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 147, No. 2,** febrero 2018. 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 147**, **No. 2**, February 2018. Printing 500. The authors are responsible for the contents of their articles. Printed in Mexico City, in the IPN Graphic Workshop – Publication Office.

Volume 147(2)

Hybrid Intelligent Systems

José Alberto Hernández Aguilar
Julio César Ponce Gallegos
Edgar Gonzalo Cossío Franco
Carlos Alberto Ochoa Ortiz (eds.)

Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2018

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 prestigious journal "Research in Computing Science" presents selected papers that discuss Hybrid Intelligent Systems (HIS) and its applications. Papers were carefully chosen by the editorial board on the basis of the at least two blinded reviews by the members of the reviewing committee and additional expert reviewers. The criteria taken into account were: 1) originality, 2) scientific contribution to the field, 3) soundness, and 4) technical quality of the papers. It is worth noting that 50% of received papers for this special issue were rejected.

The volume contains 11 papers covering various aspects of HIS in the fields of 1) Simulation and Smart Cities, 2) Knowledge Discovery Database, and 3) Metaheuristics and bioinspired algorithms. All selected papers discuss hot topics relevant for the academic and scientific community, due its potential applications in a wide variety of organizations.

Five papers are about simulation and Smart Cities: The first one presents a prototype position steering wheel control (SWC) designed for its implementation in cars that do not have this system, genetic algorithms (GA) are used to tune a PID controller connected to direct current (DC) motor for SWC that uses orientation of lines on the road. Second paper present an approach to solve the capacitated vehicle routing problem (CVRP) in a share bicycle system (SBS), besides construction of instances paper shows how Matlab was applied in the formulation of solution. Third paper presents the design and implementation of a CVRP simulator, authors propose the use of a genetic algorithm (GA) for the optimization of route and a three-tier web-based system to upload a CVRP instance and the selection of a Metaheuristic. For testing the simulator, ten different .vrp instances were processed with GA and compared it with the nearest neighbor algorithm (NNA) solutions. Fourth paper discusses the systematic review of papers centered in technological research adapted for the access of the information and the communication of people with hearing loss or deaf; under Kitchenham´s methodology, 350 papers published since 2013 until May 2017 were analyzed in order to create a technological-social model to promote the improvement of communication between deaf and hearing people. In paper fifth, authors discusses concepts and tools from Smarter Cities, Smart Contracts, and Blockchain, and presents a proposal to apply these tools in the daily operations of International Business Machines (IBM)'s supply chain to reduce transaction times radically by implementing agile practices.

KDD (Knowledge Discovery Databases) is another hot topic of discussion in HIS. The first paper of this section discusses the use of visualization techniques for the representation of multivariate analysis applied for study of student engagement at universities of careers in technology in the south of México, the instrument used in this research was the UWES-S (Utrecht Work Engagement Scale for Students). The

second paper aims to identify the learning strategies employed by university students; and represent them through graphic techniques of multivariate analysis; students are from the Administration and Administrative Computing careers. The instrument used was the Inventory of Strategies of Learning and Motivational Orientation (EDAOM). Third paper presents an Intelligent Tutoring System (ITS) that uses speech recognition in Spanish language to contribute to the study of the networks communications, paper discusses alternative and innovative techniques of human computer interaction (HCI) and the recognition with learning features on the use and details about internetworking terms.

Metaheuristics and Bioinspired algorithms are another fashion topic. The first paper of this section proposed a fuzzy inference system that allows planning and scheduling production using a Mamdani inference system tuned with a Genetic Algorithms, in this paper is proposed the use of fuzzy logic as an alternative for classical methodologies that require complex mathematical models or high precision data, that is hybridized with Genetic Algorithms (GA) for adjusting the required parameters. Second paper discusses the implementation of a heuristic algorithm for drone swarm auto-organization with the purpose of applying for wildfire alert and detection. Finally, third paper presents a proposal for the resolution of the assignment schedule problem (ASP) in the educational institutions by using the metaheuristic Gray Wolf Optimizer (GWO), this paper presents, as a use case, its application to a public university, a private university and a higher education institute.

We would like to thank Mexican Society for Artificial Intelligence (Sociedad Mexicana de Inteligencia Artificial), and MICAI 2017 Committee for all the support provided for the publication of this special volume.

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

José Alberto Hernández Aguilar
Julio César Ponce Gallegos
Edgar Gonzalo Cossío Franco
Carlos Alberto Ochoa Ortiz
Guest Editors
Mexico, February 2018

ISSN 1870-4069

Table of Contents

Page

ISSN 1870-4069

Steering Wheel Control in Lane Departure Warning System

Maximiliano Hernández García Rojas[1], Humberto Velasco Arellano[1], David Ubach González[2], Martín Montes Rivera[1], Marving Omar Aguilar Justo[3]

[1] Universidad Politécnica de Aguascalientes, Ciencias en Ingeniería, Aguascalientes, Mexico

[2] Universidad Politécnica de Aguascalientes, Ingeniería en Electrónica, Aguascalientes, Mexico

[3] Universidad Veracruzana, Veracruz, Mexico

{mc160006, mc150008, up140172}@alumnos.upa.edu.mx, martin.montes@upa.edu.mx, marvaguilar@uv.mx

Abstract. Advanced driver assistance systems (ADAS) are important to prevent road accidents, one of those are lane invasions. Nowadays optical and control systems are applied in new cars to maintain vehicles centered in a lane, nevertheless old vehicles cannot access to these new assistant systems. In this paper we present a position steering wheel control (SWC) in a lane detection prototype designed for its implementation in cars that do not have installed this system. The proposed technique uses a self-tuning PID controller connected to a DC motor for SWC that uses orientation of lines on the road. Lines are detected by applying Sobel filter convolution and Hough transform, which are image processing techniques that allow obtaining orientation. The PID controller is automatically tuned since this assistant would be installed in different cars. GA is used since it has shown good results parametrizing well known structured problems without requiring mathematical models or techniques that demand control experts. The main contributions of this research are a SCW alternative capable of been installed in any old vehicle and the algorithm implemented for using orientation as input data as the control reference with a self-tuning PID controller.

Keywords: steering wheel control, genetic algorithm, advanced driver assistance systems.

1 Introduction

In the car industry, one of the most important issues is user safety. In order to reduce current accidents and protecting the life of drivers, researchers have been looking for

driver assistants. The most advance of them evaluate the traffic situations, using sensors or cameras, and support the driver to avoid damages [16].

The research for improving intelligent vehicles using advanced driver assistance systems (ADAS), includes the lane departure warning (LDW) system, which alarms the driver when the car is out of the line markings and apply the brakes to help him to move the vehicle [23].

It took the tool of LDW to be applied in not manned vehicles to control the direction and management across the real world to improve the vehicle interaction, for this reason was important to guaranty a position steering wheel control (SWC), to modify the vehicle trajectory closing the control loop in the LDW.

In this work, we propose an SWC based in orientation of lines on the road, where an automatic PID controller is tuned by a Genetic Algorithm (GA) that adapts the controller to any vehicle without deep knowledge in the mathematical background.

Self-tuning PID controller is required, because of the variation when changing different vehicles, is here where de adaptability and evolution in a GA allows getting a tuning method for any car [20].

The selection of a GA as tuning method it is the advantage that represents at the moment to find results in an optimization problem, where the solution presents a nonlinear behavior and in a multi-objective solution, gives a robustness method [20].

2 Related Works

In the last years, the LDW has been investigated, and some important contributions explored the safety field in science, like the investigation in [15], reporting comparison between the line detection system in a virtual video and a real video which proved that the real video and the natural noise increase the effectiveness in detection.

In 2015 was report a comparison between two different ways of alarm in LDW, one activating the breaker and the other modifying the steering wheel position, they discussed that the difference in simulated experiments was no profitable [4], at the time was propose the LDW using Hough transform, reporting a 3% of false alarm, reducing the standard in that year [10]. In the following years the investigation focus in use of different filters [16, 23, 14], the application of observers [7, 21] and effectiveness evaluations [24], but in all cases, the application of LDW is just for alarm and to our best knowledge is no suitable for non-manned vehicles, on the other hand, the SWC patents describe the application and uses [1, 11, 13], where the application most significant is in a truck where was consider velocity of reaction to avoid insures [12], the review in [3], supports that non-manned vehicles have enough study with this technique.

Other works focuses on testing different lines filtering or edge detection techniques, as well as different types of line identification techniques [8, 13, 14], but only generates an alert to the driver as an output.

Works such as [1, 3, 11], use automatic steering wheel control systems on vehicles, but these are applied on modified vehicles and cannot be applied on other vehicles without a modification.

In this paper we have included a way that helps to optimize the control processes with genetic algorithm, this is proposed because this technique has presented important contributions in diverse areas [18], for example: reduction of electric circuits [6], recovering of unknown parameters in oscillators [9], multi criteria decision taking [22], and in the control area, has been show important results like the use of a genetic algorithm for obtaining the PID parameters of a electrohydraulic servo control system [2].

The research in [17], follows the optimization of PI gains in a controller and compares the differences between the results of a traditional genetic algorithm and a proposed micro genetic algorithm.

3 Theoretical Framework

Control systems are used to maintain process conditions at their desired values by manipulating certain process variables to adjust the interest variables [5]. Tuning method is an indispensable step, when the controller is adjusted to have a required behavior for the interest variable, there are several methods for adjusting the conventional PID controller but in this paper the Ziegler and Nichols method will be used, since it is one of the most popular tuning methods [19].

3.1 Ziegler and Nichols

Ziegler and Nichols proposed some tuning rules for PID controller based on an experimental response, their main objective was to adjust the controllers without knowledge of the mathematical representation of the systems [19]. The result was the definition of two methods, in this case, it is explained the ultimate gain because it will be compare with the results achieved with a GA.

3.2 Ultimate Gain

This method is used when the system presents a second or greater order response, in that case the PID controller parameters are set for eliminating the integrative and derivative effects, then the proportional gain is increased in order to generate a sustain oscillation, at this point the value of gain is named critical gain, and de period that exists in the oscillation is name critical period. Then the P, I and D gains are calculated as in Table 1.

Table 1. Tuning values for ultimate gain method.

Controller	k_p	τ_i	τ_d
P	$0.5\ K_{cr}$	∞	0
PI	$0.45\ K_{cr}$	$0.5\ P_{cr}$	0
PID	$0.6\ K_{cr}$	$0.5\ P_{cr}$	$0.125\ P_{cr}$

3.3 Digital Control

The control application is a studied science with too many branches, one of those is the digital control, and it has a big impact at moment to use digital circuits in computing, just hire the characteristics of this control need an indispensable a well defined sample period to guaranty the constant function, because it does not matter if the system works as fast as possible, the behavior is no continued [8].

The important advantages of digital control are the high flexibility and the capability to control a greater number of systems and optimize the time and space that is consumed by the controller, but it represents a higher price and is required special knowledge to manipulate them [19]. A general block diagram of a digital control loop is shown in Fig. 1, where is possible to see the transformation of the signals to be modified by a digital controller.

