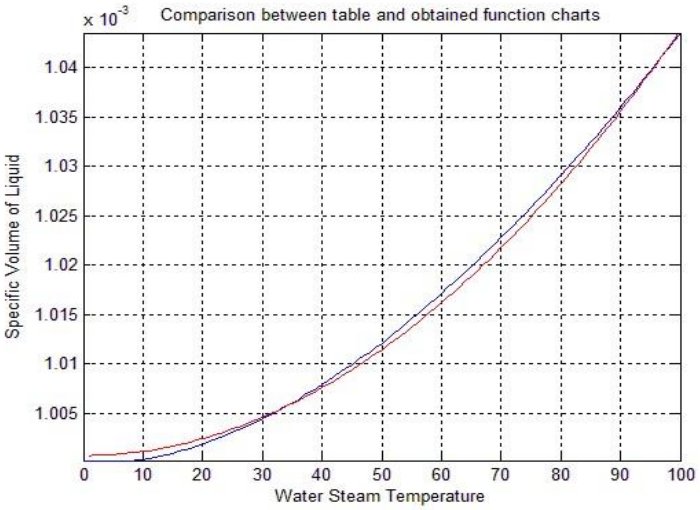


Fig. 11. Comparison between table data and obtained equation for v_f model.

h_f synthesized equation obtains and average error of 0.095211 units, which is 0.045% relative error.

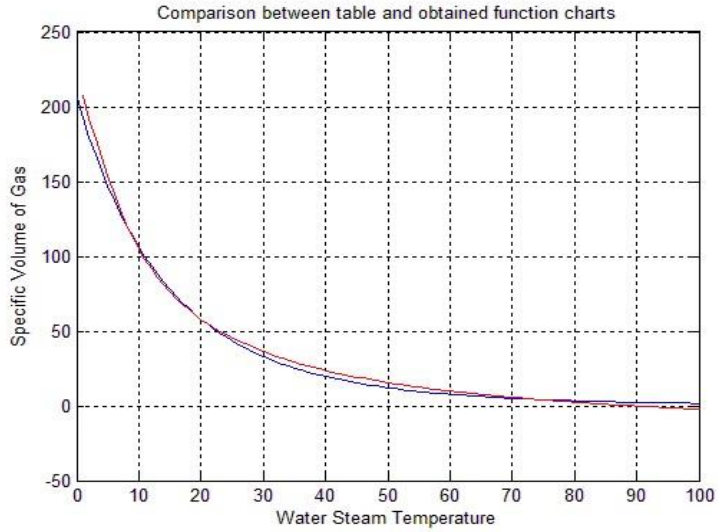


Fig. 12. Comparison between table data and obtained equation for v_g model.

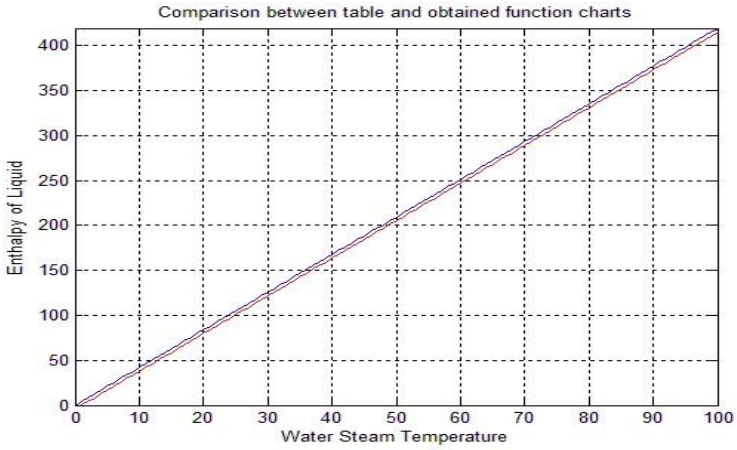


Fig. 13. Comparison between table data and obtained equation for h_f model.

h_g synthesized equation obtains and average error of 0.569151 units, which is 0.0218% relative error.

Charts of comparison between data tables and obtained equations are shown from Fig. 10 to 14, this results show that error levels are almost imperceptible when plotting data

information in a table. Curves obtained from tables are shown in blue and curves of obtained equations in red.

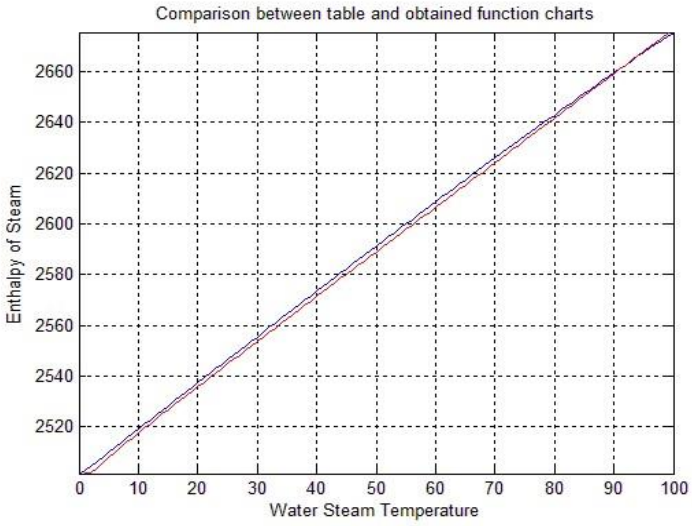


Fig. 14. Comparison between table data and obtained equation for h_g model.

Obtained equations using GP for water and water steam tables are shown from equation (5) to (9).

$$P = \frac{3.16T(3T-7)(T+3)}{10^7} \tag{5}$$

$$v_f = 0.0084T\left(\frac{5.1202T}{10^7} - \frac{\frac{1.1947}{10^6}}{T} + \frac{1.024}{10^6}\right) + 0.001001 \tag{6}$$

$$v_g = \frac{6T-1938}{24(2)^T+23} - \frac{0.02083(1081T-95511)}{T+8} \tag{7}$$

$$h_f = 4.1845T + 0.18371 \tag{8}$$

$$h_g = 1.75T + \frac{1.75T-17.4}{T+\frac{2}{T}+5} + 2502 \tag{9}$$

7 Conclusions

The proposed algorithm takes advantage of GP for synthesizing equations and applying it successfully for obtaining equations that model variables in water and water steam table. This algorithm also shows that linear GP with equation structure produce good results for synthesizing equations and therefore structures of equations of water and water steam tables are available for optimizing with GA in future work, which could improve its precision.

Using the obtained equations it is possible to solve easily and faster the Thermodynamics exercises with a small error of precision (which could be improved using optimizing techniques with the obtained equation structures) compared with traditional methodology, since tables are not required any more and interpolation it is not necessary with this equations.

References

1. Arlene, F.: How to Conduct Surveys: A Step by Step Guide. SAGE Publications. 4th Edition (2009)
2. Leithold, L.: The Calculus. Oxford University Press, 7th edition (1998)
3. Çegel, Y. A., Boles, M. A.: Termodinámica. Mc-Graw Hill, 7th edition (2012)
4. Maldonado, L. C., Chapou, J. L., Arcos, J. M., Granados, J.: Termodinámica: metodología para un mejor aprendizaje. Rev. Cub. Fís, Vol. 27, No. 2A, pp. 125–129 (2010)
5. Chau-Chyun, C, Mathias, P. M.: Applied Thermodynamics for Process Modeling. AIChE Journal, Vol. 48, No. 2, pp. 194–200 (2002)
6. Chengping, Z.: Using Excels Data Table and Chart Tools Effectively in Finance Courses. Journal of Accounting and Finance, Vol. 15, No. 7 (2015)
7. Gelbukh, A.: Enriquecimiento parcial de la consulta para la recuperación de información de alta calidad y flexibilidad. Research in Computing Science, Vol. 73, pp. 17–44 (2014)
8. Weise, T.: Global Optimization Algorithms: Theory and Application. e-book obtained from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.8184&rep=rep1&type=pdf (2009)
9. Luis, A., Prospero, C., Mejía, M., Ruiz, J., Vela, V.: Aprendizaje incremental basado en población como buena alternativa al uso de algoritmos genéticos. Research in Computing Science, Vol. 116, pp. 51–64 (2016)
10. Pérez-Gaspar, L. A., Caballero-Morales, S. O., Trujillo-Romero, F.: Integración de optimización evolutiva para el reconocimiento de emociones en voz. Research in Computing Science, Vol. 93, pp. 9–21 (2015)
11. Poli, R., Langdon, W. B., McPhee, N. F.: A Field Guide to Genetic Programming. e-book obtained from: http://www0.cs.ucl.ac.uk/staff/wlangdon/ftp/papers/poli08 fieldguide.pdf (2008)

12. Rodríguez-Flores, F., Flores-Pulido, L., Dávila de la Rosa, E.: Inteligencia de negocios y minería de datos aplicado a la industria refresquera. Research in Computing Science, Vol. 126, pp. 63–71 (2016)
13. León-Barranco, A., Saucedo-Lozada, S. N., Avendaño-Jimenez, I. Y., Martínez-Leyva, R., Carcaño-Rivera, L. A.: Business Intelligence in Educational Institutions. Research in Computing Science, Vol. 96, pp. 43–53 (2015)
14. Méndez, D.: The influence of Teaching Methodologies in the Learning of Thermodynamics in Secondary Education. Journal of Baltic Science Education, Vol. 12, No. 1, pp. 59–72 (2013)
15. Ostchega, Y., Dillon, C., Prineas, R. J., McDowell, M., Carroll, M.: Tables for the selection of correct blood pressure cuff size based on self-reported height and weight and estimating equations for mid-arm circumference: data from the US National Health and Nutrition Examination Survey. Journal of Human Hypertension, Vol. 20, pp. 15–22 (2006)
16. Department of Statistics, Ministry of Trade & Industry, Republic of Singapore: Singapore Supply and use, and Input-Output Tables. Department of Statics Singapore (2014)
17. Mignini, L., Champaneria, R., Mishanina, E., Khan, K. S., EBM-CONNECT: Graphical displays for effective reporting of evidence quality tables in research syntheses. Reproductive Health, Vol. 13–21 (2016)
18. Ross, B. J.: Using Genetic Programming to Synthesize Monotonic Stochastic Processes. CI '07 Proceedings of the Third IASTED International Conference on Computational Intelligence, pp. 71–78 (2007)
19. Baumes, L. A., Blansch, A., Serna, P., Tchougang, A., Lachiche, N., Collet, P., Corma, A.: Using Genetic Programming for an Advanced Performance Assessment of Industrially Relevant Heterogeneous Catalysts. Materials and Manufacturing Processes, Vol. 24, pp. 282–292 (2009)
20. Moreno, J. A., Cárdenas, E. G., Obregón, N.: A Novel Genetic Programming Heuristic Supported on Genetic Algorithms and Average Mutual Information as Fitness Evaluation Metric Some Applications from Hydraulics. XVIII International Conference on Water Resources (2010)
21. Karahan, I. H., Ozdemir, R., Erkayman, B.: A comparison of genetic programming and neural networks; new formulations forelectrical resistivity of ZnFe alloys. Appl Phys A Materials Science and Processing, Vol. 113, pp. 459–476 (2013)
22. Oltean, M., Grosan, C.: Using traceless genetic programming for solving multi-objective optimization problems. Journal of Experimental & Theoretical Artificial Intelligence 3, Vol. 19, pp. 227–248 (2007)
23. Darwin, C.: The Origin of Species. John Murray, penguin classics, 1985 edition (1859)
24. Montes-Rivera, M., Paz-Ramos, M. A., Orozco-Mora, J. L., Ochoa-Zezzatti, C. A.: An Alternative Structure to Replace the Tree-Based Individuals in a Genetic Programming Algorithm for Decoupling MIMO Systems. 6th MICAI Workshop on Hybrid Intelligent Systems (2013)
25. Montes-Rivera, M., Paz-Ramos, M. A., Orozco-Mora, J. L.: Generador de bloques desacoplantes empleando programación genética: Aplicado al desacoplamiento de un sistema de riego y un proceso de transporte de papel. Editorial Académica Española (2014)
26. Bäck, T.: Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms. Oxford University Press (1996)

ISSN 1870-4069

Implementation of a Reactive Model for Responding to a Trembling Earthquake: A Perspective from Virtual Reality and Multiagent Systems

Miguel Pérez-Ramírez[1], Benjamin Eddie Zayas-Pérez[1], Rodrigo Beltran-Labra[2], Diego Márquez-Salazar[2], Carlos Alberto Ochoa-Zezzatti[3]

[1] National Institute of Electricity and Clean Energies, Cuernavaca, Morelos, Mexico

[2] Polytechnic University of the State of Morelos, Jiutepec, Morelos, Mexico

[3] Autonomous University of Juarez city, Ciudad Juárez, Chihuahua, Mexico

{mperez, zayas}@iie.org.mx, {blro130524, mado140055}@upemor.edu.mx, alberto.ochoa@uacj.mx

Abstract. We present a virtual reality simulation of an evacuation caused by a trembling earthquake in a secondary school. The simulation is based on artificial intelligence for which a mathematical model is implemented that controls the behavior of agents that represent the school students in the virtual stage. The simulation allows the interaction with the agents on stage by using the gyroscope of a smartphone while roaming around the stage and thus observing the behavior of the agents when the human stampede occurs. The simulation can be executed either on a computer or on a smartphone. Once again, the importance of artificial intelligence to prevent losses of human lives is demonstrated.

Keywords: Artificial Intelligence, Virtual Reality, Simulation, Multi-agents Systems, Paranoia, Behavioral Patterns.

1 Introduction

We can have an interactive virtual world behind the screen of a computer, however, the challenge is to make that world to looks real and feels like such. Virtual reality (VR) has been evolving throughout time and has been accepted as a viable technique in different fields such as visualization [5], simulation, design and research. Our senses allow us to experience the reality we live in, VR aims to allow interaction with a scenario in third dimension where this reality is emulated even if this is not available for being dangerous

or physically inaccessible; people will be even able to see and interact within physically inexistent virtual worlds, but from their perspective these will be perceived as real. VR has a wide variety of applications in a number of different fields, such as medicine, architecture, design, sports, humanitarian logistics and even games, just to name a few of them. It is within humanitarian logistics that this research is based on. Here we describe how we developed a VR simulator for a humanitarian logistics, where we attempt to make predictions in the unpredictable natural disasters such as earthquakes in order to reduce the possibility of human losses and injuries.

Specifically, by integrating Virtual Reality and Artificial Intelligence, we seek to find out those specific obstacles that could prevent the opportune evacuation of students and teachers from a three-level building of a secondary school, and observe at the same time the behavior of all agents involved in real time.

2 Materials and Methods

For the development of this research, a Smartphone Samsung J7 was used together with a computer with a Operating System X Yosemite version 10.10.5 (Intel Core i5 processor, 6MB in Cache and 3.20 GHz, graphics card NVIDIA GeForce GTX 1070 with 8GB of RAM).

In order to model the structure of the building of the secondary school, we took as a reference the plane defined in a previous work [2], which was modeled and programmed (Fig. 1) in the framework "Menge" [6].

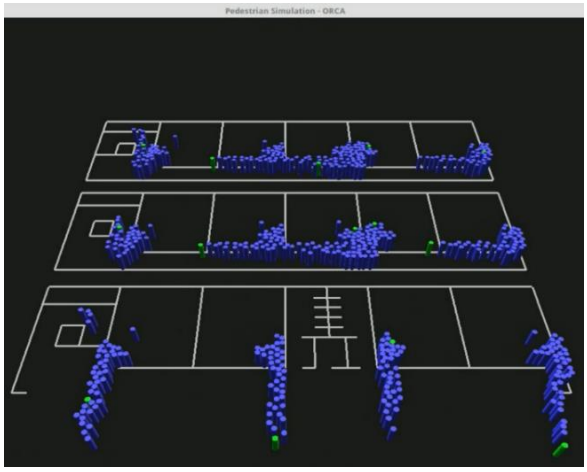


Fig. 1. Simulation of evacuation of the structure using multi-agents system.

The methodology used to calculate the pedestrian speed is made up of several sub models that are used to estimate the level of paranoia, the density in the classroom, and

the impact of the socio-cultural training of the student [2]. The equation that calculates the level of paranoia considers the following variables: noise (R), temperature (Temp), lack of sleep (Fs), intensity of hunger (Af), density of people in the classroom (D) and adjustment variable (α=4).

Two variables are considered to calculate the density of students: living room surface (S) and the total amount of people in the classroom (T).

$$D = \left(\frac{T}{S}\right) \tag{1}$$

Equation that calculates the decibel scale from number:

$$dB = (R * 90) + 30 \tag{2}$$

Equation that calculates the number scale from the decibel:

$$R = \frac{-30 + dB}{90} \tag{3}$$

The lack of sleep is defined on a scale of 0 to 1 and is calculated by taking into account sleepiness as the maximum value of the scale and 8 hours or more of sleep as the lowest value. The equation that defines the sleep-failure scale was designed in such a way as to represent a person's fatigue exponentially.

$$Y = \begin{cases} 1 \; if \; x = 0 \\ x > 0 \rightarrow \dfrac{1}{x} - \alpha \end{cases} \tag{4}$$

$$Domain = [0, \infty) \quad , \quad Range = \; (0\,, 1]$$

For the intensity of hunger are taken three states: fasting, breakfast and lunch. When the student is in a state of fasting, it is harder to react in a timely manner because his body does not have the necessary nutrients to perform optimally. By contrast, if the student breakfasted before the accident, can physically respond faster; even more, if the student breakfasted and then had lunch, she/he is in a good condition to face obstacles that are presented to evacuate the building [2]. For hunger states, we have the following values:

- Fasting = 1.00; Breakfast = 0.50; Lunch = 0.01

Equation that calculates the paranoia:

$$P = \frac{(R + Temp + Fs + Af)^D}{\alpha} \tag{5}$$

For the equation that defines the socio-cultural model the following factors are taken: level of altruism (A), gender (g) and maturity (m).

The maturity factor takes into account the maturity of mental development of the student, which is usually is determined by age.

Equation of sociocultural model:

$$C = A * m * g \tag{6}$$

Finally, to carry out the calculation of the pedestrian velocity in each student the following factors are used to evaluate the probability of vital loss: earthquake magnitude (G), level of the structure (N), index of paranoia (P), time (t), sociocultural model (C) and adjustment variable (X=0.3).

The magnitude of the earthquake is not adjusted or represented by any other value, rather than the original Richter scale, therefore in decimal and can range from 4.0 to 9.0 on the Richter scale. It takes 4.0 as initial reference value because it is a scale in which the earthquake can be felt. Since there is no record of any major 8.6 earthquake in the history of our country (México), it is taken as value limit reference 9.0 [2]. The level of the structure can take the values 1, 2 or 3 depending on the level of the structure where the student is located. The index of paranoia is calculated using equation (5) and their possible values range are from 0 to 1. The time is measured in seconds and takes into account the range of 0-90 seconds, this is because it is the maximum time evacuation before the structure collapses.

In order to calculate the socio-cultural level of the student, we apply the equation (6) and as the same as paranoia, values range are from 0 to 1.

Equation that calculates the pedestrian speed:

$$V(t) = \left[\frac{G + N}{P(t)} * C * X\right] \tag{7}$$

2.1 Displacement with Gyroscope

The gyroscope installed on the Smartphone measures the movements of the player and makes our device follow the gestures and movements (Fig. 2).

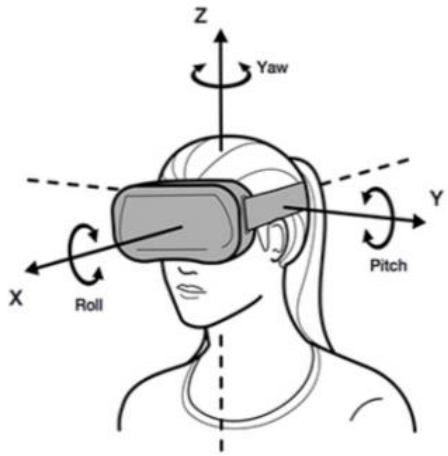


Fig. 2. Movements detected by the gyroscope [8].

A movement algorithm is proposed, so that the player can move around the stage regardless of the terrain type (Fig. 3a); the main objective of this algorithm is to detect whether the player is touching the ground or not (Fig. 3b).

```
// **** Controles Del Teclado **** \\
if (RotacionDelTeclado)
{
    x += Input.GetAxis("Horizontal") * Velocidad * Time.deltaTime;
    y -= Input.GetAxis("Vertical") * Velocidad * Time.deltaTime;
}
#if !UNITY_ANDROID && !UNITY_IOS || UNITY_EDITOR

//**** Controles Del Mouse ****\\
if (RotacionDelMouse && Input.GetMouseButton(0))
{
    x += Input.GetAxis("Mouse X") * Velocidad * Time.deltaTime *
    SencibilidadDeArrastre;
    y -= Input.GetAxis("Mouse Y") * 1.5f * Velocidad * Time.deltaTime *
    SencibilidadDeArrastre;
}
#endif
// **** Controles Táctiles **** \\
if (RotacionDeArrastre && Input.touchCount == 1)
{
    Touch f0 = Input.GetTouch(0);
    Vector3 f0Delta2 = new Vector3(f0.deltaPosition.x, -f0.deltaPosition.y, 0);
    x += Mathf.Deg2Rad * f0Delta2.x * SencibilidadDeArrastre * 10;
    y += Mathf.Deg2Rad * f0Delta2.y * SencibilidadDeArrastre * 10;
}

y = ClampAngle(y, -85, 85);
Quaternion rotation = Quaternion.Euler(y, x, 0.0f);
transform.rotation = rotation;
```

(a)

```
void Update ()
{
if (CaminarCuandoMiramos && !CaminarCuandoPulsamos && !EstaCaminando && VisionVR.transform.eulerA
    {
        EstaCaminando = true;
    }
else if (CaminarCuandoMiramos && !CaminarCuandoPulsamos && EstaCaminando && (VisionVR.transform.e
    {
        EstaCaminando = false;
    }
if (EstaCaminando)
    {
        Caminar();
    }
if (CongelarLaPosicionY)
    {
        transform.position = new Vector3(transform.position.x, CompensarY, transform.position.z);
    }
}

public void Caminar()
{
    Vector3 Direccion = new Vector3(VisionVR.transform.forward.x, 0, VisionVR.transform.forward.z
    Quaternion Rotacion = Quaternion.Euler(new Vector3(0, transform.rotation.eulerAngles.y, 0));
    transform.Translate(Rotacion * Direccion);
}
```

(b)

Fig. 3. (a) Script that contains the movement algorithm for the player, written in C#; (b); script that defines the walk of the player, written in C#.

The controls are defined in the scripts, so the player can walk on the stage. The movement of the player is defined as follows: walk forward (down); stop player (up), rotate left (left) and rotate right (right). The control movements work to handle with mouse or with Smartphone.

The implementation with Google VR SDK in the simulation gets the virtual view for the player (Fig. 4). The Google VR SDK for Unity provides additional features like spatialized audio, Daydream controller support, utilities and samples [3].

Fig. 4. The Cardboard mode allows the user to see the panorama in a Cardboard device like any other Cardboard app [4].

3 Problem Statement

The problem addressed in this research is to develop a more realistic visual environment of the evacuation of a three-level structure in a secondary school. We aim to have a more accurate observation of the behavior of the agents when they are in a state of human stampede. By using the methodology that controls the behavior of the agents used by Beltran et al. (2016) in their article "Characterization of an undulatory earthquake in a three levels building of a high school using multi-agent systems" characterizes the agents as students of the school and they are applied the model in the simulation in third dimension using "Unity". The information gathered from this study will be used to ensure that the personnel in charge of security in structures determine the factors that hinder the evacuation of the building with the exact tracking of the trajectory of each of the students. This intends to save as many human lives as possible.

3.1 Models of the Interfaces

"Unity" is an image processing and 3D development toolkit [1] that allows us to represent the scenario of the secondary school in a more real and accurate way, for that reason we take as reference the model developed in the previous work constructed in three dimensions (Fig. 5).

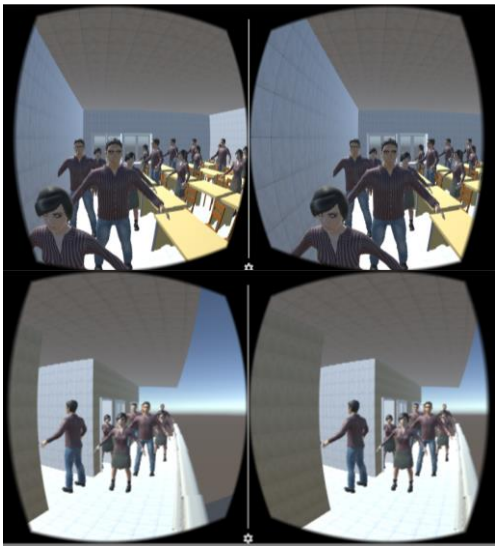
Fig. 5. Representation of the three-level structure using Unity 3D.

MakeHuman is a free and open source software for the creation of more realistic third-dimensional humans [7] to characterize high school students (Fig. 6).

The player can observe a virtual scenario as shown in Fig. 7 and 8, in which she/he can walk through the entire structure and observe the behavior of the agents.

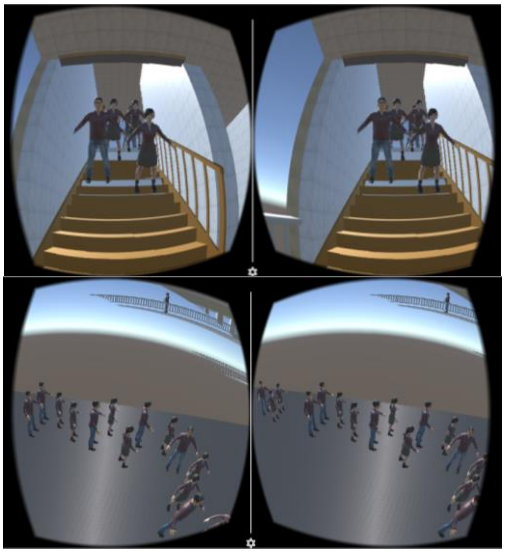
(a) (b)

Fig. 6. (a) Characterization of students' faces in MakeHuman. (b) Characterization of the body of the students in MakeHuman.



(a) (b)

Fig. 7. (a) Students starting to leave the classroom implementing Virtual Reality. (b) Evacuation seen from the corridor of the third level of the structure.



(a) (b)

Fig. 8. (a) Seen from the stairs of the second floor of the structure, this place usually occurs the human stampede. (b) View of the first students going out to the patio.

4 Design of Experiments

Once the pedestrian speed model has been mathematically tested, it was encoded and implemented in an application created in Java to generate random values (Fig. 9). These values are used to characterize the behavior of each of the students in the building. For instance, it can be seen that the velocity of student number 4 is zero at the second 20, which means that the student remained paralyzed or trapped within the structure and is interpreted as a human down.

According to the simulation, all agents spend 1 minute with 28 seconds to evacuate the structure. In order to verify to what extent the pedestrian speed model approached the real life situation with respect to the students' behavior, students were trained to perform a simulacrum, assuming that the alarm alerted them in the opportune moment. When observing the behavior of the students and later comparing it with the simulation, many similarities were detected. The most notable is the way to go down the stairs and the moment of the human stampede.

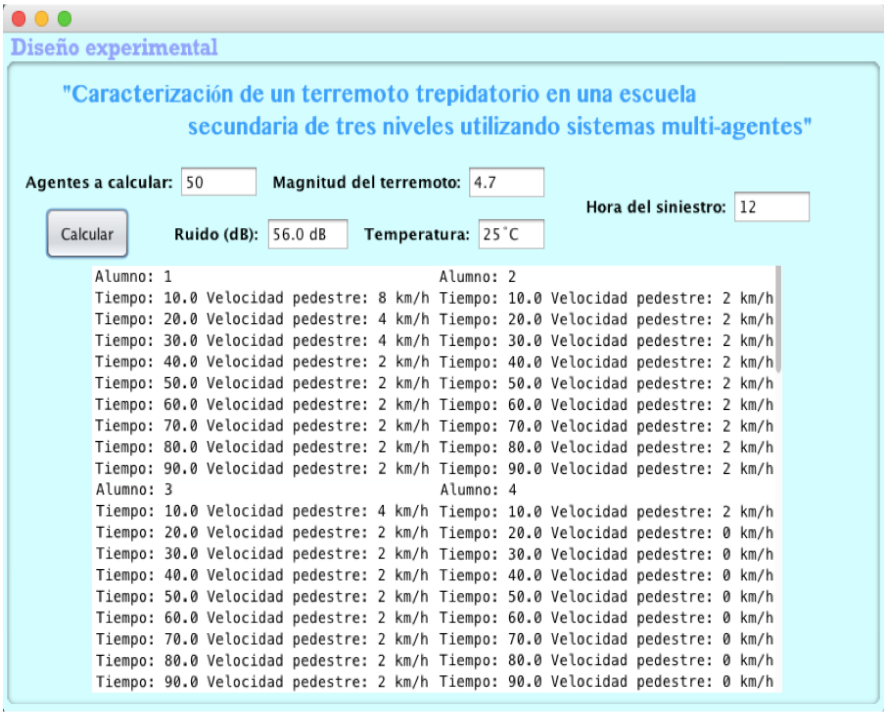


Fig. 9. Random data generation with application created in Java.

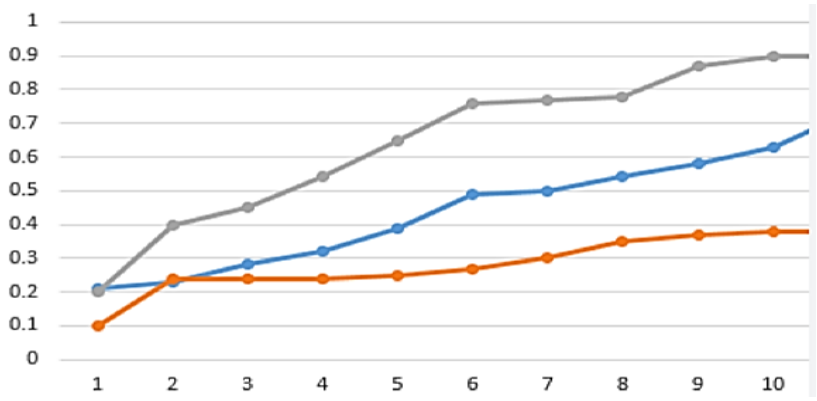


Fig. 10. The graph shows the behavior of three different students. The X axis represents the time in seconds, while the Y axis represents the pedestrian speed scale. The student represented by the orange line had a level of paranoia greater than the others while the student represented by the gray line had a lower level of paranoia, where the pedestrian speed is not very affected.

It should be noted that an accuracy of slightly more than 90% in the behavior of the agents in the simulations was obtained since in the simulacrum the students took 1 minute with 33 seconds to evacuate the building and the agents in the simulation take 1 minute with 28 seconds, a difference of 5 seconds.

5 Conclusions and Future Research

The implementation of the VR in this research has been used to detect in a more real way the possible factors that influence the safe evacuation of the building since the players can be moved throughout the building, even more, it can be perceived certain emotions similar to the adrenaline experienced in a real situation, this simulator seeks to face the challenge of saving the greatest number of human lives and eliminate obstacles to optimize evacuation routes.

As future work, it is intended that the mathematical model developed in this research can be used in different scenarios of catastrophes or natural unexpected events or even provoked, in which an increment in paranoia is caused by the situation and the environment. Where Paranoia is the main factor that intervenes in the process of escape and human movement routes, the model and the methodology can be adapted to diverse situations in which it is necessary to analyze and to propose solutions of humanitarian logistics in addition; it can be also used to perform security tests of existing properties. In near future, we are planning to apply this methodology in different scenarios with augmented reality.

Acknowledgments. Dr. Francisco Javier Bermúdez Alarcón (Director General in State Coordination of Civil Protection of the state of Morelos), for the supervision and validation of the investigation. Lic. Luz Elena Espino García (Director in General Secondary School # 4, "Valentín Gómez Farías", Cuernavaca, Morelos), for the attentions provided to carry out the simulacra in the institution. Eng. Peter Savier Oropeza. (Polytechnic University of Morelos), for the technological support offered in this research.

References

1. Autodesk Inc.: Autodesk. Autocad 2017, Website: http://www.autodesk.mx/campaigns/ (2016)
2. Beltran, R., Márquez, D., Ochoa-Zezzatti, A.: Characterization of an ondulatory earthquake in a three levels high school using multi-agent systems. In Humanitarian Logistics based on Multiagents Systems for the prevention of a plethora of disasters related with crowds, Mexico, Espinosa, ISBN: 978-136-553, pp. 29–41 (2016)
3. Google Developers.: Overview Android/NDK iOS Unity Unreal Get Started VR View Controller Support Spatial Audio VR Entry Flow VR Android Manifest VR Video Samples and Tutorials Downloads Release Notes Content Features Google VR SDK for Unity (2017)
4. Google Developers.: VR view Sample Code. Website: https://developers.google.com/vr/android/samples/vrview (2016)
5. Johnston, L. E.: A Graphical Input Device and Shape Description Interpretation Routines. In Memorandum to Prof. Mann, Massachusetts Institute of Technology, Department of Mechanical Engineering, Cambridge, Mass (1960)
6. Van den Berg, J P., et al.: Optical Reciprocal Collision Avoidance. University of North Carolina, Website: http:// gamma.cs.unc.edu/ORCA (2010)
7. MakeHuman team.: MakeHuman for education. MakeHumanTM, Website: http://www.makehuman.org (2016)
8. Olivera, C.: Augmented Reality and Virtual Reality: The king is naked! Creative Commoms, Website: http://carlosoliverablog.com/2016/08/realidad-aumentada-y-realidad-virtual/ (2016)

ISSN 1870-4069

Characterization of a Passenger Vessel Sinking and Finally Foundering in Lacustrine Environment

Peter Oropeza, Diego Hurtado, Edgar Gil, Pablo Maldonado, Miguel Ruiz

Polytechnic University of the State of Morelos, Morelos, Mexico

{mruizj}@live.com.mx

Abstract. This research considers a simulating crowds into a passenger vessel Evacuation in the face of fire and the Effect of Human Behavior and Building Characteristics on Evacuation, the most crucial aspect of a passenger`s vessel safety in the face of fire is the possibility of safe and efficient escape. The study of attitudes and behavior of passengers in relation to reactions to alarms, regulations, emergency signals, smoke and fire. An important precondition is that its fire safety facilities enable independent and adequate fire response performances by the passenger's vessel. In practice, it appears that the measures currently required by law do not always provide the support that people in burning cruiser's need and because of lack of signs. In such circumstances, timely and effective evacuation of all passengers and crew will be extremely important, and failure to evacuate in time may lead to catastrophic consequences. Consequently, understanding how individuals behave in the case of fire and fire evacuation is essential if we are to bring fire safety measures into line with occupants' needs during an incident. This paper contains a review of the available literature on human behavior in a fire so far as ship safety is concerned. The findings are presented as an overview of the critical factors which determine occupants' fire response performances, namely the characteristics of fire, human beings and ship.

Keywords: Multi-agent System, Fire response performance, Crowd Simulation, Evacuation, Human behavior under fire, Safety regulations.

1 Introduction

On the model of the passenger vessel related with evacuation in features face of fire using software simulation and the effect of human behavior and characteristics of a safe evacuation, The goal of this research is concerned that many of the actions taken by those responsible for the safety of passengers on a vessel are based on incorrect stipulations of how passengers will behave on an sinister in this case of fire aboard, Consequently the decision to raise the alarm and begin preliminary evacuation preparations is often delayed with resultant disastrous consequences. The different levels of organizations in case of an

emergency are the level of operations, the level of communication and coordination, the level of design, the level of culture & morality, and the level of social structures & organizations [9]. In every case should result in reactive to improve safety and efficiency in form of implementing a system of finding the shortest paths between departures in vessel decks for each passenger using a bracelet and implementing and a signaling system using leds and lasers, depending of the case of the risk in the vessel. One of the reasons for this is that even though a good training and procedures are essential to safe and efficient operations, human error will only be minimized but no eliminated, we therefore need to improve the emergency departures and the organization culture depends of the latent cause of erroneous actions because of lacks of signals. "The 1995 amendments to the International Convention on Standards of Training, Certification and Watch keeping for Seafarers, 1978, (STCW 95)", outline specific requirements for safety-related training. Included in this training is the requirement that all persons having responsibility for the safety of passengers in an emergency on board a passenger vessel undertake training in crisis management and human behavior in emergencies. The factors that impacts is how the crew were trained and even the passengers training through exercises on board vessel, to design new corridors with some kind of lighting that can help the passengers to guide them to the departures and lasers depending of the danger of the corridor, leisure areas and accommodation space, and to implement of a safety plan to prevent disasters. The factors such as delayed reaction to alarms, the walking speed depending of the type of person , is different the walking speed of a adult of old man , the fact that usually every passenger try to find his family members before being evacuated and problems with instructions from the crew, lacks of signals, the language of the signals or even panic, these humans factors problems should be addressed as part of the evaluation of efficiency of evacuation .The others parts that the project carried out experimental studies of the safety and affiance of lighting corridors and emergency led signs and the efficiency of different techniques of evocation like lifeboats, lifejackets, fire suppression system of the vessel and slide and fall system [3]. Usually some passengers will refuse to leave a particular place where they had an agreement to meet again with relatives. Other passengers will refuse to go directly to the assembly station because they want to search the ship to find their children, spouse or other relatives. These parameters are taken into account in evacuation simulation models. When the fire occurs, the smoke reduces the visibility and it will put and added pressure on evacuation because of the reduced time available of a secure scape and the hard way to find the closest and safest departures, and the visibility. Under these circumstances panic can also arise, if passengers get trapped in closed areas or if they judge that there is enough time for orderly escape because of the reduced visibility and finding the shortest and secure paths between exits in vessel decks. Usually the vessels passengers are guided towards and verbally instructed to go to the nearest exit without delay and under these circumstances, evacuation paths are not an issue. Technical assistance for way finding by means of graphic and acoustic signs or through the architectural design becomes important in large buildings and passengers ship, where the spatial layout is more complex [4,5]. Here we find many different sings like maps and networks that are usually used to direct occupants and passengers under normal

conditions of way finding. Way finding systems and acoustic sings as technical support for way finding during evacuation, especially in cases of smoke and in fire [7,8].

2 Methodology

2.1 Pedestrian Forces

The forces that accelerate a pedestrian are the result of internal (will) and external (collision, signals, etc.) forces [13]. An alert individual will try to avoid collisions before they happen, resulting in vanishing external forces. However, if the pedestrian density increases, collision forces will appear and the ability to move freely will be impaired. The difference between external and internal forces is admittedly not too rigorous, and only serves descriptive purposes in the present context. The forces considered are:

Internal forces: Will force, Pedestrian collision avoidance forces and Obstacle/wall avoidance forces.

External forces: Pedestrian contact forces and Obstacle/wall contact forces.

Additionally, we consider: Kinematic constraints: Motion inhibition forces.

We denote by will force the force that will accelerate (or decelerate) a pedestrian to achieve the velocity it desires. Depends by factors such as height (H), weight (W), footwear (F), drunkenness (D), inability (I), fatigue (FA). Given a desired velocity vd and the current velocity v, this force will be of the form:

VT = age * displacement (H, W, F, D, I, FA).