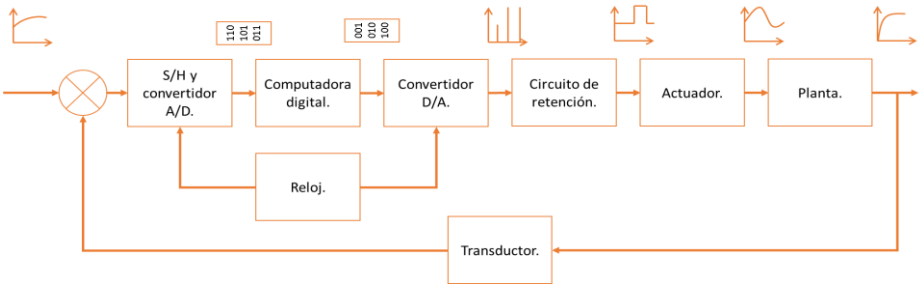


Fig. 1. Digital control close loop [16].

3.4 Genetic Algorithm

Evolutionary algorithms are population-based metaheuristic optimization algorithms that use biology-inspired mechanisms, in order to refine a set of solution candidates iteratively, the cycle of evolutionary algorithms follows the Darwin laws that are described in Fig. 2 [25].

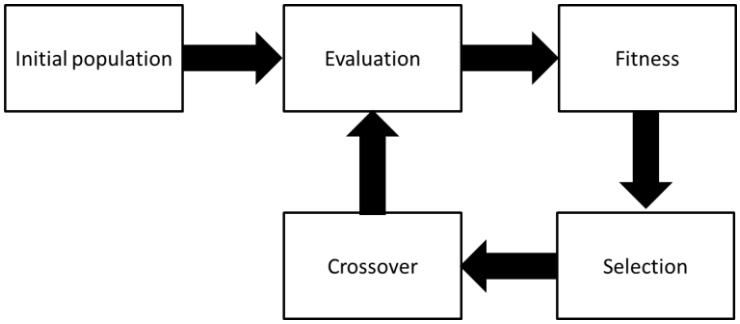


Fig. 2. Evolutionary Algorithms behavior.

As is shown in Fig. 2, the operations in a genetic algorithm have a defined process, first the initial population is random initialized for obtaining a high diversity of possible solutions, the evaluation consists in the analysis of each individual by submitting it in the process, in order to measure its quality for solving the problem. The fitness is the assignation of a numerical value depending on the behavior presented. The selection is a process where 2 individuals are selected according to its fitness and finally, the crossover take that 2 individuals to generate new members combining their genes [25].

4 Methodology

The methodology explains how it is implemented the control to get important results that could contribute measuring the differences between a conventional controller and a GA tuning method to guaranty the control and its adaptation in different cars for the SWC and the LDW systems.

4.1 Control

Ziegler and Nichols tuning control method is applied in the digital PID controller described in equation 1:

$$U\,(z) = \; Kp\left[1 + \frac{T}{ti\,(1 - z_{-1})} + \frac{td(1 - z^{-1})}{T}\right], \tag{1}$$

which is represented as the transfer function for the GA like in equation 2:

$$\frac{U\,(z)}{E\,(z)} = a + \frac{b}{1 - z_{-1}} + c(1 \; - \; z^{-1}), \tag{2}$$

where the auxiliary variables are described in equation 3:

$$a \; = \; Kp, \;\; b = \frac{KpT}{ti}, \;\; c = \frac{Kptd}{T}. \tag{3}$$

The input signal is the error between reference angle and the position angle of the motor, and the output is the duty cycle of a PWM that rotates the motor.

To determine the rotation direction of the motor, the sign of the output is used, if negative then turns to the left and if positive it turns to the right. Then the absolute value is used as duty for a PWM output.

Once the controller structure it is built, it is necessary to adjust the gains. For this propose the Ziegler-Nichols method of critical gain for systems with sustained oscillations was used. This method allows obtaining the gains k_p, τ_i and τ_d for the PID controller without knowing the model of the plant.

First, the integral and derivative part are eliminated, leaving a proportional controller. The k_p gain of this controller is modified until the system oscillates steadily as shown in Fig. 3.

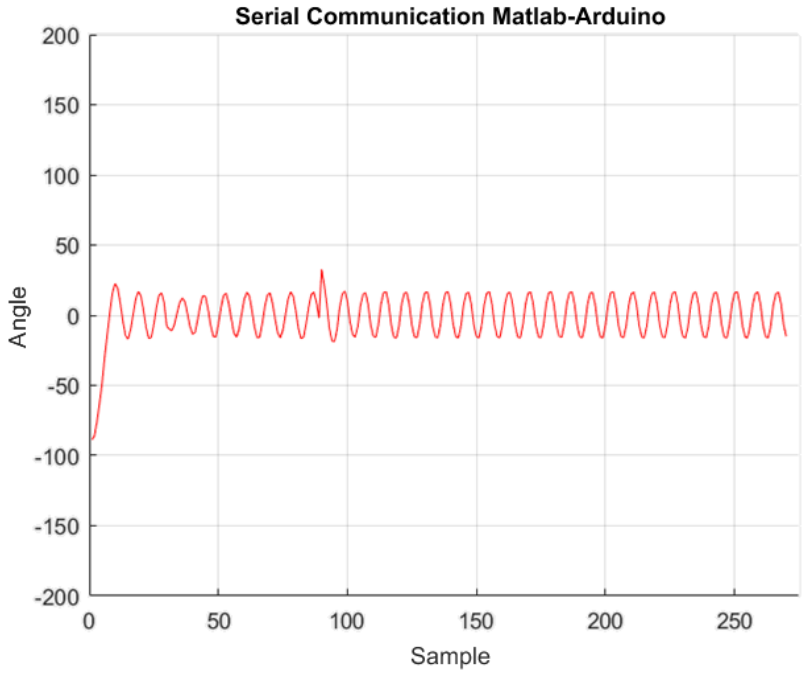


Fig. 3. System sustained oscillations.

The continual oscillation is achieved with a gain of 9.85, this value is the critical gain, and the period of the oscillations is the critical period with a value of 0.55 seconds.

Now applying the Ziegler-Nichols tuning formulas to get the gains, there are 3 different alternatives: classic, with little overshoot and no overshoot at all like is shown in Table 2.

Table 2. Ziegler and Nichols tuning results.

Controller	k_p	τ_i	τ_d
PID classic	$0.59K_{cr}$	$0.50P_{cr}$	$0.125P_{cr}$
PID controlled overshot	$0.33K_{cr}$	$0.50P_{cr}$	$0.33P_{cr}$
PID no overshot	$0.20K_{cr}$	$0.50P_{cr}$	$0.33P_{cr}$

Once s k_p, τ_i and τ_d are calculated, then are entered to the controller to verify its result.

4.2 Genetic Algorithm

The Genetic Algorithm (GA), methodology implemented to the Control is the following explained:

To initialize the GA is necessary to give some important information, the values used are show in Table 3.

Table 3. Initial Conditions.

Initial condition	**Value**
Seed	3
Population size	20 individuals
Genes number	3
Alleles number	4
Tournament size	3
Generations	100
Mutation percentage	10%
Mutation numbers	600
Mutations per generation	6
Serial communication velocity	11500 bauds

Then using the data obtain with Ziegler and Nichols tune, was propose the restriction to obtain the value for k_p, τ_i and τ_d variables avoiding making unstable the response, those values are show in Table 4.

Table 4. PID values.

Gains			
Kp		**Ti**	
Resolution	220	Resolution	2200
Max. Value	4.6546	Max. Value	0.46546
Min. Value	0.00454	Min. Value	0.000454

The initial population was randomly initialized; the individuals have a structure as show in equation 4:

$$pobl(1,1) = \{0100101101, 1010101100, 1101001010\}. \quad (4)$$

Objective function is generated following the algorithm 1 that describes the experiments to realize for measuring the behavior of each individual.

Data: pobl
Result: error
$pobl = \{0100101101, 1010101100, 1101001010\};$
for $i = 0; i <= 4; i = i + 1$ **do**
  $asa[i] = pobl[i];$
  $ese[i] = pobl[i + 5];$
  $isi[i] = pobl[i + 10];$
end
for $i = 0; i < 272; i = i + 1$ **do**
  $int\ \ s0 = digitalRead(22)\ \ s1 = digitalRead(24)\ \ s2 = digitalRead(26)$
  $s3 = digitalRead(28)\ \ s4 = digitalRead(30)\ \ s5 = digitalRead(32)$
  $s6 = digitalRead(34);$
  $s7 = digitalRead(36)\ \ s8 = digitalRead(38)\ \ s9 = digitalRead(40)$
  $s10 = digitalRead(42)\ \ s11 = digitalRead(44);$
  $floaterror = (s0 * 1 + s1 * 2 + s2 * 4 + s3 * 8 + s4 * 16 + s5 * 32 + s6 *$
  $64 + s7 * 128 + s8 * 256 + s9 * 512 + s10 * 1024 + s11 * 2048) * .087;$
  $it = (b * e) + ita;$
  $td = c * (e - ea);$
  $ut = (a * e) + td + it;$
end

Algorithm 1. Fitness evaluation algorithm.

With the evaluated individuals the proposed algorithm is applied with the well-known conventional operations, Fig. 4 shows the tournament flow diagram that apply for the selection process.

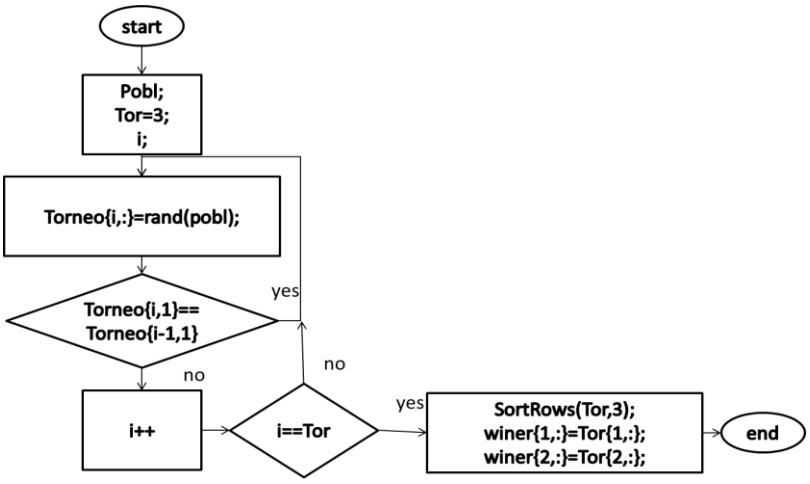


Fig. 4. Tournament flow diagram for selection process.

For the crossover and mutation process the usual algorithm was applied having one crossover point and a mutation in only one allele by gene.

The winner of all the process gives the parameters (P, I, D), which are implemented in the PID controller.

5 Results

First was apply the classic PID controller and was plot the step response that appear in Fig. 5, where de PID variable values was $k_p = 5.815$, $\tau_i = 0.275$ and $\tau_d = 0.068$. As it is shown the response have a continue oscillation approaching the reference value.

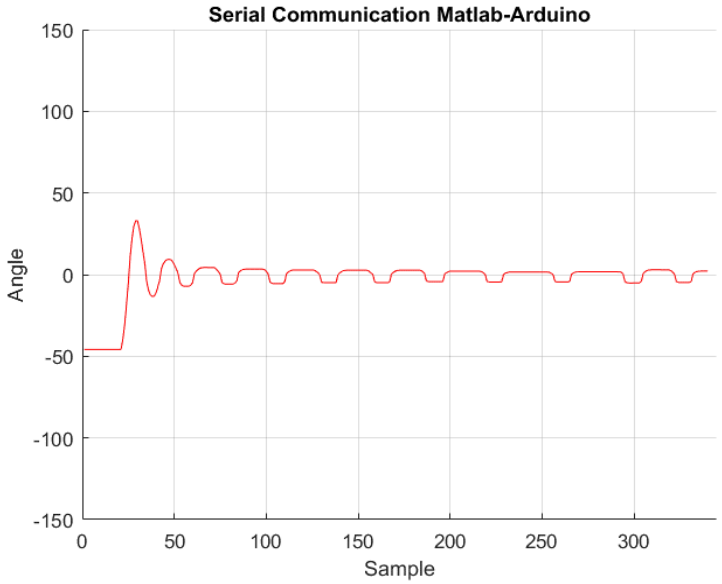


Fig. 5. Step response classic PID tune.

Then the process for controller gains with no overshooting was applied in a similar experiment to obtain the step response show in Fig. 6, in this case the establishment time greater than in other cases but the oscillation do not disappear, in this case the PID variable values are $k_p = 1.97$, $\tau_i = 0.275$ and $\tau_d = 0.1815$.

However, it was not possible to perform the tuning by this method, since the system oscillated in all the cases. The overshooting case was omitted, because as is know if the overshoot is increase the system oscillates. Unfortunately, the Ziegler-Nichols method of critical earnings cannot present results for all control cases.

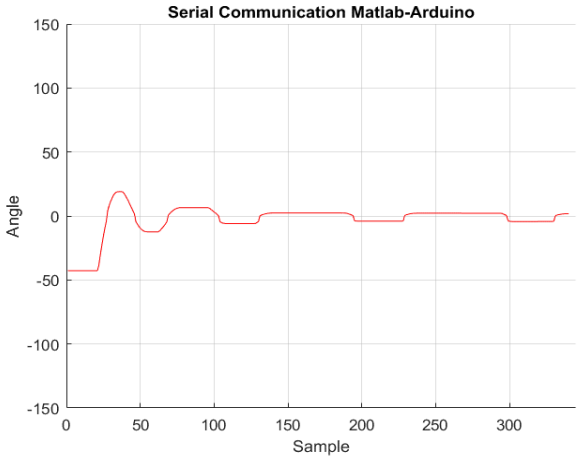


Fig. 6. Step response no overshooting PID tune.

To prove the functionality, there are three height parameters predetermined as height desired where the GA Control achieve this with the next voltage according to the set point.

After the genetic algorithm tunes the PID parameters the obtained results for PID control are given as $k_p = 38.34$, $\tau_i = 0.6575$ and $\tau_d = 4.2725$, showing a behavior faster than the Ziegler and Nichols tune for all three cases, Fig. 7 shows the step response where it has an overshooting bigger than other controllers but have no oscillation approaching the reference.

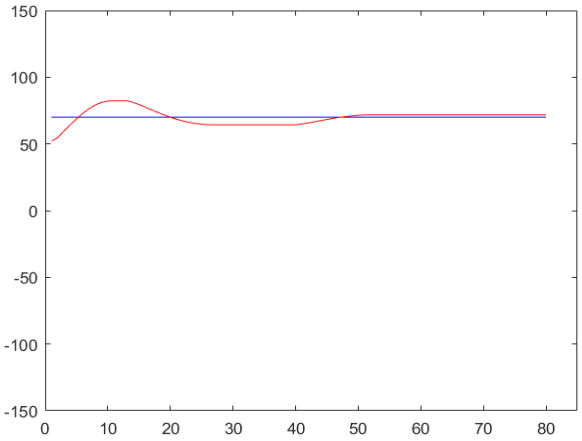


Fig. 7. PID response with GA tuning.

To get this results the GA has an evolution represented by the optimization of the population, Fig. 8 shows this performance for the tuning process, the best element on the blue line have an evolution in the first generations and find the best local result, on the other hand the diversity of the population disappears then of the 350 generations, that prove that our algorithm has enough diversity and the result is not a lake suboptimal.

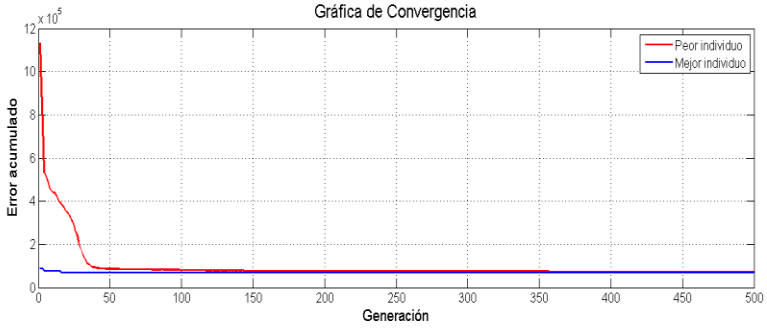


Fig. 8. GA population evolution.

6 Conclusion

This paper presents a comparison between PID controllers tuned with Ziegler and Nichols method and a using a Genetic Algorithms in a SCW system that takes as input the required orientation for the vehicle which is computed according to the detected lines, this paper shows that it is possible to obtain a self-tuning PID controller with better results than classical tuning techniques for a SWC assistant system.

The study of this control gives as a result information about the use of GA to tune a PID controller, showing the comparison between a well-known method that is Ziegler and Nichols tune, both have the capability to get a controller without knowledge about the system but, de GA have a better performance.

The PID has the capability to evolve and tune by itself the gains, was important to delimit the universe with the possible real values, to avoid get a lot of individuals that do not present an important behavior in the system.

The experiments need more time to develop and use a bigger population in more generations, to get a response with a better behavior.

In the future work, the controller will modify the reference of the warning get from the LDW to make an autonomous decision in an ADAS, raising the security conduction and performance. It is important to compare the actual control results with other optimization metaheuristic like PSO algorithms to find which the most proper tuning method is.

References

1. Akatsuka, H., Kataoka, M., Matsumoto, T., Iida, H.: Electric power steering system with motor controller. pp. 933–971 (2016)
2. Aly, A.A.: Pid parameters optimization using genetic algorithm technique for electrohydraulic servo control system. Intelligent Control and Automation, 2(02), pp. 69 (2011)
3. Behensky, M.L., Moncrief, R.L., Durfey, E.J., Loper III, M.H.: Control device such as a steering wheel for video vehicle simulator with realistic feedback forces (1991)
4. Dobres, J., Reimer, B., Chahine, N.: The Effect of Font Weight and Rendering System on Glance-Based Text Legibility. Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicle Applications (2016)
5. Coughanowr, D.R., Koppel, L.B., et al.: Process systems analysis and control, 3 (1965)
6. Cuautle, T.E., Duarte, M.A., Reyes, C.A., Reyes, G.: Automatic synthesis of electronic circuits using genetic algorithms. Computación y Sistemas, 10(3) pp. 217–229 (2007)
7. Dahmani, H., Chadli, M., Rabhi, A., Hajjaji, A.: Vehicle dynamics and road curvature estimation for lane departure warning system using robust fuzzy observers: experimental validation. Vehicle Syst Dyn, 53(8), pp. 1135–1149 (2015)
8. Ezeta, R.F.d.B.: Análisis y diseño de sistemas de control digital. McGraw-Hill Interamericana (2013)
9. Flores, F., Aguilar, C., Suárez, M. S., Cuevas de la Rosa, F.: An identification genetic algorithm for a family of Duffing's system. Computación y Sistemas, 7(2), pp. 102–112 (2003)
10. Gaikwad, V., Lokhande, S.: Lane departure identification for advanced driver assistance. Ieee T Intell Transp 16(2), pp. 910–918 (2015)
11. Handy, J.P.: Steering-wheel control. US Patent 1,132,771(1915)
12. Hickman, J., Guo, F., Camden, M., Hanowski, R., Medina, A., Mabry, J.: Efficacy of roll stability control and lane departure warning systems using carrier-collected data. J Safety Res 52, pp. 59–63 (2015)
13. Jones, M.B.: Steering wheel control attachment apparatus. US Patent 5,520,071 (1996)
14. Kortli, Y., Marzougui, M., Atri, M.: Efficient implementation of a real-time lane departure warning system. pp. 1–6 (2016)
15. Li, W., Gong, X., Wang, Y., Liu, P.: A lane marking detection and tracking algorithm based on sub-regions. pp. 68–73 (2014)
16. Mahajan, R., Patil, A.: Lane departure warning system. International Journal of Engineering Education (2015)
17. Mejía, J.F.G., Sanchez, J.C.S., Fuentes, A.A.F., Martínez, J.A.P., Reyes, C.E.T.: Sintonización de un controlador proporcional-integral derivativo aplicado a una celda termoeléctrica: Una comparación entre algoritmos genéticos. Research in Computing Science 94, pp. 179–192 (2015)
18. Momani, S., Abo-Hammour, Z.S., Alsmadi, O.M.: Solution of inverse kinematics problem using genetic algorithms. Applied Mathematics & Information Sciences 10(1), pp. 225 (2016)
19. Ogata, K.: Sistemas de control en tiempo discreto. Pearson educación (1996)
20. Prospero, L.A.C., Mejía-Lavalle, M., Ascencio, J.R., Rincón, V.V.V.: Aprendizaje incremental basado en población como buena alternativa al uso de algoritmos genéticos. Research in Computing Science 116, pp. 51–64 (2016)
21. Reagan, I., Anne, M.: Observed activation status of lane departure warning and forward collision warning of Honda vehicles at dealership service centers. Traffic Inj Prev, 17(8) (2016)

22. L.J.: A genetic algorithm application for individual and group multicriteria decision making. Computación y Sistemas (2000)
23. Son, J., Yoo, H., Kim, S., Sohn, K.: Real-time illumination invariant lane detection for lane departure warning system. Expert Syst Appl 42(4), 1816–1824 (2015)
Sternlund, S., Strandroth, J., Rizzi, M., Lie, A., Tingvall, C.: The effectiveness of lane departure warning systems—A reduction in real-world passenger car injury crashes (2016)
24. Weise, T.: Global optimization algorithms theory and application (2009)

ISSN 1870-4069

Solving the Capacitated Vehicle Routing Problem in a Shared Bicycle System of a Smart City

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

[1] Universidad Enrique Díaz de León, Guadalajara, Jalisco, Mexico

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

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

[4] Universidad Autónoma de Aguascalientes, Aguascalientes, Mexico

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

Abstract. A shared bicycle system (SBC) is a set of elements that compose an ecosystem in which a user travels in relatively short traverses from one point to another and their trajectory times are improved, exercised and treated of non-motorized vehicles, no Contaminates the environment. This ecosystem consists of bicycles, geographically distributed stations (ports), a card-charging system. In the present investigation with respect to the problem of Vehicle Routing with Capacity (CVRP) to determine a better route. MATLAB was used for the implementation of the CVRP algorithm. As a result, an improvement in the distribution of bicycles in the ports is shown.