The modeling aspect is included in the function of the effects of "group binding" (finding family members before being evacuated), which we denote by produce force the force that will accelerate - decelerate a pedestrian or the crown to achieve the velocity it desires. Because some passengers will refuse to leave a particular place, where they had an agreement to meet again with relatives, other passengers will refuse to go directly to the assembly station because they want to search the ship to find their children, spouse or other relatives. We can thus obtain the speed of walking by measuring the time required to reach a percentage (e.g. 40%) of the desired velocity, starting from rest. This speed is typically in the range of 17 km per hour but obviously depends on the pedestrian's current state of fitness and stress, fatigue, drunkenness, inability, group binding, footwear, weight, height, desire to reach a goal, climate, signals, and noise, among others.

In this case, we have an equation:

$$v(ti) = \frac{E * K(ti) * \frac{\sum_{j=1}^{n} j}{n} \lambda}{C} \simeq 1, \quad (1)$$

where

E: Age,
K: Levels of care,
K= (a+b)/2,
a: level of drunkenness,
b: level of fatigue,
J: attribute of the agent,
I: Agent,
t: Time,
n: number of attributes per agent,
λ: inability,
C: Number of relatives.

2.2 Sociocultural Model

Sociocultural model is defined as:

$$F\ reaction⟦= (P^{*}\ \lambda^{*}T^{*}U^{*}V^{*}A(x))⟧^{\wedge}s, \tag{2}$$

$$\lambda= (q+n)/2, \tag{3}$$

where:
Religiosity =P,
Education level= Q,
Education = λ,
Values= R,
Language = S,
Social level= T,
Region= V,
I: The universe of existing languages,
S: The set of languages that dominates the person,
A(x): The person dominates the language the native language of the vessel,
K= Optimal Value,
S=Level of solidarity.

2.3 The Shortest and Secure Path

We use graphs to find the shortest paths or routes from starting point to a final destination between departures in vessel decks, in order to represent the shortest path, we use graphs. A graph is mathematical abstract object, to represent the way of the passenger in the vessel, we use vertices and edges, which Edges connects pairs of vertices in this case exits. Where vertices represent exits and edges represent routes that connect the exits. Along the routes of the graph it is possible to walk by moving from one vertex to other vertices. Depending

on whether or not one can walk along the routes by both sides or by only one side determines if the graph is a directed graph or an undirected graph, in this case is a directed graph. In addition, lengths of routes are called weights, and the weights are used for calculating the shortest path from routes point to another routes. There exist different types of algorithms that solve the shortest path problem. However, only several of the most popular conventional shortest path algorithms adapts to our model and is based in Dijkstra's Algorithm [10]. The algorithm's that are going to be discussed in this paper: Dijkstra's Algorithm, Floyd-Warshall Algorithm, Bellman-Ford Algorithm and Genetic Algorithm. We proposed an algorithm obtained by extending the Dijkstra's algorithm [10]. In this paper we use it for solving the single source shortest path problem.

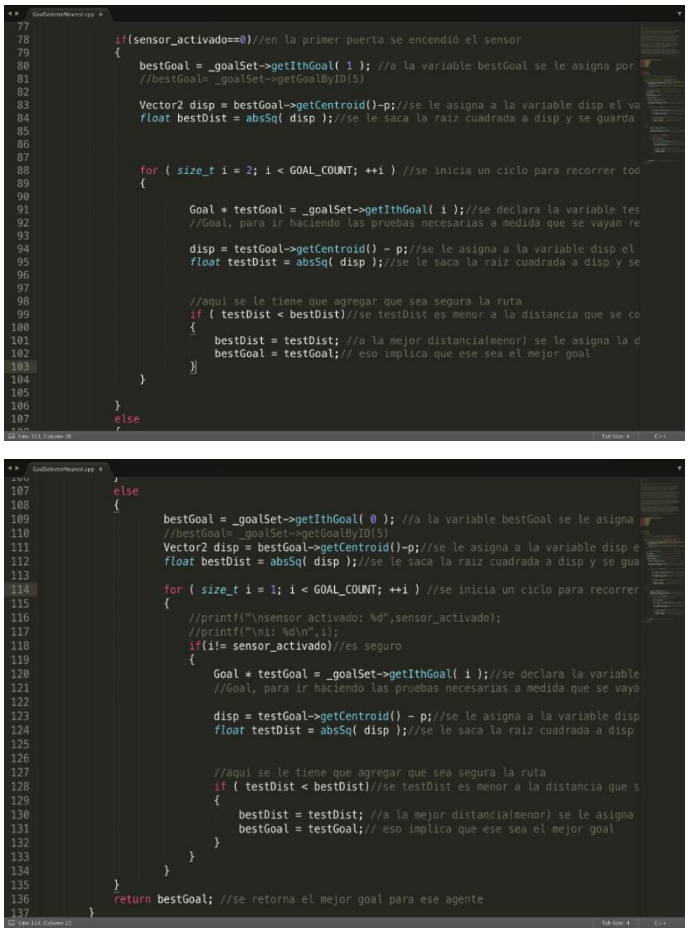


Fig. 1. Shows a variant of the Dijkstra's algorithm, which is written in C++, finding the shortest and secure path function.

2.4 Mathematical Model of Spread of Fire inside a Passenger Vessel

The development of a mathematical model for predicting rate of fire spread and intensity to a wide range of fire inside a passenger vessel is presented from the conceptual stage through evaluation, variables as Oxidizing, fuel, wind velocity, relative humidity, slope, and temperature or activation energy were all recognized as producing important effects of the spread of fire. These factors were studied and correlated to some form of fire behavior.

1. Fuel: Any material layers to release energy when oxidized violently with heat release (Any flammable material).
2. Oxidizing 2- (oxidizing agent) is an agent that can oxidize a fuel and in doing this boils down to itself (in our case is oxygen, the air is composed of about 21 % oxygen)
3. Temperature or activation energy: this is the amount of heat needed to start the combustion of a material.

REACTION FREES CHAIN: A process that allows for continuity and spread of fire giving off heat that is transmitted to fuel combustion and continuing as long as the supply of fuel and oxidizer is maintained.

$$CR_{(ti)} = (F * H * O * T) * WV, \quad (4)$$

F → Fuel.
O → oxidizer.
T →Temperature.
CR → Chain reaction.
H →moisture.
WV → wind velocity.

This equation will be used to go to adding or subtracting the coordinate values of the state to be a final this way when an agent is reached by the fire area die example:

CR= 0.9,
Min_x= -2.5 -0.9 = -3.4.
CR= 1.5,
Min_y= -1.0 -1-5 = -2.5.
CR= 1.0,
Max_x= 2.0 + 1.0 = 3.0.
CR= 0.7,
Max_y= 3.6 + 0.7 = 4.3.

3 Problem Statement

The overall problem is addressed, is better problem addressed in this dissertation is to develop a methodology for emergency evacuation in the face of fire on a passenger's vessel

planning that considers to help and management options of scape by choosing the shortest path for each passenger in order to facilitate the mobility of evacuees to the safety. The methodology is manifested in the development of a system of finding the shortest paths for each passenger. The objective of this research was to review current emergency evacuation plans and procedures for a passenger vessel. And also make recommendations for the use of advanced technologies where appropriate to improve evacuation during a sinister. The study did not consider an earthquake which are essential parts of a sinister emergency management system. This information could allow officials determine which areas need to be evacuated and when the call for mandatory evacuation is needed. Accurate tracking of a fire aboard reduces the chances of the decision-makers to make false alarms. Repeated false alarms will have a negative effect on the population as they start losing faith in the decision-makers. Thus, false alarms should be avoided.

4 Models of the Interfaces

The following figures show the interface modules.

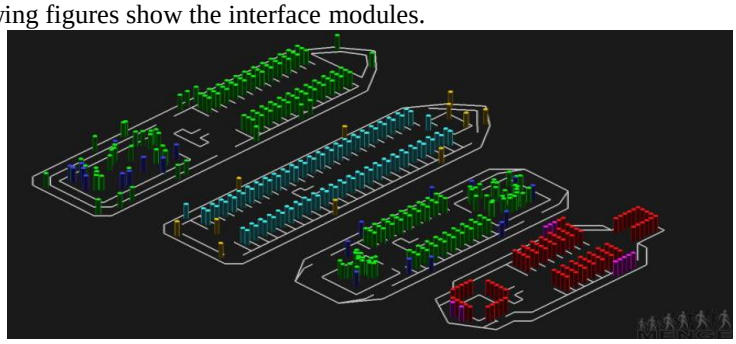


Fig. 2. Implementation in Menge: Main screen simulator contains the view of the 4 levels of the boat deck and the passengers for each level, using a simulation framework Menge.

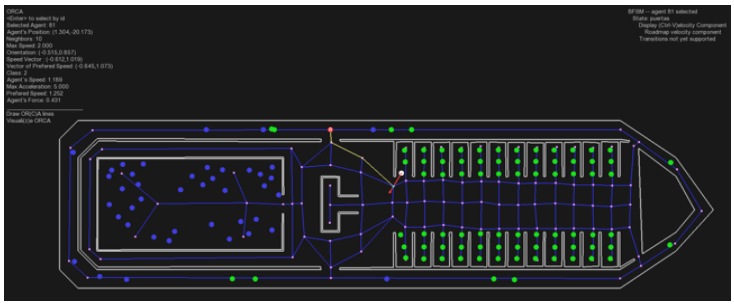


Fig. 3. Implementation in Menge: The screen shows the implantation of the system finding the shortest paths between departures in vessel decks for each passenger and the departures is free of danger.

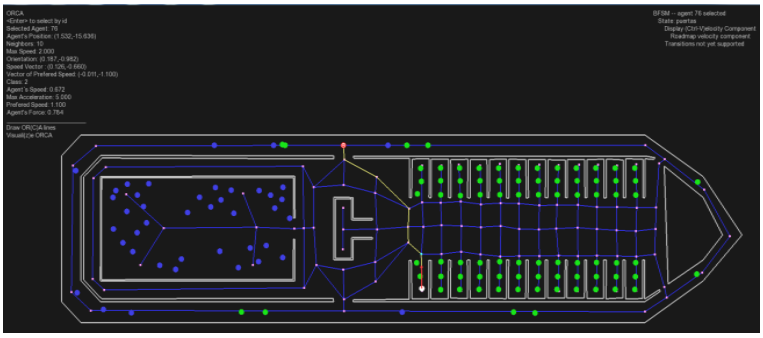


Fig. 4. Implementation in Menge: The screen shows the recalculation of the departures to the selected passenger, due to the state of danger occasioned in his appropriate departure.

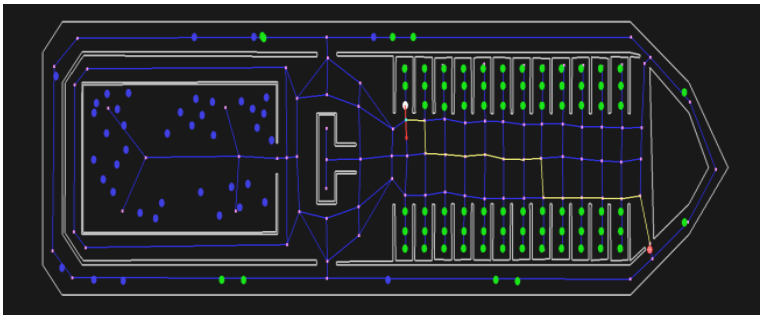


Fig. 5. Implementation in Menge: The screen shows how the selected passenger is obtaining the farthest route to his departure, due to the passenger is using a random evacuation route.

Fig. 6. Implementation in Unity, we can visualize these results in real time with the 3D viewer using the software unity.

5 Design of Experiments (DOE)

It is worth mentioning that every simulation, the position of each passenger is the same. The total of passengers aboard in the vessel is of 443 passengers. Using the shortest path, achievement saves the 100 percent of the capacity crowd, in which they were made 25 simulations with 36 seconds on average to evacuate.

Fig. 7. Passenger flow at the vessel decks in the exercise and in the simulation counting (Y axis: number of passengers who still staying in a state of danger and X axis: Numbers of simulations. The result of the simulation selecting one departure of evacuation in a random way: red, the result of the simulation using the shortest and secure path system: Blue.

Afterwards were made 25 simulations in mode in which each agent selects one departure of evacuation in a random way in the same 36 seconds, in order to view the results in that way, it was found that on average of 39 passengers who were in the state of danger (in the inner part of the boat covers) which is equivalent to the 8.8 percent of the capacity crowd; in other words, is achieved only allocating the 91 percent of the capacity crowd into safety. Using that, every agent chooses their own evacuation path without knowing if it will lead them to risk or safety, while we were using the method of the shortest path of evacuation for each agent, is achieved to save the 100 percent of the capacity crowd.

With this we prove the following hypotheses: minimize through the use of software simulation the loss of lives in a sinister even who is the passenger. The percentage of passengers that were safe in comparison to using the system of the shortest path, indicated in the paper, shows that 100% of capacity crowd on the vessel is achieved to be safe while without the use of the system only 91% of capacity crowd were save.

6 Conclusions and Future Research

This research has analyzed the risks associated with collision, fire aboard and grounding of a passenger vessel and the importance of designing for safe, swift and efficient evacuation

on board of a passenger vessel. Representing a serious threat to maritime safety and for the passenger vessel carrying a large number of people on board. This sinister have the potential to become devastating, and investigate possible events subsequent to an incident, sinking and capsizing, expected time to sink, etc.

In order to study fire aboard, grounding and collision scenarios, we have been developed models, to improve the evacuation performance. The research that we reported here sought to prove the following hypotheses: minimize through the use of software simulation the loss of lives in a sinister even who is the passenger, I mean the socioeconomic status and the socioeconomic class stratification. Unfortunately, due to socioeconomic diversity among the passengers, the layout of the ship, the evacuation is not as fluid as one would expect. These factors coupled together may have led directly impacting the survivability of a disaster. The research proposed intends on determining to what extent the socioeconomic factors played in the low mortality rate of a passenger vessel.

Those in charge of an emergency on a passenger ship believe that: panic is a natural occurrence in an emergency, sounding an alarm, such as a fire alarm, will cause panic, as a result of the above, the alarm should be delayed until absolutely necessary, the research also sought to prove that the above assumptions and beliefs are generally held as a result of media reporting of incidents rather than actual participation in emergencies. Finally, some risk control options are proposed that might reduce the risk of collision, fire aboard and grounding of passenger vessel by implementing a system of finding the shortest paths for each passenger. The information obtained from the simulations and the research will however have to be evaluated against general knowledge about crowd behavior and specific knowledge about in a situation behavior in emergencies and finding the shortest paths between exits in vessel decks. Our evaluation has shown a significant improvement in the evacuation rates when using a system of finding the shortest paths between exits in vessel decks for each passenger using a bracelet. Only a relatively small percentage of agents with good skills yields the best evacuation rates. We can visualize these results in real time with either our simple 2D using the software Menge or 3D.

Acknowledgments. This study has been supported by the Barcelona Supercomputing Center, located at the Torre Girona chapel and the National Prevention Center of disasters (CENAPRED) , would like to thank Dr. Alberto Ochoa Zezzati, Research Professor, Autonomous university of Juarez city, who inspired us to the research.

References

1. Arnskiv, M. M.: Waydinfing Systems, A proposal for new wayfinding system of passenger ferries. MEP desing report, Lyngby: DML (1999)
2. Emergency Management Australia: Flood Warning: An Australian Guide. Mount Macedon, Victoria (1995)
3. Human factors in ship design & operation II. RINA, London (2002)

4. May, M.: Wayfinding, Passenger behavior and maritime emergencies sings. In: Dick de Waard (ed.), Proceedings of the annual Meetings of the human factors and ergonomics society – Europe, Chapter November 7-9, Torino (2001)
5. May, M., Jorgensen, H. D., Ostergaard, E.: Panning of mustering on board m/f Kronprins Frederik. MEPdesing Report, Lyngby: DMI & Scandlines (2000)
6. Passini, P.: Wayfinding in architecture. Van Nostrand, New York (1992)
7. Perry, R, W.: A model of evacuation compliance behavior. In: Dynes, R. R & Tierney K. J (eds.), Disasters, collective behavior and social organization, University of Delaware Press, Newark (1994)
8. Withington, D.: Localisable, Alarms. In: Stanton, N. A. & Edworthy, J. (eds.), Human factors in auditory Warnings, Ashgate, Aldershot (1999)
9. Withington, D.: Use of Directional sound for ship Evacuation. In: 8th Word conference of Emergency Management (TIEMS 2001), Oslo, pp. 19–21 (2001)
10. Chamero, J.: Dijkstra's Algorithm Discrete Structures & Algorithms. (2006)
11. Helbing, D., et al.: Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. Transportation science, Vol. 39, No. 1, pp. 1–24 (2005)
12. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. Physical review, Vol. 51, No. 5 (1995)
13. Helbing, D., et al.: Simulation of pedestrian crowds in normal and evacuation situations. Pedestrian and evacuation dynamics, Vol. 21, No. 2, pp. 21–58 (2002)

ISSN 1870-4069

Implementation of an Algorithm for the Transfer of Citrus Using an Intelligent Model for Trains

Fernando García-Isidro[1], Julissa Martinez-Lugo[1], Alberto Ochoa[2], José Alberto Hernández Aguilar[3]

[1] Instituto Tecnológico de Superior de Naranjos, Mexico

[2] Maestría en Cómputo Aplicado, UACJ, Mexico

[3] FCAeI, Universidad Autónoma del Estado de Morelos, Mexico

Abstract: We consider the implementation of an algorithm for the transfer of citrus using an intelligent model for trains, this research is performed based on the problem of packing in containers (Bin Packing), so that objects or boxes should be accommodated in a finite number of cubes considering to minimize the use of containers. The problem of packing containers will be understood and made according to the space of the container, the calculation will be made by maximizing the number of items - or boxes of citrus- that can be stored safely, the dimensions of the container shall also be addressed, taking into account that all have the same capacity and size, an algorithm based on MATLAB was used for the arrangement of objects in containers with containers crossing name is used.

Keywords: Container, Packaging, Trains, Capacity, Dimension, Objects, Bin Packing, MATLAB.

1 Introduction

For its simplicity and combinatorial nature of the problem of packaging container makes it an NP- hard problem, that is to say, that there is no algorithm that can solve all instances of the problem in a given polynomial time with respect to the number of input objects.

The problem of packing containers has great relevance for transporting, loading trains, trucks and cargo vehicles with weight capacity, distribution of tasks in homogeneous processes, and organization objects using a computer through blocks.

In order to solve the problem for a large number of objects it has been established several very simple algorithms such as:

Next fit (AS): In this algorithm objects are listed, the first is assigned to the container number 1, the rest are assigned to the current container if fit, otherwise is assigned the next container, the new container becomes the current container.

First-fit algorithm (PA): In this algorithm, the objects are listed as containers, therefore the first object is assigned to the container 1, the rest of the objects are assigned to another container where they fit, and if the objects do not fit in any container which are assigned a new one is created.

Best Fit algorithm (MA): This algorithm is a modification of the AS for which the current object to the container having the minimum space available to store the object is assigned. In case of a tie between two or more containers, the one with the lowest risk is chosen.

In order to improve these three types of algorithms as a first step you can rearrange the objects depending on the space they occupy in descending order.

A typical example of the problem and its solution according Lodi A, S Martello, Vigo D are shown in Figures 1 and 2 respectively.

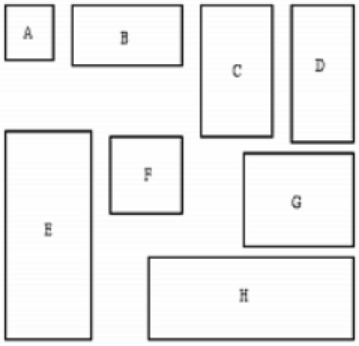


Fig. 1. Input.

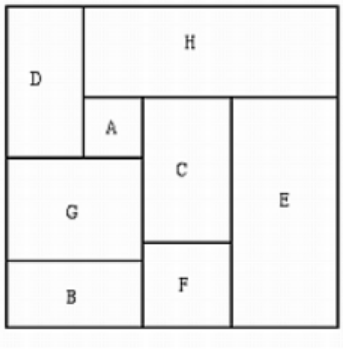


Fig. 2. Output.

Fig. 1. shows the objects (A, B, C, D, E, F, G, H) and observed that there are spaces between objects, on the other hand, Figure 2 shows a solution for this type of objects. [4]

For all types of wheeled vehicles able to run on a railway track are known as trains or rolling stock, which aims to transport different types of cargo.

Man's need to transport large amounts of cargo led him to build machines with great traction. George Stephenson built the first locomotive on 25 July of 1814, which later led to a railroad. At first locomotive it was used in coalmines where the drag of 40 tones was shown at a speed of 40 kilometers per hour.

The railroad was the product of the industrial revolution originated in England in the eighteenth and nineteenth centuries. The evolution of the locomotive has been remarkable, and its use has not only been for freight, but also for the transport of persons, both large and small distances.

The iron transport is the transport used worldwide. It offers a variety of different types of rail freight cars in boxcars, gondolas, hoppers, and trailer, duplex and triplex tank cars.

The vans are used to transport products that require protection from the environment without temperature control.

Otherwise boxcars transport perishable goods with temperature control; gondolier transport materials that do not require protection from the environment; the hoppers are used for the transportation of industrial products such as coal, scrap metal, etc.

Loads and dimensions are stipulated in international standards [6].

Fig. 3. Railroad transportation [6].

The speed of a freight train is about 70 kilometers per hour, is fixed by the general direction of railways for which takes two conditions [6].

The first: is established for each class of locomotive and is based by building system.

The second: is indicated by the state of the roads and the conditions of protection thereof.

The rate may be increased only with special authorization from the general direction of railroads.

Data on some containers to store items, restrictions of use and maximum total load

Table 1. Specifies the maximum load and dimensions of some transports containers currently on the market [6].

Name	**Length**	**Width**	**Height**
12’ steel dry box	3.431 m.	2.352 m.	2.395 m.
20’ steel dry box	5.900 m.	2.352 m.	2.395 m.
40’ steel dry box	12.022 m.	2.352 m.	2.395 m.
Name		**Maximum load**	
20’ dry		47,782 kg.	
40’ dry		58,058 kg.	
40’ high cube		57,550 kg.	

Fig. 4. Container [6].

2 Kind of Wagons

Boxcars: used to protect the goods from bad weather, theft and vandalism, the most used is the sliding walls.

Wagons edges: they are open boxes above and are used for transporting wood, scrap metal, etc.

Platforms: are used to transport vehicles, are considered high speed reaching 220 km/h, its structure is aluminum and not carbon steel.

Cage: used for transporting livestock.

Containerships: flatcars are having fasteners for containers.

Hoppers: are used to transport solid bulk cargo as aggregates, coal, minerals, etc. [1]

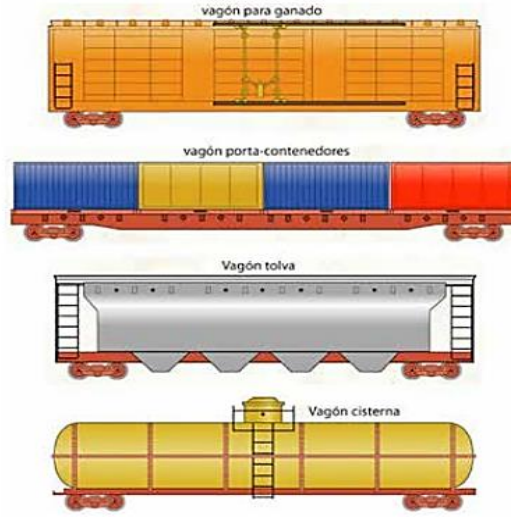


Fig. 5. Type of wagons.

3 Mode and Transport

Transport modes are systems for mobilization, used in the transfer of goods from one source point to a destination.

The means of transport are the different physical elements used in the modes of transfer of goods to distribute in diverse location with far distance.

All modes and means of transport of goods, have advantages and disadvantages which must counter the logistics manager, to select the appropriate transport to the type and amount of cargo to be sent.

Table 2. Mode and meaning of transport train.

Modes	media
Railroad	railway train
	They are special equipment with sealed wagons, chairs, and container type wagons. Usually charge bulk goods such as coal, grain, and platforms are also used for loading containers.

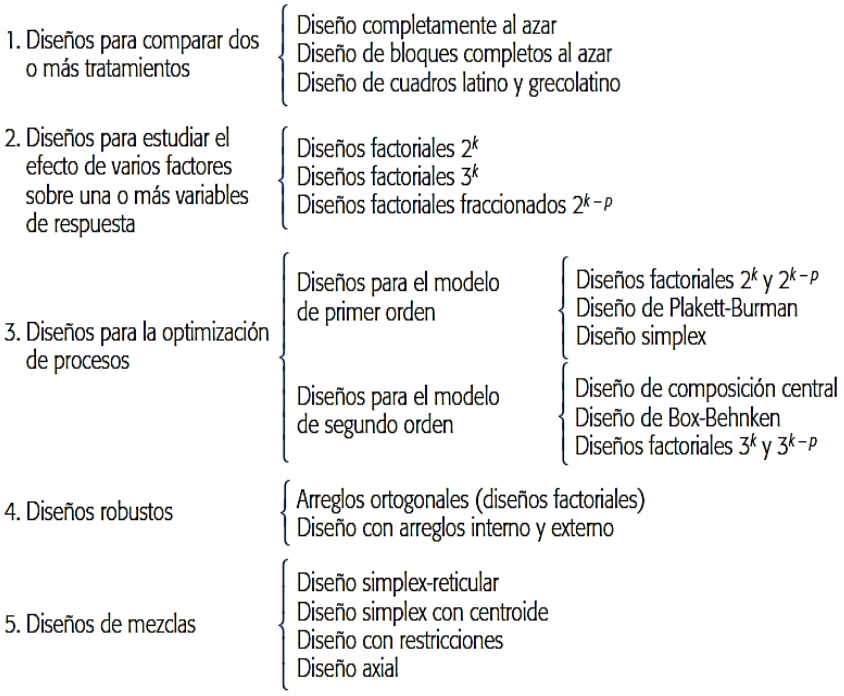


Fig. 6. Classification of experimental designs [8].

An algorithm for the optimization problem is represented by a primary quantity of items to save (multiplied by the size of occupying, being of the same type) , this result is subtracted from the total size of the container, if the space subtracts the container and if there are more items on hold, you choose to fill the container without triple the weight of all items, the weight of the container.

According [3] design of experiments is to plan and perform a set of tests in order to generate data that , when analyzed statistically , provide objective evidence to answer the questions raised by the experimenter situation.

4 Process for Obtaining Citrus Products

Harvesting methods are the same as those used for fruit eaten fresh , fruit is cut from the tree using ladders and is manually selected , the fruit is placed in baskets that can hold from 80 to 100 kilograms , to be carried to the transport vehicle [5].

Table 3. Major producers of citrus municipalities in Veracruz per hectare.

County	Naranja (Ha.)	Lima persa (Ha.)	Mandarina (Ha.)	Tangerina (Ha.)	Toronja (Ha.)	Total (Ha.)
TEMAPACHE, VER.	42,626	53	2,994	2,812	571	48,955
MARTINEZ DE LA TORRE, VER	14,423	16,049	6	644	2,790	33,912
TIHUATLAN, VER.	13,783	270	1,118	813	82	16,066
ATZALAN, VER.	4,840	4,383	55	2,264	210	11,752
TUXPAN, VER.	4,791	69	2,389	2,399	150	9,798
BENITO JUAREZ, VER.	8,000	0	0	0	0	8,000
CASTILLO DE TEAYO, VER	8,675	13	106	178	57	9,029
PAPANTLA, VER.	9,870	664	160	50	203	10,947
GUTIERREZ ZAMORA, VER	8,050	70	200	100	335	8,755
TECOLUTLA, VER.	6,376	441	41	151	265	7,275
MISANTLA, VER.	3,610	1,012	120	1,400	175	6,317
CHICONTEPEC, VER.	3,646	0	0	0	0	3,646
CAZONES DE HERRERA, VER.	3,191	16	58	189	250	3,704
TAMIAHUA, VER.	1,911	20	439	332	4	2,705
TLAPACOYAN, VER.	2,257	0	0	0	0	2,257

Table 4. Main municipalities of citrus growers in Veracruz per tonne.

County	Naranja (Ton.)	Lima persa (Ton.)	Mandarina (Ton.)	Tangerina (Ton.)	Toronja (Ton.)	Total (Ton.)
TEMAPACHE, VER.	701,679	105	59,875	39,339	12,346	813,344
MARTINEZ DE LA TORRE, VER	185,358	173,789	60	8,356	76,093	443,656
TIHUATLAN, VER.	192,962	3,100	22,360	16,260	0	234,682
ATZALAN, VER.	59,865	46,540	750	26,598	4,902	138,655
TUXPAN, VER.	57,492	1,021	38,224	28,788	3,000	128,525
BENITO JUAREZ, VER.	128,000	0	0	0	0	128,000
CASTILLO DE TEAYO, VER	121,450	91	1,900	3,560	0	127,001
PAPANTLA, VER.	98,400	5,500	0	0	4,000	107,900
GUTIERREZ ZAMORA, VER	75,800	750	2,500	1,200	3,800	84,050
TECOLUTLA, VER.	71,520	4,350	550	1,800	4,750	82,970
MISANTLA, VER.	34,397	11,955	864	19,880	3,430	70,526
CHICONTEPEC, VER.	58,336	0	0	0	0	58,336
CAZONES DE HERRERA, VER.	51,056	150	1,200	3,780	0	56,186
TAMIAHUA, VER.	26,754	296	7,016	3,981	80	38,127
TLAPACOYAN, VER.	33,339	0	0	0	0	33,339

Fig. 7. Citrus production in Mexico.

Veracruz leading citrus production in Mexico with a volume of 2.9 million tons. Followed by Tamaulipas with 638 thousand tons, and Mexican lime -producing states, Colima and Michoacan with 590,000 and 462,000 tons respectively [5].

Table 5. States citrus production in Mexico.

No	ESTADO	PRODUCCIÓN (ton.)	No	ESTADO	PRODUCCIÓN (ton.)
1	Veracruz	2,987,973.65	15	Quintana Roo	42,044.45
2	Tamaulipas	638,262.49	16	Jalisco	36,000.58
3	Colima	590,372.49	17	Baja California Sur	32,624.72
4	Michoacán	462,999.49	18	Chiapas	20,438.12
5	San Luís Potosí	436,915.80	19	Nayarit	16,022.84
6	Nuevo león	399,739.22	20	Sinaloa	14,344.80
7	Oaxaca	253,270.70	21	Baja California	11,318.94
8	Yucatán	251,073.34	22	Morelos	6,765.90
9	Puebla	225,595.90	23	Durango	2,420.70
10	Sonora	191,208.30	24	México	1,804.90
11	Tabasco	141,300.00	25	Querétaro	1,415.50
12	Guerrero	85,816.56	26	Zacatecas	260.00
13	Hidalgo	58,731.05	27	Aguascalientes	124.00
14	Campeche	58,267.56	28	Guanajuato	44.60

5 Problem Packaging of Objects in Containers: Bin Packing

The problem of distribution of objects in containers, in English Bin Packing is a classic problem of NP- hard combinatorial optimization . Because Bin Packing is a highly complex problem can not be solved big cases using an exact algorithm . An optimal solution can be found by considering all ways to make a partition of n objects in a subset of size n or smaller.

6 Algorithms for Optimization Problems and Generic Packaging Containers

$$O(g(x)) = \left\{ \begin{array}{l} F(x): \text{Exists } c, x_0 > 0 \text{ such as} \\ \forall x \geq x_0: 0 \leq |f(x)| \leq c\,|g(x)| \end{array} \right\}$$

A container f (x) belongs to items g (x) , when there is a positive constant c such that from a quantity of the items x0 , f (x) not on passing containers (x) , means that the function f is less aga from a given value by a constant factor.

Generic algorithm (GA) .
Assuming a search space x, and the function:
f: x → R
The general problem is to min f
X ε x

Where x is the vector of decision variables and f is the objective function. This methodology consists of three major phases . The first is the generation population (chromosomes) ; the second is the recombination of these chromosomes or crossover criterion and the last mutation . Finding the best solution it is guided by the results and evaluation of the objective function f for each of the individuals generated . Finally , based on this evaluation, the chromosome with higher value of f represents the best solution [4].

```
Elegir una población inicial de cromosomas;
 mientras (la condición no es satisfecha) hacer
    Repetir
        si (condición de cruce es satisfecha) entonces
        {Seleccionar los cromosomas padres;
         Elegir parámetros de cruce;
         mostrar cruce};
        si (condición de mutación es satisfecha) entonces
        {elegir puntos de mutación;
        mostrar mutación}
        evaluar funcion objetivo
        hasta que se crean hijos suficientes;
      Seleccionar nueva población;
    fin mientras.
```

Fig. 8. Description of the operation of AG [4].

Program in Matlab of the generic algorithm for the problem of packaging in containers. By using a container crossing.

```
function [pob bins]=cruce(numind,nuevabins,nuevapob,pesos,capacidad)
Cruce de contenedores
bjacm=[];
 for i3=1:size(listap,2)
 [listac indc]=sort((hijo(1:numbinsaux,numobj+2))',2,'ascend');
 for i4=1:size(listac,2)
 if (listac(i4)+listap(i3)) < capacidad
 hijo(indc(i4),hijo(indc(i4),numobj+1)+1)=vecobj(indp(i3));
 hijo(indc(i4),numobj+1)=hijo(indc(i4),numobj+1)+1;
 hijo(indc(i4),numobj+2)=hijo(indc(i4),numobj+2)+listap(i3);
 objacm=[objacm vecobj(indp(i3))];
 listac(i5)=listac(i5)+listap(i3);
```

```
break
end
end
end
vecobj=setdiff(vecobj,objacm);
if size(vecobj,2)~=0
indbin=1;
numbinsaux=numbinsaux+1;
for i3=1:size(vecobj,2)
if (hijo(numbinsaux,numobj+2)+pesos(vecobj(i3)))>capacidad
numbinsaux=numbinsaux+1;
indbin=1;
end
hijo(numbinsaux,indbin)=vecobj(i3);
hijo(numbinsaux,numobj+1)=indbin;
hijo(numbinsaux,numobj+2)=hijo(numbinsaux,numobj+2)+pesos(vecobj(i3));
indbin=indbin+1;
end
end
nuevapob(:,:,conthijo)=hijo;
nuevabins(conthijo)=numbinsaux;
conthijo=conthijo+1;
end
end
pob=nuevapob;
bins=nuevabins;
```

In this code is explain each iteration related with the correct location of objects in a transport model by train.

Table 6. Iteration 1.Testing generic algorithm: Crossing container.

Number of executions	Data	Result	% Total	Probability Overload
1	0,1,1,1,1	39.8	37.1	0.374
2	1,0,0,0,1	24.4	22.8	0.226
3	1,1,0,0,0	22.4	20.9	0.208
4	0,1,0,0,0	20.4	19.0	0.189
Sum		107	99.8	0.998

Table 7. Executions in Matlab program.

>> [pob bins]=cruce(0,1,1,1,1) pob = 1 bins = 1	**>> [pob bins]=cruce(1,0,0,0,1) pob = 0 bins = 0**	**>> [pob bins]=cruce(0,1,0,0,0) pob = 0 bins = 1**	**>> [pob bins]=cruce(1,1,0,0,0) pob = 0 bins = 1**

Table 8. Iteration 2.

Number of executions	Data	Result	% Total	Load problems
1	1,1,0,0,0	22.1	31.9	0.319
2	0,1,0,0,0	20.4	28.9	0.289
3	1,0,0,1,1	18.4	26.1	0.261
4	0,1,1,0,1	9.8	13.9	0.139
Suma		70.4	100.8	1.008

Table 9. Executions in Matlab program.

>> [pob bins]=cruce(1,1,0,0,0) pob = 0 bins = 1	**[pob bins]=cruce(0,1,0,0,0) pob = 0 bins = 1**	**[pob bins]=cruce(1,0,0,1,1) pob = 0 bins = 0**	**[pob bins]=cruce(0,1,1,0,1) pob = 1 bins = 1**

Table 10. Iteration 3.

Number of executions	Data	Result	% Total	Load problems
1	1,0,0,0,1	24.4	22.8	0.228
2	0,1,1,1,1	39.8	37.1	0.371
3	1,1,0,0,0	22.4	20.9	0.209
4	0,1,0,0,0	20.4	19.0	0.190
Suma		107	99.8	0.998

Table 11. Executions in Matlab program.

>> [pob bins]=cruce(1,0,0,0,1) **pob =** **0** **bins =** **0**	**[pob bins]=cruce(0,1,1,1,1)** **pob =** **1** **bins =** **1**	**[pob bins]=cruce(1,1,0,0,0)** **pob =** **0** **bins =** **1**	**[pob bins]=cruce(0,1,0,0,0)** **pob =** **0** **bins =** **1**

Table 12. Iteration 4.

Number of executions	Data	Result	% Total	Load problems
1	1,1,0,0,0	22.4	31.5	0.315
2	0,1,0,0,0	20.4	28.7	0.287
3	1,0,0,1,1	18.4	25.9	0.259
4	0,1,1,0,1	9.8	13.8	0.138
Suma		71	99.9	0.999

Table 13. Executions in Matlab program.

[pob bins]=cruce(1,1,0,0,0) **pob =** **0** **bins =** **1**	**[pob bins]=cruce(0,1,0,0,0)** **pob =** **0** **bins =** **1**	**[pob bins]=cruce(1,0,0,1,1)** **pob =** **0** **bins =** **0**	**[pob bins]=cruce(0,1,1,0,1)** **pob =** **1** **bins =** **1**

Table 14. Iteration 5.

Number of executions	Data	Result	% Total	Load problems
1	1,0,0,0,1	24.4	22.8	0.374
2	0,1,1,1,1	39.8	37.1	0.226
3	1,1,0,0,0	22.4	20.9	0.208
4	0,1,0,0,0	20.4	19.0	0.189
Suma		107	99.8	0.998

Table 15. Executions in Matlab program.

>> [pob bins]=cruce(1,0,0,0,1) **pob =** **0** **bins =** **0**	**>> [pob bins]=cruce(0,1,1,1,1)** **pob =** **1** **bins =** **1**	**>> [pob bins]=cruce(1,1,0,0,0)** **pob =** **0** **bins =** **1**	**>> [pob bins]=cruce(0,1,0,0,0)** **pob =** **0** **bins =** **1**

Table 16. Capacity and containers are listed in the following table.

	Container 1	Container 2	Container 3	Container 4	Container 5
Capacity (tons)	2	2.4	3	4	4.4

The capacity of all containers is determined as follows :

(1, 1, 1, 1, 1) = 41.8 (1, 1, 1, 1, 0) = 29.4 (1, 1, 1, 0, 1) = 29.8
(1, 1, 1, 0, 0) = 7.4 (1, 1, 0, 1, 1) = 38.8 (1, 1, 0, 1, 0) = 8.4
(1, 1, 0, 0, 1) = 8.8 (1, 1, 0, 0, 0) = 22.4 (1, 0, 1, 1, 1) = 39.4
(1, 0, 1, 1, 0) = 9 (1, 0, 1, 0, 1) = 9.4 (1, 0, 1, 0, 0) = 23
(1, 0, 0, 1, 1) = 18.4 (1, 0, 0, 1, 0) = 24 (1, 0, 0, 0, 1) = 24.4
(1, 0, 0, 0, 0) = 20 (0, 0, 0, 0, 0) = 18 (0, 1, 1, 1, 1) = 39.8
(0, 1, 1, 1, 0) = 9.4 (0, 1, 1, 0, 1) = 9.8 (0, 1, 1, 0, 0) = 23.4
(0, 1, 0, 1, 1) = 18.8 (0, 1, 0, 1, 0) = 24.4 (0, 1, 0, 0, 1) = 24.8
(0, 1, 0, 0, 0) = 20.4 (0, 0, 1, 1, 1) = 19.4 (0, 0, 1, 1, 0) = 25
(0, 0, 1, 0, 1) = 25.4 (0, 0, 1, 0, 0) = 21 (0, 0, 0, 1, 1) = 34.4
(0, 0, 0, 1, 0) = 22 (0, 0, 0, 0, 1) = 22.4

Results are calculated as follows:

The values 18 and 8 only considered the containers 4 and 5 have a storage capacity of relatively equal [8].

Examples:
(1, 1, 1, 1, 1)= 2+2.2+3+4+4.4+18+8=41.8,
(0, 1, 0, 1, 0)=2,4+4+18=24.8,
(0, 0, 1, 1, 0)=3+4+18=25.

7 Conclusions and Future Work

The results obtained on the problems of packaging containers experimentally demonstrate that combinatorial problems that are untreatable or difficult to solve with an exhaustive

search or classical techniques, can be solved satisfactorily with evolutionary algorithms that are modified appropriately.

An advantage of these processes is that it is not necessary to know the description of the problem, a way to solve the problem is with the implementation of the generic algorithm, but is complicated adapting techniques to find an optimal solution, and thats where the generic algorithms have found an important area of solution.

In future work we should be studied more chromosomes of the cell or humans to solve problems with the help of artificial intelligence.

References

1. Berbey, A., et al.: Trenes: Material rodante del transporte ferroviario.
2. Caliz, E.: Validación de la calidad de las soluciones obtenidas en la tesis de maestría optimización integral del problema de carga de contenedores tridimensionales (2012)
3. Gutiérrez, A., et al.: Análisis y diseño de experimentos. Tercera edición, Editorial Mc Graw Hill (2013)
4. Hernández, N., et al.: Introducción a Matlab para Resolver Problemas de Ingeniería Aplicando Algoritmos Genéticos (2013)
5. Licona, J.: Estudio de Mercado para identificación de necesidades de infraestructura logística para la comercialización de jugo de cítricos en Veracruz (2009)
6. Mata, G., et al.: Un problema de tipo Bin Packing (2011)
7. Reglamento General de Ferrocarriles. Aprobado por Decreto Nº 90.325 del 12 de Setiembre de 1936 y actualizado al 31 de Diciembre de 1995
8. Vidal, A.: Algoritmos heurísticos de optimización. Universidad de Santiago de Compostela, Facultad de Matemáticas Master en Técnicas Estadísticas (2013)

ISSN 1870-4069

Improving Travels of the Public Transport System of Guadalajara Using ACO

Edgar Gonzalo Cossio Franco[1], José Alberto Hernández Aguilar[2], Alberto Ochoa-Zezzatti[3], Julio César Ponce Gallegos[4]

[1] Universidad Da Vinci/Universidad Enrique Díaz de León, Mexico

[2] Universidad Autónoma del Estado de Morelos, Mexico

[3] Universidad Autónoma de Ciudad Juárez, Mexico

[4] Universidad Autónoma de Aguascalientes, Mexico

kofrran@gmail.com, alberto_doctor@hotmail.com, alberto.ochoa@uacj.mx, julk_cpg@hotmail.com

Abstract. This paper presents the application of ant algorithm to improve public transport route in the metropolitan area of Guadalajara. We use a methodology with an eight-step sequence in which it is possible to: 1. Identify the problem, 2. Identify the sources of information, 3. Prepare the census to collect the data. 4. Raise the information, 5. Analyze the data, 6. Treat the information, 7. Implement the algorithm and 8. Analyze the results. We use a methodology with an eight-step sequence in which it is possible to: 1. Identify the problem, 2. Identify the sources of information, 3. Prepare the census to collect the data. 4. Raise the information, 5. Analyze the data, 6. Treat the information, 7. Implement the algorithm and 8. Analyze the results. The contribution is to improve the routing Public transport in the ZMG through the implementation of Bio-inspired algorithms that will shorten the distance and the time of Travel from origin to destination.

Keywords: C++, Instances, Traffic, Ant Colony Optimization.

1 Introduction

In the last population census of 2010, conducted by the INEGI, identified in the area of mobility growth rates in terms of vehicle fleet it has been increasing: the rate of motorcycles increased by 30% annually between 1990 and 2010, from 16,000 to 177,000. In the same period, the rate grew to 7.31% cars, and trucks and cargo vans to 7.66%. The lower rate is that of passenger buses -3.49% - certainly closer to the growth of housing in the state. ZMG concentrated two-thirds of the vehicle fleet with 65%.

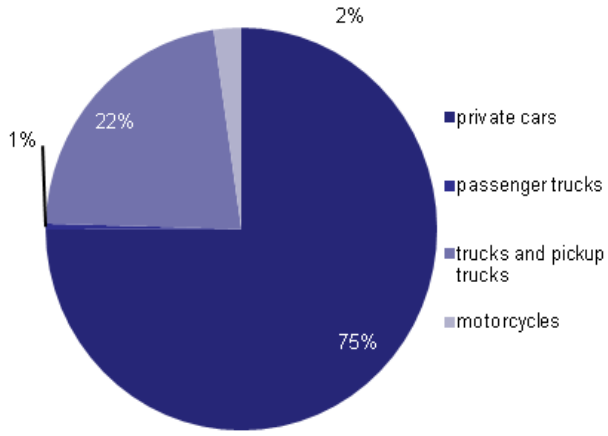


Fig. 1. ZMG Vehicle fleet [1].

In the figure above 4 types of transport are shown. It is identified as the predominant percentage use of private cars and the least dominant is represented by passenger trucks [1].

A bio-inspired algorithm based on ant colony is proposed for the optimization of the routing problem.

According to the literature reviewed, the following elements were found:

- Application of ant algorithm

The ant algorithm is a powerful implementation to solve complex problems; is based on the application of heuristics and probability for decision making [2].

It is possible to use ACO to estimate the traffic flow paths, as established in [3], where the application is shown to a specific case of Havana.

ACO is used in pharmaceutical industry to analyzing proteins for the study of AIDS [4].

Various routing optimization systems can be found in the market using bio-inspired algorithms, among which the following can be found:

- SOSACO

It is a search engine for social networks. The algorithm identifies new paths easier without breaking the original graphs [5]. This intelligence is useful in the application of GPS systems and in the precision they offer [6].

- MIDACO

Mixed Integer Distributed Ant Colony Optimization is an innovative optimization software for continuous and combinatorial problems. A special feature of MIDACO is its parallelization capability. The MIDACO parallelization option is based on the concept of reverse communication and can be used in all common distributed computing support

architectures (including HPC and GPU). This function aims to parallel the evaluation (expensive CPU time) of the objective functions and constraints. It is designed for time consuming applications, where a single evaluation requires seconds, minutes or even hours [7].

- Vehicle routing problem

The original assumption of VRP is the problem of vehicle routes of a city; the problem is in the distribution of routes, times and load for each unit. The goal is to achieve [8]:

- that all nodes are visited exactly by a vehicle once,
- that all routes begin and end at a depot,
- For each unit, the load does not exceed the capacity of vehicles,
- For each unit, the distance does not exceed the given limit.

VRP extends its variants to suit different road problems [9]:

- CVRP (Capacitated Vehicle Routing Problem)

The units have uniform capacity and must meet customer demands known for a single product from a common warehouse at a minimum cost of transit. That is, CVRP is like VRP with the additional limitation that each unit must equal the uniform capacity of a single product.

- VRPTW (Vehicle Routing Problem with Time Windows)

Each of the clients, as well as the deposit, has several time windows of delivery or delivery [10].

2 General Objective

In Jalisco the number of vehicles has grown over the last thirty years by 7.29% per year, this amount is above the percentages of population and housing; motorcycles have 30%, cars with 7.66% and finally public transportation with 3.49%.

The ZMG contains two-thirds of total vehicles with 65% [11]. One of the main problems, when talking about mobility in public transport, is the time it takes to reach a unit from the origin to the destination, caused by ignorance of the route, inaccessibility to roads, climate issues, traffic load (peak hour / hour valley).

Our goal is to implement ACO to a problem of public transport in the metropolitan area of Guadalajara with the aim of improving travel.

3 Process

As defined at the beginning of the present study, the proposed methodology is composed of eight stages which are shown in the Figure 2.

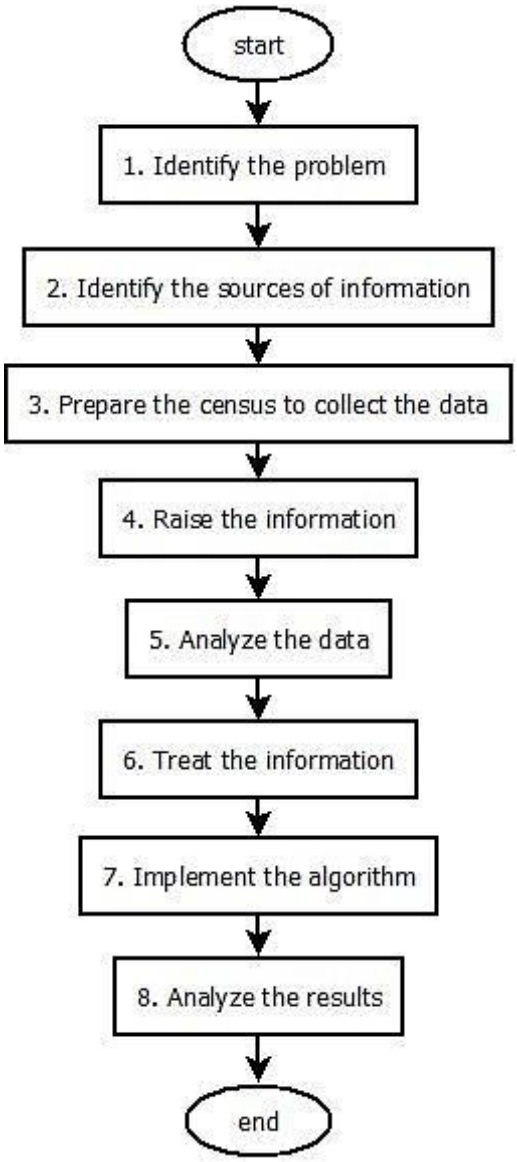


Fig. 2. Process.

The total distance is calculated, the number of routes for each execution and the number of vehicles, with the inputs 25 cities (nodes).

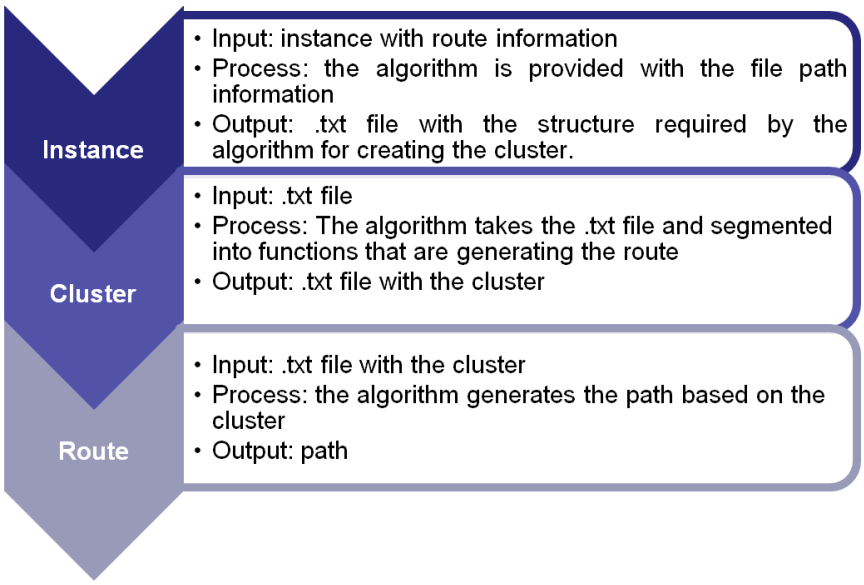


Fig. 3. Stages of processing.

The process consists of three elements; an instance as input, clusters are created and finally the route is created.

4 Experiment

The complete path consists of the following data:

The following data is used in the experiment:

- 33 units,
- 71 drivers,
- 37 points,
- Daily average per unit:
  - Passengers: 528,
  - Laps: 4,
  - Time of turns: 50 minutes.

The experiment consists of:

1 Identify points along the route.

Table 1. Control points.

trip	number	punto	lat	long	X	Y
going	1	Base 214	20.63128	-103.272	95	73
	2	Farm GDL	20.63564	-103.276	92	68
	3	Rio Nilo	20.63904	-103.277	92	63
	4	Curva Malecon	20.64367	-103.278	92	56
	5	Templo Malecón	20.64762	-103.274	94	50
	6	Pizzería 61	20.65312	-103.28	90	43
	7	Bolería	20.6552	-103.294	80	40
	8	Ramón Val-Pensador	20.65748	-103.309	70	37
	9	Arena Jalisco	20.66469	-103.317	64	26
	10	Cuartel colorado	20.67049	-103.334	52	18
	11	Zapatería prado	20.67298	-103.342	46	15
	12	Pque. Sn. Fco.	20.67357	-103.348	42	14
	13	8 de julio- la paz	20.67216	-103.353	38	16
	14	Paztolsa	20.66566	-103.36	34	26
	15	Lab. Piza	20.66386	-103.365	30	28
	16	Retorno 61	20.658	-103.378	20	36
	17	Isla Gomer	20.63682	-103.392	11	66
	18	Reyes Heroles	20.62875	-103.391	12	77
	19	Zansibar-Imss	20.62631	-103.391	11	82
	20	Contestación	20.62345	-103.4	6	84

trip	number	punto	lat	long	X	Y
return	1	contestación	20.62345	-103.3999	6	84
	2	Zansibar-Imss	20.62631	-103.391	11	82
	3	Reyes Heroles	20.62875	-103.391	12	77
	4	Isla Gomer	20.6368	-103.392	11	66
	5	Retorno 61 Enrique D/Washing	20.65792	-103.377	20	36
	6		20.66463	-103.364	32	33
	7	Paztolsa	20.66651	-103.359	34	26
	8	Pque. Sn. Fco.	20.67286	-103.347	44	22
	9	Med 56	20.67099	-103.341	48	25
	10	Cv Ruiz Sanchez	20.66268	-103.32	63	37
	11	Vinos la playa	20.65814	-103.308	72	42
	12	Mercado Osos	20.65281	-103.294	82	49
	13	Templo Malecón	20.64762	-103.274	94	50
	14	Subida malecón	20.64169	-103.277	94	65
	15	Santa Rosalia- Mal	20.63678	-103.279	93	72
	16	Farm GDL	20.63516	-103.275	92	68
	17	Base 214	20.63128	-103.272	95	73

2 Identify the number of passengers moving from one point to another.

Table 2. Instances.

Number	X	Y	Demand
1	95	73	14399
2	92	68	9763
3	92	63	9892
4	92	56	5029
5	94	50	16833
6	90	43	23515
7	80	40	25165
8	70	37	13695
9	64	26	30269
10	52	18	12624
11	46	15	9037
12	42	14	21949
13	38	16	30900
14	34	26	29200
15	30	28	31100
16	20	36	33461
17	11	66	6754
18	12	77	5020
19	11	82	4794
20	6	84	0

Number	X	Y	Demand
21	6	84	0
22	11	82	45261
23	12	77	16684
24	11	66	48247
25	20	36	147
26	32	33	59589
27	34	26	40230
28	44	22	36416
29	48	25	20500
30	63	37	16673
31	72	42	21461
32	82	49	10546
33	94	50	16833
34	94	65	10300
35	93	72	7200
36	92	68	928
37	95	73	0

For the experiment, three factors were selected: Shifts, Units and Days to determine the route. The turns are identified in three moments: M, T and N where M = Morning (8 a.m.), T = Late (12 p.m.) and N = Night (6 p.m.).

The selection of the shifts seeks to identify which route is the most demanded and in which shift. Three units (r5, r12 and r23) are analyzed. Finally we consider the days of the

week from Monday to Friday (d1, d2, d3, d4 and d5) where d1 = Monday, d2 = Tuesday, d3 = Wednesday, d4 = Thursday and d5 = Friday.

The Figure 4 shows the analysis mentioned.

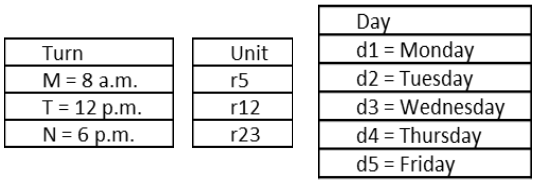


Fig. 4. Factors for analysis.

The next step is to prepare the matrix in Excel with the data for analysis in Statgraphics Centurion XVI. The result is shown in Image 4:

	U5				
	d1	d2	d3	d4	d5
M	335	288	286	307	285
T	205	294	291	284	149
N	202	128	101	156	178

	U12				
	d1	d2	d3	d4	d5
M	207	32	246	299	292
T	84	168	135	113	113
N	252	341	246	0	130

	U23				
	d1	d2	d3	d4	d5
M	292	332	286	228	164
T	255	295	258	261	205
N	207	189	183	164	0

Fig. 5. Data for the Matrix.

The data in each table represent the sum of the loads that are defined by time as the case for M, T and N.

In total there are 45 records identified by time, unit, day and capacity.

The matrix generated with the above information is shown in Table 3:

Table 3. Result matrix.

time	Unit	day	capacity
T	r23	3	258
M	r12	5	292
T	r12	1	84
T	r23	2	295
T	r23	1	255
M	r5	2	288
M	r23	3	286
T	r5	2	294
N	r5	2	128
T	r12	3	135

N	r12	4	0
N	r12	3	246
M	r5	4	307
M	r12	3	246
T	r23	5	205
N	r5	4	156
N	r5	3	101
T	r5	1	205
T	r12	5	113
N	r23	2	189
T	r23	4	261
N	r23	5	0
N	r23	1	207
M	r5	3	286
T	r5	5	149
N	r5	5	178
T	r5	3	291
N	r5	1	202
T	r12	4	113
M	r12	1	207
N	r12	5	130
M	r12	4	299
T	r5	4	284
N	r12	1	252
M	r12	2	32
M	r23	2	332
M	r5	1	335
M	r23	1	292
N	r23	3	183
M	r23	5	164
N	r23	4	164
N	r12	2	341
T	r12	2	168
M	r5	5	285
M	r23	4	228

The experiment seeks to analyze the variable capacity and identify which shift has the highest demand. The summary for the capacity variable is shown in Table 4:

Table 4. Result matrix.

Count	45
Average	210.356
Median	207.0
Variance	7722.69
Standard deviation	87.8788
Coeff. of variation	41.7763%
Minimum	0
Maximum	341.0
Range	341.0

The capacity variation is shown in Figure 6:

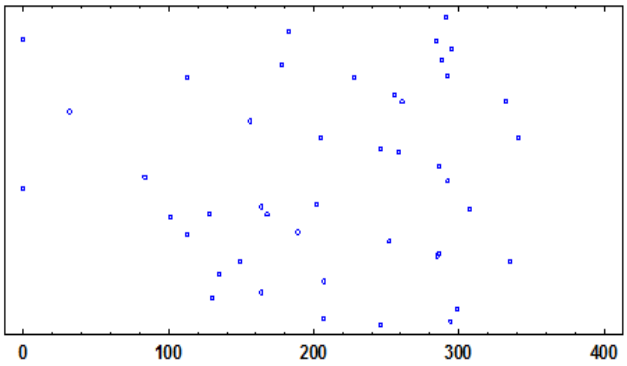


Fig. 6. Capacity variation.

In the figure 7 you can see in which shift there is more load:

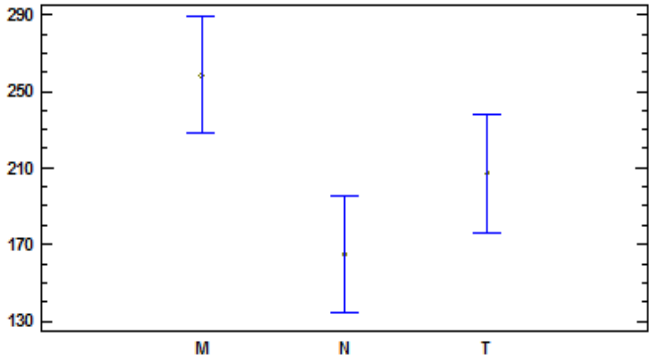


Fig. 7. Shifts.

The graph shows that the lap where most charges are in the M, approximately 290, followed by T, it is the turn of 12 pm with 200 and less load is N, that is to say that at 8 pm people do not even require one unity.

The Figure 8 shows a larger capacity per shift, identifying M with greater capacity for unit r5 while for the same turn of unit r12 is the one that has less demand.

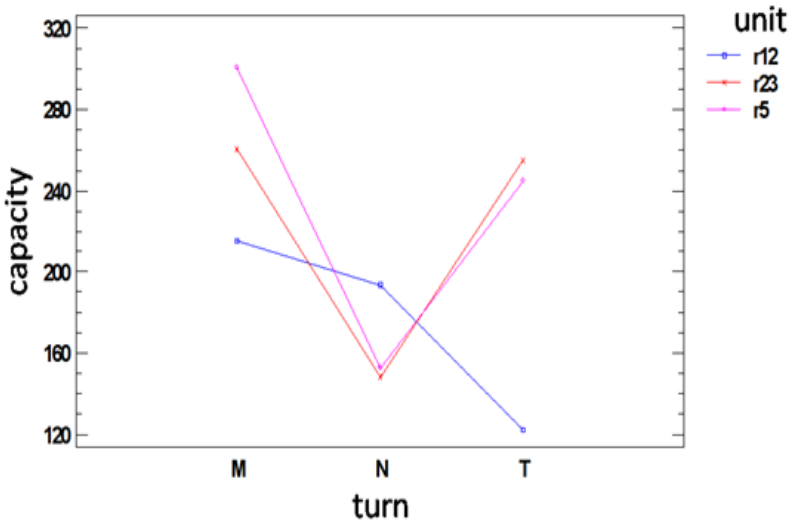


Fig. 8. Capabilities Per Turn.

We take the unit that has the most demand in the shift that more request and with greater capacity as case of application of the algorithm to determine an improvement. The case is for unit 5, at time M on day 1, i.e. unit 5 has more demand and therefore moves more people on Mondays; In the Table 5 the analysis can be appreciated:

Table 5. Result matrix.

time	unit	day	capacity
M	**r5**	**1**	**335**
M	r5	4	307
M	r5	2	288
M	r5	3	286
M	r5	5	285

Points by passing the unit M are 12, as shown in the Table 6:

Table 6. Crossing points.

Demand	Point	unit	time
13	PAZTOLSA	5	04/04/2016 08:03
47	8DEJULIOLAPAZ	5	04/04/2016 08:10
10	PQUE.SN.FCO	5	04/04/2016 08:13
27	ARENA COLISEO(LUCHA)	5	04/04/2016 08:16
45	CRUZ VERDE RUIZ SANCHEZ	5	04/04/2016 08:22
11	VINOS LA PLAYA	5	04/04/2016 08:26
38	MERCADO DE LOS OSOS	5	04/04/2016 08:32
41	TEMPLOMALECON	5	04/04/2016 08:40
35	SUBIDAMALECON	5	04/04/2016 08:43
9	STA,ROSALIA MAELCON	5	04/04/2016 08:45
35	FARMACIA GUADALAJARA	5	04/04/2016 08:47
24	BASE214	5	04/04/2016 08:51

The instance generated from the previous points is shown in Table 7:

Table 7. Instance IM12.

number	x	y	demand
1	34	26	13
2	38	16	47
3	42	14	10
4	64	26	27
5	63	37	45
6	72	42	11
7	82	49	38
8	94	50	41
9	94	65	35
10	93	72	9
11	92	68	35
12	95	73	24

3 Apply the algorithm [12].

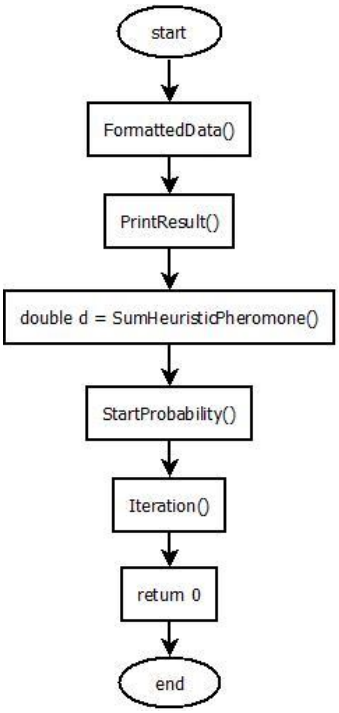


Fig. 9. The algorithm.

The main method consists of 5 functions in which processing algorithm is performed, as shown previously.

5 Results

The results derived from the execution of the algorithm with the instance IM12 are shown in the Table 8:

Table 8. IM12 results.

cluster	cost	iterations	route					
cluster-r5-clu1-3.vrp	81	9	0	2	1	0		
cluster-r5-clu2-4.vrp	128	14	0	2	3	1	0	
cluster-r5-clu3-3.vrp	145	100	0	1	2	0		
cluster-r5-clu4-5.vrp	170	8	0	1	4	2	3	0

A cluster is generated by each array that creates the algorithm; It was decided to section in cluster to facilitate the analysis to a point more granular and not so general.

The cluster column contains the generated instances, the cost column is the result of the algorithm processing, the iterations column contains the cycles and finally the route column contains the suggested route.

The following images show the results of the execution

cluster-r5-clu1-3.vrp

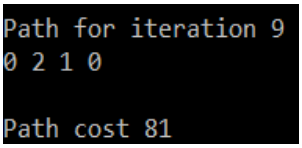


Fig. 10. Execution result cluster-r5-clu1-3.vrp.

cluster-r5-clu2-4.vrp

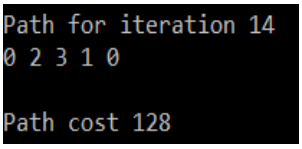


Fig. 11. Execution result cluster-r5-clu2-4.vrp.

cluster-r5-clu3-3.vrp

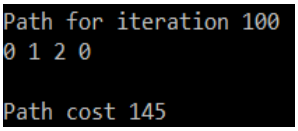


Fig. 12. Execution result cluster-r5-clu3-3.vrp.

cluster-r5-clu4-5.vrp

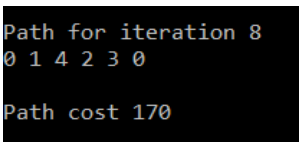


Fig. 13. Execution result cluster-r5-clu4-5.vrp.

6 Conclusions

It is possible to improve the conditions under which units of public transport work by applying the ant colony algorithm.

The sequential ACO algorithm was implemented and the results were promising in terms of performance in response times and distribution of the units at different points and this suggests a reduction in the units needed to cover the sections.

It is possible to visualize the route geographically since the 37 control points of the route are identified and located.

This work shown the importance of real instances to solve specific problems supported by information technologies and metaheuristics.

7 Future Work

Future work is expected to apply parallelism using CUDA to further improve the algorithm and apply the present approach to MATLAB and Smart Cities.

References

1. SEMOV. Available at: http://semov.jalisco.gob.mx (2016)
2. Fox, B., Xiang, W., Lee, H. P.: Industrial applications of the ant colony optimization algorithm. The International Journal of Advanced Manufacturing Technology, 31(7), pp. 805–814 (2006)
3. Bolufé-Röhler, A., Otero-Pereira, J. M., Fiol-González, S.: Traffic Flow Estimation Using Ant Colony Optimization Algorithms. Computación y Sistemas, pp. 37–50 (2014)
4. Grau, I., Nápoles, G.: Mutating HIV Protease Protein Using Ant Colony Optimization and Fuzzy Cognitive Maps: Drug Susceptibility Analysis. Computación y Sistemas, pp. 51–63 (2014)
5. Rivero, J. C.: Using the ACO algorithm for path searches in social networks.
6. UC3M. Available at: http://portal.uc3m.es/portal/page/portal/actualidad_cientifica/noticias/buscador_redes_sociales (2016)
7. MIDACO-SOLVER. Available at: http://www.midaco-solver.com (2015)
8. Bullnheimer, B., Hartl, R., Strauss, C.: An improved Ant System algorithm for the Vehicle Routing Problem. Annals of Operations Research, 89, pp. 319–328 (1999)
9. NEO. Available at: http://neo.lcc.uma.es/vrp/vrp-flavors (2015)
10. Blocho, M., Czech, Z. J.: Parallel Processing and Applied Mathematics. 9th International Conference, PPAM 2011, Torun, Poland (2012)
11. Gutiérrez-Pulido, H.: Diez problemas de la población de Jalisco: una perspectiva sociodemográfica. Gobierno de Jalisco (2010)
12. Md. Rabiul Islam. Available at: https://github.com/rabiulhasibul/ACO-for-CVRP (2016)

ISSN 1870-4069

Simulation and Application of Algorithms CVRP to Optimize the Transport of Mineral Metallic and Nonmetallic by Rail to Export Level

Edna Cruz[1], Alberto Ochoa[2], José Alberto Hernández Aguilar[3], Eliud De La Cruz[1]

[1] Instituto Tecnológico Superior de Naranjos, Veracruz, Mexico

[2] Maestría en Cómputo Aplicado, UACJ, Ciudad Juárez, Mexico

[3] FCAeI, Universidad Autónoma del Estado de Morelos, Morelos, Mexico

Abstract. Metallic and nonmetallic minerals produced at the State of Puebla, for convenience, can be transported in containers by train to a seaport cargo to be exported to other countries; for transporting minerals by railways, it must analyze which is the most optimal route to bring the product, taking into account various factors involved through each route in order to get a greater benefit for the transportation of minerals by train. This article compares two metaheuristic algorithms applied to Capacitated Vehicle Routing Problem (CVRP), in order to determine which algorithm gives better optimization solutions that help to make the best route for the transfer. As a final result, the implementation of an Ant Colony optimization algorithm was more successful in the runtime that a Genetic Algorithm; because it is slow to find an optimal solution among all generations.

Keywords: Metallic and nonmetallic, Export, Container train, Seaport, Metaheuristic Algorithms, Capacitated Vehicle Routing Problem (CVRP).

1 Introduction

The mining sector is one of the main sources of economy in Mexico, where the industrial sector, which provides social benefits that put Mexico in the main destination for investment in mineral exports in developed in 24 of the 32 states of the Republic Latin America and the fourth worldwide [1].

Mexico is among the top 10 in the production of 18 minerals [1]; it ranks second in the production of metallic minerals, the exploitation of silver and bismuth; fifth in lead; sixth in zinc, molybdenum and cadmium; and thirteenth place in iron production worldwide. In addition to exploiting non-metallic minerals, in which ranks second as a producer of fluorite, fifth in diatomite, sixth in barite, graphite and plaster, and eleventh place in feldspar, according to INEGI, 2014, explain in [4].

The state of Puebla is located in the east central area of Mexico and is strategically located within four physiographic regions; Neo Volcanic Axis, the Sierra Madre del Sur, the Sierra Madre Oriental and the Gulf Coastal Plain. This makes the state a major producer mostly of non-metallic minerals [2].

Table 1. Volume of mining production 2012- 2013 in the state of Puebla (tons).

MINERALS	2012	2013
METALLIC		
Iron	6, 925 t	6, 925 t
NON-METALLC		
Limestone	5,220,082 t	4,437,070 t
Sand	4,292,000 t	3,651,563 t
Gravel	2,007,008 t	1,705,957 t
Calcite	1,567,678 t	1,506,767 t
Clay	555,000 t	471,750 t
Cast	258,980 t	220,133 t
Feldspar	352,308 t	138,279 t

Documentary Management Control and Strategic Indicators, Ministry of Economy, National Institute of Statistics and Geography, S.H.C.P.

Minerals that the state of Puebla exploits are minerals such as feldspar, Mexico exports it mainly to Venezuela, United States, Colombia and Chile, the plaster is exported to countries: United States, Ecuador, Costa Rica and Venezuela. Inside the exploitation of metallic minerals export of iron is characterized to countries such as Hong Kong, Japan, the United States and Canada. To carry out the export of these minerals from Puebla, it requires subsequent transfer to a nearby seaport, since most of the products exported by our country are transported by sea, due to the high load capacity and adaptability to transport all kinds of products [2, 3].

2 Problem to Resolve

Minerals like Iron, feldspar and gypsum exploited by the state of Puebla, are on the list of major minerals exported our country [2]; so as the need to exploit the mining sector of the State of Puebla to transport their minerals by railways to a nearby seaport where they can be exported. To do this is necessary to consider various factors that influence the transfer from Puebla to the seaport in order to find the best route for transporting minerals.

3 Justification

In the railway system in Mexico we can find intermodal railways corridors for maritime type where each route has its weight restrictions.

a) Intermodal maritime corridors. b) Railways classified by weight capacity.

Fig. 1. Railways with maritime intermodal corridors and symbology weight capacity on lines [4].

In this investigation will be compared the routes from the train station in Puebla to the station near the seaport in Veracruz Port by train railways and the second route to the train station at the seaport of Altamira, Tamaulipas; by metaheuristics techniques taking into account the characteristics of minerals to be transported, in this case: Iron, Feldspar and Cast to choose the right type container train to use. In the course of the two routes, the train passes through different stations to which we will call nodes; with the help of algorithms, routes will be analyzed to compare their benefits.

Table 2. Characteristics of metallic and nonmetallic minerals.

METALLIC MINERAL	TECHNICAL DATA.	USES
Iron.	It is soft, with a metallic sheen, of steely gray to black. It corrodes in air humid conditions or elevated temperatures that causing an oxidation reaction.	Steel, coating , chemicals, agrochemicals, pharmaceuticals, cosmetics, paints, water treatment, construction, refining, electrical, electronics, textiles, ceramics, glass, transport, mining, agriculture, among others.

NON-METALLIC MINERALS	TECHNICAL DATA	USES
Feldspar.	They belong to a group of minerals forming aluminum silicates combined in three forms: potassium, sodium and calcium. Its color can range from white to colorless and completely melted at 1,530 ° C.	As aggregate in glass, ceramics, pottery, enamels for ceramics, white ceramics, paints, coatings, plastics, rubber, sealants, adhesives, welding electrodes, abrasives, substitutes.
Plaster	It is a dehydrated calcium sulfate (CaSO4 2H20), presented in tabular crystals peel able sheet, generally colorless	Construction, chemicals, agriculture, chemistry, writing chalks, ceramics, medicine, dental, food, mining works, foundry, water treatment, paper, paintings, ornaments.

Market 2014. General Mining Coordination profile. Directorate General of Mining Development.

According to the characteristics of the mineral, we implemented the wagon called nacelle, specialized to transport goods as ore, coal, metals, scrap, etc. [4].

Table 3. Characteristics of the car used for transporting Fierro, Feldspar and Plaster mineral [4].

Nacelle bulk	
Length	15.8 meters.
Length with couplings	17.4 meters
Height	2.87 meters.
Capacity	90 a 100 tons
Net Weight Without Charge	29.7 Tons

4 Methodology

The logistics involved in the transport of minerals by rail for export can be modeled by CVRP (Capacitated Vehicle Routing Problem) with metaheuristic techniques to analyze a series of routes that indicate which algorithm is suitable to adapt to each route optimizing transportation time and feasibility of it. In the following figures we show the maritime intermodal corridors to compare.

Route 1 b) Route 2

Fig. 2. Routes a1 y b2 of maritime intermodal corridors to seaport.

In the literature it is mentioned that the optimization model CVRP, is used to determine the optimal routes for distribution of the limited load transportation to different geographical locations; It is an optimization problem NP- Hard type related to the Bin Packing Problem (BPP) when assigning loads to vehicles and also a Traveling Salesman Problem (TSP) that suggests the best route to follow [5].

In our case to analyze, it will be a starting point passing through different nodes to reach the seaport where the trainload is exported. For each issue, special restrictions are presented and as a result, it has been developed variants of VRP in order to obtain a better approximation most adaptable solution for each case. To compare the performance and optimization of each route will address the metaheuristic techniques reported in the literature for the vehicle routing problem with limited load.

4.1 Genetic Algorithms (GA)

Genetic algorithms are a class of evolutionary algorithms suitable for search and optimization problems where the main objective is to find a set of parameters that minimize or maximize an adaptive function; these algorithms operate on a population or set of solutions presented as binary strings or chromosomes. The function of the executed algorithm is to cross of individuals with greater ability to improve the population and eliminate individuals of lesser capacity, to find the best chromosome that is the solution of the problem [6, 7]. The evolution of the population is performed with the application of genetic operators: selection, recombination and mutation.

In the literature there are some basic concepts to better understanding of genetic algorithms [6]:

- Chromosome: binary string representing an individual or solution, where each element in the chain is called gene.

- Population: infinite set of chromosomes.
- Fitness: criterion that assesses the quality of a chromosome. A higher fitness, better solution and more likely to survive and transmit their characteristics to their offspring.
- Crossing: operation by which new offspring are produced from two parent chromosomes randomly selected.
- Mutation: operation where selected at random and change one or more genes on chromosome; it occurs with low probability.

An example taken from works of Sait and Youssef (1999), a genetic algorithm is defined as follows [6]:

Be $P = \{x^1, x^2, ...\} \subseteq \Omega$ the population.

While not stopping criterion is met:

Select as parents $\{x^i, x^j, ...\} \in P$ with probability proportional to $a(x^i)$ y $a(x^j)$

Parents crossing $\{x^i, x^j\}$ to generate the set O of k children $O = \{x_1^{i,j}, x_2^{i,j}, ..., x_k^{i,j}\}$

For each child generated:

Generate mutation probability P_m

Do that this child enters to P with likely fitness.

If the child enters to P, do while

Randomize, with probability inversely proportional to the fitness, an element that leave P, to be replaced by the new element.

End

End

End

Differences can occur between genetic algorithms when are adapted to a problem of optimization that tries to attack, because the parameters that are handled are not always the same, as well as there are basic fundamentals that genetic algorithms can be presented as representation and adaptive function, Selection Methods where the proportional selection comes, the selection algorithm by Ranking, tournament selection method; recombination operator and mutation.

Genetic Algorithm for TSP

These algorithms applied to the TSP, in execution represent a solution as a sequence of nodes through permutations (1,…, *n*) while operators crossover and mutation may generate *n*-tuples that are not permutations. This is why it is important to define the crossover and mutation operators during encoding.

4.2 Ant Colony Optimization (ACO)

The metaheuristics algorithms of Ant Colony Optimization (ACO) were inspired by the behavior of real ant colonies by their structural organization and behavior that allows them to have good communication. Ants on their way from the starting point to the food supply, according to studies, pheromones are deposited on the ground leaving a trail to the rest of the colony; when more is the concentration of ants on a route, the greater the likelihood that another ant follow. This type of algorithms where a set of actors who take the role of "ants", work together as a colony and aims to indicate the best route to follow through pheromones.