Keywords: smart city, shared bicycle system, MATLAB, CVRP.

1 Introduction

Smart city is defined as that which intelligently integrates economy, people, mobility, the environment, government and sustainability [1]. In the case of the present investigation will focus the attention in the branch of the mobility since it is by means of which the (SBC), is specified. Figure 1 shows the mobility ecosystem of a smart city.

Within the mobility block there is a great concern to project more efficient plans of displacement of the people known as Sustainable Urban Mobility Plans (SUMP) whose objective is to align each of the blocks to guarantee a better quality of life of the people [2].

Fig. 1. Smart city ecosystem.

The National Institute of Statistics and Geography (INEGI For its acronym in Spanish), considers as of 2015 in its intercensal survey to the bicycle as a means of transfer to the place of work [3]; For Jalisco, eight indicators were captured in the survey; A) Truck, taxi, bus or bus, b) Truck, taxi, bus or bus, c) Private vehicle, d) Work transport, e) Bicycle, f) Walking, g) Other and h) Not specified. Figure 2 shows the percentages by which people move to their place of work. In total, the survey found that in Jalisco 2,854,085 people use some means to move to their place of work.

Under this scheme, considering that 5.46 is equivalent to 155,833 people using bicycles to go to their place of work then it is speaking that the fact of having a SBC is of vital importance to respond to the demand of citizenship.

The Capacitated Vehicle Routing Problem (CVRP), derives from the Vehicle Routing Problem (VRP); Was created by G. B. Dantzig and J. H. Ramser and the objective is to establish a strategy where cars to decrease the consumption of gasoline [4], as well as the distance between points that visit the agent during its travel along the path of distribution. Figure 3 shows the VRP.

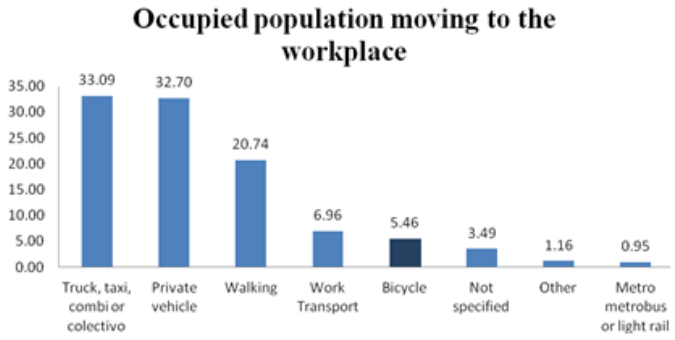


Fig. 2. Type of transport.

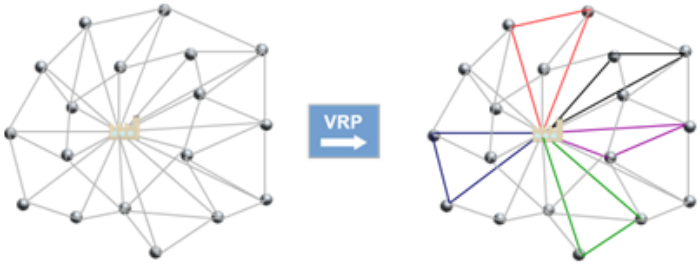


Fig. 3. The VRP.

An SBC can be modeled as in CVRP since they have identified the control points, load, discharge and step frequencies.

The VRP could be used but it was decided to use the variation with capacities (CVRP) since only the VRP is in charge of ensuring that all the points in a route are visited at least once. In the case of the present investigation, there is the variant of capacity and load per unit from point to point.

The mathematical model of the CVRP is shown below [13]:

$$\sum_{i \in V} \sum_{j \in V} c_{ij} x_{ij}$$

$$\sum_{i \in V} x_{ij} = 1 \qquad j \in V \setminus \{0\}$$

$$\sum_{j \in V} x_{ij} = 1 \qquad i \in V \setminus \{0\}$$

$$\sum_{i \in V} x_{i0} = K$$

$$\sum_{j \in V} x_{0j} = K$$

$$x_{ij} \in \{0,1\} \qquad \mathbf{i}, \mathbf{j} \in V$$

$$\sum_{i \notin S} \sum_{j \in S} xij \geq r(S), \qquad \forall S \subset V \setminus \{0\}, \quad S \neq \phi$$

2 Description of the Problem

The present research focuses on the city of Guadalajara, Jalisco, which is considered the first smart city in the world since 2013 [5]. There are four more in the world: Wuxi in China, Trento in Italy, Casablanca in Morocco and Kansas City in the United States.

Despite the benefits of the smart city ecosystem, there are areas of opportunity for further improvement because, as new paradigms are concerned, the ecosystem cannot yet be considered an axiom.

Among the problems encountered in an SBC, specifically in the mobility block, it is found that when the user arrives at a station he expects to always find a unit available to take and move to his destination; Usually what happens is that the stations are limited to 10 to 18 bicycles, depending on the geographic location of the station as it is located, it is the demand for example the number of units that are destined in the station of the center Of the city with respect to one that is in the periphery, where the quantity is smaller.

In that scenario the user must wait for a new unit to arrive or move to another nearby station. Another problem that is detected is the maintenance of the units derived from wear and tear by natural use. As long as it is possible to identify routes where the units are less traveled in terms of mileage, it will be possible to prolong the units in good condition.

Figure 4 shows the points and the flow of the SBC that represent the complete route of [6] for Guadalajara in 2015.

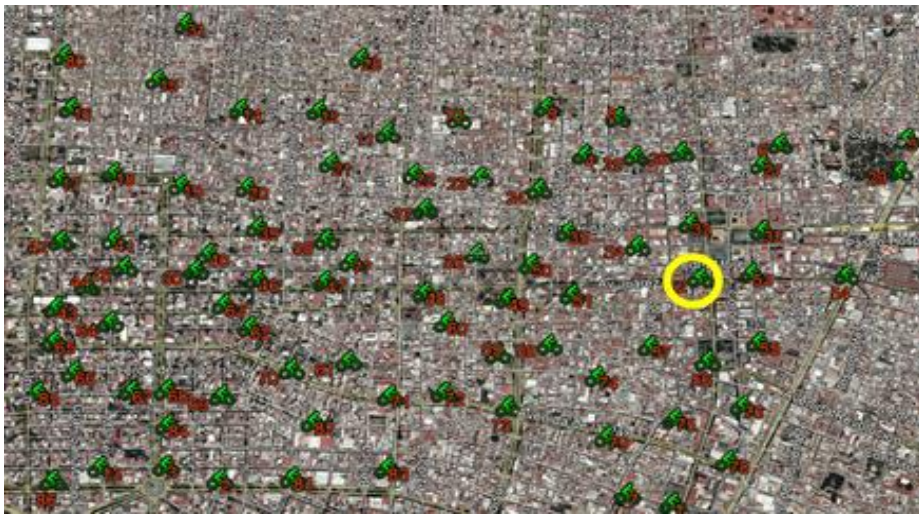


Fig. 4. SBC path.

From the complete route shown in Figure 4, there are 86 points in total, in what is known as the first square of the city, the points go from *Calzada Independencia* to *Avenida Union* and from *Manuel Acuña* to *Avenida Niños Heroes*.

The VRP is a problem of Nonlinear Programming but its combinatorial nature and characteristics can be represented by a graph and on that basis describe the objective function and restrictions.

Given a graph G(V, E) where V is the vertexes set and E is the edges set, the VRP is formulated as follows:

V = {v0, v1, … , vn} is a vertexes set, where:

v0 is considered as a depot.

And V'=V \{v0} be used as the set of n costumers.

A= {(vi, vj) | vi, vj V; i j} is an arc set
C is a matrix of costs (positives) or distances cij between customers vi and vj.
d is a vector of the customer demands.
Ri is the route for vehicle i
m is the number or vehicles with the same characteristics. One route is assigned to each vehicle.
When the problem is symmetric, it means, cij=cji for all (vi, vj) A, then E= {(vi, vj)| vi, vj
V; i< j}.

With each vertex vi in V' is associated a quantity qi of some goods to be delivered by a vehicle. The VRP thus consists of determining a set of m vehicle routes of minimal total cost, starting and ending at a depot, such that every vertex in V' is visited exactly once by one vehicle. For easy computation, it can be defined b(V)=[(Σvi v di)/C], an obvious lower bound on the number of trucks needed to service the customers in set V. If is considered a service time ∂i (time needed to unload all goods), required by a vehicle to unload the quantity qi at vi. It is required that the total duration of any vehicle route (travel plus service times) may not surpass a given bound D, so, in this context the cost cij is taken to be the travel time between the cities. The VRP defined above is NP-hard. The feasible solution of it can be represented as follows:

a partition R1,...,Rm of V; and
a permutation σi of Ri U 0 specifying the order of the customers on route i.

The cost of a given route (Ri={v0, v1, … , vm+1 }), where vi V and v0=vm+1=0 (0 denotes the depot), is given by:

$$C(R_i)= \sum_{i=0}^{m} C_{i,i+1} + \sum_{i=1}^{m} \partial_i \quad (1)$$

A route Ri is feasible if the vehicle stops exactly once in each customer and the total duration of the route does not exceed a pre-specified bound D: C(Ri)≤D.
Finally, the cost of the problem solution S is:

$$F_{VRP}(S)=\sum_{i=0}^{m} C(R_i) \quad (2)$$

3 Related Work

In [7], it is shown that efforts have been made in Singapore to support the rearrangement and distribution of bicycles. They argue that for this type of system (SBC), to be useful, the correct number of bicycles should be deployed in the right places, as this affects the rate of use of bicycles and how bicycles circulate within the system.

In [8], mechanisms based on multi-agent systems are proposed to identify the bicycle crossing points as well as better distribution of the same.

A work was done by [9], where the problem of shared bicycles with ACO was attacked.

In [10], open data applied in Washington D.C. to the system of shared bicycles with the purpose of locating the terminals according to the best satisfaction of the users.

4 The Proposal

As part of the strategy of the present investigation to address the above problem is taken as a point of reference a port (52), of 86 (as shown in Figure 4). The geographical location of this point is 16 de Septiembre Av. (Alcalde) and Juárez Av. (Vallarta Av.) Figure 5 shows the point and in Figure 6 the depot is shown.

Fig. 5. Location of the port.

Fig. 6. Depot.

This port was chosen for being the most demanded of all the SBC according to their number of trips. According to [6], the total number of bicycles for this point is 16, while for other ports the total is between 8 and 10, according to their point of location and demand. In addition, the number of trips is higher, as shown in Table 1.

Table 1. SBC.

key	bikes	travels
GDL-52	16	2054
GDL-48	10	2028
GDL-49	22	2020
GDL-50	16	1873
GDL-73	13	1830
GDL-32	10	1687
GDL-63	16	1679
GDL-54	16	1595
GDL-09	13	1510
GDL-19	10	1474
GDL-64	10	1468
GDL-56	10	1440
GDL-71	7	1424
GDL-61	10	1415
GDL-62	10	1335
GDL-84	13	1253
GDL-60	16	1162
GDL-65	10	1114
GDL-80	10	1093
GDL-36	13	1057
GDL-17	10	1020
GDL-15	10	1013
GDL-14	10	992
GDL-40	10	987
GDL-43	10	971
GDL-51	10	952
GDL-47	10	943
GDL-28	7	923
GDL-33	10	920
GDL-31	7	920
GDL-78	7	911
GDL-03	10	904
GDL-26	10	885
GDL-10	7	883
GDL-46	7	852
GDL-70	10	843
GDL-42	7	835
GDL-79	7	802
GDL-72	10	798
GDL-01	7	782
GDL-77	7	753
GDL-05	7	726
GDL-86	10	725
GDL-08	10	709
GDL-68	10	696
GDL-25	10	694
GDL-16	10	690
GDL-07	10	672
GDL-20	10	665
GDL-39	10	658
GDL-44	10	652
GDL-82	7	615
GDL-34	7	610
GDL-45	16	595
GDL-21	10	588
GDL-37	10	577
GDL-81	7	562
GDL-85	10	554
GDL-57	10	547
GDL-59	10	518
GDL-76	10	517
GDL-13	7	510
GDL-29	10	504
GDL-58	7	500
GDL-02	10	492
GDL-06	7	485
GDL-22	10	474
GDL-27	7	474
GDL-66	10	465
GDL-38	10	463
GDL-12	7	444
GDL-04	7	430
GDL-41	10	426
GDL-11	7	418
GDL-83	7	416
GDL-35	10	413
GDL-53	10	409
GDL-55	10	402
GDL-23	10	379
GDL-67	10	378
GDL-74	7	376
GDL-18	7	360
GDL-24	10	358
GDL-69	10	351
GDL-75	7	340
GDL-30	16	0

Once the data set is identified, the trips are counted according to their location and thus a sample is created for the analysis of the behavior of the SBC. An analysis of the trips made by each user (with the objective of quantifying the trips per unit), is created the instance to solve the CVRP which is shown in Table 2.

Table 2. Instance.

Id	X	Y	D
1	61	65	2
2	25	52	2
3	56	41	3
4	29	54	1
5	44	48	2
6	31	55	1
7	48	42	1
8	42	39	3
9	47	36	1
10	56	40	3
11	56	40	3
12	59	40	1
13	56	51	1
14	42	39	3
15	27	61	3
16	51	44	2

Id	X	Y	D
17	51	62	1
18	50	70	1
19	27	61	3
20	25	52	2
21	39	51	1
22	48	53	1
23	33	69	1
24	69	54	1
25	51	44	2
26	38	66	1
27	55	35	3
28	55	35	3
29	25	79	2
30	25	79	2
31	72	38	1
32	42	39	3

Id	X	Y	D
33	45	48	2
34	71	42	1
35	55	35	3
36	27	61	3
37	61	65	2
38	65	40	2
39	65	40	2
40	41	65	1

Where Id is the identifier of each of the points, X, Y are the location coordinates within the plane and D is the load of each of the points, which are determined by the instance obtained from the SBC. D is determined by the trips made to node 52 at a given time. Figure 7 shows the process of the proposal.

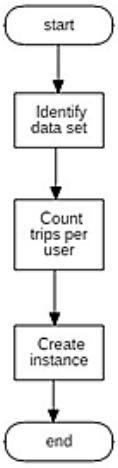


Fig. 7. Process.

5 Methodology

The process of Figure 7 is explained below.

Identify data set. At this point the information analysis is performed; are located the busiest points and limits that covers the route as well as bicycles per point.
The point 52 of the set of nodes is identified given that it is the one that has the most affluence as shown in Table 1 with 2,054 trips in total.

Count trips per users. Once the set of data to be analyzed has been identified, a count is made of the users who go to what point to determine the most requested and at what time. Create instance. When the instance is created it is necessary to give it a format as shown in Table 2. Prior to the execution of the algorithm it is necessary to have an Id (node / ant), X (position x in the plane), Y (position y in the plane) and D (point-to-point loading).

For the implementation of this work we used the Nearest Neighbor Algorithm (NNA), algorithm proposed by Sas Wahid Hamzad [11].

The process to create the instance with the obtained data is observed in [12], where instances were compared to determine the optimal solution of given instances.

As part of the strategy, the points that make up the route are identified, which are shown in figure 8.

Fig. 8. The Grid.

Tracing the route is possible to identify a complete network in which the routes made by the units are long.

6 Results

The result is constructed from the execution of the algorithm in MATLAB from which the graph shown in Figure 9 is derived. Where each list represents the points that each bicycle is suggested to visit.

As the result of the execution of the algorithm with the instance, it was possible to distribute the points in two segments, the first one by visiting:

1>37>17>18>40>26>22>13>24>34>38>39>12>3>10>11>16>25>7>33>5>21>6>4>15>23>1

and the second by visiting:

1>31>27>28>35>9>8>14>32>2>20>19>36>29>30>1

The total area was 3.90 km2.

The result is shown in Figure 9.

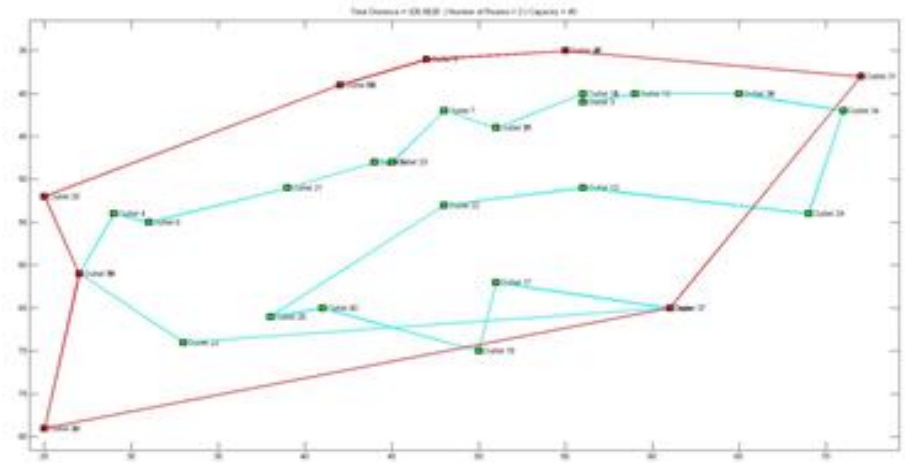


Fig. 9. Result.

7 Conclusions

After the experimentation of the present work, it is observed that through the application of computational techniques it is possible to improve a SBC in terms of distance and time and derived from it is possible to keep the units (bikes) in good condition since the lesser use is given to each one of them, the time of life is prolonged for more time.

8 Future Work

As a future work, it is proposed to submit the SBC instance to solve the CVRP in a parallel environment in order to observe the behavior of both schemes, that is, sequential and parallel.

It is also expected to be able to determine, by geo-positioning, safe routes according to crime rates.

References

1. IEEE: http://smartcities.ieee.org/about.html (2017)
2. INNOPRO: http://www.innopro.es/pdfs/libro_blanco_smart_cities.pdf (2017)
3. INEGI: http://www.beta.inegi.org.mx/proyectos/enchogares/especiales/intercensal/default.html =1(2017)
4. Dantzig, R.: The truck dispatching problem. Management science, pp. 80–91 (1959)
5. INFORMADOR: http://www.informador.com.mx/economia/2015/622072/6/guadalajara-la-primera-ciudad-inteligente.htm (2015)
6. JALISCO: https://datos.jalisco.gob.mx/dataset/catalogo-georreferenciado-de-terminales-mibici (2015)
7. Jia, S. et al.: Models for Effective Deployment and Redistribution of Bicycles Within Public Bicycle-Sharing Systems. Operations Research 201361(6), pp. 1346–1359 (2013)
8. Méndez, L. et al.: Multi-agent System for the Control and Monitoring of Shared Bicycle Fleets. International Symposium on Ambient Intelligence (2017)
9. Cashous, W., et al.: Ant Colony Optimization Applied to the Bike Sharing Problem. (2015)
10. Longbiao, Ch. et al.: Bike sharing station placement leveraging heterogeneous urban open data. (2015)
11. Matlab: https://www.mathworks.com/matlabcentral/profile/authors/3883690-sas-wahid-hamzah (2016)
12. Cossio-Franco, E., Hernández-Aguilar, J., Ochoa-Zezzatti, A., Ponce-Gallegos, J. Comparison between instances to solve the CVRP. International Journal of Combinatorial Optimization Problems and Informatics, 9(2), pp. 41–54 (2018)
13. Ponce, J.C., Recio, A., Ochoa, A., Ornelas, F., Padilla, A., Torres, A.: Implementación de un algoritmo de colonia de hormigas en CUDA aplicado al problema de ruteo de vehículos in Aplicaciones Modernas de Optimización, la experiencia entre cuerpos Académicos (2014)

ISSN 1870-4069

Design and Implementation of a CVRP Simulator Using Genetic Algorithms

Iván Ramírez Cortes[1], José Alberto Hernández Aguilar[2,3], Miguel Ángel Ruiz[1], Marco Antonio Cruz-Chavez[2], Gustavo Arroyo-Figueroa[3]

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

[2] Autonomous University of the State of Morelos, Cuernavaca, Morelos, Mexico

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

jose_hernandez@uaem.mx

Abstract. We discuss the design and implementation of a CVRP simulator, for this purpose firstly, we discuss briefly CVRP and its taxonomy, later we discuss in depth the genetic algorithm (GA), further we present the design and implementation of a three-tier web-based system to upload a CVRP instance and select a Metaheuristic for its processing. For probing our simulator, we calculate the best route(s), processing ten different .vrp instances, with our GA and compare it with the nearest neighbor algorithm (NNA). Preliminary results show genetic algorithm generates some best routes, but processing time is higher regarding nearest neighbor algorithm.

Keywords: CVRP, simulation, genetic algorithm, nearest neighbor algorithm, client-server technology.

1 Introduction

Research Problem

It has been identified that in the state of Morelos, Mexico, small and medium businesses need a logistics in real time to solve routing problems, which requires the implementation of a plotter of routes to represent real-world instances of the CVRP (Capacitated Vehicle Routing Problem) that meets following features:

Operate Under Three Tiers Client-Server Architecture

Allows the reading of. vrp files from the client-side, send the request to the server, select the algorithm or heuristic goal to use (for example nearest neighbor, ant's colony, genetic). Call the corresponding method and display in real time the results of the progress in the process of optimization of the route. Is required the use of java running

on a web platform for the implementation of the plotter of routes for the CVRP by intelligent algorithms, since the server will be mounted on open source software.

Justification

The research group UAEMOR CA124 "Operations research and computer science", participated in the convocation of the PRODEP 2015 for the creation of academic networks, derived from this convocation was obtained financial support for the implementation of several projects, within these projects was financed the corresponding plotter for the solution of the CVRP, which for its complete solution includes not only the estimate of the calculation for routes but its real-time plotting.

At present, the lack of simulators that graph in real time this type of problems is very wide, and it is even more difficult to find simulators operating via the client-server architecture.

Hypothesis

Hi. Using a graphing tool for problems CVRP type, operating under the client-server architecture, allows determining if genetic algorithms obtained better results than the nearest neighbor algorithm.

Ho. Using a graphing tool for problems CVRP type, operating under the client-server architecture, does not allow determining if genetic algorithms do not obtain better results than the nearest neighbor algorithm.