The ant colony optimization gives various solutions by iterative procedures where each artificial colony casts a solution at each iteration, respectively until an end condition. The first such algorithm was applied to the traveling salesman problem (TSP) to obtain optimal results [8, 9].

Travelling Salesman Problem (TSP) applied to Ant Colony Optimization (ACO)

Ants provide solutions probabilistically with pheromones that are track; for the case of TSP the following formulas represent one rule to follow:

$$p_{ij}^{k}(t) = \frac{\left[\tau_{ij}(t)\right]^{\alpha}.\left[\eta_{ij}\right]^{\beta}}{\sum_{l \in N_i^k}\left[\tau_{il}(t)\right]^{\alpha}.\left[\eta_{lj}\right]^{\beta}} \quad \text{With} \quad j \in N_i^k, \tag{1}$$

where $p_{ij}^{k}(t)$ is the probability that an iteration *t* algorithm, the ant *k*, currently located in the city *i*, choose the city *j* as next stop. N_i^k it is the set of cities not yet visited by the ant k. $\tau_{ij}(t)$ is the cumulative amount of pheromone on the arc (*i*,*j*) of the network in iteration t. η_{ij} is the heuristic information for which in the case of TSP, the inverse of the distance between the cities is used i and j. α and β are two parameters of the algorithm that can be edited [9].

4.3 Design of experiment

The optimization algorithms based on Ant Colony are intended to represent various agents using random methods, while sharing information routes that are traveled.

In genetic algorithms it is simulated genetic evolution where in each generation are selected the fittest and are crossed to generate better-adapted descendants; considering that mutations are also generated in the process.

The following tests will be done with the corresponding algorithms to define which algorithm is best suited to the route optimization solution.

Testing by Genetic Algorithm

With a genetic algorithm executed by a Java applet , testing for route optimization are performed in the first test was conducted between 10 nodes that simulate the cities featured in the journey by train from Puebla to the port of Altamira. In the second test simulate 7 nodes representing the number of stations that are in the path of the state of Puebla to the port of Altamira.

Test 1 by Genetic Algorithm for TSP:

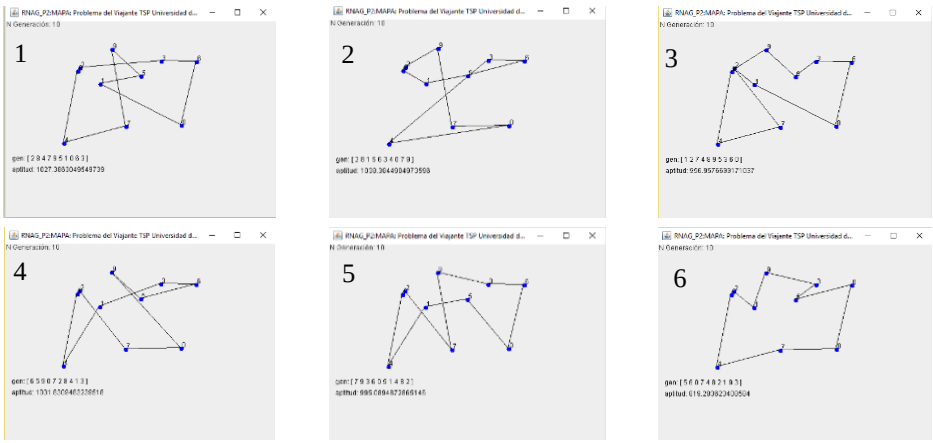


Fig. 3. Simulation screens of train traveling with 10 nodes using the Genetic Algorithm for TSP with a subprogram in java [10].

In the test parameters 1, the simulation of ten nodes are with mutation probability of 10% , the probability of crossing 60% and a maximum number generations of 10. Six tests which showed different routes to reach the optimum were made.

Test 2 by Genetic Algorithm for TSP:

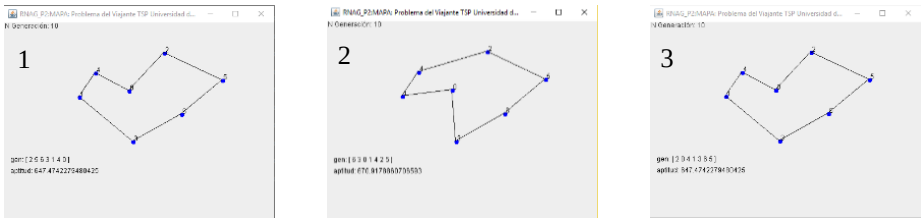


Fig. 4. Simulation screens of train traveling with 7 nodes using Genetic Algorithm for TSP with a subprogram in java

In this second test, only the number of nodes were modified to visit , just as with a population size of eleven , mutation probability 10 %, probability of crossing with 60% and the maximum number of generations of 10.

Testing by Ant Colony Optimization Algorithm

With an ACO algorithm for TSP executed on a Java applet, two tests performed, where the first presents one route with 10 nodes representing cities that appear in the journey by train from Puebla to the port of Altamira. In the second test are 7 nodes representing the number of stations that exist in the path of the state of Puebla to the port of Altamira, they were simulated.

Test 1 by Ant Colony Optimization algorithm for TSP:

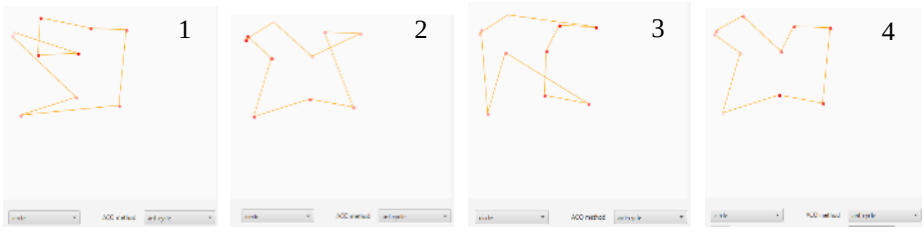


Fig. 5. Simulation screens train travel with 10 nodes using the algorithm Ant Colony Optimization for TSP with Java applet [11].

Tests conducted with 10 nodes representing cities that are in the path of the train from Puebla to the port of Altamira. Pheromone initial value on each line of 2.00 and ants value of 2.00 to select the next node.

Test 2 by Ant Colony Optimization algorithm for TSP:

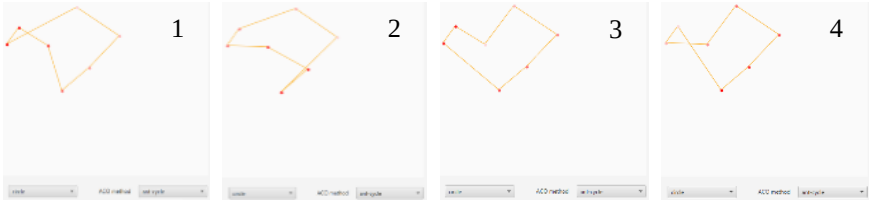


Fig. 6. Simulation screens, train travel with 7 nodes using the algorithm Ant Colony Optimization for TSP with a subprogram in java.

Tests conducted with 7 nodes representing the stations that are in the path of the train from Puebla to the port of Altamira. Pheromone initial value on each line of 2.00 and a value of 2.00 for ants to select the next node.

5 Results

To compare the efficiency of two algorithms, two tests were conducted, the first was performed with a map of 10 nodes of which the location is the same for the different executions. In the second test it was performed with a map with 7 nodes, where they remained in the same location for each run.

Table 4. Results shown in the execution of the AG and ACO algorithms for Travelling Salesman Problem, Own source made with data obtained with the implementation of algorithms in java applets [10, 11].

Algorithm	E1	E2	E3	E4	E5	E6
	Tests with 10 nodes					
AG	Route 1	Route 2	Route 3	Route 4	Route 5	Route optimal
ACO	Route 1	Route 2	Route 3	Route optimal		
Tests with 7 nodes						
AG	Route optimal	Route optimal	Route optimal			
ACO	Route 1	Route 2	Route 3	Route optimal		

In the first test between 10 nodes with Genetic Algorithm for TSP, it was observed that five executions were carried out before finding the optimal route and with the second test between seven nodes, genetic algorithm was more effective to provide the optimal route in the first test.

In the first test with 10 nodes using the ant colony algorithm, it was observed that three executions were carried out before finding the optimal route proving to be more efficient

than GA algorithm. With the second test between seven nodes, 3 tests were also performed before displaying the optimum route.

6 Analysis and Discussion

In the algorithm Ant Colony Optimization by Ant cycle model; the ants represent the train carrying minerals that are dispersed among the nodes representing the cities that exist between the state of Puebla and Altamira. With the application of an algorithm of this kind, it is privileged by the local update of pheromones used to detect the busiest and most convenient routes for transport.

The genetic algorithm, its optimization technique is based on the evolution to random selection of the most optimal routes to carry out the transport, by creating new generations where again the most suitable route is chosen to create a new generation with the nearest solutions; in this way it is like in every generation are left the most optimal routes.

7 Conclusions and Future Research

Each optimization problem contains certain specific parameters; to implement an optimization algorithm for some problems is necessary to identify the components that carry the same, for choosing which metaheuristic technique is closer to solution. In conclusion, an algorithm of Ant Colony Optimization it is more adaptable for providing solutions to Travelling Salesman Problem, because this approach shows is good to find the routes closest to go, in less runtime. For the genetic algorithm, when exist fewer nodes on the route, yields better results; however, as more nodes or cities are added and increases the complexity, the ACO algorithms have better runtime.

This paper presents two TSP algorithms, the optimization algorithms Ant Colony and Genetic Algorithms, in order to analyze which is best algorithm that can adapt to transportation of minerals by railways destined to two important seaports in Mexico to export. In future work they could be analyzed the loading and unloading time with the algorithms discussed in this paper, as well as analyzes the nodes where it is convenient to exchange load at a station that is on the train route.

References

1. The importance of mining in Mexico. Available at : http://www.industriaminera mexicana.com/2013/02
2. Mexico, D. F., Salazar, D. Q.: National Institute of Statistics and Geography.
3. Bravo-Diaz, M., Maritima, I.: How to Export seaborne.
4. Ferromex. Available at: http://www.ferromex.com.mx
5. Orrego-Cardozo, J. P.: Solution to the problem of vehicle routing with limited capacity "PTRC" through heuristic scanning and the implementation of the genetic algorithm Chu-Beasley (2013)

6. Gonzalez-Vargas, G., González-Aristizabal, F.: Metaheuristics applied to vehicle routing. A case study: Part 1: formulation of the problem. Engineering and Research, 26(3), pp. 149–156 (2006)
7. Valencia, E.: Optimization through genetic algorithms. In: Annals of the Institute of Engineers of Chile, 109(2), pp. 83–92 (1997)
8. Aparicio-Guirao, D.: Application of ant algorithms for solving a problem balancing robotic assembly lines (2012)
9. Boats, L., Rodriguez, V., Alvarez, M. J., Robusté, F.: Algorithm based on ant colony optimization by solving the problem of freight from multiple sources to multiple destinations. Santander, Spain (2002)
10. Arceti, A.: Traveling Salesman Problem (Genetic Algorithm). Available at: https://soporte900.wordpress.com/2011/11/13/proproblema-del-viajante-tsp/
11. Birger, M.: Ant Colony Optimization Demo. Available at : https://github.com/kusha/ant-colony

ISSN 1870-4069

Decreased Business Uncertainty by Using Bayesian Networks for the Paradigm Shift in Business Simulator

Alberto Ochoa[1], Miguel Ruiz-Jaimes[2], Sandra Leon[2], Yadira Toledo[2], Iván Ramírez[2]

[1] Juarez City University, Ciudad Juárez, Mexico

[2] Universidad Politécnica del Estado de Morelos, Morelos, Mexico

Abstract. Most small businesses related with Textile Industry or clothes in Latin America fail in his first two years this due to lack of follow-appropriate decisions to bring out various problems, which is why the business simulators help lessen the burden of analyzing the right decision-making in future scenarios and have a high complexity due to exogenous events. To implement a business Simulator for SMEs and in particular in the marketing segment, it is necessary to know the main variables that affect the decisions that should be taken as necessities in any business, and especially when it's an SME that is about to begin its operations. Strategies as the use of Bayesian networks are associated with the behaviors are linked with the variables and scenarios that can be a presenter during the life of the business.

Keywords: Business simulator, Exogenous variables, Marketing to detail, Decision based on strategies.

1 Introduction

On the model of the marketing, variables shall be taken into consideration more important in making decisions, and which are also linked to the different modules that must have the Simulator business for the University's incubator. Business Simulator encompasses each of the disciplines of management, however, the investigation of the document in specific fits in the field of marketing, which is the focus of the module to develop. About the module of marketing should be considered and specified that it is of utmost importance since the marketing consists of a set of principles and practices that are carried out with the aim of increasing trade, and especially the demand. The concept also makes reference to the study of the procedures and resources that pursue the same aim. The marketing involves the analysis of commercial enterprises management. Its intention is to withhold and retain customers' actuals having an organization, while trying to add new buyers.

2 Methodology Used: Bayesian Networks

Bayesian networks may be constructed by applying a learning algorithm to the data set that allows learning a structure. The greatest potential for them resides in the ability to combine expert knowledge with empirical data.

Other benefits of these networks include:
- Ability to reflect causal relationships,
- Running the model with incomplete data and assumptions,
- Forward and backward reasoning,
- Modelling uncertainty,
- Combination of quantitative and qualitative data.

We could define a Bayesian network as a set of variables coded probabilistically, a graphical structure that connects these variables in terms of relationships of conditional independence and a set of distributions susceptible of conditional probability that can be modified based on evidence by means of Theorem of Bayes' theorem. One of the most important advantages of Bayesian networks is that they can represent both the quantitative aspect of a problem as its qualitative aspect. The theoretical support of the qualitative dimension in Bayesian networks is provided by the graph theory. Graph theory tries to create graphical models (graphs) that represent the elements of the problem in a holistic sense and was introduced by Euler to solve the Seven Bridges of Königsberg.

Advantages of Bayesian Networks

The most important advantage offered by Bayesian networks over other methods of multivariate analysis that can represent unison qualitative and quantitative dimension of a problem in an intelligible graphical environment. Another important advantage of Bayesian networks is that they can work with lost data efficiently, which in practice is desirable (e.g. highlighted by [2.3.4]). However, a disadvantage of missing cases is that if a state of the variable appears the probability for this is zero. Also, reduce the envelope setting data [6] and as Bayesian statistical technique, combining prior knowledge about the problem that we study with experimental data [11]. Bayesian networks also allow you to discover the underlying causal structure in a dataset (as is analyzed in: [4, 7, 9, 11, 16]). This means we can build a probabilistic graphical model from a database containing a set of observations on a set of variables. Bayesian networks have advantages over classical rule-based expert systems are to be used when, for example, for decision-making. First, Bayesian networks represent all information in a single format (probabilistic and graphic) making simple interpretations, allow recant conclusions obtained previously and are no longer reasonable in light of new evidence, they give us a general problem vision, generate a set of ordered alternatives and facilitates the explanation of findings [8]. On the other hand, when we build a Bayesian from expert knowledge to use in counseling in a decision-making network, assigning probabilities (or generation) is simpler [7, 12]. In addition, Bayesian networks allow working with concepts of decision theory as value or expected value against decision

problems [7, 8, 13]. In the context of inference, Bayesian networks allow bidirectional inferences; that is, from effects to causes and from causes to effects. And what is more interesting, they allow performing abductive inferences; that is, find the best explanation for a set of data [9, 11]. However, the most important advantage of a Bayesian network in the field of statistical inference is derived from its ability to perform local computations [14, 19]. This property allows updating probabilities that are carried out efficiently without having to calculate all possible combinations between all levels of the variables (which would mean an exponential increase in the calculations as the number of variables in the model or number of states by variable) when incorporated determined knowledge to the graphic structure. Thanks to this property derived from the principles of dependence and conditional independence, Bayesian networks are tools particularly suited for tasks that require rapid and continuous updating control processes, e.g. industrial. Another important advantage of Bayesian networks is that they can work with lost data efficiently, which in practice is desirable (e.g. highlighted by [9, 11, 15, 17]). However, a disadvantage of missing cases is that if a state of the variable appears the probability for this is zero. Also, reduce the envelope setting data [9] and as Bayesian statistical technique, combine the foreknowledge that we respect to the problem of study with experimental data [16]. In [18], Bayesian networks are statistical oriented tools for causal modeling. A Bayesian network is defined by two main components: a qualitative based on graph theory and other quantitative based on probability theory. These tools have been used for prediction, classification and diagnosis through the use of Bayes' theorem as a heuristic updater belief propagation algorithms and probability. However, although the survey methodology has been used to illustrate the automatic construction of Bayesian networks, there are not many jobs of this kind. We show the process of building Bayesian networks from data from a survey of entrepreneurial attitudes in a sample of university students. Bayesian networks (also known as probabilistic causal networks, Bayesian expert systems, probabilistic expert systems, causal networks, belief networks and influence diagrams) are statistical tools oriented modeling, qualitative dimension of a Bayesian network is based on the theory graphs. Although the definition of graph may vary according to the authors (e.g. [13, 15, 16, 17, 20]), a graph can be defined as a set of nodes (variables or vertices) and a set of edges (or links). However, a Bayesian network is a special type of graph, or what is known as a directed acyclic graph (or DAG). In a DAG edges or bonds linking the variables they are oriented and are plotted as arrows. The directed links indicate a Bayesian network causal influence or dependence between variables. Thus, the graph A → B indicates direct influence on B; or what is the same, that B depends on A. In causal terms, we would say that "A causes B" or "B is the effect of A". In a Bayesian network encodes the graphical relationships of dependence and conditional independence between variables that are used to expedite the update probability and statistical inference. In the quantitative dimension, a Bayesian network consists of three major elements: the concept of probability as a degree of subjective belief on the occurrence of an event, Bayes' theorem as a heuristic updater beliefs and a set of conditional probability functions. The probability can be understood, at least four ways: classically, as relative frequency when trials tend to infinity, axiomatically or subjectively. The first three are compatible and have been encompassed under the name

of classical perspective of probability or frequentist and often counter the fourth has been called Bayes (i.e. [2, 7, 9, 11, 13, 18]). Bayesian networks use Bayes' theorem to update the model of probabilities. For example, suppose we have a model that represents several variables and each has an associated probability of occurrence in each of their states. If we know the state of one Bayes theorem it recalculates all probabilities the model to be consistent with this evidence and spreads throughout the network via the links between nodes. Finally, for the Bayes theorem to operate must exist, one for each variable, a set of probability functions encoding relations' probabilistic dependence of each variable and all the variables on which it depends. In particular, there must be a table of conditional probability (TCP) representing the probability of each state variable depending on the variables on which it depends. Computer programs that work with Bayesian networks are designed so they can learn the graphical structure and the implicit stochastic structure in a dataset. That is, build a (qualitative and quantitative) Bayesian network from a database. There are different procedures aimed at achieving this objective (see, for example, [3, 8, 10]), and although many have used survey data to illustrate the automatic construction of Bayesian networks not many networks built from surveys. Finally, indicate that Bayesian networks are a type of statistical tool rarely used in survey research and we believe that this methodology could benefit from its use, either through automatic creation procedures, or through expert judgment [4].

Construction of Bayesian network from data

There are several techniques for building Bayesian networks from a database. In this research, we have used the algorithm K2 [3] and stochastic variable neighborhood search ([16]). We passed briefly describe these algorithms. See Figure 1.

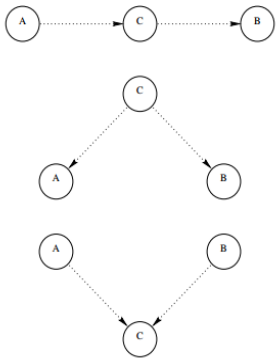


Fig. 1. Graphic characterization of the concept of d-separation

Applications of Bayesian networks to software engineering

Bayesian networks are becoming increasingly popular in engineering, artificial intelligence and statistical time. They have been successfully applied to domains such as medicine, risk

assessment, vision, diagnostics systems and networks, fraud detection, spam, etc. [11]. In software engineering Bayesian networks, have been used in different areas such as:

Estimation of effort and quality. In [4] is provide a critical review of the methods available in the literature for predicting defects, arguing that models based on size and complexity cannot predict efficiently. The authors state that the incorporation of Bayesian networks taking into account such factors as the ability of programmers / analysts, complexity of design and the methods and procedures used generates more accurate predictions. They have shown that is possible to introduce process attributes and product attributes in a Bayesian network. [18]) compared the predictions of Naïve Bayes classifier to estimate the effort of software projects with other techniques of data mining and regression trees and neural networks, showing that such a simple technique as naïve Bayes has the potential to it is used as a possible technique viable estimate

• Testing software. Author of [20] discuss how Bayesian networks can be used as a testing tool, with questions like "what if ..." that helps managers and staff responsible for testing in decision-making.

• Reliability. Fenton et al. have carried out several projects, such as DATUM (Dependability Assessment of safety critical systems through the Unification of Measurable) [9], SERENE (Safety and Risk Evaluation using Bayesian Networks) in which networks have been developed to estimate the reliability and security of systems in various fields [7].

• Graphics and user interaction interfaces. As an example, the Lumiere project [12] of Microsoft investigated ways to improve the interaction between users and software using Bayesian networks. He served as the basis for Microsoft Office wizard when the user used the aid, since the inference engine considers the Bayesian network user actions, application events and user profiles. They have also been used as diagnostic aides for detecting printing problems, etc.

Properties and limitations of Bayesian networks

Bayesian networks have a number of features that make them suitable for software engineering:

• Graphic representation. By their nature, Bayesian networks provide a graphical representation of the explicit dependency relationships domain. Generally, the variables in software engineering such as effort or cost are influenced by many factors. Bayesian networks allow us to model complex systems allowing us to understand the causal relationships graph displaying them through.

• Qualitative and quantitative modeling. Bayesian networks are formed by a qualitative component, the graph, and a quantitative part, the probability tables, which allow using objective criteria (for example, using completed projects) and subjective (e.g. using beliefs domain experts.

• Bidirectional Inference. Bayesian inference networks can do both ways, i.e., the input variables can be used to predict the output variables and vice versa. Setting the output variables with desired values, it is possible to predict which values of the input variables

allow such output. For example, using inference forward, one can predict the final number of defects based on variables such as project size, complexity, design effort, effort in testing, etc. On the contrary, we could fix a number of defects and effort needed to predict which satisfies said output.

• Sensitivity analysis. Given a set of evidence, Bayesian networks allow easily calculate the sensitivity of certain variables, simply by modifying the evidence.

• Uncertainty. Bayesian networks can model degrees of certainty rather than exact values. Therefore, allow modeling uncertainty effectively and explicitly way, so they can make good predictions with incomplete information. In fact, [11] estimates claim that software engineering is a probabilistic assessment of a future event, and is therefore the reason that project managers do not obtain good results.

• Values of confidence. The output of a Bayesian network is a probability distribution instead of unique values. This type of information can be used to measure the confidence we can place in the output of the Bayesian network, which is essential if the model will be used in decision-making. For example, in variable low, medium, and high states, the Bayesian network estimates the probability of each of the states [16].

Applications

Utilization	Research Field
Bayesian networks are important for psychology in two ways: economically and scientifically. Economically Bayesian networks could generate a market of "elicitation of probabilities" oriented development of expert systems where psychology could play a key role. The task for psychology in this regard would guide the experts on how to make appropriate judgments likely trying to overcome the biases that people usually incurred when performing such tasks.	Psychology
A particularly case: Prostate cancer is the most common malignancy among men over 50 years of age and the second leading cause of cancer death (lung cancer is first). The probability of recovery depends on the stage of the cancer (whether it is located exactly in the prostate or has spread to other parts of the body) and the patient's health in general. Therefore, it is important to diagnose at an early stage. However, symptoms of prostate cancer are very similar to those of benign prostatic hyperplasia (BPH) or other prostate problems, so it is easy to confuse. Therefore, it is useful to have a tool that helps the doctor, especially general practitioners, to make a differential diagnosis of possible diseases based on their probabilities. The use of tools to make the design of a Bayesian network is paramount, is why referring to the above introduction, we will talk about some of them.	Medicine

3 Problem Statement

In an economic system is the growth engine of the company, in different sizes, ratings, and economic activity in the preceding paragraphs is justified because in this case study interested SMEs. However, they present some difficulties according to [11, 20]:

- Lack of organization and management.
- Insufficient technology for new challenges.
- Staff turnover due to lack of selection processes.
- Disability credit for lack of collateral.
- Problems capital.
- Business model and inefficient operation.
- Limited to financing at reasonable rates of interest Access.
- Do not reinvest their profits for improving machinery and equipment.

The problem is multifactorial and it depends on is staying within its respective industry, for example, specialized studies of the National Commission for the Protection and Defense of Financial Services Users (CONDUSEF) indicate that only half of existing SMEs in the country accounted survive for eighteen months [20], in this regard the difficulties financial reasons are skewed because 80 percent of SMEs who die for lack of financial management [11].

The effect disappears an SME, it is certainly a failure for the entrepreneur, but the effect is greater for the family that depends on this and goes without a source of income. The National Employment Service in its statistics records mention that for an unemployed person to find work again, time is of one to six months. Of course without considering that this should be permanent and well paid, the latter under the present conditions of our economy can be a Utopia.

According to [13], there are many factors that put their stability and survival at risk, with the absence of a strategic business plan that primarily impacts the growth and permanence of an SME. Therefore, some of the mistakes that commonly make are:

- Lack of institutionalization of their organizational values
- Debts and obligations assumed by shareholders in a personal capacity
- Inadequate segregation of duties
- Technological Backwardness
- Lack of reinvestment of profits in working capital
- Lack of trained personnel
- Lousy service to its customers
- And, therefore, problems with the tax authorities

However, we cannot generalize, because at the time no doubt the particular reality of each company is different and depending on their economic activity to play, so it is of

importance to make a diagnosis to identify the factors of business competition according to the industry you choose. One factor resulting in lack of success of a business is the case of entrepreneurs who start in reverse, i.e., they put the business based on what they know or like to do, without considering whether that product or service will cover a need market, it sometimes ends up winning the dealer, who identified the need to market, rather than the producer [11].

Also, the lack of business, i.e., many start their business because they have no other, and that prevents them gauge the integrity and scope of its business. Adapting to changes in the environment is another key difficulty of SMEs does not have a vision or reading systematic risk. Among the factors leading to delay a determination, especially in SMEs, it emphasizes paying little attention to the behavior of markets and the environment [18]. The ability of the owners to adapt to changes in their environment does not depend on the size of the company, strategically it should go out to his vision and mission to become more competitive and stay in their economic activity.

And according to [8], SMEs can also have certain advantages over others:

• Personalized service: As a small company, worries and occupies more customer satisfaction, usually a worker treats all customer purchase process.

• Improved access to market information: For the size and proximity of your target audience, get first-hand information about political and economic developments that could affect it so you can improve your processes.

• Flexibility tastes and preferences: You can easily adapt to the requirements of their customers or simply patronizing your customers.

• Commitment of employees: Having a single organization, it is easy for workers to contribute new ideas for production or management that facilitate the operation of the company and that these ideas are applied causing workers feel motivated. Unlike large companies is difficult a change in structure.

• Better handling of complaints: As personalized attention, you can try a worker to resolve the claim or complaint from a customer, ensuring the satisfaction of this.

• Improved ability to correct errors: Because of the size and structure of these businesses, it is easier to correct errors they have.

• Faster and timely decisions: For its simple and little hierarchical structure, the ability to make decisions is more optimal and timely.

• Allow the creation of value and in the future better decision-making.

It is the result of a profound and serene reflection by the investigator after reviewing the relevant literature (theoretical and empirical background) detail and internalized the main concepts and theoretical propositions that allow you to formulate clearly and domain the problem to be solved with the investigation. The growth of the national economy as through System Cyclical indicators in the last 35 years and considering the respective federal governments has the following behavior:

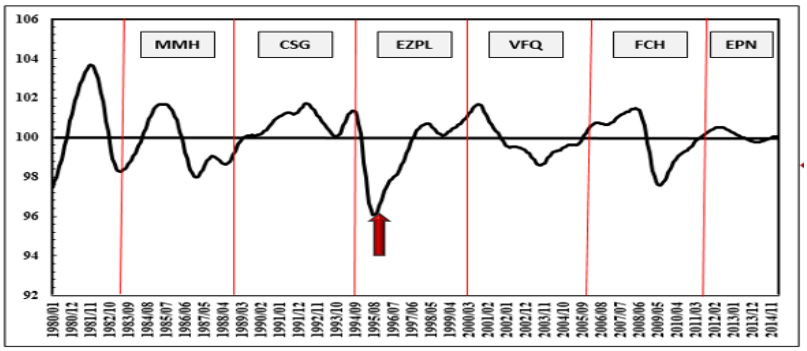


Fig. 2. The growth cycle in Mexico. Cyclical system Indicators: 1980 to 2015. Source: Taken from: The cycle 1980.1-2015.1 national economic growth in the context of the economic policy of neoliberal [11].

Note for Fig. 2:
MMH.- Miguel de la Madrid Hurtado,
CSG.- Carlos Salinas de Gortari,
EZPL.- Ernesto Zedillo Ponce de León,
VFQ.- Vicente Fox Quesada,
FCH.- Felipe Calderón Hinojosa,
Enrique Peña Nieto EPN.

Models of the interfaces

In the Simulator's business, as first instance will be specifically in the module of marketing the General information of the company as the name, the commercial rotation, many employees and above all the initial investment, this with the purpose of having a control according to the strategies and options that are available as a tool.

Fig. 3. Main screen simulator contains 4 modules that mimic the behavior of the company, which in turn contains sub-modules where the player will take the decisions of the company

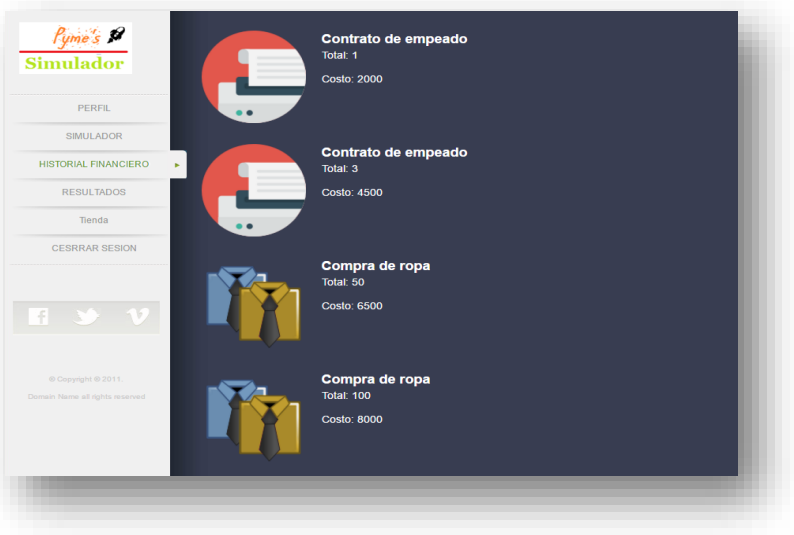


Fig. 4. History of all movements, purchases, contracts, and monitoring the management of the company

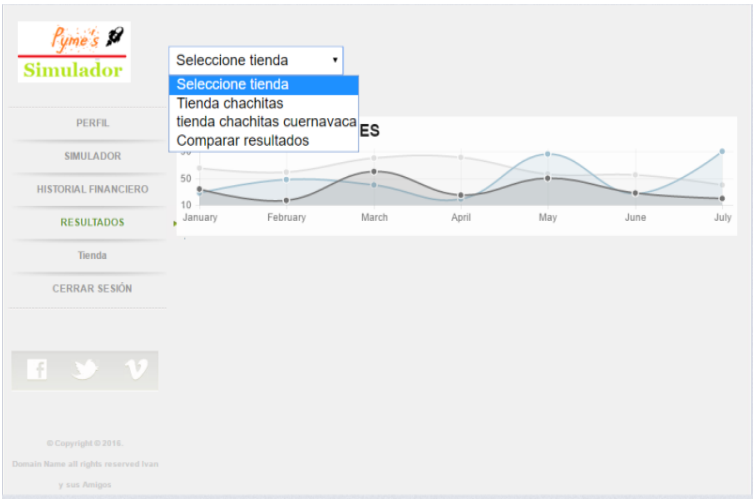


Fig. 5. May compare the results of different companies, either individually analyze the result of decisions by a graph.

The simulators are mostly computer programs that are using a programming language. The algorithm of the program design are models that are starting to define a number of

relevant internal, external variables and the possible interactions between these, which should allow to simulate the operation [3].

Therefore, it is important to have well-defined strategies, so that there are no variables that are not used in the design of the program. Module marketing that is aimed at promoting and advertising the business serve as a simulator guide before starting operations and that this allowed mistakes, while they still cost.

4 Conclusions and Future Research

Module marketing that is aimed at promoting and advertising the business serve as a simulator guide before starting operations and that this allowed mistakes, while they still cost. A business simulator is a learning tool that is intended to play an economic, financial or business system. This is given a virtually, allowing you to participate in the decision-making process and in the process of management of a company. Business Simulator gives guideline to learn without any risk, but is empirical, acquired knowledge of what is and is not for. In [1] is considered that a business simulation is a creation of a business for the purpose of experiment and learn real environment through new experiences, where the participant assumes a role and must make decisions that impact on the environment, the situation and the results of the company as is analyzed in [12]. Therefore, comes to the conclusion that is indispensable in the field of the incubator of the diverse government systems, where in addition to a theorist of the SME study, in a simulator that lasts approximately 8 weeks' equivalent to 4 years (real time), since they are in practice knowledge and aware of the consequences of decisions taken without any risk.

The advantages of simulators are:

• You can learn the result of decisions taken in a prior period. This leads to the competitive advantage of information before acting.

• Allows you to know the effect of implemented marketing strategies.

• Allows you to measure and evaluate the results achieved in the performance together with each of the modules.

• Develops the ability to make decisions, carry out strategies and correct unforeseen.

• Allows you to apply and develop knowledge in the field.

• Facilitates the panorama of the scenarios that may occur in the company.

Precisely in the marketing module you can interact with the strategies of low-cost, since a SME to initiate operations must have a backup of at least 4 months of salaries, expenses, etc. While it is unveiled, and attract customers, however, not so should be put to one side the advertising and promotion. To the study analyzed un [10] state that: The marketing concept is based on the definition proposed by the United States Marketing Association (American Marketing Association), which expresses that it is the process of planning and executing the conception, price, promotion and distribution of ideas, goods and services to create exchanges that satisfy individual and organizational objectives. The key is the

concept of Exchange, in which someone renounces something to receive anything you need or want, which should happen as follows:

a) Have at least two parties involved.
(b) Each party must have something of value to the other.
(c) Each party must be able to communicate with each other and provide goods or services to such party.
(d) Each party must be free to accept or reject the offers of the other.
(e) Each party must believe that it is appropriate or desirable to deal with the other, as is shown in [13].

The job of marketing is essential today and is useful in business simulators.

References

1. Blanque, S., Odriozola, J.: Simuladores: Herramienta de apoyo para el aprendizaje del marketing. Facultad de Ciencias Económicas, Universidad Nacional del Nordeste, Buenos Aires, Argentina, pp. 2–8 (2010)
2. Cowell, R. G.: Parameter learning from incomplete data for Bayesian networks. In: AISTATS (1999)
3. Edwards, W.: Hailfinder. Tools for and experiences with Bayesian normative modelling. American Psychologist, 53, 416–428 (1998)
4. Edwards, W., Fasolo, B.: Decision technology. Annual Review of Psychology, 52, 581–606 (2001)
5. González, E., Cernuzzi, L.: Apoyando el aprendizaje de habilidades empresariales mediante la utilización de un simulador. En: J. Sánchez (Ed.), Nuevas Ideas en Informática Educativa, Volumen V, pp. 8–19, 23–35, Santiago de Chile (2009)
6. Harary, F.: Graph Theory. Addison-Wesley, Reading, MA (1969)
7. Heckerman, D.: A Bayesian approach to learning causal networks. In: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, pp. 285–295, Morgan Kaufmann Publishers Inc. (1995)
8. Howard, J., Sheth, J.: Prespectives in Consumer Behavior. Grenvich, Scott Foresman (1969)
9. Huete, J. F.: Sistemas expertos probabilísticos: modelos gráficos (1998)
10. Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., ... Gerstein, M.: A Bayesian networks approach for predicting protein-protein interactions from genomic data. science, 302(5644), 449–453 (2003)
11. Lamb, C., Hair, J., McDaniel, C.: Principles of marketing. 2ª ed., Publishing Co. USA: South-Western (1994)
12. Maciá, M. A., Barbero, M. I., Pérez-Llantada, M. C., Vila, E.: Psicología y teoría de la decisión: aplicaciones. Madrid: Universidad Nacional de Educación a Distancia (1990)
13. Mussons, J., Comajuncosa, J.: Nuevas técnicas de formación e investigación empresarial mediante el desarrollo de un simulador de negocios complejo en entorno real. Universidad Politécnica de Cataluña, Campus Nord/Manresa, Barcelona, Cataluña, España (2005)
14. Nadkarni, S., Shenoy, P. P.: A causal mapping approach to constructing Bayesian networks. Decision support systems, 38(2), 259–281 (2004)
15. Neapolitan, E. R., Morris, S.: Probabilistic modelling with bayesian networks. The SAGE handbook of quantitative methodology for the social sciences, 371–390 (2004)

16. Palomo, M.: Los procesos de gestión y la problemática de las Pymes. Ingenierías Vol. XXVIII No. 8, pp. 25–31 (2005)
17. Pearl, J.: Bayesian networks, causal inference and knowledge discovery. UCLA Cognitive Systems Laboratory, Technical Report (2001)
18. Puga, J. L., García, J. G., De la Fuente Sánchez, L., De la Fuente Solana, E. I.: Bayesian nets as modelling in psychology. Anales de Psicología/Annals of Psychology, 23(2), 307–316 (2007)
19. Santesmases, M.: Marketing, conceptos y estrategias. 4ª. ed., Pirámide. Madrid, España (1999)
20. Spirtes, P., Glymour, C., Scheines, R., Kauffman, S., Aimale, V., Wimberly, F.: Constructing Bayesian network models of gene expression networks from microarray data (2000)

ISSN 1870-4069

Implementation of a Security Model for Malware Based on Artificial Immune System

Santiago Yip Ortuño[1], José Alberto Hernández Aguilar[1], Carlos Alberto Ochoa Ortiz[2]

[1] Autonomous University of Morelos State, Cuernavaca, Morelos, Mexico

[2] Autonomous University of Ciudad Juárez, Chihuahua, Mexico

alberto.ochoa@uacj.mx
totoy.yip@gmail.com, jose_hernandez@uaem.mx (corresponding author)

Abstract. This research discusses intrusion detection systems based on computer networks and a model for the detection of malware using artificial immune system (AIS). The SIA has three main theories: the clonal selection, negative selection and network theory. This work used the ClonalG algorithm developed by Castro & Timmis (2002) [5] and implemented in Weka 3.6.4 for the intrusions detection in the KDD 1999 database. Preliminary results indicate good results, since was obtained 77.92% accuracy in the classification of threats using CLONALG algorithm, and 92.69% of accuracy by using CLONALG and feature selection of a total of 494,021 processed registers.