Scope and Limitations

We analyze the problem and abstract its requirements in four modules:

Module of reading of .vrp instances

- The system will be able to read files of .vrp type previous proper validation.

Module of parameters selection

- Type of Technology to use for the server process: sequential, parallel.
- Select the type of algorithm to simulate (genetic algorithm).

Module of nodes graphing in real time

- Plots the local optimal routes regularly, showing where the user will be able to select the best-desired route.

Module of Graphical Export

- Once calculated the best route, it could be exported in JPG or PNG format, as well as the summary of information (total distance and processing time).

Methodology for Development: Extreme Programming

XP is an agile methodology for software development, which basically consists in strict adherence to a set of rules that are focused on the needs of the client in order to achieve

a good quality product in a short time, focused on enhancing relationships as key to the success of software development.

The philosophy of XP is to satisfy the customer's needs, integrating it as part of the development team. In all the iterations of this cycle, both the client and the program learn.

Structure of Document

In section one we describe the problem at hands, section two discuss related work, including VRP theory, its taxonomy and main approaches to solve it, we focus in genetic algorithm metaheuristics as a method of solution. Later we discuss CVRP mathematical model. In section three, we present the implementation of the CVRP simulator based on the client-server technology; we present pseudo code of main functions and include segments of Matlab code created to implement a genetic algorithm to solve CVRP instances. Finally, we present preliminary results, conclusions, and future work.

2 Related Work

2.1 VRP

The Vehicle Routing Problem (VRP), is used to determine a set of routes for a fleet of vehicles that are based on one or more deposits or warehouses, to satisfy the demand of multiple customers geographically dispersed proposed by Hillier and Lieberman [10]. According [14], importance of VRP is due:

> *"The transport of goods in urban environments plays a very important role in the sustainable development of a city, as high levels of movement of goods occur within cities".*

In Figure 1, we show the models that originate the Travel Salesman problem (TSP), from which derived the first VRP models.

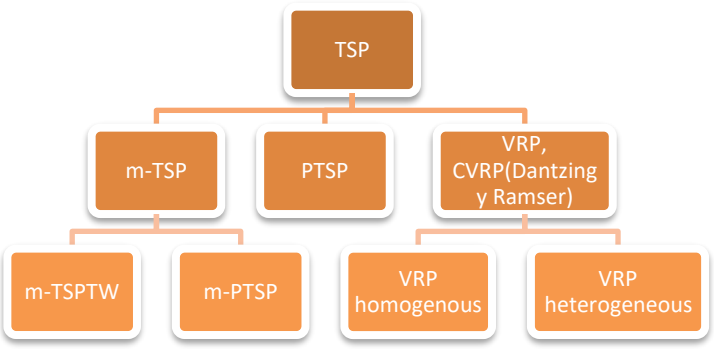


Fig. 1. Models originating VRP problem [13].

Given TSP and VRP are NP-hard problems [8]; there are several approaches to try to solve them. In figure 2, we show the main methods of solution of VRP, this figure shows its taxonomy.

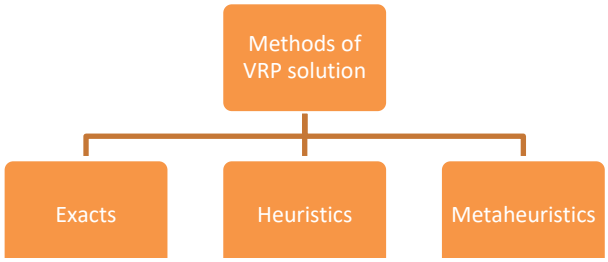


Fig. 2. Methods of solution to the VRP [13].

Exact Methods

In Figure 3, we show the different exact methods to give a solution to the VRP.

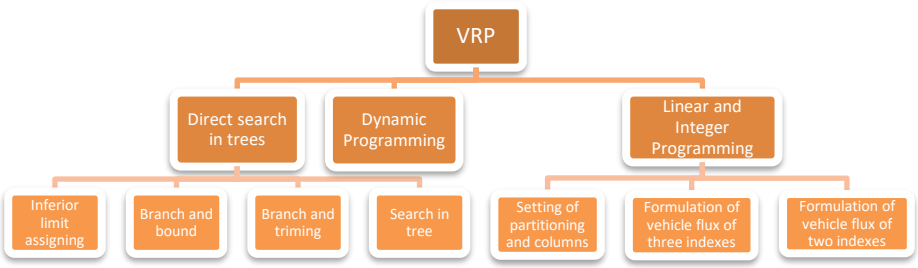


Fig. 3. Classification of exact methods [13].

Heuristics

The heuristics are procedures that provide acceptable quality solutions limited through an exploration of the search space [4]. These methods are based on routes that contain a single node to find the best pair (node, route), that represents the best intersection [11]. Figure 4 shows the classification of the heuristics in constructive methods, methods of two phases and heuristics of improvement.

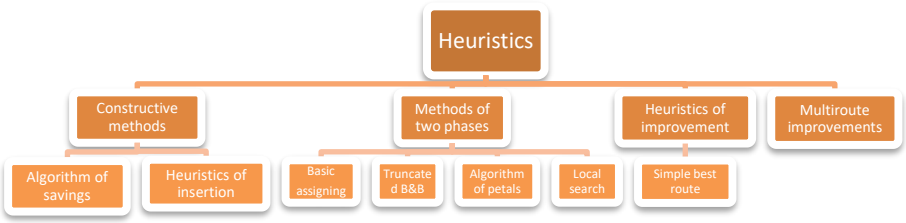


Fig. 4. Classification of Heuristics [13].

Metaheuristics

Most of these methods of solution were developed in the 1990s; one of its features is that procedures of search try to find acceptable solutions [5]. In figure 5, we show the methods of solution that consist of Simulated Annealing, Taboo Search, Neural Networks, Genetic algorithms, ant colony algorithms and search in neighborhoods.

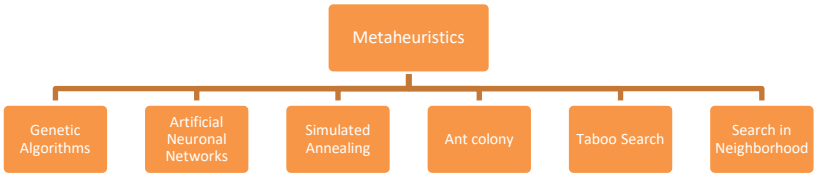


Fig. 5. Classification of Metaheuristics [13].

2.2 Genetic Algorithm

Genetic Algorithms (AGs) are adaptive methods that can be used to solve problems arising from the search, optimization [13] and decision making [15]. They are based in the genetic process in living organisms [2, 7]. It is based on the principles of the laws of the natural life proposed by Darwin. A genetic algorithm according to [9], basically consists of 1) An initial population, which can be generated in a random way, 2) Calculation of the fitness function, 3) Selection based on the fitness of the population, 4) Generate a new population through cross and mutation, 5) Generate a cycle of the above until the function of unemployment is true, and 6) get the best population of individuals.

Pseudo Code of Genetic Algorithm [9]

```
BEGIN A /*Simple Genetic Algorithm */
```

```
        Generate an initial population.
        Compute the fitness function of each individual.
        WHILE NOT Finished DO
      BEGIN to /*produce new generation*/
        FOR POPULATION SIZE / 2 DO
        BEGIN /*Reproductive cycle*/
            Select two individuals of the previous generation,
            for the crossing (probability of selection pro-
            portional the evaluation function of the individ-
            ual).
            Cross with a certain probability the two individ-
      uals obtaining two descendants.
            Mutate The two descendants with a certain proba-
      bility.
            Compute the evaluation function of the two mutated
      descendants.
            Insert the two mutated descendants in the new gen-
      eration.
          END
          IF        the population has converged, THEN
              Completed = True
       END
  END
```

Mathematical Model for the CVRP

The CVRP has as an objective function which purpose is to minimize the costs of the vehicles in the trajectories. Table 1 displays a list of the indexes and variables that occupies the mathematical model of the CVRP.

Table 1. Indexes and variables of CVRP model.

Nomenclature	Description	Nomenclature	Description
I	The node of the departure of the vehicle.	Cij	Cost of transport of the node i to node j
J	The node of the departure of the vehicle.	Dj	Demand in the node j
d	Vehicle to use (1,2,3,…k)	U	Resource capacity k
x_{ij}^k	Customer Demand	N	Number of clients
Y_{ij}	If it is equal to 1, the vehicle k is assigned to the arc from node i to node j. It is equal to 0 otherwise		

The mathematical model of CVRP routing, according to [1, 12], is:

$$Minimize \sum_{(i,j)\in A} c_{ij} * y_{ij}. \quad (1)$$

Subject to:

$$R1 \sum_{1 \le k \le K} x_{ij}^{k} = y_{ij} \; ; \; \forall \; \boldsymbol{i, j},$$

$$R2 \sum_{1 \le j \le n} \boldsymbol{y_{ij} = 1} \; ; \; \forall \; \boldsymbol{i},$$

$$R3 \sum_{1 \le i \le n} \boldsymbol{y_{ij} = 1} \; ; \; \forall \; \boldsymbol{j},$$

$$R4 \sum_{1 \le j \le n} \boldsymbol{y_{0j} = k},$$

$$R5 \sum_{1 \le j \le n} \boldsymbol{y_{i0} = k},$$

$$R6 \sum_{1 \le i \le n} \sum_{1 \le j \le n} \boldsymbol{d_i * x_{ij}^{k} \le u} \; ; \; \forall \; \boldsymbol{k},$$

$$R7 \sum_{i \in Q} \sum_{j \in Q} \boldsymbol{y_{ij} \le |Q| - 1}; \; \forall \; \boldsymbol{subset \; of \; Q \; of} \; \{1, 2, \dots, n\},$$

$$\boldsymbol{R8 \; k \le K},$$

$$R9 \; \boldsymbol{y_{ij}} \in \{0, 1\} \; ; \; \forall \; (\boldsymbol{i, j}) \; \in \boldsymbol{A},$$

$$R10 \; \boldsymbol{x_{ij}^{k}} \in \{0, 1\} \; ; \; \forall \; (\boldsymbol{i, j}) \; \in \boldsymbol{A}, \forall \; \boldsymbol{k}.$$

A is a set, which is defined as A={(i,j) : yij=1}, i.e. the set of edges of the graph with the possibility of making a single travel from node i to node j. Paths should begin in the node 0 and conclude at the same. Each arc (i, j)∈A, i≠j, the set has a cost Cij. Each of the vehicles has the same charge q, and each customer has a demand di, i ∈ C, for each of the clients, also is to assume that all data are integers known and not negative [6]. The restrictions in table 2 are intended to design a set of Minimum cost routes, one for each vehicle, so that: a) Be met exactly once to each client, b) Each path starts and ends in the tank, and c) Respect the capacity constraints of the vehicles.

Table 2. Description of elements.