Keywords: Artificial Immune System, ClonalG, Intrusion Detection System, Security model.

1 Introduction

1.1 Information Technologies and Security

The growth of information technology has generated a change in the world; currently the use of technology is an obligation to achieve competition, both for companies and organizations of all kinds. However, despite all the benefits that these technologies offer there are many threats, which? Among all those who manipulate ICTs, which eleven materialized cause irreparable damage, ranging from damage to the image of an entity or person, millionaire losses and even loss of freedom or endangering human lives. Some of the best-known cases include Sony Pictures, Home Depot [13], Celebgate, Stuxnet [12] and Ransomware [14].

For these reasons technologies have emerged that help reduce the risks of using these, within these tools are the intrusion detection systems [4]. This research proposes the

evaluation of the intrusion detection system with a bioinspired heuristic known as Artificial Immune System (AIS) [2].

Problem at Hands

Can an artificial immune system (AIS) detect and report intrusions in an institution?

Hypothesis

H. Through an artificial immune system is possible to detect and report intrusions in an institution.

H0. Through an artificial immune system is not possible to detect and report intrusions in an institution.

Structure of the document, in Section 2 we discuss the theoretical framework, in section three the methodology used for this research as well as the description of the algorithm used (CLONALG), the preprocessing of the database, and the results of an experiment using the KDD cup 1999 database [18]. Finally, we present the conclusions, future work and our conclusions.

2 Review of Literature

2.1 Information Technologies and Security

Whitman and Herbert (2011) in his work "Principles of information security" the define information security as: "The Protection of Information and its critical elements including the systems and the hardware that use stored and transmitted the information" [15].

For that reason, the security of the information includes all those mechanisms, controls, devices, best practices, etc., that ensure the 3 basic aspects of the information [8]:

- Availability: The information is available when it is required.
- Integrity: The information should not suffer any type of alteration; the modifications to be made shall be solely done by processes or mechanisms known in the treatment of it.
- Confidentiality: the information will be available to the persons entitled to it.

At the same time, according to [15] an intrusion is defined as:

"The satisfactory accesses to an information system in order to disrupt, modify, remove or damage the information or integrity of the same".

2.2 Intrusion Detection System

Due to the exponential growth of the threats have emerged different types of tools that allow us to their detection and mitigation, among these are the physical and logical tools; within the physical tools there are: Firewalls, content filtering, intrusion prevention system (IPS), among others; within the logic are: Antivirus, Intrusion Detection Systems (IDS), etc. Among of the tools the IDS have had enough popularity to the companies why? An IDS consists of procedures that react to detect patterns of intrusion, this includes all those actions taken by an organization when an intrusion is detected [15] due is difficult to know what an attacker will do as explained in [24]. For this reason, the security of the information includes all those mechanisms, controls, devices, good Practices, etc. that ensure the 3 basic aspects of the information [8] which as mentioned in [25] represents a challenge for cyber law in Mexico and abroad.

2.3 Artificial Immune System

The biological immune system is a collection of molecules with highly evolved procedures that allow the identification and elimination of any substance foreign to the body that protects [2, 3], the artificial immune system is a simulation of the biological functioning of the immune system to perform specific tasks [5, 6].

The artificial immune system [10-11] has three main theories:

Clonal Selection:

- According to the work reported by [1], the clonal selection theory was proposed by Frank Macfarlane Burnet in 1959 in his work "The clonal selection theory of acquired immunity", the main idea of this theory states that cells are able to recognize antigens will be those that will proliferate.

Negative Selection

- Kim & Bentley (2001) in their work “An Evaluation of Negative Selection in an Artificial Immune System”, quoted that the theory of negative selection was proposed by Stephanie Forrest in the year 1994 in his work "Self - no self-Discrimination in a Computer", where the antibodies generated by the immune system reacts only against the antigens by omitting any action against its own cells [16].

Network Theory

- According to [1] the theory of networks in the artificial immune systems was proposed by Niels K. Jerne in 1974 in his work: "Toward a Network Theory of the Immune System", in establishing a network of antibodies that recognize antigens.

3 Methodology

The proposed methodology for this research is called KDD (knowledge data discovery); is based on the works of [7, 17] we briefly explain it in the following figure.

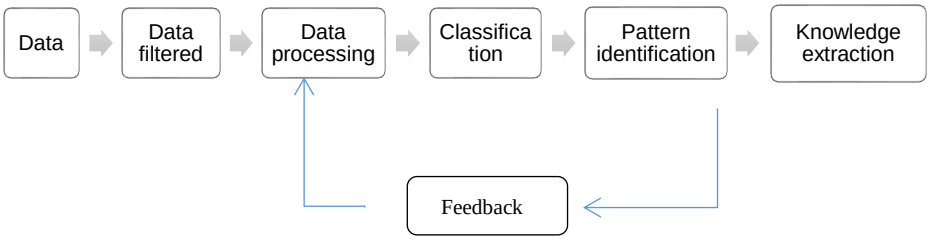


Fig. 1. Cycle generic data mining [7].

1) The test databases to be used in this research are: 1) the available on the Internet for the KDD CUP 1999 tournament [18] (10% of 4 million records).
2) The data will be processed in Weka ClassAlgoss for validation, and presented in tables and graphs.
3) We use SIA algorithms available in the WEKA 3.6.4. [19], the algorithm used in this research is based on the work of [5].
4) According to [1] were run existing SIA algorithms in intrusion detection. For this research is proposed ClonalG:

ClonalG Algorithm [5]

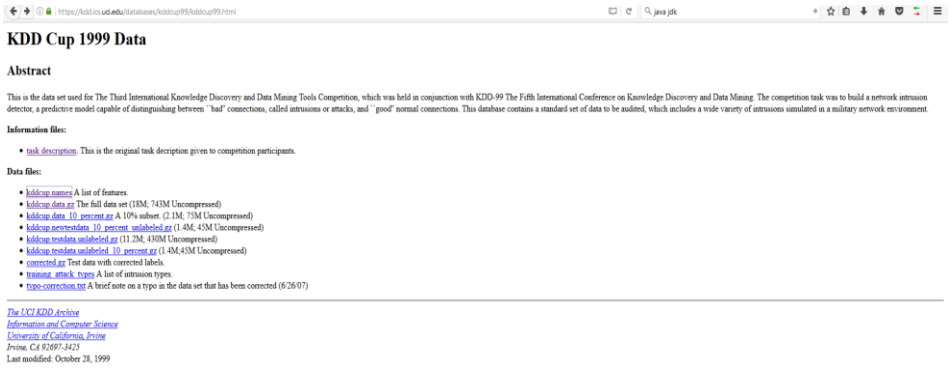
1. Beginning: Initial population created in a random way (P)
2. Antigen presentation: for each pattern, do:

   a. Fitness function: present to the P population and determine its affinity for each element in the P population;
   b. Clonal Selection: select n1 elements with highest fitness function of P and generate clones of these individuals in a proportional way according affinity to antigen: highest the fitness, the greatest number of copies and vice versa;
   c. Maturity of fitness: mutate all these copies with a ratio inversely proportional regarding its fitness function according initial pattern: More fitness, mutation ratio is lesser and vice versa. Add these mutated individuals to P and select the best individual to preserve him as the memory "m" of antigen.
   d. Dynamic goal: replace n2 number of individuals with less fitness function (randomly generated) by the newest;

3. Cycle: Repeat the step 2 until certain criterion is obtained.
5 The classification of the data shall be made using the ClonalG algorithm.
6) The results obtained will be compared among themselves and with other experiments reported in the literature.

Experiment

For the testing of this research was used to the database of the KDD cup 1999, in the competition that year of the KDD is proposed to build a model for the detection with the ability to distinguish between "good" and "bad" connections, the database consists of about 4 million records with 42 attributes each, see Figures 2-3.



KDD Cup 1999 Data

Abstract

This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 The Fifth International Conference on Knowledge Discovery and Data Mining. The competition task was to build a network intrusion detector, a predictive model capable of distinguishing between ``bad'' connections, called intrusions or attacks, and ``good'' normal connections. This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment.

Information files:

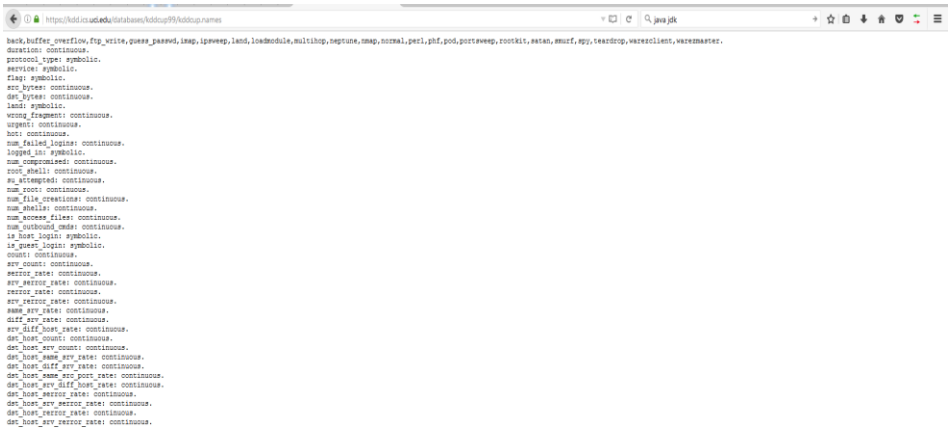
- task description. This is the original task decription given to competition participants.

Data files:

- kddcup.names A list of features.
- kddcup.data.gz The full data set (18M; 743M Uncompressed)
- kddcup.data_10_percent.gz A 10% subset. (2.1M; 75M Uncompressed)
- kddcup.newtestdata_10_percent_unlabeled.gz (1.4M; 45M Uncompressed)
- kddcup.testdata.unlabeled.gz (11.2M; 430M Uncompressed)
- kddcup.testdata.unlabeled_10_percent.gz (1.4M;45M Uncompressed)
- corrected.gz Test data with corrected labels.
- training_attack_types A list of intrusion types.
- typo-correction.txt A brief note on a typo in the data set that has been corrected (6/26/07)

The UCI KDD Archive
Information and Computer Science
University of California, Irvine
Irvine, CA 92697-3425
Last modified: October 28, 1999

Fig. 2. KDD cup data base [18].



```
back,buffer_overflow,ftp_write,guess_passwd,imap,ipsweep,land,loadmodule,multihop,neptune,nmap,normal,perl,phf,pod,portsweep,rootkit,satan,smurf,spy,teardrop,warezclient,warezmaster.
duration: continuous.
protocol_type: symbolic.
service: symbolic.
flag: symbolic.
src_bytes: continuous.
dst_bytes: continuous.
land: symbolic.
wrong_fragment: continuous.
urgent: continuous.
hot: continuous.
num_failed_logins: continuous.
logged_in: symbolic.
num_compromised: continuous.
root_shell: continuous.
su_attempted: continuous.
num_root: continuous.
num_file_creations: continuous.
num_shells: continuous.
num_access_files: continuous.
num_outbound_cmds: continuous.
is_host_login: symbolic.
is_guest_login: symbolic.
count: continuous.
srv_count: continuous.
serror_rate: continuous.
srv_serror_rate: continuous.
rerror_rate: continuous.
srv_rerror_rate: continuous.
same_srv_rate: continuous.
diff_srv_rate: continuous.
srv_diff_host_rate: continuous.
dst_host_count: continuous.
dst_host_srv_count: continuous.
dst_host_same_srv_rate: continuous.
dst_host_diff_srv_rate: continuous.
dst_host_same_src_port_rate: continuous.
dst_host_srv_diff_host_rate: continuous.
dst_host_serror_rate: continuous.
dst_host_srv_serror_rate: continuous.
dst_host_rerror_rate: continuous.
dst_host_srv_rerror_rate: continuous.
```

Fig. 3. Available attributes in KDD cup data base.

Figure 2 shows the web page where was downloaded the database, in figure 3 are displayed the names of the 42 attributes of the database.

Subsequently, the database was imported through the Excel tool, as shown in figure 4.

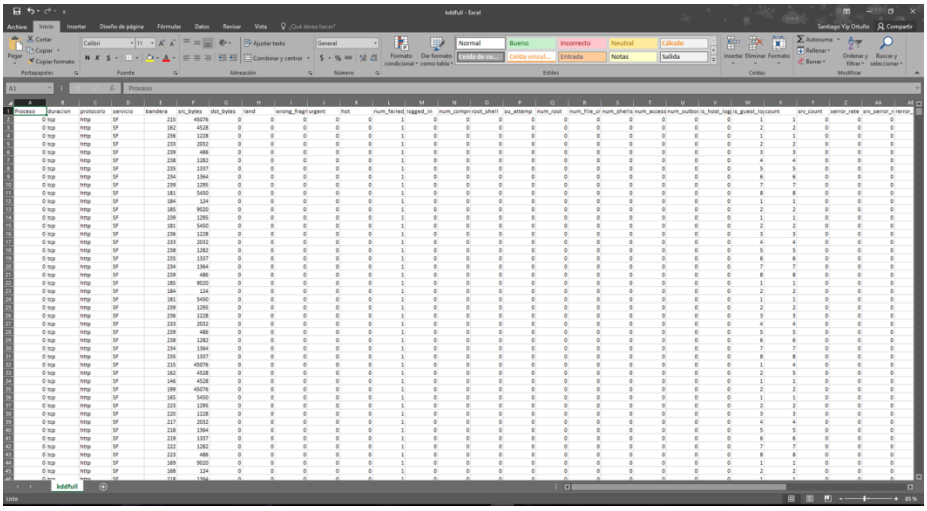


Fig. 4. Cleaning of the information in Excel.

Once preprocessed the BD, we proceeded to the analysis in Weka 3.6.4 running in Linux platform, initially we started the ranking of all the 42 attributes achieving an accuracy of 77%, with an execution time of 91.8 seconds per model see figures 5-6.

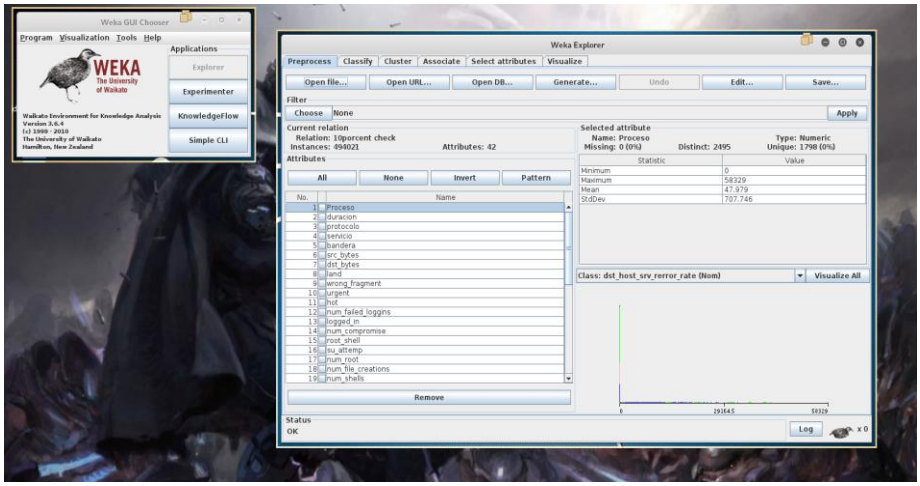


Fig. 5. Load data in Weka.

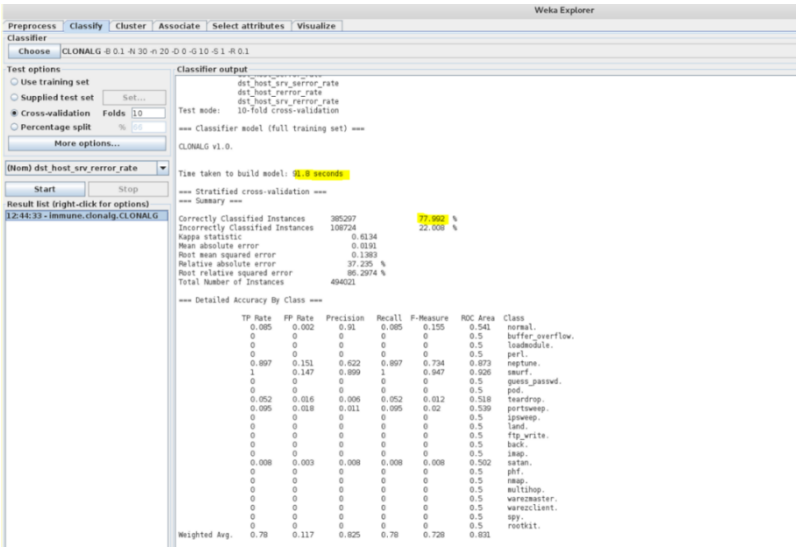


Fig. 6. Classification using CLONAL G, cross-validation 10 Folds.

In order to optimize the execution time and improve the accuracy, a selection feature process was applied, the filter that reached a better accuracy was the GeneticSearch CfsSubsetEval, which allows discarding 34 attributes and improving the accuracy of 77% to 92%, with a running time of 34.18 seconds per model, see Figures 7-8.

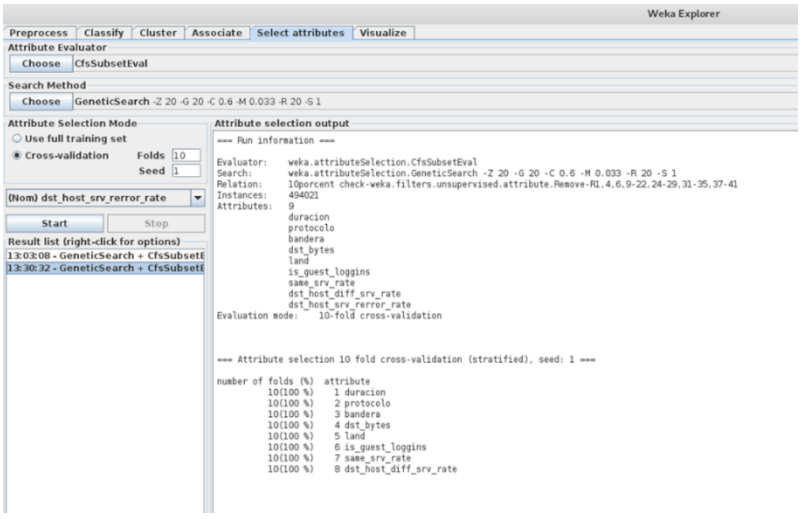


Fig. 7. Feature selection process.

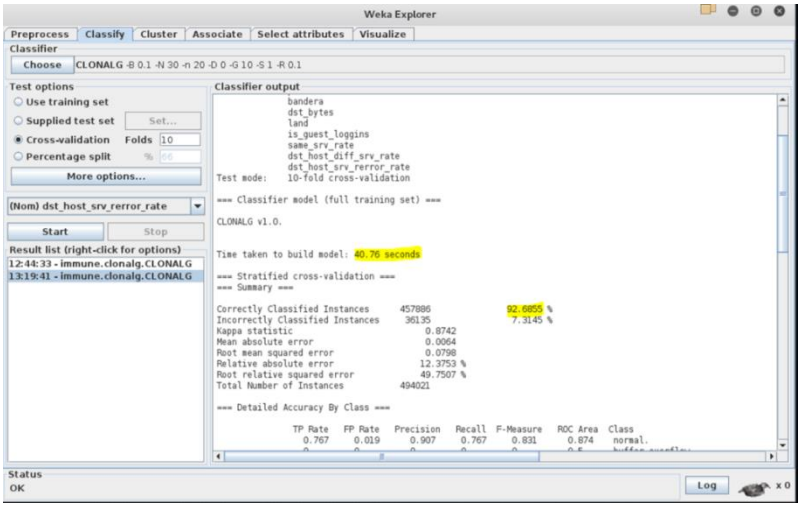


Fig. 8. Data Base Clonal G KDD cup with feature selection.

4 Results and Discussion

In this section are summarized - in a table- a comparative analysis of the experiments carried out in this research, in columns 3 and 4 are shown precision and time for classification process by using CLONALG on a full set of database fields; in columns 4 and 5 Precision and time by using CLONALG and Feature selection process is presented. Columns 6 and 7 show improvements in precision and precision in time.

Table 1. Comparative Analysis for the two models of classification generated.

Database 494021 Records	Precision Clonal G	Time CG (s)	Precision CG + FS	Time CG+FS (s)	Improvement Precision	Improvement Time (s)
KDD cup 1999	77.92%	91.8	92.69%	34.18 Add to	14.77%	51.04

As table 1 show by using feature selection, processing time decrease and precision increase substantially. The use of artificial immune system through the use of the ClonalG algorithm for classification of traffic in a network yielded very good results 92.69% of accuracy. Yan and Yu (2006) in his work "AINIDS: an immune-based network intrusion detection system" obtained 88 per cent of accuracy in classification [20]. Xiaojie Jinquan Zeng et al. (2009) in their work "A self-adaptive negative selection algorithm used for anomaly detection" obtained 88% of correct classification [21]. Itzhak Levin (2000) in the

KDD-99 Classifier Learning Contest [18] obtained 92% of classification [22]. Our model provides a better performance than those related works.

5 Conclusions and Future Work

Artificial Immune System is a promising technology to identify malware and intrusions in computer networks, therefore our hypothesis is true. Artificial Immune System is very important due its adaptive nature regarding other available technologies.

Preliminary results show AIS provides an acceptable technology to identify intrusions regarding similar technologies. Feature selection reduces substantially the time to carry out classification process and increase precision.

The results presented here are not in any way final, they represent only the first steps for the generation of a robust model for intrusion detection system. We want to increase the capabilities of this approach to detect intrusion in smart grids. We would like to compare different SIA algorithms available in literature and with respect traditional algorithms like J48 or a priori. As future work we plan to apply the methodology described above with actual instances of an institute that experience IDS and include comparisons with other well-known techniques like Bayesian Networks [23]. We would like to implement a prototype system in R Language as described in [9].

References

1. Al-Enezi, J. R., Abbod, M. F., lsharhan, S.: Artificial Immune Systems - Models, Algorithms and Applications. Academic Research Publishing Agency (2010)
2. Bachmayer, S.: Artificial Immune Systems. Department of Computer Science, University of Helsinki (2008)
3. Dasgupta, D., Ji, Z., González, F.: Artificial immune system (AIS) research in the last five years. IEEE Congress on Evolutionary Computation, 1, pp. 123–130 (2003)
4. Dario-Duke, N., Chavarro-Porras, J. C., Moreno-Laverde, R.: Smart security. Scientia Et Technica , 1(35) (2007)
5. Castro, L. N., Timmis, J.: Artificial Immune Systems: A New Computational Intelligence Approach. Springer Science & Business Media (2002)
6. Farmer, J. D., Packard, N. H., Perelson, A. S.: The immune system, Adaptation, and Machine Learning. Elsevier Science Publishers B.V., pp. 197–204 (1986)
7. Dave, J., Jian, P., Micheline, K.: Data mining: concepts and techniques. Elsevier (2011)
8. ISO: The portal of ISO 27001 in Spanish. Obtained from what is a http://www.iso27000. es/sgsi.html ISMS? (2012)
9. Torgo, L., Torgo, L.: Data mining with R: learning with case studies. Boca Raton, FL:: Chapman & Hall/CRC (2011)
10. Zum-Herrenhaus, M., Schommer, C.: Security Analysis in Internet Traffic Through Artificial Immune Systems. INTERREG IIIC/e-Bird, Workshop Trustworthy Software, pp. 1–9 (2006)
11. Zum-Herrenhaus, M., Schommer, C.: Healthy-security analysis in internet traffic through artificial immune systems. ArXiv preprint arXiv:0805.0909 (2008)

12. Symantec: El gusano Stuxnet. Obtained from: El gusano Stuxnet, Available at: http://www.symantec.com/es/mx/page.jsp?id=stuxnet (2010)
13. Neal, D.: Home Depot confirms 53 million email addresses stolen in recent hack. Available at: http://www.v3.co.uk/v3-uk/news/2380100/home-depot-confirms-53-million-email-addresses-stolen-in-recent-hack (2014)
14. Kaspersky, K. S. N.: Report: Ransomware in 2014-2016 (2016)
15. Whitman, M. E., Herbert, M. J.: Principles of information security. Cengage Learning (2011)
16. Kim, J., Bentley, P. J.: An Evaluation of Negative Selection in an Artificial Immune System. In: Proceedings of GECCO, pp. 1330–1337 (2001)
17. Hernández-Aguilar, J. A., Burlak, G., Lara, B.: Diseño e Implementación de un Sistema de Evaluación Remota con Seguridad Avanzada para Universidades Utilizando Minería de Datos. Computación y Sistemas, 13(4), pp. 463–473 (2010)
18. Cup, K. D. D.: Dataset: Available at: http://kdd. ics. uci.edu/databases/kddcup99/kddcup99. html 72 (1999)
19. Weka. Available at: https://sourceforge.net/projects/weka/files/weka-3-6/3.6.4/ (2017)
20. Yan, Q., Yu, J.: AINIDS: an immune-based network intrusion detection system. In: Defense and Security Symposium, International Society for Optics and Photonics, pp. 62410U–62410U (2006)
21. Jinquan, Z., Xiaojie, L., Tao, L., Caiming, L., Lingxi, P., Feixian, S.: A self-adaptive negative selection algorithm used for anomaly detection. Progress in natural Science, 19(2), pp. 261–266 (2009)
22. Rojas-Gonzalez, I., García-Gallardo, J.: Bayesian Network Application on Information Security. I. P. Nacional, Ed., Research in Computing Science, 51, pp. 87–98 (2010)
23. Pimentel, J. C. L., Monroy, R.: Formal support to security protocol development: a survey. Computación y Sistemas, 12(1), pp. 89–108 (2008)
24. Argüelles-Arellano, M. D. C.: Challenges of Cyber Law in Mexico. Computación y Sistemas, 20(4), pp. 827–831 (2016)

ISSN 1870-4069

Design of a Technology Management Infrastructure for Large Volumes of Data in an Intelligent Power Network

Guillermo Escobedo Briones, Norma E. Jácome Grajales, Gustavo Arroyo-Figueroa

National Institute of Electricity and Clean Energies, Cuernavaca, Morelos, Mexico

{gescobedo, nejacome, garroyo}@iie.org.mx

Abstract. In Mexico, high - priority energy generation target, associated with clean energy sources, have been established through the promotion of other abundant natural resources such as solar or wind energy. In addition to this, there is a global trend supported by the development of renewable energies that is leading the electrical systems to integrate distributed generation capacities. This new operation perspective of the electrical network is called: "Smart Grid". The objective of this research is to present a technological infrastructure for the management of large volumes of information through Big Data tools to support the integration of renewable energy. The infrastructure includes a methodological architecture for the acquisition, processing, storage, management, analysis, monitoring and forecast of large amounts of data. The development of a Big Data application for the analysis and monitoring of the information generated by photovoltaic systems is included as a case study. The goal is to have timely information to make better decisions to improve the integration of renewable energy in the Smart Grid.

Keywords: Big Data, Smart Grid, Renewable Energy, Distributed Generation, Photovoltaic Systems, Electric Power Utility, Information Systems, Data Analytics.

1 Introduction

As part of efforts to promote energy efficiency and sustainability and reduce dependence on fossil fuels as a primary energy source, Mexico has increased its efforts to promote the use of renewable energy sources and clean technologies to generate electricity. However, this effort is just beginning. It has a legal goal to increase the percentage of non-fossil energy in the portfolio of primary energy sources for electricity generation by at least 35% by 2024. There is also a global trend, supported by the development of renewable energies, cogeneration in situ and the development of communication infrastructure and information technology, which is leading electrical systems to integrate capabilities of distributed generation in small capacities, but in a lot of points. This is a phenomenon that is changing the way in which electrical networks are designed and operated. This forces to consider a

wide and competitive portfolio of projects, as well as bold steps to achieve this legal goal, which includes removing barriers and promoting renewable energy sources [1].

The growing demand for clean and sustainable electricity and the strong dependence of modern life to have a high quality power supply, reliable and continuous, has forced conventional electric systems to incorporate distributed generation with renewable energy sources. Here are some reasons that gave rise to the concept of Smart Grid (SG). In addition, the high level of incident solar radiation in Mexico is somehow favoring the installation of Photovoltaic Systems interconnected with the power grid (PG) for the production of electricity at the point of consumption. Likewise, given the benefits that the user and the power grid potentially have by the use of this technology, there will be more users of electrical service motivated to invest in it [2].

Moreover, considering that the information technology and communications play an important role in the use, transfer of knowledge and clean energy supply, Big Data is a trend in Mexican companies, in order to obtain value of data and maintain a competitive edge in the business. Big Data will allow to process and analyze large amounts of information in order to generate projects that benefit organizations to better understanding of their environment, innovate and improve their products or processes and define new business models.

Therefore, this work integrates, processes and analyzes information generated by photovoltaic cells sensors with Big Data, concentrating the necessary concepts for the understanding and implementation of projects related to the subject. In the same way, it seeks to create an infrastructure or platform that can be used in SG future projects.

2 Challenges of the Traditional Electrical System

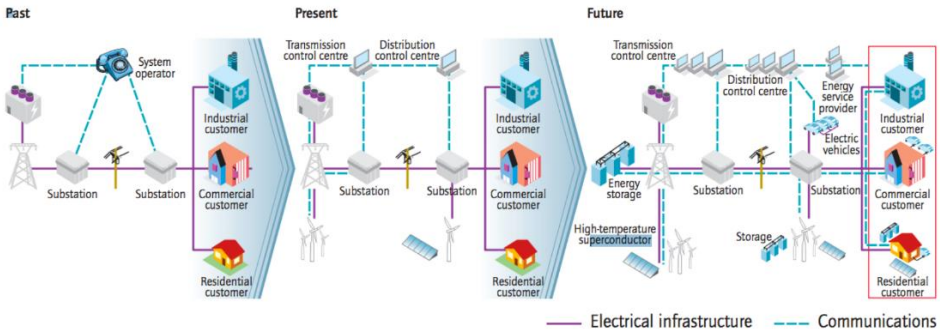


Fig. 1. IEA Smart Grid [4].

The traditional power grid requires changes in terms of restructuring, with the establishment of an electricity market that encourages competition and efforts to ensure energy efficiency in a context of environmental sustainability. These challenges have led

governments, electricity companies and scientific community to find solutions that enable efficient, reliable and responsible energy use, appealing to an optimal and flexible design of the electrical network. This modernization of the electricity network is known as the Smart Grid [3].

Figure 1 depicts a smart grid taken from the Smart Grid Technology Roadmap of the International Energy Agency (IEA) published in 2011. In this figure the evolution experienced by the electricity sector is shown. The vision corresponds to an electrical network in constant development, in real-time, with a two-way flow of energy and information, between the power generator, the network operator and the end consumers.

3 Integration of Distributed Generation Through Renewable Resources in the Smart Grid

The general objectives of SG are reducing the negative impact on the environment, reduce dependence on non-renewable natural resources and increase energy security through the diversification of distributed generation sources with renewable energies. SG technology can significantly reduce the barriers to large-scale integration of renewable resources in the electrical network [5].

The IEA estimates, in their study "New Policies Scenario", that globally, the deployment of renewable resources can save about 4.1 Gigatons of CO2 emissions in 2035 in contrast with the emitted in 2010 [6].

The main challenge lies in the amount, type and location where renewable resources can be interconnected with the transmission and distribution systems and end consumers. This integration task of distributed generation from renewable resources in the SG requires careful and systematic approach.

Distributed generation technologies can be divided into two big groups: those that use fossil fuels and those that use renewable energy resources. In the first group there are internal combustion engines, micro gas turbines, Stirling engines and fuel cells. In the second, it can be found the generation with biomass, wind turbines, tidal power, geothermal and photovoltaic cells [7].

In this work, it is consider the analysis and monitoring of the information generated by photovoltaic cells.

4 Operation of a Photovoltaic System

There are two types of photovoltaic systems, autonomous systems and interconnected with (PG) power supply systems. It is important to note that for purposes of this work only PGs will be considered.

Figure 2 illustrates the operation of a PG. Sunlight is converted into direct current electric power by solar cells. This direct current passes through the different components until the

investor turns it into alternative, which can provide domestic consumption or be sending directly to the mains.

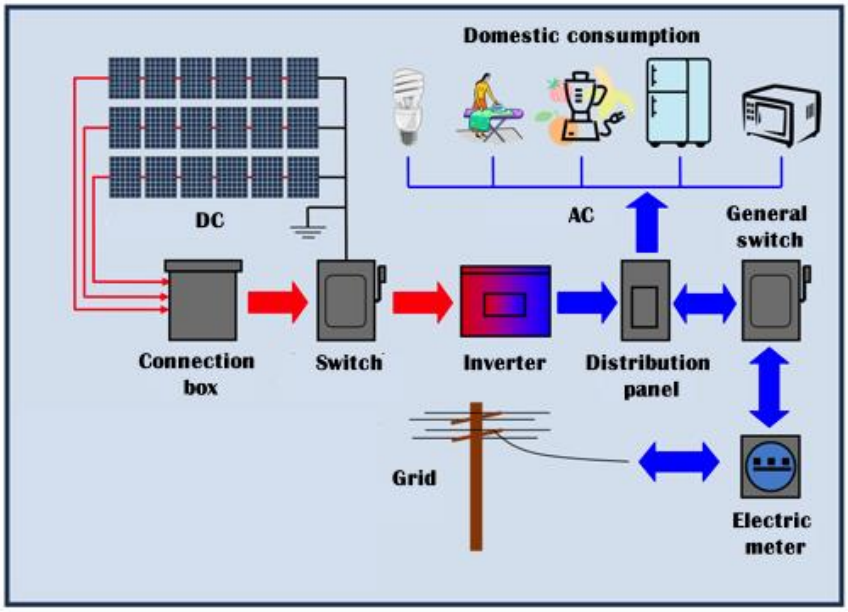


Fig. 2. Operation of a PG [8].

5 Handling Large Volumes of Data from Distributed Generation Sources in the Smart Grid

To face the enormous challenge of handling large volumes of data from distributed generation sources in the SG, there are different technologies that fall under the term Big Data. These technologies allow obtaining of value through the intelligent interpretation of data. However, these strategies to obtain that potential for smart grid in Mexico are still at an early stage of development.

Big Data will enable to process and analyze large amounts of information in order to generate projects that benefit the institutions and organizations to achieve a better understanding of their environment, innovate and improve their products or processes and define new business models.

In our modern and digital society, create digital data at anytime and anywhere has become an unnoticed everyday task. The digitalization era has led to an explosion of data. As an example: it is estimated that since the beginning of mankind until 2003, it has been generated about 5 million gigabytes of data. In 2011, the same amount of data was created

almost every two days. By 2013, the same amount of data was created every 10 minutes [9]. Hence the need to convert data into information to generate knowledge as a platform to make better decisions.

This work integrates, processes and analyzes information generated by photovoltaic systems with Big Data, concentrating the necessary concepts for the understanding and implementation of projects related to the subject, in the same way it seeks to create a technologic platform that can be used in future SG projects.

6 Technologic Platform

6.1 Defining Big Data Architecture

According to Mysore, Khupat, & Jain [10] Big Data architecture consists on different dimensions and logical layers. Thus, the layers provide an approach of components of the organization with specific functions. In Figure. 3 Big Data architecture developed in this paper is presented.

6.2 Analysis of Hardware and Software Infrastructure for Big Data

With the emergence of the Big Data concept, it has been created a variety of tools and products to support the management and exploitation of information. Therefore, the tool selection process becomes a key factor for the viability and success of a Big Data solution. Technologies and tools of Big Data can be classified according to their functions and features:

- Processing. Technologies dedicated to mass data processing.
- Real time data acquisition technologies (streaming).
- Technologies for information storage.
- Data visualization and analytics.
- Tools for statistical development.
- Extraction, transformation and loading (ETL) tools.
- Frameworks, which include an integration of different components for acquisition, processing, analysis and visualization of data.

In Figure 4 an overview map of Big Data tools classified according to their functionality is shown. Open source tools are included [11].

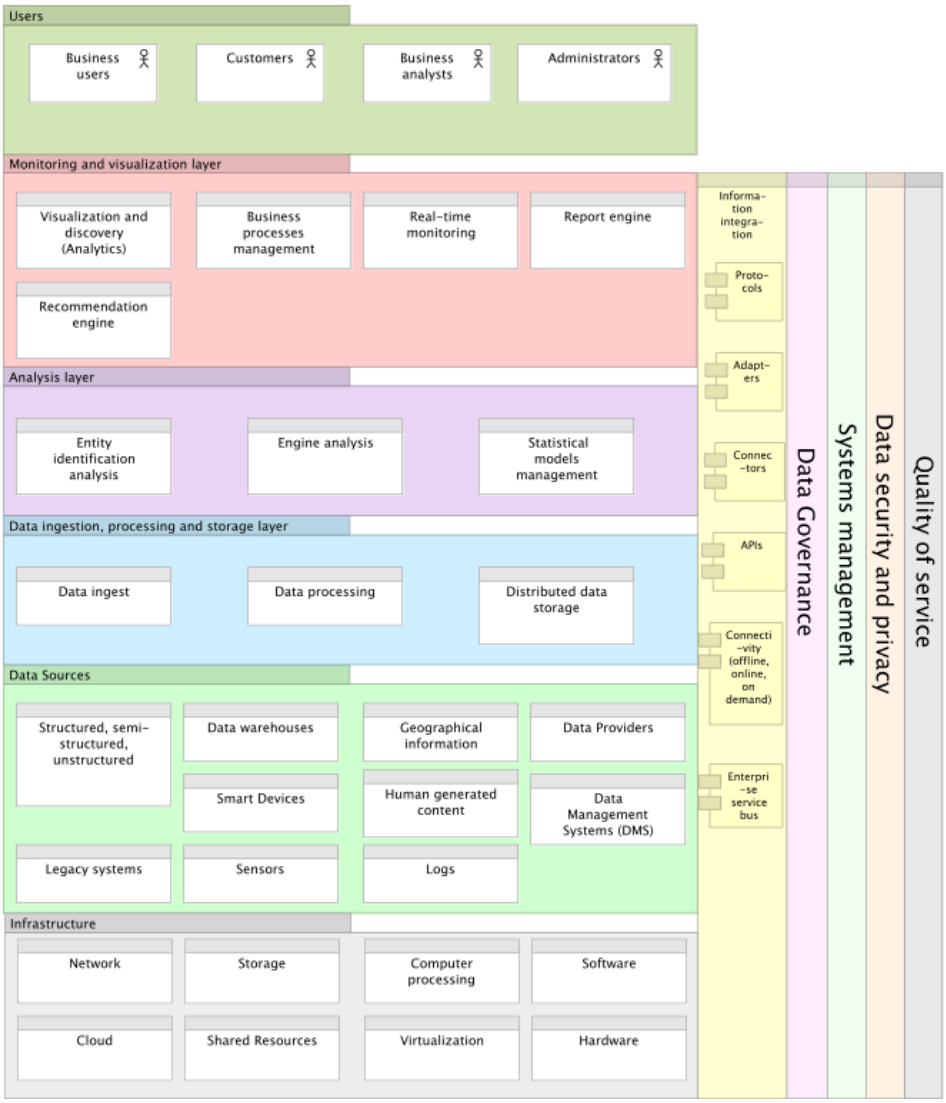


Fig. 3. Components of the logical and vertical layers.

For the selection and evaluation of the required platform. For this work the following criteria was considered:

- Open source platforms.
- Free Software Licenses.

- Real-time processing.
- Distributed processing.
- Distributed storage.
- Scalability.
- Simplicity in writing programs.
- Easy integration with other tools.
- Hardware Resources.

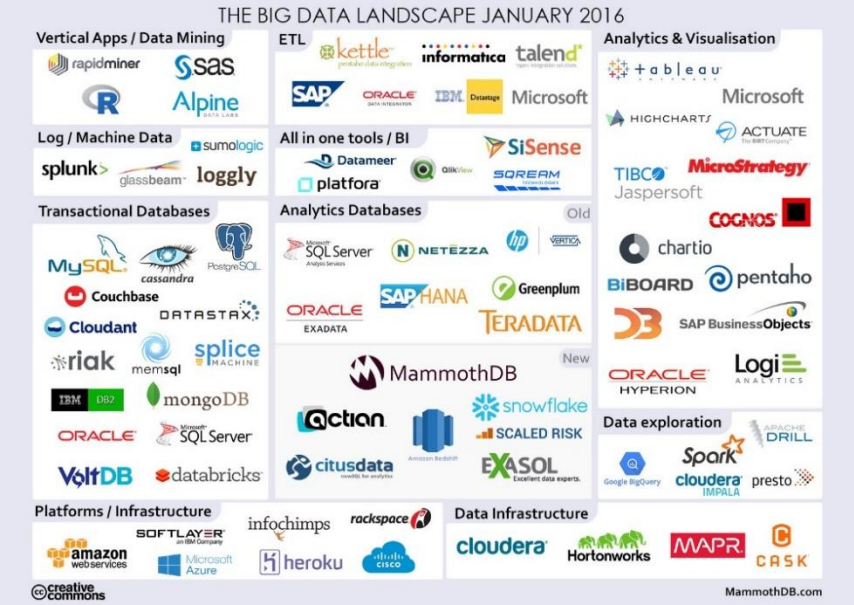


Fig. 4. Big Data map [12].

As a result of the evaluation of Big Data platforms and according to the need of real time PG information monitoring, the Spark platform was chosen as the ideal option according to its free licensing, real-time processing, speed, scalability, simplicity in creating programs and ease of integration with other components.

7 Methodology and Case Study

The objective of Big Data is to provide new knowledge to the company from the processing and analysis of information. This new knowledge will help to support the business decision-making. This section presents the results obtained from applying Big Data in the integration of renewable energies into the Grid.

The main benefits of the implementation of Big Data technologies are:

- Make better decisions with timely and reliable information.
- Show detailed information of the process of generating electricity with renewable energy through the analysis of historical and real-time information.
- Allow the data delivery in a dynamic and flexible way through control panels.
- As a result of the above, improve the the integration of renewable energies in the Smart Grid.

The case study considers the construction of a Big Data infrastructure for the ingestion (acquisition), processing, storage, analysis (descriptive analytics) and monitoring (real time) of the information generated by photovoltaic systems.

The building of the infrastructure is based on existing technologies and available open source tools.

7.1 Architecture of the Photovoltaic System Interconnected to the Grid

In this section it is described the architecture of a Photovoltaic System interconnected to the grid (PG) that is installed at the CFE's National Training Center (CENAC) of Celaya, Guanajuato; as well as the different variables that are monitoring to measure its operating performance.

Figure 5 shows the current Celaya's PG architecture. There are described the main components of the system, the way it performs the interconnection with the CFE's distribution network and the process of obtaining data from sensors.

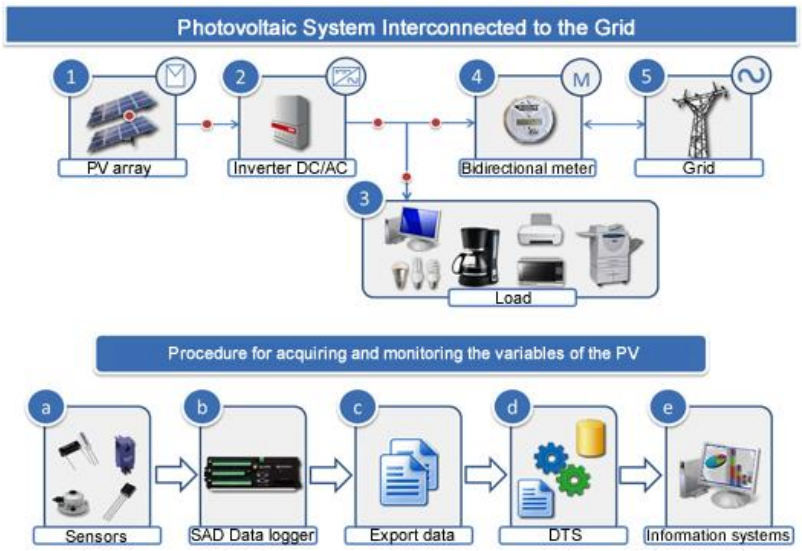


Fig. 5. Architecture of the PG [13].

Sunlight is converted into direct current electricity by the photovoltaic array. This direct current passes through the different components until the investor turns it into alternative, which can provide domestic consumption or send it directly to the mains. There is a data acquisition system (SAD) for the collection and storage of information. The SAD measures electrical signals from sensors (located in different components of the system) at a sampling rate determined and stores it in a flat file. This file is transported (exported) for processing data that will be monitored and analyzed by different users.

Importantly, the data stored in the SAD for each of the variables were scheduled according to the average value of 10 minutes of acquired values each10 seconds [13].

According to experts of the Electrical Research Institute of Mexico (INEEL) Management of Renewable Energy, the implementation of the PG, for its operation monitoring and performance evaluation is subject to international practices and recommendations (IEC 61215, IEC 61730, UL 1703) for the execution of this type of activity in particular.

7.2 System of Photovoltaic System Measurements

For measuring the operating performance of photovoltaic systems 22 basic variables (obtained directly from the SAD), which are listed below, are monitored:

- Three climatic variables: ambient temperature, solar irradiance and irradiance on the horizontal plane of the array.
- In each PG array: current, voltage, power and temperature.
- To the inverter output current, active power and reactive power.
- In each of the grid lines (L1 and L2): active power and reactive power, both input and output, as well as the voltage between lines.

These variables are documented in the specification of requirements PG Celaya [13].

From the measured variables directly from SAD other variables that provide important additional information to measure the operational performance of PG are calculated. These derived variables are classified according to four groups: solar radiation, power, performance indicators and environmental benefits.

Finally, also have varying plant (area of the settlement, azimuth array tilt array) module (technology, brand, and model module), inverter (technology, make, and model) and the PG system itself (type of system, installation type, type of arrangement). They describe technical specifications of these components.

7.3 Implementation of the Case Study

Logical layers group the various components that integrate the Big Data architecture of the prototype: infrastructure, data sources, ingestion layer, processing, storage, analysis and monitoring and visualization layer. Figure 6 shows the implementation of the Big Data

architecture proposed for this work, as well as the complete flow of information. In general, the flow of information consists of the following steps:

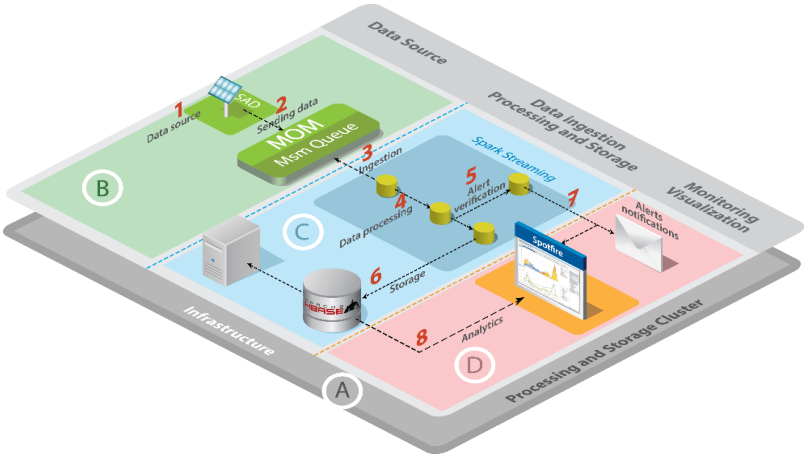
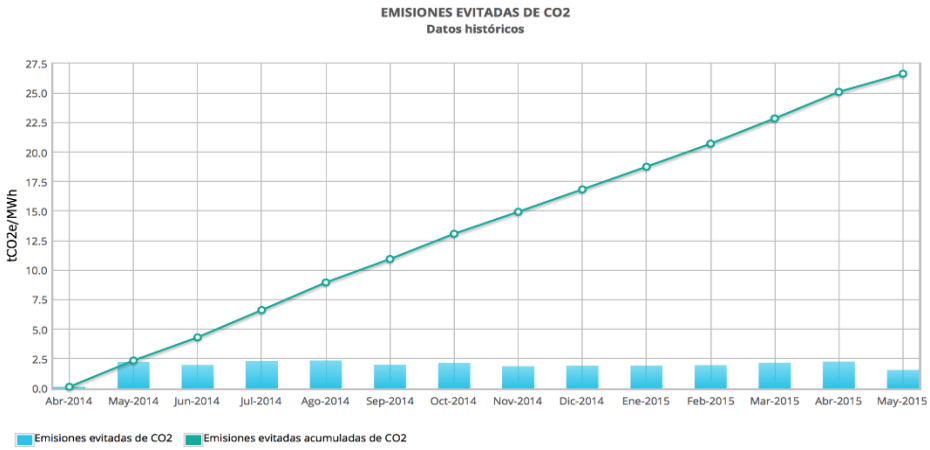


Fig. 6. Implementation of the architecture applied to the photovoltaic system.

1. **SAD-S**: simulates the data generated by the SAD in real time. As a requirement to validate and test the Big Data infrastructure, it was necessary to develop a simulator to generate real-time data and produce data for each of the 22 variables (power, current, voltage, temperature, irradiance, etc.) according to the average value of 10 minutes of measurements acquired every 10 seconds. Also, estimated failure and recovery times are simulated to generate abnormal operation alerts.
2. **Sending data**: the simulator sends the data to the MOM, Message Oriented Middleware. The messaging broker is used as an application to exchange information between the SAD simulator (SAD-S) and Spark Streaming.
3. **Ingestion**: Spark Streaming acquires the MOM data and places it in the memory of the processing cluster. The ingestion process is initiated by establishing a real-time context in order to receive continuous input data streams (DStream). In this way the data can be divided into batches to be processed by the Spark engine.
4. **Data processing**: once the data is placed in memory, Spark Streaming performs the processing of the data through the algorithms written in the "Scala" language. These algorithms calculate statistics, such as sums and averages about the measured variables of the photovoltaic system.
5. **Alert verification**: during the data processing, the measurements of the variables simulated by the SAD-S are validated to detect a failure event in a photovoltaic system (failure simulation of the electric power supply). In case that the variable measurements are out of range, an alert message, which must be distributed by the alert notification component, is created.

6. **Storage**: the processed information is sent to the HBase database, which relies on distributed HDFS storage. Internally, the data storage is managed by the Master through the RegionServers, which store the data in files and send it to the HDFS. These files are divided into one or more blocks to be stored in a set of DataNodes [14].
7. **Alerts notifications**: alerts notifications are responsible for coordinating the distribution and publication trough different media of the alerts generated. For this work the alerts are notified through the dashboards and via e-mail to the operators of the grid.
8. **Analytics**: dashboards and visualization boards provide users with the ability to understand, search and navigate data. This ability to create analysis through reports and dashboards allows stakeholders to make decisions and design appropriate strategies to improve the operational efficiency of photovoltaic systems. The boards were developed using the Spotfire tool, which takes the information stored in HBase.



9. **Fig. 7.** Emissions avoided CO2.

8 Results

This section shows some of the dashboards developed in this research. These dashboards include descriptive analytics and diagnostic analytics. Basically, the dashboards display the operational performance of photovoltaic systems.

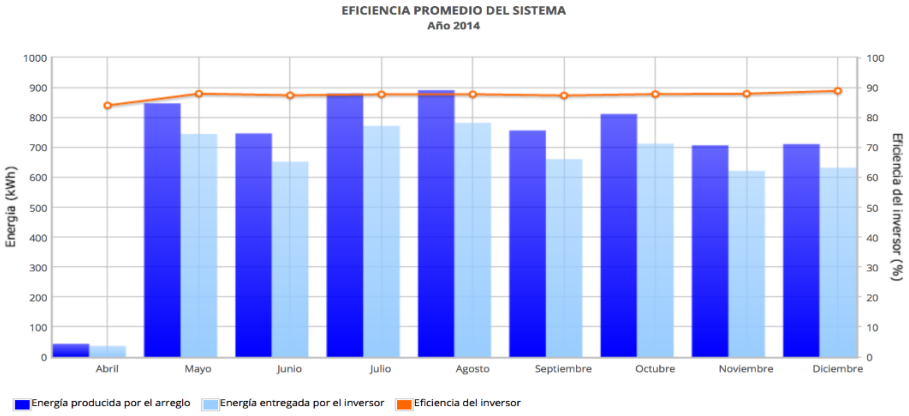
8.1 Emissions Avoided CO_2

In Figure 7 the avoided CO_2 emissions are monitored for each day, month, year and historical as well as all CO_2 emissions avoided accumulated over time. This is based on the

net energy output of the inverter of the photovoltaic system. In this case, monthly historical data is from April 2014 and a cumulative emission trends observed until May 2015.

8.2 Average System Efficiency

Figure 8 shows monthly the energy production of the photovoltaic generator and the output delivered through the inverter to the load or grid, as well as the average efficiency of the inverter.



Fig, 8. Efficiency of the photovoltaic system.

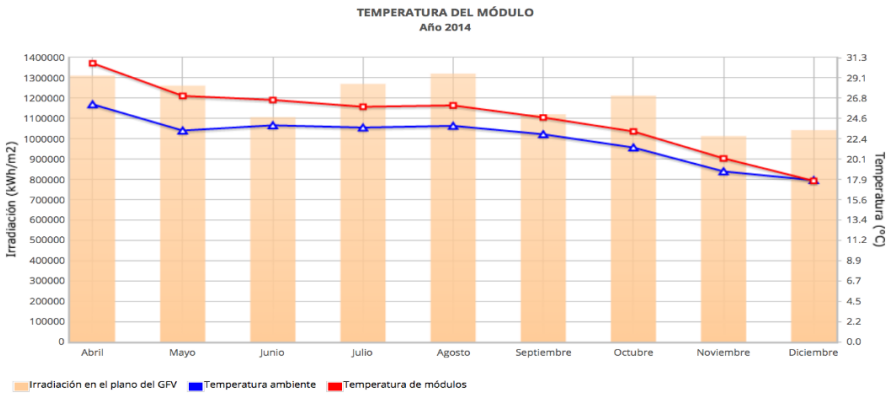


Fig. 9. Meteorological information.

8.3 Meteorological Information

In Figure 9, the meteorological information (irradiation, ambient temperature and temperature of modules) of the site where the photovoltaic system is installed is monitored.

9 Conclusions and Future Research

As a result of this work a technological infrastructure for handling large volumes of information that can be used to implement projects related to the Smart Grid. This infrastructure includes the design of architecture of Big Data for intake, processing, storage, analysis and monitoring of large volumes of data; and the identification and evaluation of the leading technology for the development of these solutions platforms.

The case study for processing data generated by photovoltaic systems interconnected with the grid, allowed to develop boards for analysis, monitoring and display of real-time information.

Consequently, this work will serve as a basis for future research related to energy efficiency, energy for sustainable development and Big Data, which are issues that concern the State and institutions like the Department of Energy, Energy Regulatory Commission, the Institute Electrical Research and the Federal Electricity Commission.

As future work, it is currently being developed for the implementation of mathematical design models predictive and prescriptive analytics, which help define solutions for forecasting demand.

References

1. SENER.:Estrategia Nacional de Energía 2013-2027. SENER, México (2013)
2. González, R., Jiménez, H. R., Huacuz, J. M., Munguía, G.: First 3-phase Grid-connected Photovoltaic System in Mexico. In: 3rd International Conference on Electrical and Electronics Engineering and XII Conference on Electrical Engineering (2006)
3. Escobedo, G.. Jácome, N.. Arroyo. G.: A Framework of Business Intelligence and Data Analytics for Smart Grid Business Intelligence for Smart Grid. In: ICEI-International Conference on Enterprise Information Systems (2016)
4. IEA: Technology Roadmap Smart Grids. Available at: https://www.iea.org/publications/freepublications/publication/smartgrids_roadmap.pdf (2011)
5. CRE: Marco Regulatorio de la Red Eléctrica Inteligente (REI) en México - Para la Comisión Reguladora de Energía. ESTA International, LLC (2014)
6. IEA: World Energy Outlook 2013. Available at: http://www.worldenergyoutlook.org/media/weowebsite/2013/WEO2013_Ch06_Renewables.pdf (2013)
7. López, J., Padilha, A., Gallego, L.: Ubicación óptima de generación distribuida en sistemas de energía eléctrica. Ingeniería y ciencia, Vol. V, No. 9, pp. 9–23 (2009)

8. Gerencia de Energías No Convencionales del IIE.: Guía de usuario. Sistemas fotovoltaicos interconectados con la red, aplicaciones de pequeña escala, Cuernavaca, Morelos, México (2010)
9. Liu, Y.: Big Data and Predictive Business Analytics. Journal of Business Forecasting, Vol. 33, No. 4, pp. 40 (2014)
10. Mysore, D., Khupat, S., Jain, S.: Big data architecture and patterns. Part 3: Understanding the architectural layers of a big data solution. Available at: http://www.ibm.com/developerworks/library/bd-archpatterns3/index.html (2013)
11. Accenture: Accenture Technology Vision 2015 Digital Business Era: Stretch Your Boundaries. Available at: https://www.accenture.com/us-en/_acnmedia/Accenture/Conversion-Assets/Microsites/Documents11/Accenture-Technology-Vision-2015.pdf (2015)
12. MammothDB: Landscape Big Data. Website: http://www.mammothdb.com/the-big-data-landscape-by-mammothdb (2016)
13. IIE: Especificación del Sistema informático para monitoreo y análisis de los datos obtenidos del sistema fotovoltaico de 5kW con la tecnología de película delgada y silicio amorfo. IIE, Gerencia de Tecnologías de la Información (2015)
14. George, L.: HBase: The Definitive Guide. First Edition, O'reilly, USA (2011)

ISSN 1870-4069

Development of a Framework for the Use of a Tool for Machine Learning and Data Mining

Jacqueline Ramos Landeros, Ma. De Lourdes Margain Fuentes

Ingeniería en Sistemas Estratégicos de Información, Universidad Politécnica de Aguascalientes, Aguascalientes, Mexico

{up130142, lourdes.margain}@upa.edu.mx

Abstract. The "Universidad Politécnica de Aguascalientes" is an institution who's interested in quality education, ensuring the control of the subjects that are imparted in a long term in the mayor. Taking this on account, the design and development of a Moodle platform focused on data mining was planted, this subject is currently given to ninth quarter students. In order to re-inforce the knowledge that the course has to offer, it was decided to utilize as a support tool "Weka", it is classified as an "intelligent software". It provides a sustenance in different areas, such as, Marketing, Manufacturing, Health, Energy, Finance, Medicine, inter alia, for its application it must be taken in account the type of assignment it is wished to realize. The purpose of this course is for the student to obtain the necessary knowledge in this field and during this process the material developed can be of use as a tool for this.

Keywords: Data Mining, Automatic Leaning, Business Intelligence, Knowledge Discovery in Databases, Software Weka, Software Engineering.

1 Introduction

At present the task to improve the Access to information that is given to us by the companies is gaining more strength, especially in modern business, where process based in the recourse of information extraction is mainly required, whereby because of the huge workplace and the amount of information, that is called Data Base, is necessary to count on new methods of data processing, new technologies that facilitate the process of search and extraction of knowledge to service companies in the taking of decisions that benefit its performance.
A way to achieve said results are situations or states in which an enterprise pretends to achieve data mining, it is of great importance in the working world, because it allows the obtaining of knowledge based on the data that are found stored, through the process of implicit information extraction, previously known and potentially useful.

The objective of the course is to facilitate the support material for both students and teachers in the learning of the data mining subject, a compilation of a varied set of information, in combination with one of its most famous tools (Weka).

2 Design Foundations

In order to model and develop the course a Learning Management System (LMS) was used, it provides a backing for both educators and pupils in the visual teaching and learning process also it is used to create, endorse, administrate, store, distribute and manage the activities in a visual way.

The task to create the contents for the courses is developed using a Learning Content Management System (LCMS).

Fig. 1. Learning Management System (LMS) Model.

Students that learn on their own and at their rhythm are alone and are completely independent, while the e-learning courses facilitated or directed by an instructor offer different levels of support from tutors, instructors and collaboration between scholars. The e-learning courses often employ both approaches, but in order to be brief and practical, it is easier to examine them in separated ways.

2.1 Moodle Platform

The use of the e-learning platform of Moodle compared with other systems implies that it is made based on the constructivist social pedagogy, where communication has a relevant space in the way of knowledge construction. Being the objective of generating an enriching learning experience [1].

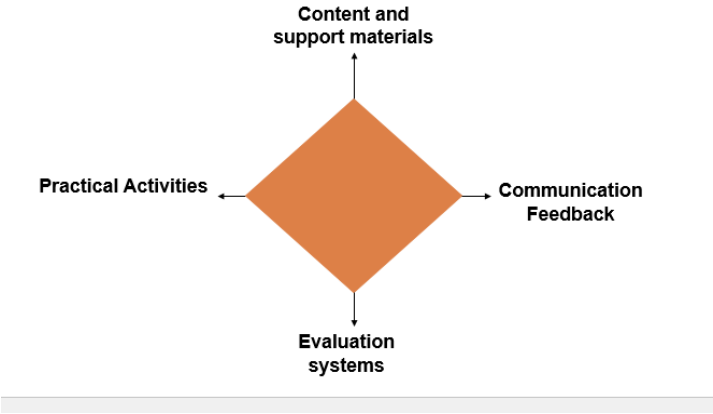


Fig. 2. Moodle work system.

The philosophy planned by Moodle includes a constructive approximation based on the social environment of the education, emphasizing that students (and not only teachers) may contribute in the educative experience in many methods.

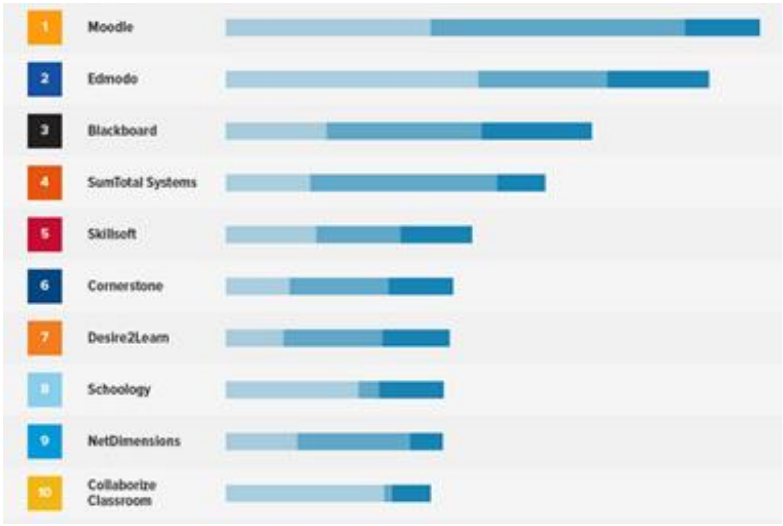


Fig. 3. Comparison between the most used LMS platforms [2].

It can be seen in the previous image how the Moodle platform is the most used among the different systems that may be found nowadays, due to its stability and intuitive interface that permits the user to develop the learning process successfully.

2.2 Methodology

The model that was used as a methodology for the development of the course was the ADDIE model. Is a framework that lists generic processes that instructional designers and training developers use [2]. It represents a descriptive guideline for building effective training and performance support tools in five phases (Fig. 4).

Fig. 4. ADDIE Model [3].

Analysis phase: The analysis phase clarifies the instructional problems and objectives, and identifies the learning environments, learner´s existing knowledge and skills.

Dsign phase: The design phase deals with learning objectives, assessment instruments, excercises, content, subject matter analysis, lesson planning, and media selection.

Development phase: In the devolpment phase, instructional designers and developers crate and assemble content assets blueprint in the design phase.

Implementation phase: The implementation phase develops procedures for training facilitators an learners.

Evaluation phase: The evaluation phase consists of two aspects: formative and summative, the first one is present in each stage of the ADDIE process, while summative evaluation is conducted on finished instructional programs or products.

3 Organization and Components of the Course

The organization of the course twirls around of sections that divide the contents and activities based on its main function, impart the student the background necessary to the apprenticeship in the Data Mining field. It is observed in Fig. 4 the course structure.

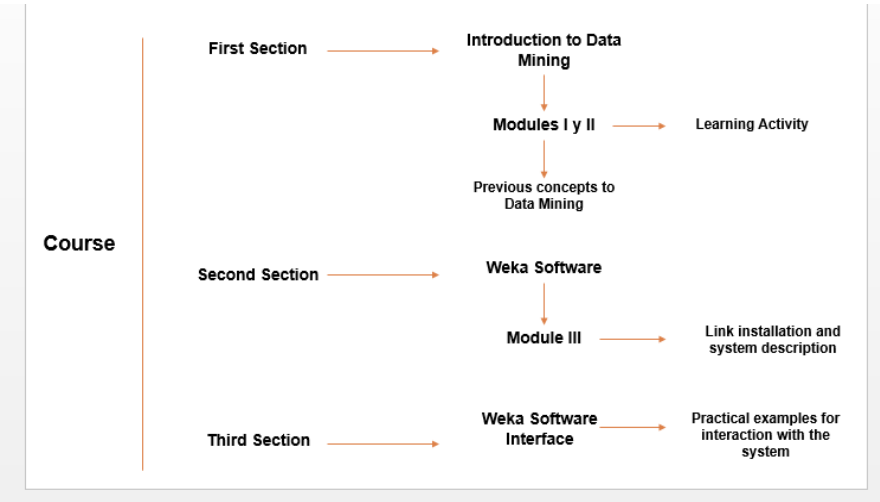


Fig. 5. Course Structure.

3.1 Description of the Sections of the Course

First Section:

The first section was established specially for the student to comprehend the background concepts of Data Mining and therefore continue with the learning process, which was made in two modules, which are detailed next:

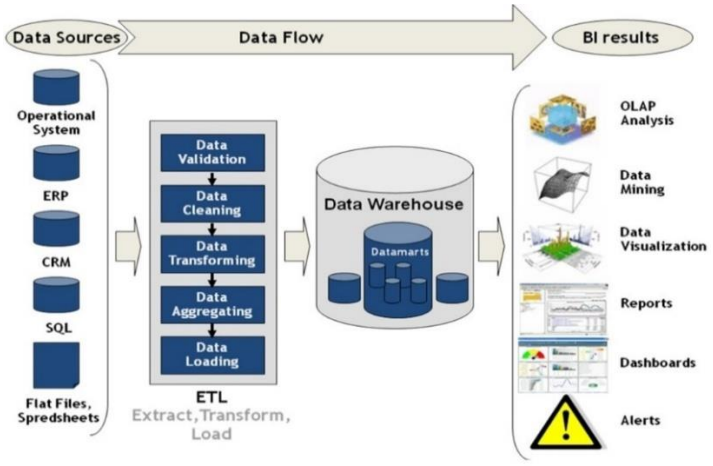


Fig. 6. BI Scheme [5].

Module I:

The first step to Data Mining is to comprehend the business, this means to determine the objectives in order to decide the goals of the Data Mining thus the Business Intelligent subject was investigated, and as a result BI scheme, it is an essential process that is obtained as a result of the sources and data flows to achieve the process of Data Mining.

Business Intelligence:

Business Intelligence is the ability to transform data in information, and information in knowledge, in a way that it can improve the decision making in business. [5]

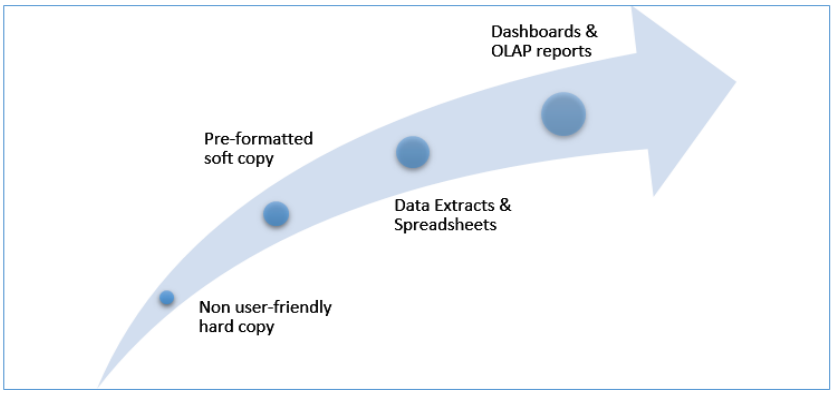


Fig. 7. The Evolution of Business Intelligence [5].

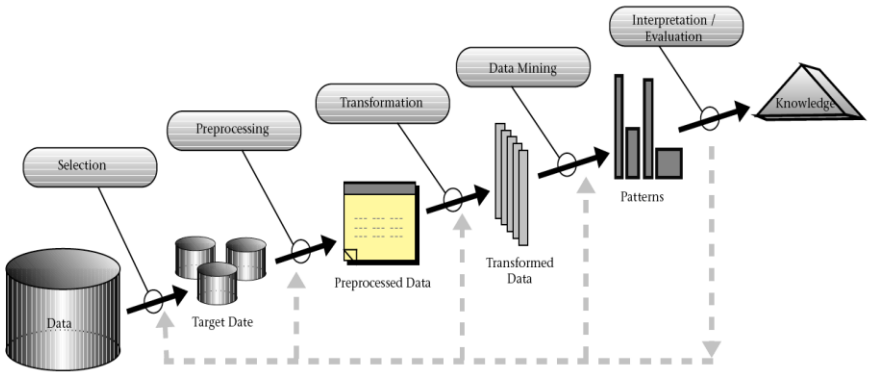


Fig. 8. KDD process.

Knowledge Discovery in Databases (KDD):

With the necessity of being able to handle large quantities of data, a study field arises, named Knowledge Discovery in Databases (KDD).

It refers to "the non-trivial process of discover knowledge and information potentially useful in the contained data inside an Information Warehouse. It is not an automatic process, it is an iterative process that explores thoroughly extremely large volumes of data to determine relations. It is a process that consist in using methods of Data Mining (algorithms) to extract (identify) what is considered as knowledge" [5]

Data Mining:

In simple terms it is about a data exploitation method and information extraction that transforms it into useful knowledge to help the decision making in an organization through the determination of patterns and models. It arises to comprehend the content in a Data Warehouse.

Module II

The second module is deeply detailed about the preparation of the data from its output up until its integration.

Output and Data Set Description

This topic is produced by the phase of data preparation, that will be used in order to model or the main Project analysis.
Describing the data set that will be used in the modeling and the main project analysis.

Data Selection and Recollection

Once the Data of interest are collected, an explorer (in this case the user) may decide what type of pattern will be found. The kind of knowledge that is desired to extract will marc clearly the technic of Data Mining used to decide which data will be used for the analysis.

Data Cleaning

It is a fundamental process in the data migration, its main objective is the quality of data obtained at the end of migration. Data Cleaning is particularly important when data come from heterogeneous sources (different sources) that may not share the same scheme of data or can´t represent the same real entity in different ways.
The process of cleaning constitutes great part of the estate of transformation of the data during Data Warehouse construction.
This duty includes the construction of data preparation operations such as the production of derivate attributes or the entry of new registers, or the transformation of values for existent attributes.

Outputs Derived Attributes

The derived attributes are new attributes that are built from one or more existing attributes in the same registry.
To clarify what's said before, example: area=length*width

Generated Registry

In this step the newly generated registry are described.
Example: Create registry for women that bought lipstick color brown, thus there is no reason to obtain such registry in the raw data, but in order to model this might have sense to represent explicitly the fact that certain women have not made a purchase.

Data Integration

This is the method by which the information is combined in multiple tables or files to create new archive or values.

Data Format

Reformatting information refers to modifications made to the data that don`t change its meaning, but may be required by the modeling tool.

Second Section (Module III)

Software Weka:

On the third module an explanation in the functionality of the Weka software is developed, nowadays different tools exist for the visualization and algorithms for the analysis of data joined in a user graphical interface to easily access its functions making use of the processing of information manifested in Data Bases, but after making said investigation, a conclusion was made that this tool is the most reliable in the market and at the same time it is at full disposal for the user.

"Weka is a software tool for the automatic learning and Data Mining designed on base Java and developed at the university of Waikato in New Zealand in 1993, this tool with the acronym (Waikato Environment for Knowledge Analysis) provides license distribution GNU-GLP or free software" [5].

Application Fields

Weka has a group of technics that can be applied with success to multiple fields, such as Marketing, Manufacturing, Health, Energy, Finance, Medicine, among other, for its application it must be taken in account the king of task is being planned to realize.

- Classification methods based on neuronal nets.
- Numeric methods manipulation over data (financial statistics).

- Classification methods based on vector support machines.
- Meta classifiers.
- Methods of decision trees implementation.
- Probability estimation methods.

Third Section (Software Interphase)

In this section a paragraph for the installation of the software is included. The sections of the software are:

- *Pre-process:* Here are included all the tools and filters to charge and manipulate the data that are going to be used.
- *Cluster:* Here various methods of grouping are integrated.
- *Associate:* includes the technics used for the association rules.
- *Select Attributes:* It contains the access options to the different technics used to reduce the amount or number of attributes.
- *Visualize:* This section allows to do the studies of behavior, using the visualization technics included in Weka.

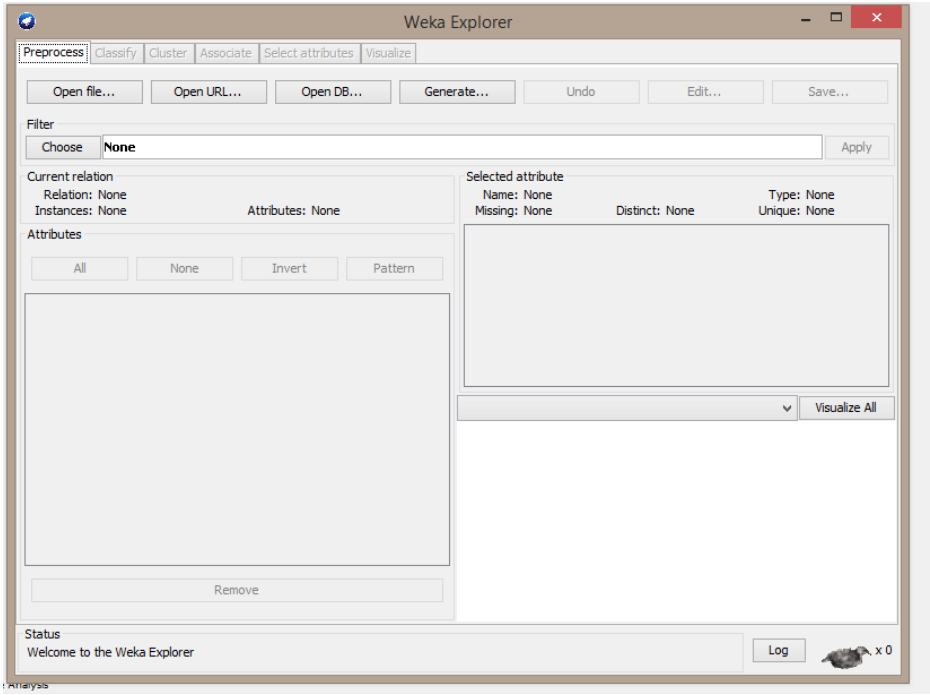


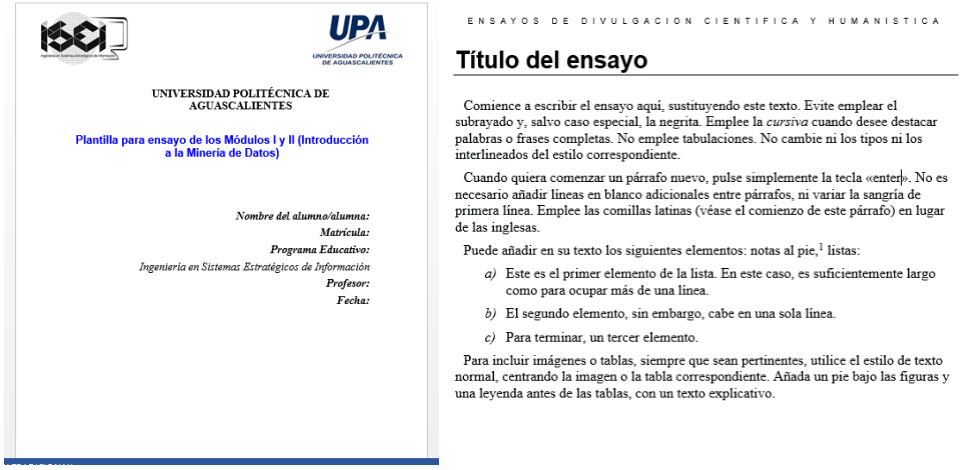
Fig. 9. Weka Software Interphase.

4 Results

To implement the course on the Moodle LMS platform the following results are obtained.

Fig. 10. Course implemented in the Moodle platform.

4.1 Exercises and Material



UNIVERSIDAD POLITÉCNICA DE AGUASCALIENTES

Plantilla para ensayo de los Módulos I y II (Introducción a la Minería de Datos)

Nombre del alumno/alumna:
Matrícula:
Programa Educativo:
Ingeniería en Sistemas Estratégicos de Información
Profesor:
Fecha:

E N S A Y O S D E D I V U L G A C I O N C I E N T I F I C A Y H U M A N I S T I C A

Título del ensayo

Comience a escribir el ensayo aquí, sustituyendo este texto. Evite emplear el subrayado y, salvo caso especial, la negrita. Emplee la *cursiva* cuando desee destacar palabras o frases completas. No emplee tabulaciones. No cambie ni los tipos ni los interlineados del estilo correspondiente.

Cuando quiera comenzar un párrafo nuevo, pulse simplemente la tecla «enter». No es necesario añadir líneas en blanco adicionales entre párrafos, ni variar la sangría de primera línea. Emplee las comillas latinas (véase el comienzo de este párrafo) en lugar de las inglesas.

Puede añadir en su texto los siguientes elementos: notas al pie,[1] listas:

a) Este es el primer elemento de la lista. En este caso, es suficientemente largo como para ocupar más de una línea.

b) El segundo elemento, sin embargo, cabe en una sola línea.

c) Para terminar, un tercer elemento.