Numbering	Description
F.O.	The objective function you want to minimize the total cost of the sum of all the travel, i.e. the distance traveled. Each customer must have an assigned vehicle as well as a sequence to achieve a minimum cost.
R1	In this restriction is indicated if the path was already covered or not will be on the path, the variable x helps to indicate whether or not to use the vehicle k in the arc i,j in the case of x=1.
R2 R3	Indicate the activation of the arc i,j by means of the variable, and what determines a path between nodes i,j, also ensures that every customer is an intermediate node of any route. That is to say, that ensures that each client is visited once by a vehicle.
R4 R5 R6	Indicate that k is the number of vehicles used in the solution and that all those who depart from the deposit must be returned to the same.
R7	Notes that each vehicle does not exceed its capacity.
R8	Monitors that the solution does not contain cycles using 1.2 nodes, …n. Otherwise, the arches to contain any cycle passing through a set of nodes Q and

	the solution would violate the constraint because the left side of the restriction would be at least |Q|.
R9	Limits the maximum number of vehicles to use up to a maximum amount.
R10	Indicate that the variables x, y are binary

3 Implementation of CRVP Simulator

The following are the modules implemented within the simulator and shown in Figure 6. System includes a three-tier architecture as the used in [17].

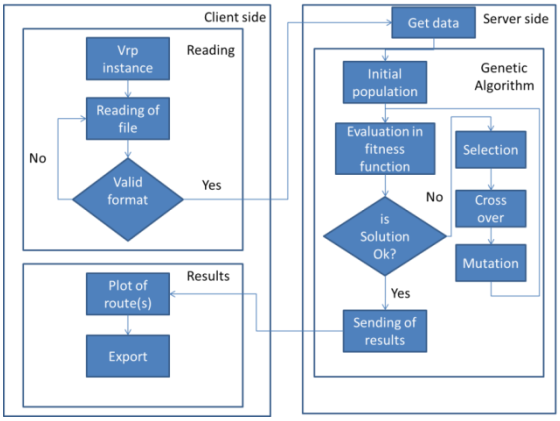


Fig. 6. Functional description of CVRP Simulator.

File Validation

For a correct operation of programs is mandatory to validate that the .vrp file has a correct structure. This function validates a .vrp extension, and description of file content: name, comment, type, dimension, edge type, weight type, and capacity.

Module of File Reading

This module is intended to validate the reading of the files of type .vrp taking into account: a) At the time of actually read a file; b) the extension is correct (.vrp); c) Correct structure, d) titles are spelled correctly.

Reading of File

As a first step, the simulator reads a file .vrp. To be able to read the file is implemented the next pseudo code using java.

Setting of Instances for Java and Matlab

This function creates instances from Java to be able to be used in Matlab. Through the use of functions to read from and write to files, the file includes an identifier, coordinates x & y and the demand, the function creates a new .m extension file. We

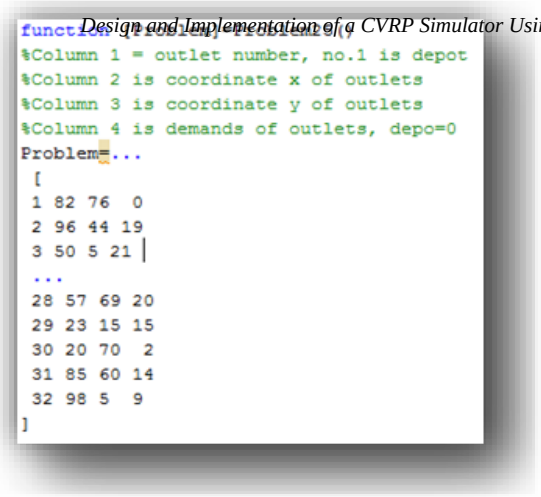


Fig. 7. Example of m file created in Java and compatible with Matlab.

decide to use Matlab due its flexibility to code algorithms and capacity for doing simulations as shown in [18, 19].

Server Side Module

In figure 6 is shown in the diagram corresponding to the server side, in which Matlab receives the .m file generated in Java, read the coordinates, and uses a genetic algorithm to perform calculations of routes and produce its graph interacting with Matlab.

Implementation of the Genetic Algorithm

In order to obtain the optimal routes, we implemented a genetic algorithm in Matlab, next are displayed the most important functions.

Selection by tournament

For the implementation of the selection was used tournament, Matlab code is shown next:

```
Selection by tournament Matlab code
T = round (random (2*N, S) * (5) +1;
% Tournament
[d, idx] = max (F(t),[],2); % index to determine winners
W = T(sub2ind (size(t), (1:2*N)', idx)); %winners
```

Crossover

We used one-point crossover. Matlab code is shown below:

```
One point crossover (Matlab code)
Pop2 = Pop (W(1:2: FIN),1:)% First winners variable of Pop2
P2A = Pop (W(2:2:FIN),:)% Second winners variable of Pop2
```

```
Lidx = sub2ind (size (Pop), round (random(N,1)*(5)+1)
```

Next Matlab code shows the selection process to perform the crossover:

```
Selection process to perform crossover (Matlab code)
% Selection of one point
vLidx= P2A(Lidx)*(1,G) % Point of two winners
[r,c] = find (Pop2 == vLidx) %winners 1
[d, Ord] = sort( r ) %Sort linear index
r = r (Ord); c = c (ord) % Reorder index
Lidx2 = sub2ind(size(Pop), r, c) % Convert to one index
Pop2(Lidx2) = Pop2(Lidx) % Half of crossover 1
```

Mutation

Finally, there is the stage of mutation which is shown next:

```
Mutation (Matlab code)
Idx = rand (N,1) < Muta % Selection of individuals to permute
Loc1 = sub2ind(size(Pop2), 1: N, round(random(1,N) *(5) +1)) %
Interchange of index 2
Loc 2 = sub2ind(size(Pop2), 1: N, round(random(1,N) *(5) +1)) %
Interchange of index 2
Loc2 (idx == 0) = Loc1 (idx == 0) % Probability of mutation
[Pop2 (Loc1), Pop2(Loc2) ] = deal (Pop2(Loc2), Pop2(Loc1)) % Mu-
tation
```

Plotting

Once the optimal routes are calculated is possible to plot them, by using Matlab functions, plotting each path to a different color, each graph will have a reservoir called "Depo" as well as *n* number of customers called *"C"*, a summary to display the number of routes, a total distance of travel and customers of each route.

Export

By means of the Matlab is possible to export resulting graph to two files, one in .PDF format and the other in .PNG format.

Client-side displaying

In this stage, the results are displayed on the client side, that is to say, the graphics in image format as well as exported in .PDF format.

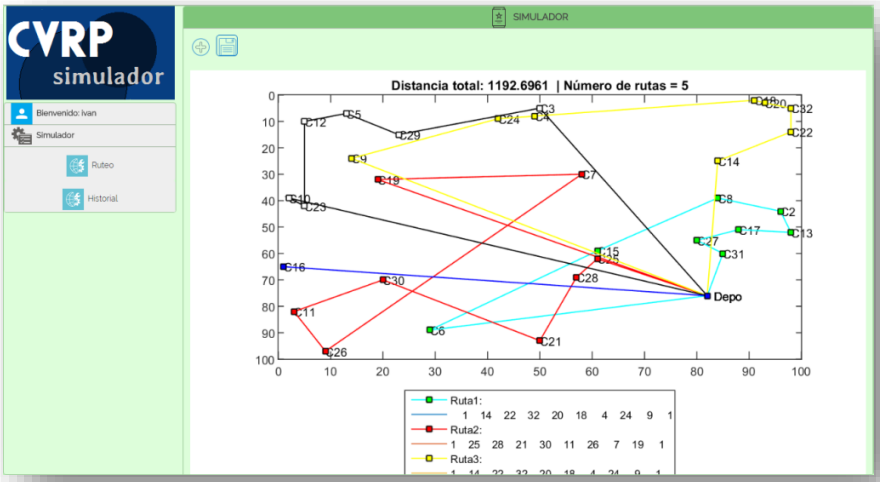


Fig. 8. Displaying of optimal routes in text and graph in client side.

4 Results and Discussion

We compared the performance of genetic algorithm approach regarding nearest neighbor algorithm for ten different .vrp instances (32, 33, 34, 36, 37, 38, 39, 44, 45 & 46). We compare resulting total distance and execution time for each instance.

Table 3. Genetic Algorithm versus nearest neighbor algorithm.

			Genetic Algorithm		Nearest Neighbor		Differences	
#	**Instance size**	**# of routes**	**Total distance**	**Execution time**	**Total distance**	**Execution time**	**Total distance**	**Execution time**
1	32	5	1192,6961	0,22122	1.042,8645	0,18527	-150	-0,03595
2	33	5	1046,4215	0,56151	968,8926	0,19729	-77,5289	-0,36422
3	34	5	1066,7455	0,21822	1035,0417	0,29952	-31,7038	**0,0813**
4	36	5	1112,45	0,21977	1047,6603	0,21634	-64,7897	-0,00343
5	37	5	1155,9072	0,25358	1080,5174	0,22331	-75,3898	-0,03027
6	38	5	1078,3103	0,22807	893,8056	0,23601	-184,5047	**0,00794**
7	39	5	1264,3359	0,13557	1181,4545	0,24701	-82,8814	**0,11144**
8	44	6	1300,0512	0,31112	1347,1924	0,29582	**47,1412**	-0,0153
9	45	6	1342,6026	0,31926	1406,3867	0,31129	**63,7841**	-0,00797
10	46	7	1350,0706	0,31857	1308,7354	0,32642	-41,3352	0,00785
						Average	-59,70398	-0,024861

For instances 44 and 45 Genetic Algorithm obtained better results in distance than the nearest neighbor, in all the other cases nearest neighbor was better. Regarding execution time, genetic algorithm performance was lower regarding NNA.

5 Conclusions and Further Work

The simulators are one of the most important parts of the optimization process, allowing the user to perform tests and avoid generation of extra expenses due errors, allowing the user to generate optimal solutions and carry them out in the industry.
A genetic algorithm allows the generation of optimal routes acceptably, however, there are other artificial intelligence algorithms that can compete, this is the reason why this simulator has open the possibility of testing different A.I. techniques.

Our future work is to implement the simulator not only for working with an algorithm, it means, if you do not get results the expected results, it will be possible to try different algorithms like ant colony, swarm particle optimization, among others. On the other hand, it aims to make the simulator to operate under a mobile platform such as Android and/or IOS. One of the most important challenging future work is making the change from graph nodes to real-world environment, that is to say, that plot really the streets where would the fleet of vehicles being working. We would to include time windows analysis in routing problems as described in [16].