Para incluir imágenes o tablas, siempre que sean pertinentes, utilice el estilo de texto normal, centrando la imagen o la tabla correspondiente. Añada un pie bajo las figuras y una leyenda antes de las tablas, con un texto explicativo.

Fig. 11. Essay template.

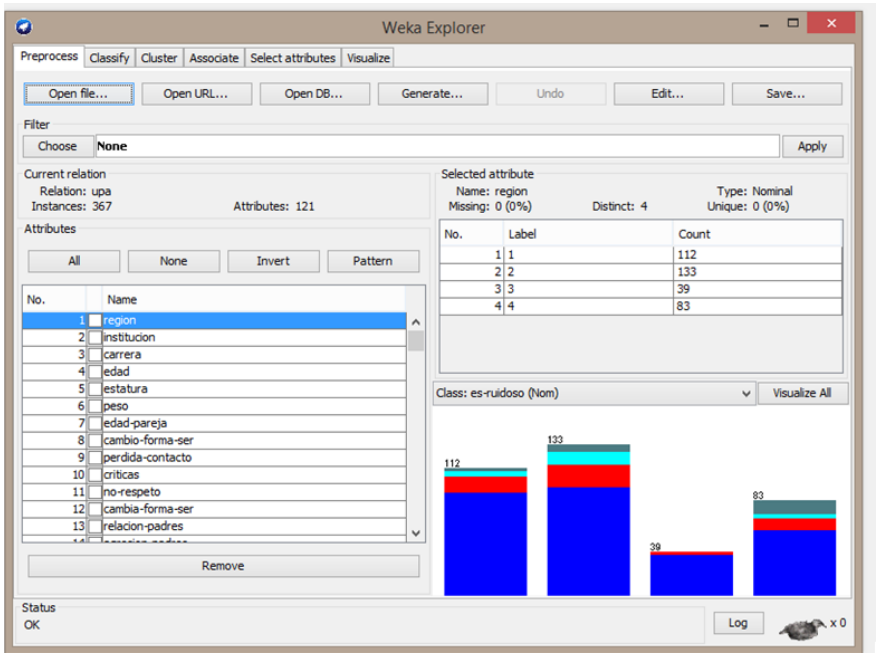


Fig. 12 Database for practice.

5 Conclusions

At the end of the realization of this material, it has been concluded that the main core of it is that the students become capable of develop pan autodidact way of learning, because in college is of great importance the search of knowledge trough the own ways of the autodidact learning. In order to achieve this it is necessary to count on the didactic material that facilitates the search of the subject in which it is going to be worked.

The methodology exposed in this material is focused in providing a line of solution in the definition of advanced terms in the Data Mining field, so that through its knowledge and application using different tools that this branch has to offer it is desired to achieve an objective which is knowledge extraction.

The information of the course has the objective of providing assistance as a tool of support in future works of the students, whereby it will be necessary to improve the content as the information is upgraded.

Acknowledgements. The development of this course would not have been possible without the help from the Dr. Ma. De Lourdes Margain Fuentes (Academic Program Director).

References

1. Sancho, J. B.: La Plataforma Educativa Moodle. Available at: http://www.fvet.uba.ar/postgrado/Moodle18_Manual_Prof_1.pdf (2007)
2. Puzziferro, M., Shelton, K.: A model for developing high-quality online courses: Integrating a systems approach with learning theory. Journal of Asynchronous Learning Networks, pp. 12, 119–136 (2008)
3. Wang, S. K., Hsu, H. Y.: Using the ADDIE model to design Second Life activities for online learners. TechTrends, 53(6), pp. 76–81 (2009)
4. Dillenbourg, P.: Collaborative learning: Cognitive and computational approaches. Advances in learning and instruction series. Elsevier Science, Inc., PO Box 945, Madison Square Station, New York, NY 10160-0757 (1999)
5. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H.: The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), pp. 10–18 (2009)
6. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI magazine, 17(3), pp. 37 (1996)
7. Liu, H., Motoda, H.: Feature extraction, construction and selection: A data mining perspective. Springer Science & Business Media, 453 (1998)
8. Black, E. W., Dawson, K., Priem, J.: Data for free: Using LMS activity logs to measure community in online courses. The Internet and Higher Education, 11(2), pp. 65–70 (2008)
9. Macfadyen, L. P., Dawson, S.: Mining LMS data to develop an early warning system for educators: A proof of concept. Computers & education, 54(2), pp. 588–599 (2010)
10. Davies, J., Graff, M.: Performance in e-learning: online participation and student grades. British Journal of Educational Technology, 36(4), pp. 657–663 (2005)
11. Nagel, L., Kotzé, T. G.: Supersizing e-learning: What a CoI survey reveals about teaching presence in a large online class. The Internet and Higher Education, 13(1), pp. 45–51 (2010)

12. Davies, J., Graff, M.: Performance in e-learning: online participation and student grades. British Journal of Educational Technology, 36(4), pp. 657–663 (2005)
13. Romero, C., Ventura, S., García, E.: Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1), pp. 368–384 (2008)
14. Macfadyen, L. P., Dawson, S.: Mining LMS data to develop an early warning system for educators: A proof of concept. Computers & education, 54(2), pp. 588–599 (2010)

ISSN 1870-4069

A Comparison Represented in the Form of Radar of University Student Engagement in Degrees in Technologies

Erika Yunuen Morales Mateos[1], José Alberto Hernández Aguilar[2], Carlos Alberto Ochoa Ortiz[3], María Arely López Garrido[1]

[1] Universidad Juárez Autónoma de Tabasco Cunduacán, Tabasco, Mexico

[2] Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, Mexico

[3] Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Chihuahua, Mexico

erika.morales@ujat.mx,jose_hernandez@uaem.mx, alberto.ochoa@uacj.mx, arely.lopez@ujat.mx

Abstract. This research work aims to raise awareness of the results of the student engagement in university students of three careers in technologies, of a university in the south of Mexico, through the development of graphics that allows displaying in an adequate manner the cases under study, facilitating the interpretation of the same. To measure the student engagement was used the instrument Utrecht Work Engagement Scale for Students (UWES-S), which is composed of the dimensions, vigor, dedication and absorption. Graphs presented in this work are a graphics array correlation, a graphical representation of radial and a chart GGE bi plot. It was observed that the student engagement is in general, for this population sample, average level, identifying as predominant the dimension of dedication.

Keywords: UWES-S, Vigor, Dedication, Absorption, Student engagement.

1 Introduction

Universities in Mexico deal today to achieve good indicators of quality operating with educational models that involve the integral formation of the students to raise the quality of their educational process. It is currently considered the student is responsible for their own learning, involving the search, selection, analysis and evaluation of information, assuming a role with greater activity in the generation of his own knowledge. Therefore, it is important that the universities students feel concerned, motivated and committed to their academic process, due now are required high levels of psychological linking with the studies, i.e. academic engagement or student engagement [1]. The psychological well-being is one of the primary elements that directly affects the behavior and performance of university

students. However, there are no recent studies to determine the situation on the behavior of the students, as well as the factors that lead to the same.

The positive psychology focuses on the process of analysis of capabilities and potential of the people, with their appearance are generated theories that allow finding elements that impact in human capabilities and at its optimal performance [2].

In the field of occupational health psychology, from the positive approach, Schaufeli and Bakker [2] have developed instruments such as the Utrecht scale of engagement in the work, (Utrecht Work Engagement Scale, UWES), which allows to measure the engagement in the work. The engagement in recent years is a concept that is very used in the business environment and consulting, appearing recently studies in the academic environment [1]. In addition, Schaufeli and Bakker [2] have developed a version of the UWES for students, thus obtaining the Student engagement, because both the workplace as an engagement to involve student activities. Students like the workers perform structured activities that imply obligations and that are aimed at achieving a goal.

In the literature is found research carried out in universities at the international level where it has been implemented the UWES-S for the student engagement. Da Wedge, Soto, Gutierrez et al. [3], developed a research at the University of Vigo in Spain, in which obtained the level of student engagement and its dimensions. In Latin America, Parada and Perez [4] applied this instrument to students of a University of Chile. In Mexico, Leon, Romero, and Olea [5] applied the UWES-S to university students of engineering of the University of Sonora, obtaining the factorial validity of the instrument and the average student engagement.

An important contribution of this work is to make known through the use of the instrument UWES-S, linking psychological with the studies, namely that so dedicated, absorbed and vigorous are students, to know the level of student engagement. These indicators will detect symptomatic levels and take appropriate decisions based on the same.

In this study are developed graphical representations to display the results obtained of UWES-S. We applied a correlational analysis that measures the linear association between variables, represented in a graphical array of correlation. In a radial representation there is a comparison of the cases, where is possible to appreciate the results of the dimensions of the student engagement that involves the UWES-S. Finally, a multivariate analysis was applied, using the graphics of type GGE biplots involving the analysis of main components, allowing observation of the cases involved in this study, corroborating the results obtained in previous analyzes. The various analyzes applied in this study have been developed using the language R as a proposal [6].

Research has been found where different techniques of data analysis have been used to obtain information about the academic performance of university students, such as knowing the factors that cause students to desertion through the use of decision trees [7], the use of techniques of grouping to know the styles of learning [8], among others. Some research involves the personal factor with academic performance, data have been obtained from formal surveys, which have been applied data analysis techniques to discover the relationships between the variables involved, such as rules of association techniques [9], Classification techniques such as trees [10]. There are currently a large variety of studies

based on the use of principal components where observations are presented on the axes formed by the first two main components [11].

2 Materials and Proposed Method

2.1 Data Description

The purpose of this research is to present results of a comparison on the student engagement that present students of different grades of a population sample belonging to a school of informatics and systems, at a university in the south of Mexico. The degrees considered for this study were degree in administrative computer (LIA), degree in computer science (LSC) and degree in information technology (LTI). For convenience a non-probabilistic guided sample was used [12]. The survey of the UWES-S was applied to groups of four students per degree taking into account two men and two women, who agreed to answer the questionnaire on a voluntary basis, obtaining a total sample of 12 students, to which we had access in the period February-August 2015.

2.2 Scale of Welfare in the Academic Context (Utrecht Work Engagement Scale for Students, UWES-S)

Salanova and Schaufeli [13] define the engagement as "a positive psychological state characterized by high levels of energy and vigor, dedication and enthusiasm for the job, as well as total absorption and concentration in the labor activity".

Schaufeli and Bakker [2] developed the questionnaire scale Utrecht of engagement in the work, (Utrecht Work Engagement Scale, UWES), which was originally created for the Dutch population and currently has been validated at the international level. The UWES-S is composed of three dimensions that comprise the engagement: vigor, dedication and absorption. The original version of 17 items, Schaufeli and Bakker [2] indicates that: "the vigor is evaluated using six items that relate to the high levels of energy and resilience, the willingness to devote efforts, not fatigued with ease, and persistence in the face of difficulties. The Devotion is assessed through five items that relate to the sense or meaning of work, to feel excited and proud of their work and feel inspired and challenged by the job. The absorption, is evaluated using six items that relate to be happily immersed in his work and present difficulty to leave, in such a way that the time passes quickly and you forget everything around you".

Subsequently Schaufeli and Bakker [2] also developed the version of Utrecht Work Engagement Scale for Students (UWES-S), where the 17 items that originally composed this questionnaire were modified using the student approach. The following are examples of the items that make up the constructs UWES-S Schaufeli and Bakker [2]: For the dimension vigor we can mention: "I feel strong and vigorous when I am studying or go to classes, I can continue studying for long periods of time when I am very resistant to cope with my tasks such as student". For the dimension dedication: "I think that my career has

meaning, I am excited with my career, my career is challenging for me." Finally, in the absorption dimension "time flies when I perform my duties as a student, I am happy when I am doing tasks related with my studies; it is difficult for me to quit my study".

The responses of the instrument are measured according to a Likert-type scale where zero means "Never" and six "always". The original internal consistency obtained by the authors for the version UWES-S of 17 items in Dutch students was to vigor, dedication and absorption of 0.63, 0.81 and 0.72 respectively; compliance with the criterion of superiority to 0.60 for an instrument of recent development [2].

2.3 Data Analysis

A descriptive analysis was applied to the data obtained for the values minimum, maximum, standard deviation and mean, corresponding to the dimensions that make up the student engagement: vigor, dedication and absorption, as well as the student engagement as a variable.

It was also applied a correlational analysis represented in a chart that measures the degree of linear association between variables. The size of the coefficient indicates the degree of association between the variables, i.e., while close to one the greatest correlation, on the contrary is less if it is closest to zero [14].

Developed radar graph representations which compares the results obtained in the different dimensions of student engagement, as well as the value obtained at the level of engagement student, this for each case presented, as well as for the population shows grouped by race. The radar charts are based on the implementation of radial axles with common origin and scales standardized. Each one of the axis represents the value of a specific indicator [15].

With the data is also implemented a chart for a multivariate analysis: GGE biplot for each case presented with regard to the engagement of students and their dimensions. In this type of graphics are used by the combined effects of genotypes (G) and interaction of genotype-environment (IGA), that facilitate the visual identification of the genotypes and the environments of the assessment. These graphics are constructed using the first two main components, derived from the breakdown of the combined effects of G+IGA. The first component, when is highly correlated with the main effect of genotype, represents the proportion of performance that is only due to the characteristics of the genotype. The second principal component represents the part of performance due to the interaction genotype-environment [16].

3 Results

In this descriptive analysis were applied calculations to the sample of this study, corresponding to the minimum and maximum values, average and standard deviation. With a sample of 12 students of three races in technologies: LIA, LSC and LTI, in Table 1, we can observe the values obtained for the student engagement and each one of the dimensions

that comprise it. The minimum value obtained in the observations presented in the absorption with 1.3, the maximum value is presented in the dedication with 6.0. The values that are displayed in the media of the dimensions from highest to lowest are for dedication 4,767, vigor 3.983 and 3.908 absorption, finally the average of this population shows in the student engagement is a 4.233.

Table 1. Descriptive statistical population study sample.

Variable	Minimum	Maximum	Mean	Standard Deviation
Vigor	1.5	5.7	3.983	1.4708
Dedication	2.2	6.0	4,767	1.2787
Absorption	1.3	5.8	3.908	1.4171
Engagement Student	2.2	5.8	4.233	1.2759

A scale was created to qualify the student engagement, i.e. a table with values set, that allow evaluating the observations of an individual or group of individuals [17]. It is also possible to use other methods to calculate the levels of engagement with the UWES-S, refer to Schaufeli and Bakker [2].

Table 2. Scale proposed to measure the UWES-S.

Category	**UWES-S**	**Vigor**	**Dedication**	**Absorption**
Very Low	Score < 2.20	Score < 2.80	Score < 2.50	Score < 2.70
Under	2.20 ≤ score < 3.30	2.80 ≤ score < 3.80	2.50 ≤ score < 3.50	2.70 ≤ score < 3.60
Middle	3.30 ≤ score < 4.70	3.80 ≤ score < 5.20	3.50 ≤ score < 4.50	3.60 ≤ score < 4.70
High	4.70 ≤ score ≤ 6.00	5.20 ≤ score ≤ 6.00	4.50 ≤ score ≤ 6.00	4.70 ≤ score < 6.00

Using the previous scale can be interpreted for this population shows that the average vigor 3.983 and 3.908 absorption, is medium, i.e. the students feel regularly vigorous and absorbed, even more dedicated because they obtain a value considered high, with a 4,767, are motivated, consider that his career has meaning, are excited and proud of their studies. Finally, the student engagement is of 4.233 indicates that present a mid-level engagement, regularly feel committed.

In a study carried out in Mexico to a group of engineering students from the University of Sonora, average values were found for vigor, dedication and absorption of 3.43, 3.99 and 3.09 respectively [5]. Therefore, when comparing the results, it is obtained that to the north and south of the country; the highest score is obtained in the dedication, followed by vigor and finally absorption. Student engagement is similar in these populations of university students.

3.1 Correlational Analysis of the Dimensions of the Student Engagement

The correlation analysis is represented by a chart of matrix where you measured the degree of association between the dimensions that comprise the student engagement of the population sample under study. The method used is the Pearson correlation coefficient, which is a measure of the linear relationship between two quantitative variables [18], the significance level is 0.01. In Figure 1 are shown that the resulting correlations between variables vigor, dedication and absorption are close to one, which indicates that there is a degree of important relationship between these variables, so that the instrument used to measure the student engagement is suitable for this population sample.

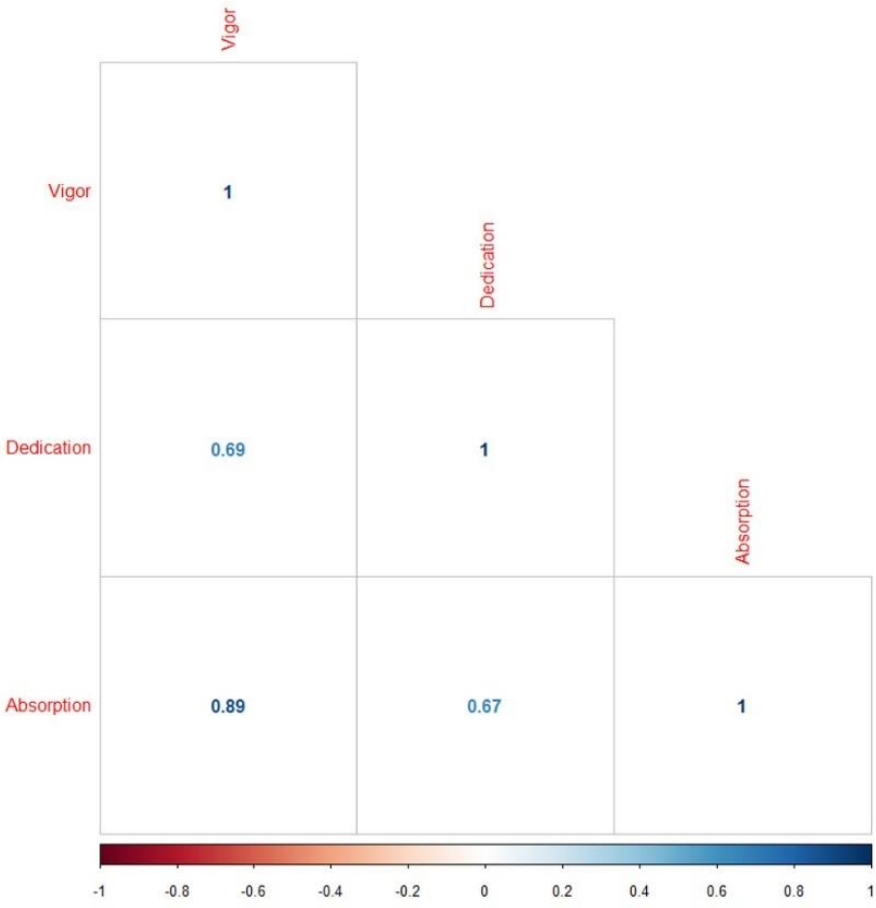


Fig. 1. Chart of correlation matrix of the dimensions vigor, dedication and absorption of the UWES-S, Source: (Self-realization, 2016).

3.2 Representation of Radar

Developed representations of radar charts where is displayed the comparison of the results of the student engagement, and their dimensions of vigor, dedication and absorption, taking into account the sample population by career and individual cases.

Figure 2 shows the result of the student engagement by careers, each one of them corresponds to an axis and in its scale presents the level obtained in the student engagement. You can see graphically that the students of LSC presented a greater engagement student, followed by the students of the career of LIA and finally the students of LTI. In the same way, it can be seen that this population shows in general highlights the level of engagement and on the contrary the lowest level as presented in the absorption.

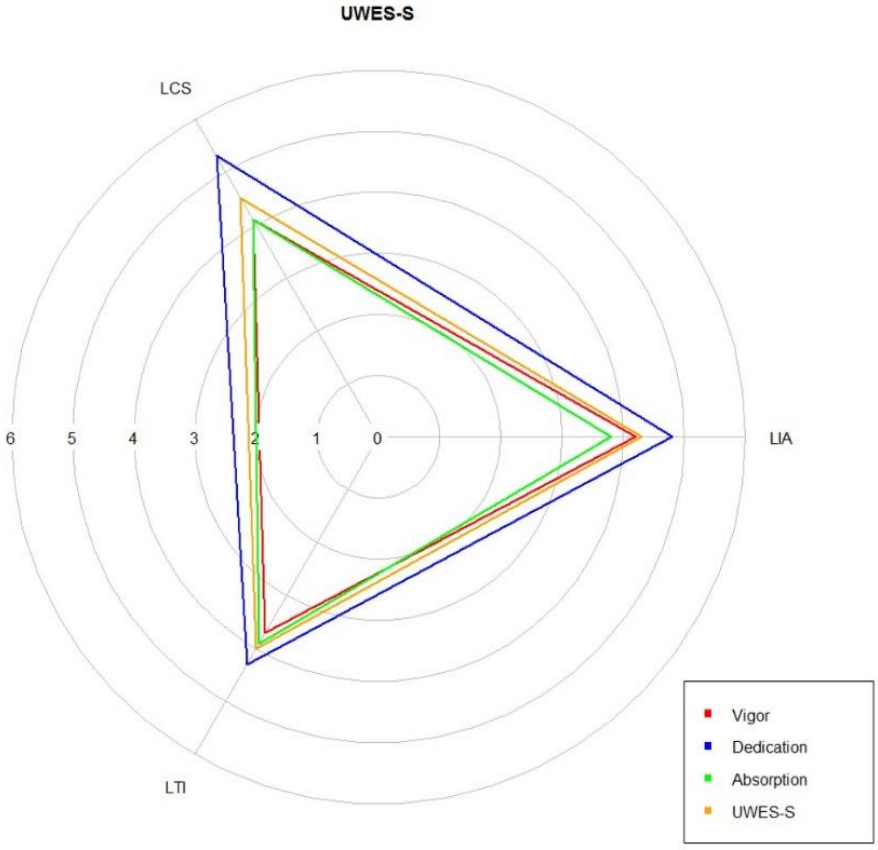


Fig. 2. Radial representation of the dimensions vigor, dedication, absorption and UWES-S, by degree, obtained from the students of the races in LIA, LSC AND LTI (Source: Self-realization, 2016).

In Figure 3, is shown the chart that corresponds to the cases presented in this study, identified by the acronym of the career to which they belong LSC, LIA and LTI, followed by the letter H or M, for men and women respectively and finally by a number which identifies them. In general, women have a level of dedication, vigor, and absorption more frequent and with values higher than men. One of the representative cases is the LSCM2, which presents the highest level of dedication of the population sample, followed by vigor and absorption, both at levels equal. Another interesting case is LIAM2 that is approximately the same levels in the student engagement and in all the dimensions of which it is composed. Two cases that behave very similar are those of two students of different races LSCH1 y LTIH2, since they present levels of vigor and absorption lower than the dedication and a level of student engagement under.

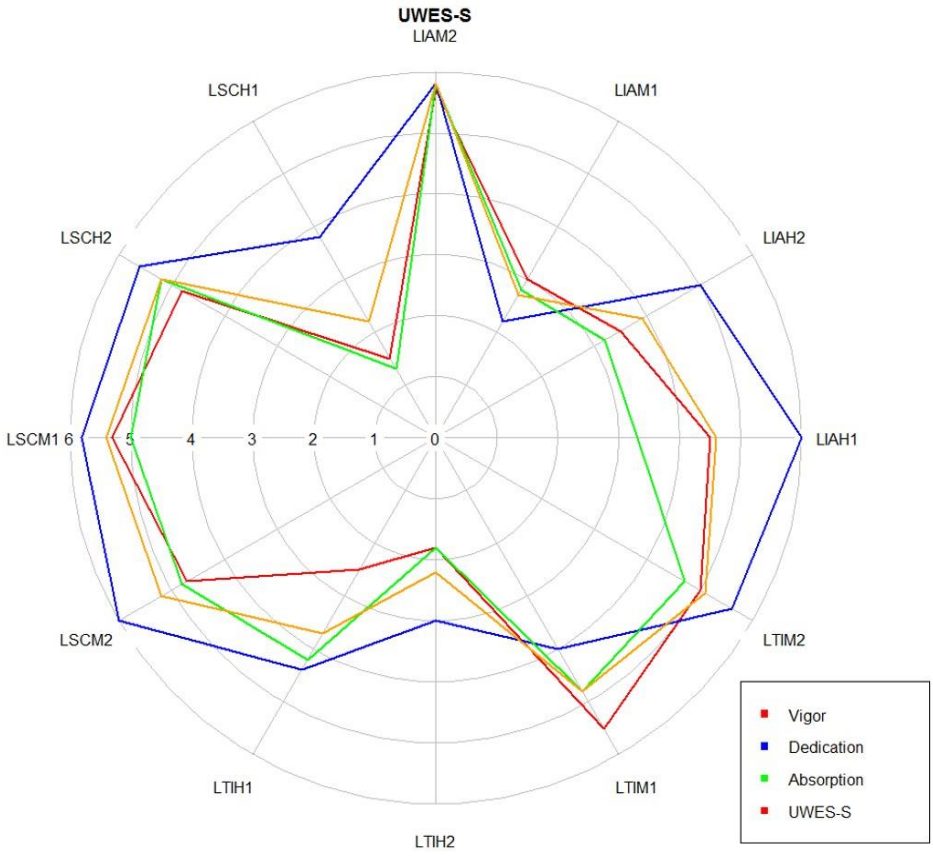


Fig. 3. Radial representation of the dimensions vigor, dedication, absorption and UWES-S, obtained by cases of students of the careers in LIA, LSC AND LTI (Source: Self-realization, 2016).

3.3 Multivariate Analysis: GGE Biplots

Presented below is a multivariate analysis through a graph of type GGE biplot for each case of this population sample, with regard to the engagement of students and their dimensions.

In Figure 4, can be observed the genotypes (G), which are the students, where they continue to identify by the career, gender and an additional number only, it also identifies to the interaction of the genotype-environment (IGA), such as the student engagement and its dimensions of vigor, dedication and absorption. The graph allows viewing the first two main components, derived from the breakdown of the combined effects of G+IGA [16]. The first principal component explaining 87.72% and the second main component the 9.08%, so the variability explained by the two shafts is greater than 50% [16], so that with a 96.8% are considered valid interpretations presented.

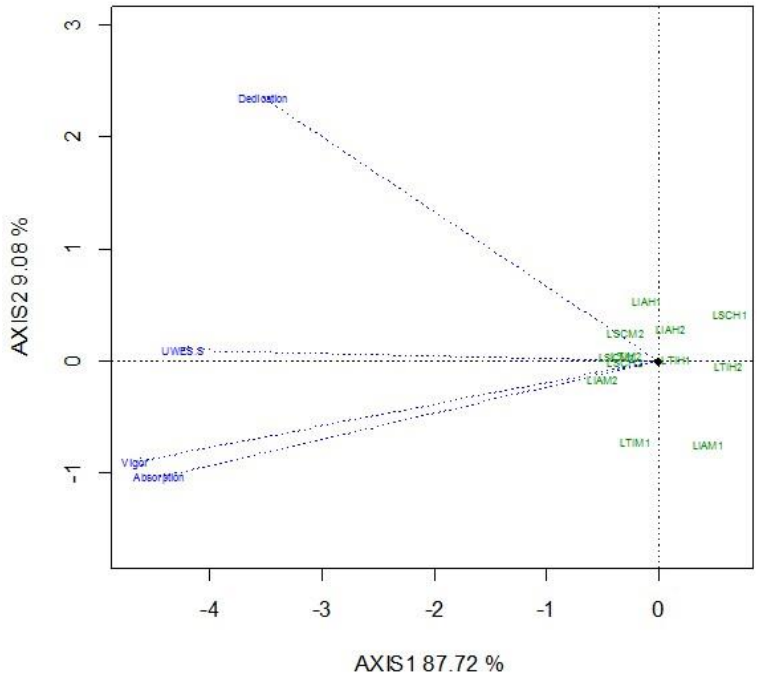


Fig. 4. GGE biplot graphical dimensions vigor, dedication and absorption and UWES-S, Source: (Self-realization, 2015).

The graphics GGE biplots are interpreted in terms of distances between elements and orientation of the axis; this allows us to get an approximate way, from the chart, the values of all the variables for each case [16]. For the description of the results have been considered the cases with most representative values (ends), and can be seen in Chart other cases with

features close to these. The variable absorption takes high values for the majority of the cases, and lowers for cases that are in the opposite direction as LSCH1 and LTIH2. The variable vigor is characterized by high values in a general way in the cases presented and low in cases LSCH1 and LTIH1. The dimension dedication has low values in cases LIAM1 and LTIH2. Considering the cases not common, LSCH1, LTIH2 and LIAM1 are characterized by low values in the student engagement (UWES-S).

4 Conclusions

In this work was given to know the university student engagement of a population sample of three degrees LIA, LSC and LTI from a University of the South of Mexico. It is important to know aspects from the psychological point of view that directly assist in the formation of the university students, and measures can be taken to support them. The student engagement is a positive psychological state that is characterized by high levels vigor, dedication and absorption. The instrument UWES-S, allows knowing the engagement student, in a practical way, obtaining successful outcomes, since it has been validated at the international level. This work has focused on presenting results of student engagement, highlighting the visual aspect, which allows you to easily identify the elements involved in the same. For this is developed various graphical representations that allow to know from different perspectives the results.

We developed a correlation analysis, represented in a graphical array correlation, where it was found that there is a correlation between all the dimensions of the student engagement. In a radial graph representation were compared the races and the cases submitted, resulting dimension of dedication is the more present in university students, they are motivated, and feel that his career has meaning. This study concludes with a multivariate analysis, using a chart GGE biplot, where are displayed all the cases involved in this study and may be observed the student engagement and its levels, on the basis of the distances.

Finally, to carry out these analyzes with graphical representation, it was found that university students feel regularly vigorous and absorbed, highlighting the dedication, are excited and proud with his career, since regarded it as a challenge. Its engagement presents a student average level, regularly feel committed to their studies.

It is proposed to continue with studies related to the student engagement, in other universities, developing graphical representations to obtain results of a set of data, from different approaches to facilitate the understanding of the results.

References

1. Casuso, M. J.: Estudio del Estrés, Engagement y Rendimiento Académico en Estudiantes Universitarios de Ciencias de la Salud. Disertación doctoral publicada, Universidad de Málaga, España (2011)

2. Schaufeli, W. B., Bakker, A.: Utrecht Work Engagement Scale (UWES). Escala de Engagement en el trabajo de Utrecht, Ocupational Health Psychology Unit: Utrech University (2003)
3. Da Cuña, I., Soto, M., Gutiérrez, M., Lantarón, E. M.: Relación entre las estrategias de aprendizaje y el engagement académico: una investigación en estudiantes universitarios. Innovagogia, pp. 648–655 (2012)
4. Parada, M., Pérez C.E.: Relación del engagement académico con características académicas y socioafectivas en estudiantes de Odontología. Educación Médica Superior, 28(2), pp. 199–215 (2014)
5. León, J. A., Romero, L. F., Olea, J.: Estudio de validez factorial del síndrome de burnout y engagement en estudiantes universitarios de ingeniería. Alternativas en Psicología, (27), pp. 42–53 (2012)
6. R Development Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing (2016)
7. González, S. B., Figueroa, J.: Discovering Relationships among Personal and Academic Factors with Academic Performance using Association Rules. Research in Computing Science 118, pp. 9–17 (2016)
8. Quintana, M., Hernández, J. E.: Análisis de la Influencia de las Inteligencias Múltiples en el Desempeño Académico Aplicando Técnicas de Minería de Datos. Research in Computing Science 67, pp. 51–61 (2013)
9. Quintana, M., Trinidad, J. C., Morales, S. J., Landassuri, V. M.: Análisis Comparativo de Algoritmos de Minería de Datos para Predecir la Deserción Escolar. Research in Computing Science 67, pp. 51–61 (2013)
10. Gudino-Penaloza, F., González-Mendoza, M., Mora-Vargas, J.: Uso de técnicas de agrupamiento en la clasificación de estilos de aprendizaje. Research in Computing Science 95, pp. 135–146 (2015)
11. Sánchez, A., Cruz-Gutierrez, V., Posada-Zamora, M. A., Torrijos, M. T., Osorio, M. A.: Estudio del análisis de componentes principales en bases de datos de calidad del aire. Research in Computing Science 120, pp. 9–19 (2016)
12. Hernández, R., Fernández, C., Baptista, M.: Metodología de la investigación. 5a ed., México, D.F., México, McGraw-Hill Interamericana (2010)
13. Salanova, M., Schaufeli, W. B.: El Engagement de los empleados un reto emergente para la dirección de recursos humanos. Estudios Financieros 261, pp. 109–138 (2004)
14. Aquije, R., Ramírez, R., Castillo, H.: Guía para la aplicación del análisis multivariado a las encuestas de hogares. Lima, Perú, INEI (2002)
15. García, S., Abad, P., Huapaya, E.: Guía para la presentación de gráficos estadísticos. Talleres de la Oficina Técnica de Administración (OTA) del Instituto Nacional de Estadística e Informática (2009)
16. Guisande, C., Vaamonde, A.: Gráficos estadísticos y mapas con R. Ediciones Díaz de Santos, España (2013)
17. Cuñer, N.: Escala de Inteligencia para Niños de Wechsler, WISC-IV. Diccionario de Psicometría. Montevideo, Uruguay (2013)
18. Nieves, A., Domínguez, F.: Probabilidad y Estadística para Ingeniería, Un enfoque moderno. MCGrawHill (2009)

ISSN 1870-4069

Comparative Study of Learning Strategies of Bachelor Students in Nursing

Mary Arely López Garrido[1], José Alberto Hernández Aguilar[2], Carlos Alberto Ochoa Ortiz[3], Erika Yunuen Morales Mateos[1], Carolina González Constantino[1]

[1] Universidad Juárez Autónoma de Tabasco Cunduacán, Tabasco, Mexico

[2] Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, Mexico

[3] Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Chihuahua, Mexico

arely.lopez@ujat.mx, jose_hernandez@uaem.mx, erika.morales@ujat.mx, alberto.ochoa@uacj.mx, carolina.gonzalez@ujat.mx

Abstract. This study aims to graphically represent a comparison between men and women of the results obtained in the EDAOM of students of Nursing Baccalaureate, for this is selected the newly admitted students in the cohorts, 2010, 2011 and 2012 that answered the EDAOM through the institutional system of tutorials, describes the sample using statistical measures of central tendency, a correlation analysis between the subscale of EDAOM and finally the representation in a radial graph of the results obtained in the EDAOM where each axis represents a subscale with the respective value for men and women. It is concluded that there is a need to strengthen the motivational guidelines where both population samples obtained low results.

Keywords: Learning strategy, Academic performance, Comparative, Radial graph.

1 Introduction

In Mexico the Institutions of Superior Education (IES) have the problem of desertion and lag and as a result student low rate of efficiency terminal, being necessary to increase the performance of the students to reduce the disapproval and the desertion and thereby better results in the indicators of utilization and efficiency terminal. To support the IES the National Association of Universities and Institutions of Higher Education (ANUIES) made the document HIGHER EDUCATION IN THE TWENTY-FIRST CENTURY, within this document is the Program of Integral Development of the students, in which he proposes that the IES implement tutoring programs and integral development with the objective of supporting the students, so that a high portion of them completed their studies in the times scheduled [1], for this reason the IES implemented and operate institutional programs of

tutorials to help you pull down the desertion, disapproval, academic lag and increasing efficiency terminal [2].

Some IES, as part of the operation of the institutional program of tutorials, to the new students answer the inventory of learning styles and motivational guidance (EDAOM), this consists of a self-assessment that the student makes on their learning strategies [3]. The results obtained reinforce the learning strategies.

Learning strategies are the conscious and intentional activities that guide the actions to be taken to reach certain learning goals [4].

Through the student's learning strategies, the causes of low or high academic achievement can be identified [5], in other words, improving learning outcomes can be achieved by incorporating assessment procedures to identify risks and determining strategies that should be encouraged [3].

We have found some studies that have used data analysis techniques to determine which subjects put at risk of dropping to a student due to failure, this information is useful to the tutor to implement learning strategies that support academic performance [6], or through some psychological test determining academic achievement [7]. Other studies employ grouping techniques to classify learning styles [8] or classification rules to know the relationship between academic and personal factors with academic performance [9].

There have been a number of investigations to determine the relationship between the EDAOM and academic performance, such is the case of the study by Gonzalez, Castaneda and Maytorena [10], in which it is evaluated to 229 degree students with the objective of obtaining a structural model for the diagnosis of self-reported strategies in undergraduate students to explain the school achievement. At the University of Magdalena Colombia conducted a descriptive-correlational study of the learning styles and averages academics, the obtained correlations are not significant in determining that there is no statistically significant relationship between the styles of learning and academic performance [11]. In 2013 another study was conducted to determine the relationship between the learning strategies and the academic performance, the study was conducted with 144 students, concludes that there is no relationship between the learning strategies and performance [12].

The different data analyses were performed using various tools. The descriptive analysis was performed with SPSS Statistics 22, obtaining the correlations between the subscale of EDOAM and the radial graph was used R.

This study aims to graphically represent a comparison of the results obtained from the EDAOM of undergraduate students in nursing, and in this way to know that learning strategies need to be strengthened.

2 Materials and Proposed Methods

2.1 Description of the Data

The objective of this study is to perform a comparison of the learning strategies employed by the students of the Nursing Baccalaureate belonging to an institution of higher education

in the southeast of Mexico, a descriptive study was performed, the participants were all newly admitted students belonging to the generational cohorts 2010, 2011 and 2012 so that the sample was not probabilistic or addressed to [13]. This sample was composed in total by 44 students, distributed in the following manner: female gender male 82% and 18%, between the ages of 18 and 30 years.

2.2 Inventory of Learning Styles and Motivational Guidance (EDAOM)

The instrument used was the Inventory of Learning Styles and Motivational Guidance, which allows to identify quickly, the self-evaluations that the students of middle and higher education make of their learning strategies and motivational guidelines to the study. The EDAOM is composed of two sections, used in this study is that of self-reported that measures the self-evaluations of students on the learning strategies and motivational guidelines from the frequency, the difficulty or ease and the quality of the results of the use of these strategies, has 89 Likert-type reagents, corresponding to 4 scales: a) styles for the acquisition of information, b) styles of information retrieval, c)styles of information processing and d) styles of metacognitive self-regulation and meta-motivational, Figure 1 shows the structure of the EDOAM [3]..Below are the strategies that evaluate the EDAOM [12]:

- The Strategies of Acquisition of the information is composed by:
  - Selective. Superficial processing strategies of what is being learned. Important information is selected in the most rapid and complete way, to understand it and to use it in new learning. They include, flipping through and quick readings of materials with the aim of finding the most important, underlining to find information quickly, reviewing when the material is difficult to understand.
  - Generative. They are strategies of deep processing of the acquired information, it adds a symbolic construction of which is being learned. Some strategies are: Analogies to relate concepts, mental mapping, structuring what is read using synoptic tables.
- The Strategies of Recovery of the information learned, allow to: pre-activate, reactivate and keep activated the information to be able to be used later. Some of these strategies are: rereading, generating questions to evaluate what is understood, remembering an image.
  - Faced with different academic tasks
  - During exams
- The Processing Strategies are related to structuring the information learned and to reason or criticize what has been learned. Some of these strategies are: grouping of concepts into categories, relating ideas together, and structuring knowledge through conceptual maps.
  - Convergent style. Reproduce the information to be learned.
  - Divergent style. Create innovative productions and think critically about what has been learned.

- The Metacognitive and Meta motivational Self-Regulation Strategies include activities that allow the student to evaluate, plan, and regulate whether or not he or she has met the learning goals:
  - Efficiency perceived
  - Internal Contingency
  - Perceived autonomy
  - Orientation to external approval
- Learning Task Strategies
  - Orientation to the task itself
  - Orientation to achievement of goals
- Strategy of Learning Materials, as to their usefulness to provide efficient learning.

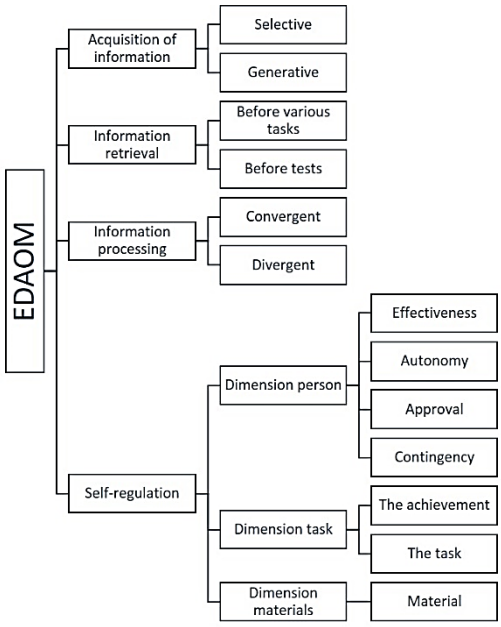


Fig. 1. Structure of the EDAOM.

2.3 Data Analysis

A descriptive analysis was applied to the data for the sample and that same applied a correlational analysis and understanding of the relationship between the subscales of EDAOM.

Finally was represented in a graph radial for a comparison of the averages of the assessment of the EDAOM between men and women.

In the descriptive analysis were applied calculations to the sample of this study, corresponding to the minimum and maximum values, as well as the central tendency statistics: median and mean. With a sample of 44 students of the Degree in Nursing, the results obtained are shown in Table 1, noting that the generative subscales, disaster recovery tests, convergent, contingency, guidance to the achievement, and material have an average greater than or equal to 76(minimum score to get a good result of the EDAOM), the minimum value what have the subscales autonomy and approval. The subscale retrieval of information before tests has a mean of 81.55 and Approval has the lowest score at 42.55.

In the sample are recorded a total of eight men and a total of 36 women, whose percentage values correspond to 18 and 82 respectively.

Table 1. Descriptive statistical sample of students of degree in Nursing.

Scales	Subscales	Media	Medium	Diversion Standard	Min	Max
Acquisition of the Information	Selective	68.59	69.00	12.006	37	91
	Generative	78.95	78.50	9,970	51	100
Retrieval of Information	Before various tasks	71.59	70.00	8.241	47	90
	Before Tests	81.55	83.00	11.619	50	100
Information processing	Convergent	76.73	77.00	13.407	46	100
	Divergent	72.77	71.00	14.595	31	97
Motivational Guidance	Effectiveness	50.86	51.00	10.973	23	71
	Contingency	79.27	79.00	7.456	64	94
	Autonomy	50.23	51.50	12.077	6	71
	Approval	42.55	41.00	16.936	14	86
	The achievement	80.84	81.50	12.844	34	100
	The task	72.80	71.00	10.529	49	94
	Material	78.59	80.00	11.639	57	100

3 Results

3.1 Correlational Analysis

Correlation analysis consists of analyzing data from a sample to know the degree of association or relationship between two or more variables in a population [14]. For known if there are associations between the subscales (Selective, Generative, Information retrieval before tasks, Retrieval of information before exams, Convergent, Divergent, Efficacy,

Contingency, Autonomy, Approval, Homework Orientation, Orientation to goal achievement and materials) of the EDAOM, a correlation analysis was performed between the variables, as shown in Figure 2, the most significant relationships are between: generative subscales and information retrieval during the tests with a level of 0.84, which allows determining that the style of information acquisition is generative in which students use critical thinking to build new knowledge thus facilitating the retrieval of information learned during exams.

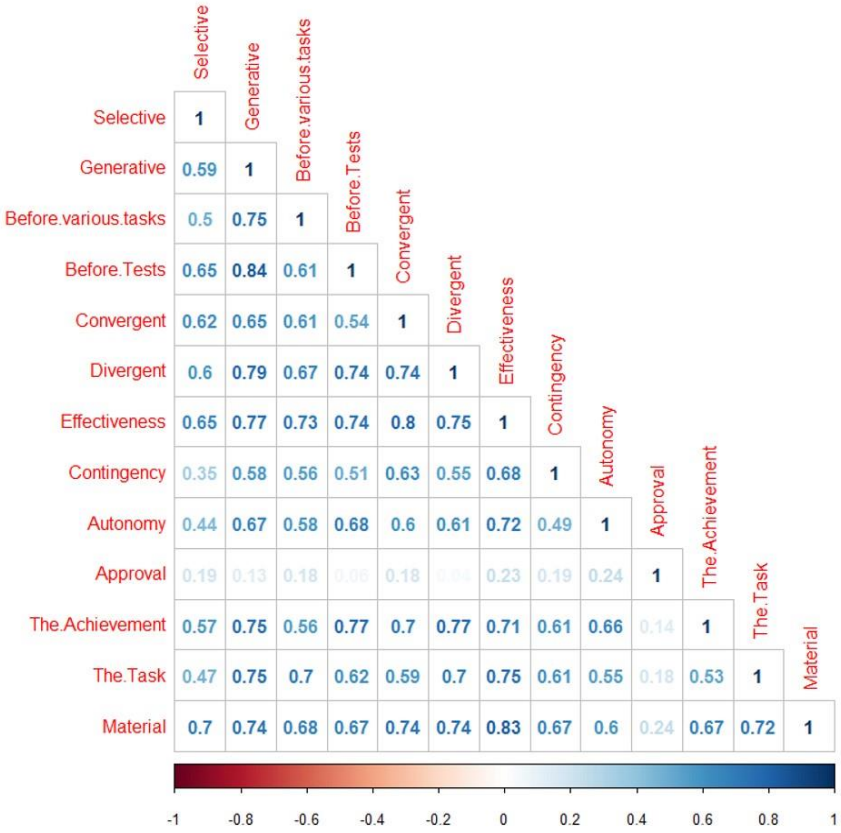


Fig. 2. Correlation of subscale of EDOAM of undergraduate students in Nursing, Source: (Self-realization, 2015).

The relationship between the convergent and perceived efficacy subscales has a level of 0.8, the student considers that it obtains effective results when using the style of processing of the convergent information which consists of reproducing the information learned.

The relationship between the perceived efficacy subscales and the material correlation level is 0.83, the student considers that it has obtained effective results when searching for and adding new material in addition to the one indicated and provided by the teacher.

3.2 Radial Representation

There are studies in which the principal components are analyzed in order to reduce the dimensionality of the data and can be represented and visualized in a more simple way [15]. The radial representation has become an effective option in the support of visualizations of high dimensionality and with large volumes of data, with increased scalability and with a better use of available space [16]. In a radar chart you can make a representation of the differences between the results obtained and the results ideals [17]. Presents from the origin of coordinates the position reached by the different variables of a study, this way can determine how dispersed or concentrated the variables are found [18].

Viewing the comparison of the results obtained from the assessment of the EDAOM of students on the Degree in Nursing is shown in Figure 3 via a radial graph. In this graph is plotted in each axis the result of each subscale of the EDAOM of men, women and the ideal minimum value. Table 2 shows the guide to interpret the results obtained from the EDAOM.

Table 2. Interpretation of EDAOM Guide.

Percentages	Interpretation
100 - 76	Good development of strategies.
75 - 56	Suggested to strengthen the learning strategies or motivational guidance. It is a regular result does not represent a critical failure.
55 - 0	Notes the need to train learning strategies or motivational guidance.

Figure 3 shows the results obtained by the students of the Bachelor in nursing which is compared by men and women and in addition the minimum percentage that indicates that there is a good development of learning strategies.

On the basis of this representation shown below is the description of the results obtained by both men and women and the result required minimum that indicates a good development of strategies:

In the subscale Selective both men and women obtained a result regulate, which indicates that requires strengthening these strategies.

In the subscale disaster recovery tasks indicate that although women obtained a better result those women, both require reinforcing these strategies.

With regard to the processing women get a converged result regulate, requiring strengthening these strategies.

In regard to the subscales Effectiveness, autonomy and approval both obtain a low result, i8ndicando that work is needed in the strengthening of these strategies.

Then described the results toward the comparison between men and women: shows that women employ more the strategy of acquisition of the selective information that is performed by selecting information fast readings more important, understanding literally what they are reading and use this knowledge when necessary in new programming.

In the subscale of information retrieval, women use strategies that allow you to retrieve the knowledge about the different academic tasks unlike men who employ more recovery strategies of the information during the tests.

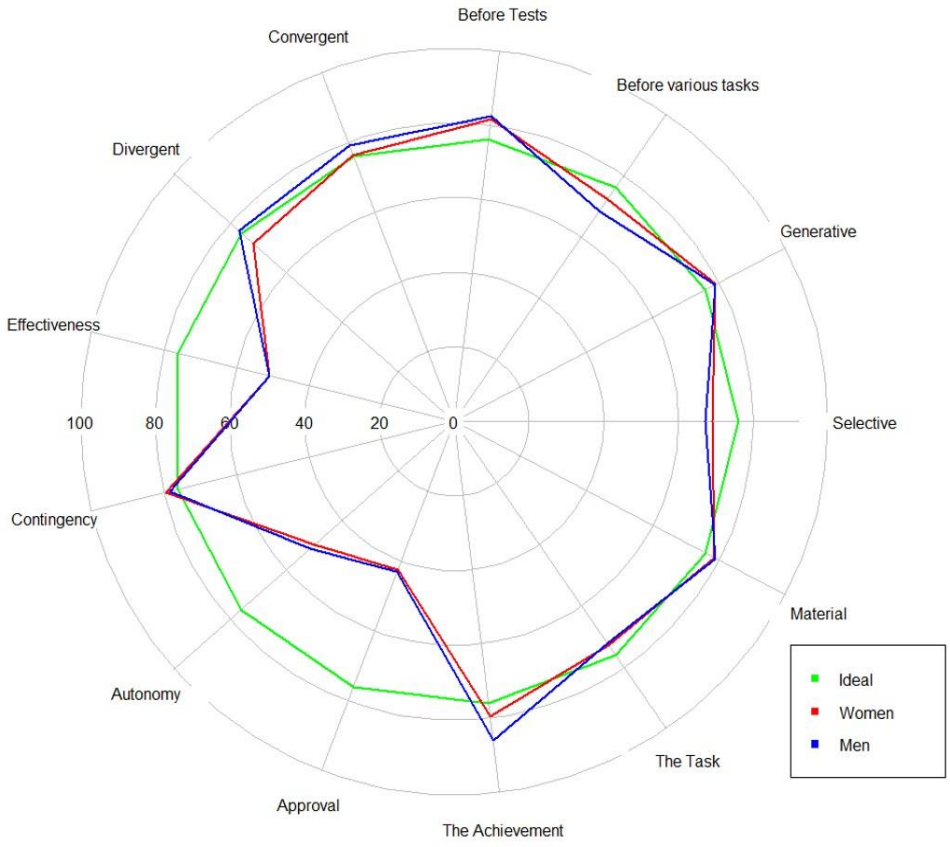


Fig. 3. Radial representation of the subscale of EDAOM of students of degree in Nursing (Source: Self-realization, 2016).

For the scale of information processing the man have a better outcome than women, it is noted that the men were a better result in the converged subscale that is used strategies that allow you to reproduce the information learned; likewise in the subscale divergent its result

is better than that of women indicating that employ strategies that allow you to create and think critically about what they have learned

In regard to the perceived autonomy to achieve their learning men obtained a better result than women.

With regard to the orientation to the achievement of goals that indicates the motivation to learn, men obtained a good result, better than women.

4 Conclusions and Future Work

Currently due to the need to improve and maintain the quality indicator efficiency terminal, which is affected by the lag caused by the disapproval, further studies are needed to know the strategies employed by the students and in this way the academic administration make the best decisions to strengthen if necessary the learning strategies. It is important that elements be used to visualize the results obtained. The implementation of the radial graph in this study allowed to show in a simple way the values obtained in the different subscales and to display the results of men and women.

Among the future work has considered the use of other graphs that let you display in a simple way the Analysis of Multivariable Data.

References

1. Romo, A.: La incorporación de los programas de tutorías en las instituciones de educación superior. Asociación Nacional de Universidades e Instituciones de Educación Superior (2004)
2. Universidad Juárez Autónoma de Tabasco: Programa Institucional de Tutorías (2003)
3. Castañeda, S.: Educación, aprendizaje y cognición. Teoría en la práctica. El Manual Moderno S.A. de C.V, pp. 277–298 (2004)
4. Valle, A., Barca, A., González, R., Núñez, J. C.: Las estrategias de aprendizaje. Revisión teórica y conceptual. Revista latinoamericana de Psicología, 31(3), pp. 425–461 (1999)
5. Beltrán, J. A.: Estrategias de aprendizaje. Revista de Educación, 332, pp. 55–73 (2003)
6. Quintana, M., Trinidad, J. C., Morales, S. J., Landassuri, V. M.: Análisis Comparativo de Algoritmos de Minería de Datos para Predecir la Deserción Escolar. Research in Computing Science 67, pp. 51–61 (2013)
7. Quintana, M., Hernández, J. E.: Análisis de la Influencia de las Inteligencias Múltiples en el Desempeño Académico Aplicando Técnicas de Minería de Datos. Research in Computing Science 67, pp. 51–61 (2013)
8. Gudino-Penaloza, F., González-Mendoza, M., Mora-Vargas, J.: Uso de técnicas de agrupamiento en la clasificación de estilos de aprendizaje. Research in Computing Science 95, pp. 135–146 (2015)
9. González, S. B., Figueroa, J.: Discovering Relationships among Personal and Academic Factors with Academic Performance using Association Rules. Research in Computing Science 118, pp. 9–17 (2016)

10. González, D., Castañeda, S., Maytorena, M.: Estilos de aprendizaje y aprovechamiento en ingresantes universitarios. Revista de Psicología de la PUCP (2000)
11. Barboza, C., Lara, R., Palmezano, A.: Estilos de aprendizaje y rendimiento académico en estudiantes universitarios. (2008)
12. Niño, M.: La Relación Estilos de Aprendizaje y Rendimiento Académico en Alumnos de una Facultad de la UANL. (2013)
13. Hernández, R., Fernández-Collado, C., Baptista, P.: Metodología de la Investigación. Mc Graw Hill (2006)
14. Nieves, A., Domínguez, F.: Probabilidad y Estadística para Ingeniería. McGraw Hill (2009)
15. Manufacturing Terms: Radar Gráfico. (2016)
16. Alvarado, J.: Visualización multidimensional, nuevas técnicas de exploración de datos y representación de información. Revista Memorias, Vol. 8, No. 14, pp. 72–84 (2010)
17. Martínez, L., Pérez, M.: Análisis de Estrategias de Gestión: aplicaciones polivalentes del gráfico radial. (2011)

ISSN 1870-4069

Intelligent Tutoring Systems based on Virtual Reality for the Electrical Domain

Miguel Pérez-Ramírez[1], Norma J. Ontiveros-Hernández[2], Carlos A. Ochoa-Ortíz[3], José A. Hernández-Aguilar[4], Benjamín E. Zayas-Pérez[1]

[1] National Institute for Electricity and Clean Energies, Cuernavaca, Mexico

[2] Tecnológico Nacional de México/Instituto Tecnológico de Zacatepec, Zacatepec de Hidalgo, Morelos, Mexico

[3]Autonomous University of Juarez City, Ciudad Juárez, Chihuahua, Mexico

[4] Autonomous University of Morelos State, Cuernavaca, Morelos, Mexico

{mperez, zayas}@iie.org.mx, norma.oh@itzacatepec.edu.mx, alberto.ochoa@uacj.mx, jose_hernandez@uaem.mx

ABSTRACT. In this paper a model to integrate the technology of tutoring systems to existing virtual reality training systems is described. The training systems are devoted to dangerous activities for energized line maintenance in the electrical domain. The systems are already in use and have impacted positively the traditional training method, which was based on classroom classes and camp training. The system augmented the availability of specialized knowledge and improved transfer of training. Nevertheless, we aim to provide further training capabilities with the salient proprieties of intelligent tutoring systems.

Keywords: Virtual reality, Intelligent tutoring systems, Training, Dangerous tasks.

1 Introduction

Power live-line maintenance refers to the work done by linemen on energized lines to carry out power restoration and to prevent power interruptions. Linemen must be well trained, since this high-risk task not only must be executed rapidly to keep time of outages as short as possible, but at the same time prevent accidents and avoid loss of human lives. Tradi-

tionally this kind of training at CFE (Spanish acronym for Federal Commission of Electricity), is accomplished through a combination of theoretical teaching at classroom and practical sessions. The former is given at job place; the latter requires workers to be moved to training camps. Although this model has been effective, there is the need to enhance training, reduce cost and provide training to a larger number of linemen. Based on previous research and application of virtual reality (VR) in the field of training [1] [2] [3], a complementary training tool was implemented. Rather than substituting the traditional method, the training tool aims to complement it and improve it, by tackling economical costs due to travel and stay expenses and making training available to a larger number of workers.

We have developed different training system devoted to maintenance of energized lines, such as ALEn^{3D} MT (for medium tension energized power lines - MTEPL), ALEn^{3D} AT (for high tension power lines [1]) and ALEn^{3D} LS (for underground power lines). We have developed also training systems for performing electrical tests to primary equipment of distribution substations (3DMaPPS) and for testing protections of distribution substations (SAMPyM3D). These VR training systems are used by CFE as tools to train high risk activities within safe virtual environments.

Even though these systems have improved the traditional training method, a new demand has arisen to integrate the well-known technology of Intelligent Tutoring Systems (ITS) to these training systems; so that, they can exhibit an intelligent and personalized behavior and they can be used to improve massive training. An ITS is conceived as a computer software designed to simulate a human tutor's behavior and guidance, but for the time being, our VR systems are only able to keep records of the progress of the students and the different courses and instructors but they do not make any decision regarding the delivery of the instructional content to specific students. Thus far, they only make sure that students learn a correct sequence of steps in a maintenance procedure (MP) and consider all the safety regulations on each step. Here we discuss about the integration of ITS technology to these systems.

The rest of the paper is organized as follows: Section 2 includes some related works. Section 3 includes the architecture of our VR training systems and how they are going to be extended to include ITS. Section 4 discuss about details of the integration of ITS. Section includes some conclusion and finally a list of references is included.

2 Related Work

Integrating the technologies of VR and ITS, is not new. This integration is approached as the building Intelligent Interactive Learning Systems (IILS) [4]. A more generic approach conceives this integration as combining artificial intelligence techniques with those of virtual environments to produce intelligent virtual environments [5]. Although putting together these two technologies in a system is considered promising, it is also considered a difficult

task since sometimes is claimed that the modular structure of an ITS clashes with the interactivity and the structure of a VR system [6]. However, a classical source of IILS, even introduces Steve, an animated pedagogical tutor who inhabits a virtual environment, monitoring it and controlling it through virtual motor actions [7]. Other applications involving pedagogical agents are reported, within the firefighters' domain [8], circuit analysis [9] and even trying to influence students' affective states by using virtual agents [10]. The USA army is a pioneer in the integration of these two technologies [11], this time they are focusing the team training by using ITS based on VR. They are doing research to understand how to select a team instructional strategy, as well as finding a balance between individual and team feedback.

Despite technical burdensome that might implicate the integration of these two technologies, the combination provides exciting opportunities for the learning of complex skills. On one hand, VR provides different benefits as a learning tool, such as visualization, active learning based on interactivity and 3D navigation among others [12]; on the other hand, ITS might contribute with some degree of flexibility to adapt instruction, and thus individualize instruction [13].

Whereas the literature of ITS is rich, what is not common is to find VR training systems (VRS) devoted to maintenance to energized power lines as is the case of our training systems. We believe that it is quite viable to add some intelligence to our VRS as is argued below.

3 Architecture of the VR Training Systems

All the systems described above have the same architecture, which include the following main modules (Fig 1.):

- *VR maintenance procedures module* (VRM). This is the main module of the system; it is in charge of management of the instructional content. Maintenance procedures are delivered in terms of 3D animations complemented with environmental audios and textual and audio explanations. This module records the progress of students in two modalities. When the user is attending a classroom course, the system is connected to a database located in a web server to record information of users, courses and trainees' progress, and when the systems is used offline for self-learning, this progress is recorded in a local database.
- *Users and courses management module* (UCMM). This module allows the creation of courses where instructors can register students. It also integrates, manages and generates reports of courses, users and users' progress, which include results of practical and theoretical evaluations.
- *Licenses module* (LM). In order to permit access to the system, a user license is needed. The license management module requires user's personal information, company adscription and computer processor serial number to control access to the system.

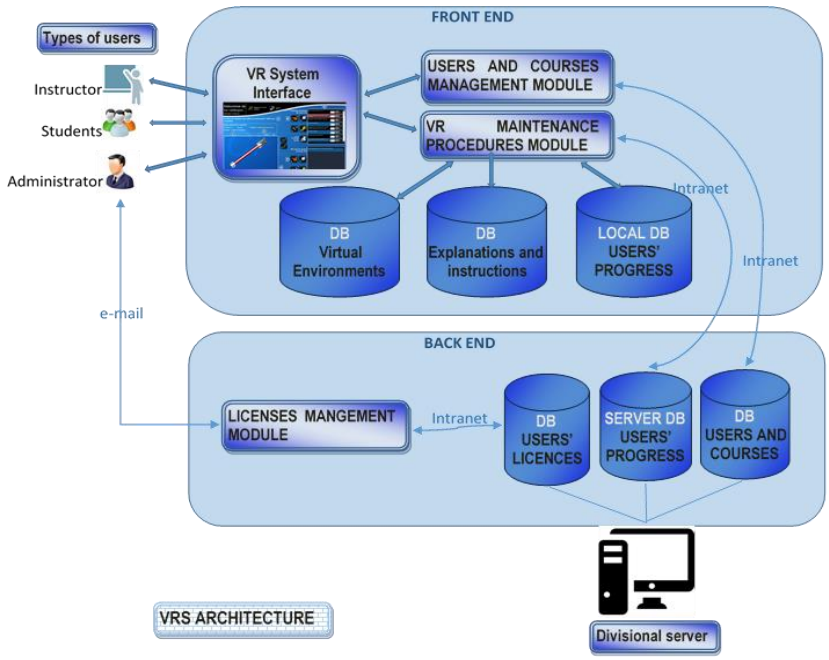


Fig. 1. Architecture of the virtual reality training systems.

3.1 Operation of the System

The operation of the systems is as follows. Once the system is installed, the trainees have to request a license, which is granted by a licenses administrator. Without license, the trainees are only able to see one maintenance procedure, just to get familiar with the system. The license is valid only for the personal computers where the system is installed. Having the license, the system is activated and the user will be able to see all the 43 maintenance procedures. The system will be able to recognize three types of users with different privileges, administrator, instructor and students. Administrators will have all the privileges to manage the system; they are usually computer science specialist who help users to install the system. There will be an administrator for each of the sixteen distribution divisions of CFE. The instructors will be able to create courses, add the student allowed to take the courses and the maintenance procedures they are going to teach on each specific course. They will use the system as a training support tools when they are training new students.

Here the VRS is connected to the CFE´s intranet so that the progress of the students is recorded in a database which can be used to generate different reports and exams. Finally, the students can use the systems in two ways, namely: a) in a classroom course programmed

by the company and guided by an instructor. Here the progress of the students can be monitored by instructor and even the administrator. These courses are scheduled by the company as part of its training programs, and b) as a self-learning tool, here the students are free to learn, practice and review any maintenance procedure they wish. The progress of the student here is recorded only in a local database which cannot be monitored by the company but only by the students themselves. The student can progress to their own pace and spend the time they wish.

3.2 Functionality of the system

The functionality of the systems has three modes namely: a) learning, b) practice and c) evaluation (Fig. 2).

Fig. 2. Main menu of our VRS for electrical tests to substations.

In the learning mode, the system takes the control and teaches each MP taking the student through the whole sequence of steps which composes the MP until it is completed. The student cannot skip any step, as the intention is to learn the correct sequence, since neglecting a step or skipping it might be translated in terms of accidents. In the practice mode, the students are allowed to jump from one step into another so that they appreciate specific details of a specific step or sequence of steps.

In the evaluation mode, there are two different types of evaluation: theoretical or practical. In the theoretical evaluation, the system allows instructors to generate exams by selecting multiple choice questions from a questions database and then are marked automatically by the system once the students answered the exams.

In the practical exam, the students are required to perform a specific MP within the virtual environment (VE) but with no help provided by the system. The VE is the same for both, theoretical and practical modes. In this mode, all mistakes as well as the progress made by students are recorded in a database.

Finally, a MP is presented in two parts. In the first part, the student must select all the tools, materials and safety gear needed to perform a MP (Fig. 3). In real work tasks, this activity corresponds to a check list to make sure all the necessary equipment is ready and in working order before moving to the site where the MP will be realized. In the second part, the VRS presents the students the sequence of steps of the MP that the student has chosen (Fig. 4).

Here, we have to emphasize that the systems only present the sequences of the MP but do not take any decision on what it should delivery and when, regarding the instructional content. It lacks the intelligence owned by ITS.

Fig. 3. Selecting tools, materials and safety gear needed to perform a MP.

Fig. 4. Virtual Environment of a step of a specific MP.

4 Requirements for VRS Systems Architecture

The point here is that we want to add to our VRS some intelligence by integrating ITS technology. Even though there exist different types of ITS, it is widely accepted that an ITS provides individualized tutoring or instruction by including three components [14;15]: a) knowledge of the domain b) knowledge of the learner and c) knowledge of teacher strategies (Fig. 5).

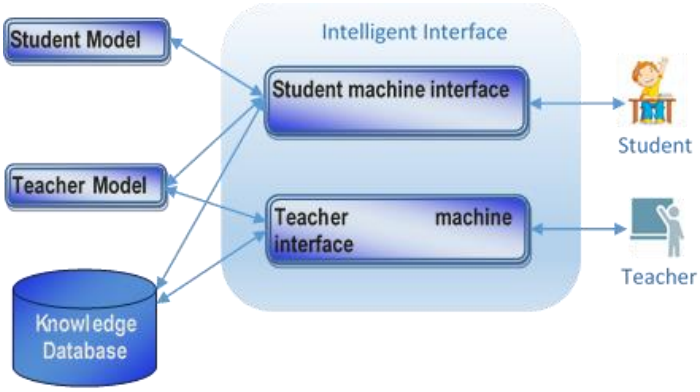


Fig. 5. Basic Structure of Intelligent Tutoring System [17].

Now personalized instruction means that the ITS must be able to do the following [15]:

- Accurately diagnose students' knowledge structures, skills, and styles.
- Diagnose using principles, rather than preprogrammed responses.
- Decide what to do next.
- Adapt instruction accordingly.
- Provide feedback.

A student learns from an ITS by providing solution to a situation or problem (Fig. 6), then the system compares the student solution with the expected one, the differences allow the ITS to update the student model and decide what kind of feedback should be delivered to the student. This is repeated until the expected solution and the one provided by the student are the same [15].

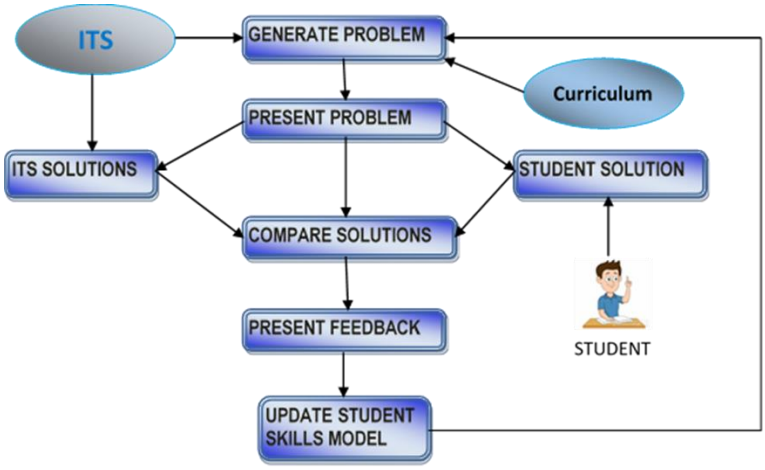


Fig. 6. How an ITS works [15].

We need to analyze and determine what changes we have to introduce in the original architecture of the VRS. In fact, analyzing our VRS, we found out that they already possess some of the components of an ITS.

1. *Knowledge of the learner*. For instance, the VRM already records the progress of the students in the users´ progress database, this is part of the ITS component, namely "knowledge of the learner". Here the VRS records different kinds of information such as the maintenance procedures (MP), that each student has learnt, but also we have more detailed information, for instance we know where each student made mistakes when

learning. The VRS also records outcomes of two types of exams, namely, a) theoretical, consisting of a list of multiple choice questions, that the VRS marks automatically, thus, here we have information of correct and incorrect answers of each student, and b) a practical exam, where the student perform a MP by his own, without any help provided by the system, here any mistake is recorded, we have records of what a student was requested to do by the system, and what the student did, wrong or correct, this recording in achieved on each step and sub step of a MP.

All this information can be helpful to know each student's academic progress, weaknesses and strengths. That is, some information already exists but it needs to be exploited.

2. *Knowledge of the domain*. We have a rich database full of knowledge of different aspects of the electrical domain on each VRS. This knowledge is stored in term of 3D animations, informative text and audio, again on each step and sub step of every MP. This knowledge came from two sources: a) specialized documents of the field and b) the expert knowledge of at least 10 experts with 20 years or more of practical experience on the field. This team of experts validated each step of each MP and they even introduced in the systems the so called "additional information," where they provided a piece of specialized knowledge, which in some cases is not described in documents, it rather came from all their years of experience in the field, that is to say, it is empirical knowledge owned by the experts.
3. *Knowledge of teacher strategies*. All the VRS were planned to follow the multidimensional approach to learning developed by Pérez and Ontiveros in 2009 [12]. The main idea in this model is to recognize that there are different factors or dimension that intervene in specific leaning processes. This includes the selection of learning theories such as Constructivism, active learning, behaviorism, and so on, providing feedback to different learning channels, even the affective aspect can be considered [16]. For the time being, the multidimensional approach is already followed by our VRS, but it is open to be enriched by integrating other elements in order to meet the requirements of an ITS. It is worth to mention that VR technology is quite flexible to create leaning contexts as has been pointed out by [12].

5 Proposed Architecture for Our VR Training Systems

Two main things need our VRS in order to get some intelligence: a) the VRS need to use the information of the student and the domain, they already posse in order to work as an ITS; and b) the integration of teaching strategies is needed. Once the VRS are complemented with these two components, they need to behave as an ITS as depicted in Fig. 6, this should still have to be programmed. Thus the new architecture of the VRS or rather Smart VRS (SVRS) would look like in Fig. 7.

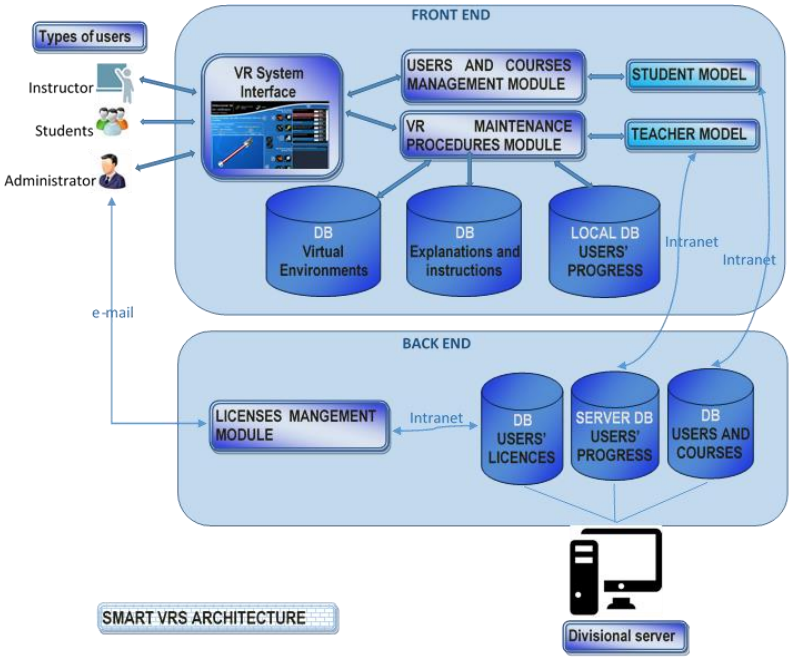


Fig. 7. Architecture of a Smart VRS.

As we can be observed, the front end and the back end of the VRS remain as they are now, the intelligence will be provided by the two new modules: the student model and the teacher model. These will operate over the data already available and will provide the strategies to exhibit an intelligent behavior when teaching. The student module might work with the data already stored in the databases, although their structure might also be modified or adapted in order to harbor any additional data required by the two new modules. There is a key fact that the teacher model needs to consider, which is relevant in the dangerous application domain, all the MP or electrical tests that they include in the domain of the VRS must be a sequence of steps. Students must learn not only the technical details of the MP but also de sequence.

We have mentioned that skipping steps magnifies the probability of accidents. For instance, a MP in energized line follows the following general structure: a) climbing up the structure (power pole), so that on each step all reference points (energized points), must be covered using isolated blankets and different kinds of covers so that the structure gets fully isolated; b) realize the objective of the MP, here again on each step the lineman must make sure that all reference points are not unnecessarily uncovered and finally, c) climbing down the structure, here all covers and isolated blankets used, are removed in the reverse order in which they were collocated in the structure, so that all the time the lineman is working on a

covered (isolated) section of the energized structure. Thus, skipping a step might mean neglecting or leaving uncovered a reference point, which would be simply a suicide. This is why the sequencing is vital in all MP or electrical tests. Thus, the teacher model needs to pay attention to sequencing of MP and provide known teaching strategies so that students are able to learn effectively these dangerous tasks.

6 Conclusions

We have analyzed and presented a modified architecture for our VRS, so that they integrate ITS technology. This endeavor is technically viable since the original architecture remains and the two new modules can be integrated without great impact to the original architecture. It has been emphasized that the teacher model must care about the sequencing of MP for safety sake.

An additional benefit that we can foresee is that the proposed architecture would apply for all our VRS since they were built under the same architecture and so they work in the same way, the main difference among them is the instructional content. Elsewhere we have also explored other possible features such as introduction of animated agents [18] and affective computing [16] that together with the technology of ITS might provide solid bases to build a second generation of VRS according to the technology roadmap delineated for VR in CFE processes [19].

References

1. Stanney, K.M.: Handbook of Virtual Environments, design implementation and applications, Lawrence Erlbaum Associates Publishers, Mahwah, NJ (2002)
2. Burdea, G.C., Coiffet, P.: Virtual Reality Technology, 2nd ed., Wiley-IEEE Press, New Brunswick, NJ (2004)
3. Pérez-Ramírez, M., Zabre-Borgaro, E., Islas Pérez, E.: Prospectiva y ruta tecnológica para el uso de la tecnología de realidad virtual en los procesos de la CFE. IIE, Cuernavaca, México (2004)
4. Lopez-Garate, M., Lozano-Rodero, A., Matey, L.: An Adaptive and Customizable Feedback System for Intelligent Interactive Learning Systems. In: Proceedings of Intelligent Tutoring Systems: 9th International Conference on Intelligent Tutoring Systems, ITS 2008, Montreal, Canada (2008)
5. Luck, M., Aylett, R.: Applying artificial intelligence to virtual reality. In: Intelligent virtual environments, Applied Artificial Intelligence, 14(1):3–32 (2000)
6. Imbert, R., et al.: A Multiagent Extension for Virtual Reality Based Intelligent Tutoring Systems. In: Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies, Niigata, Japan (2007)
7. Rickel, J., Johnson, W.L.: Intelligent Tutoring in Virtual Reality: A Preliminary Report. In: du Boulay, B., Mizoguchi, R. (eds.) Artificial Intelligence in Education. Proceedings of Eighth World Conference on AI in Education, Kobe, Japan, IOS Press, Amsterdam, p. 294 (1997)

8. Buche, C. et al.: MASCARET: pedagogical multi-agents systems for virtual environment for training. In: Proceedings of the International Conference on Cyberworlds, Singapore (2003)
9. Butz, B.P., Duarte, M., Miller, S.M.: An intelligent tutoring system for circuit analysis. IEEE Transactions on Education, 49(2): 216–223 (2006)
10. Frasson, C., Chalfoun, P.: Managing Learner's Affective States in Intelligent Tutoring Systems. In: Advances in Intelligent Tutoring Systems, series Studies in Computational Intelligence, vol. 308, pp. 339–358 (2010)
11. VRAC: Team Intelligent Tutoring Systems. Virtual Reality Applications Center. http://www.vrac.iastate.edu/research/team-intelligent-tutoring-systems/ Accessed 6 Mar 2017 (2017)
12. Pérez-Ramírez, M., Ontiveros-Hernández, N.J.: Virtual reality as a comprehensive training tool. In: Artificial Intelligence & Applications, Guanajuato, México: SMIA (2015)
13. Chad Lane, H., Johnson, W.L.: Intelligent Tutoring and Pedagogical Experience Manipulation in Virtual Learning Environments. In: Cohn, J., Nicholson, D., Schmorrow, J. (eds.) The PSI Handbook of Virtual Environments for Training and Education, Estport CT, Praeger Security International (2009)
14. Educause: 7 things you should know about Intelligent Tutoring Systems. https://net.educause.edu/ir/library/pdf/eli7098.pdf. Accessed 26 Feb 2017 (2017)
15. Thomas, E.: Intelligent Tutoring Systems (ITS). SDSU Dept. of Educational Technology. In: The Encyclopedia of Educational Technology. Bob Hoffman (General Editor), San Diego State University, http://www.etc.edu.cn/eet/eet/articles/tutoringsystem/start.htm. Accessed 10 Mar 2017 (2017)
16. Ontiveros-Hernández, N.J., Pérez-Ramírez, M., Hernández, Y.: Virtual Reality and Affective Computing for Improving Learning. Research in Computing Science 65, pp. 121–131 (2013)
17. Song, B. Yan, X.: Intelligent Tutoring System of Primary Mathematics Based on EGL and Web 2.0. Software Engineering 4(2):34–40 doi: 10.11648/j.se.20160402.15 (2016)
18. Hernández, Y., et al.: Designing empathetic animated agents for a b-learning training environment within electrical domain. Journal of Educational Technology & Society 19(2):116–131 (2016)
19. Pérez-Ramírez, M.: Ruta tecnológica de realidad virtual para el sector eléctrico. Komputer Sapiens 1(3): 1–16 (2011)

Impreso en los Talleres Gráficos
de la Dirección de Publicaciones
del Instituto Politécnico Nacional
Tresguerras 27, Centro Histórico, México, D.F.
noviembre de 2017
Printing 500 / Edición 500 ejemplares