References

1. Ahuja, R., Magnati, T., Orlin, J.: Network flows: theory, algorithms, and applications. Prentice Hall (1993)
2. Alander, J.: On optimal population size of genetic algorithms. In Proceedings CompEuro´92, Computer systems and Software Engineering, pp. 65–70 (1992)
3. Bremermann, H.: Optimization through evolution and recombination. In: Self-Organizing systems. Spartan Books (1962)
4. Butrón, A. A.: Asignación de rutas de vehículos para un sistema de recolección de residuos sólidos en la acera. Revista de Ingeniería, 13(5), pp. 5–11 (2001)
5. Contardo, C.: Formulación y solución de un problema de ruteo de vehículos con demanda variable en tiempo real, trasbordos y ventanas de tiempo. Universidad de Chile (2005)
6. Dorigo, M., Gambardella, L.: Biosystems. ELSEVIER, 43(2), pp. 73–81 (1997)
7. Fraser, A.S.: Simulation of genetic systems by automatic digital computers. I. Introduction. Aust. J. Biol. Sci., 10, pp. 484–491 (1957)
8. Garey, M., Johnson, D.: Computers and intractability. A Guide to the Theory of NP-Completeness, W. H. Freeman & Co. (1990)
9. Goldberg, D.: Genetic algorithms in Search Optimization and Machine Learning. Addison-Wesley (1989)
10. Hillier, F., Lieberman, G. J.: Introducción a la Investigación de Operaciones 8th ed. McGraw-Hill/Interamericana Editores (2006)
11. Laporte, G., Gendreau, M., Hertz, A.: An approximation algorithm for the traveling salesman problem with time windows. Operations Research, 45(4), pp. 639–641 (1998)
12. Olivera, A.: Heurísticas para problemas de ruteo de vehículos. Uruguay (2004)

13. Rocha, L., González, C., Orjuela, J.: Una revisión al estado del arte del problema de ruteo de Ingeniería: Evolución histórica y métodos de solución. Ingeniería 16(2), pp. 35–55 (2011)
14. Thompson, R.G., Van-Duin, J.H.: Vehicle routing and scheduling. Innovations in freight transportation, pp. 47–63 (2003)
15. Leyva-López, J.C.: A genetic algorithm application for individual and group multicriteria decision making. Computación y Sistemas, 4(2), pp. 183–188 (2000)
16. Pérez, N., Cuate, O., Schütze, O., Alvarado, A.: Integración de las preferencias de los usuarios en el proceso de toma de decisiones para problemas de optimización de muchos objetivos discretos. Computación y Sistemas, 20(4), pp. 589–607 (2016)
17. Hernández-Aguilar, J. A., Burlak, G., Lara, B.: Design and Implementation of an Advanced Security Remote Assessment System for Universities Using Data Mining. Computación y Sistemas, 13(4) (2010)
18. García-Mendoza, C.V., Juárez, M., et. al: Filtro estimador por deconvolución y pseudoinversa: descripción e implementación recursiva. Computación y Sistemas, 18(4), pp. 833–840 (2014)
19. Naredo, E., Duarte-Villaseñor, M.A., García-Ortega, M.D.J., Vázquez-López, C. E., Trujillo, L., Siordia, O.S.: Novelty Search for the Synthesis of Current Followers. Computación y Sistemas, 20(4), pp. 609–621 (2016)

ISSN 1870-4069

Augmented Reality System to Promote the Inclusion of Deaf People in Smart Cities

César Lozano Diaz[1], Rocio Maciel M.[1], Manuel Larios V.[1], C. Alberto Ochoa[2]

[1] University of Guadalajara Smart Cities Innovation Center, Zapopan, Mexico

[2] Universidad Autónoma de Ciudad Juárez, Mexico

{clozano, rmaciel,vmlarios}@cucea.udg.mx, alberto.ochoa@uacj.mx

Abstract. The use of technological tools for the inclusion of people with auditory disabilities today is booming, with a significant advance in the close caption for television, the cochlear implant, dictionaries of sign languages apps (SL) in apps, translators of voice-to-text or channels of videos that promote the learning of SL, but without a visible impact in the communities of deaf people or society in general, is still not perceived this technology being adopted everywhere to promote equal opportunities and the well-being of people with disabilities in the Smart cities, due to many of the developments do not pursue the research, usability analysis or the creation of models that help to understand, replicate, improve, or build strategies to ensure an inclusion based on technology and the immersion to the deaf culture by listeners The systematic review to research conducted in this topic helps us understand the implications and challenges faced To follow with innovation in the area of technologies for inclusion and understand that not only it is providing technology for a deaf person, but to create an environment that promotes culture to improve communication between inhabitants of a Smart cities. For this reason the present article is a systematic revision centered in technological research adapted for the access of the information and the communication of people with hearing loss or deaf; presented under Kitchenham´s methodology; the information is searched and classified through the analysis of clusters k-means which summary is presented under the selection and classification of 350 articles published since 2013 until May 2017 with looks towards the analysis of methodologies or models that promote the inclusion of deaf people in society with the usage of technological tools in order to start the guidelines to create a technological-social model to promote the improvement of communication between deaf and hearing people.

Keywords: inclusion, augmented reality, translation from voice to text, deaf people.

1 Introduction

The constant evolution of portable devices and emerging technologies have allowed development of applications that promote the inclusion of deaf people, or with hearing loss in school environments and urban environments, all of them developed to promote a better quality life.

Bouzid et al. in [1], affirm as the technologies become smaller and cost reduction, there is a high probability that deaf people have access to school education with the assistance of technology, through the adoption of proper technologies; However, from the perspective of deaf people community this connection is still far away, according to Ruiz [2], the results of his research had confirmed that the social inclusion in the school method still has not been effective at all, the lack of replicable models and the fact the sign language is different in every country had complicated the process; this leads the people interested in the subject to question: Are there sufficient research that allows creating replicable models for the inclusion of deaf people?, Is there sufficient technology to guarantee the access of information to deaf people?, Is it all said in themes of inclusion for deaf people?, and last, Is it appropriate continue develop improvement in existing technologies or create new technologies to promote its inclusion?. Therefore, this systematic revision is looking to provide information that allows studying the possible solutions to these previous questions like a validation for the develop of future research in the matter.

Now it has become indispensable for the future of Smart cities guarantee the access to the information to its population with special needs; improving the accessibility to public places will absolutely determinate the life conditions of the people, the accessibility standards should be comprehensive across all barriers [3], cities were its population will have a higher number of deaf people, according to the OMS [4] exists 360 million people in all around the world with hearing loss, of which 32 million are children, this quantity has been increasing due to genetic causes, complications during labor, infectious diseases, chronic infections in the ear, the usage of some medicines, the exposure to excessive noise to mention some of facts, hence it is necessary completely know where the research is located and the technologies in this matter to continue the development of these technologies guarantying the smart living.

2 Methodology

This study was conducted as the systematic revision of literature using proposed directives by Kitchenham [5]. In this case, the objective of this review is to evaluate the research referred to the inclusion of deaf people and technologies developed for them.

The steps in the revision method of systematic literature will be documented as the following:

2.1 Research Questions

Starting from the principal questions of the research, which lead us to consider general questions:

RQ1. Which technologies and infrastructure exists in the cities for the inclusion of deaf people in the different areas of a smart city?

This question is related to the importance on knowing that has been done and is being done with the technology in support of deaf people in the cities, schools or business. To answer this will contribute to analyze if it is still relevant for further research on the topic.

RQ2. Has it been worked with RA as the way to promote in the deaf people community the access to the communication?

Through this analysis examines the relevance of developing applications with augmented reality (AR) for the access to the spoken information and the teaching and learning of language of sign (SL). Due to the fact that in the AR is a technology that allows you to increase the information that an individual can obtain for himself to interact with the physical world, to provide more data than that which cannot be registered through the senses.

RQ3. How will this impact the metrics in the school of a smart city, in the inclusion of deaf people?

The answer to these questions takes relevance in assessing the importance of continue to develop improvement in existing technologies or create new technologies to promote the access of education, work and communication, helping the incorporation into the labor force to the thousands of people who have been excluded from many of these specialized environments, if this occurs, the impact on the economic growth of a smart cities will be benefited.

The analysis of these questions generates the opportunity to separate the variables that allow building the key words used in the first phase, with the purpose of building a search matrix to use a clusters analysis. Table 1.

Table 1. Search terms (The search is done with the keywords and their possible synonyms).

Key works	Synonym
Inclusive **education**	Inclusive
Augmented reality	AR
Virtual reality	VR
Translation from Voice to text	Subtitle, translation
People deaf	
Sign Language	LSM
Voice recognition	
Inclusion educational model	Model, inclusion
Virtual avatar	assistant

The methodology selected to analyze the usability of the technology developed is a qualitative type, following the method research-action emancipatory; considering that

this method is closely compromised with the transformation of the organization and social practice. In which is proposed a critical process of intervention/reflexion, practical, of action and change, of ethic and compromise to the community service.

The analysis of usability, it is done in the University of Guadalajara, between a sample of 4 deaf students, and in the case of the serious game a sample will be done with 100 hearing students in the language center of CUCEA (Economic and Management Sciences).

3 Search Strategy

3.1 Sample Protocol and Registration

The revision protocol is being performed using the data bases, IEEE, EBSCO, Jstore, Sage Journals, the key words will be used under the criterion:

Word in abstract / just a Journal/ period 2013 - 2017 / language. Based on this, the research matrix will be built. (This one can be viewed in this web page link http://bit.ly/2sqeBI3).

This brings the total of 3,743 relevant and timeless articles in the four data bases mentioned previously. This quantity is high. In advance research, if we cannot make that by searching using key words, it means that the article is centered on this theme. Hence the need of using Cluster analysis, which allows us to find the whole possible homogenous groups, gathered by their probable similarity, by measuring their Euclidean distances. (The distance matrix is constructed using the number of documents found in each database. Distances and iterations are calculated using the SPSS software).

The dendrogram (Figure 1), obtained after having applied the k-means with five interactions, helps create a second search by groups in which they were called:

a) Emerging Technology,

b) Means of communication,

c) Possible means,

d) Education.

Through the K-means iterative partition, the compound groups are examined in each partition to be subdivided and give way to new groups. The decision on the cluster number is represented by the evolution of the mergers. After that, groups are separated to form search chains, reducing the probability of finding duplicate investigations in each database, due to the groups formed.

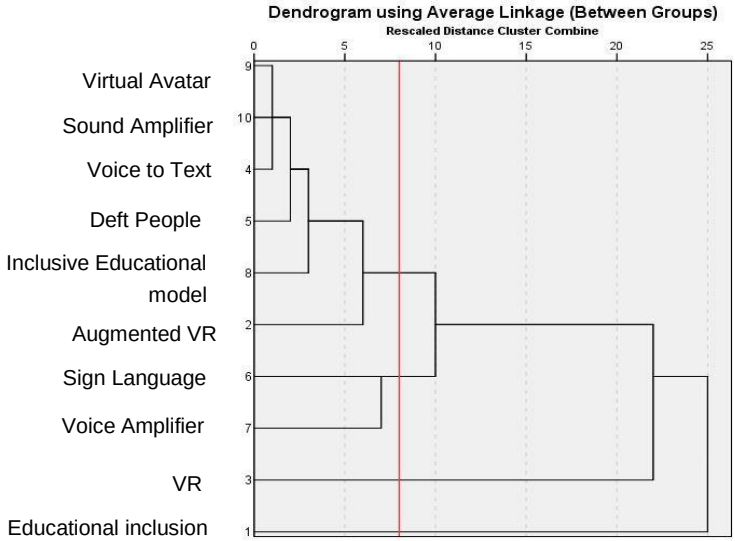


Fig. 1. Dendrogram: Graphical representation of the hierarchical classification, obtained by applying the k-means cluster method in SPSS.

3.2 Search Strategy

The criterion for the second search is based on the formed groups, creating search starting with the combination of their terms and logic connecting AND" and "OR". Table 2.

Table 2. Formed chains based on groups formed in the dendrogram.

	Search chains
1	((Virtual avatar) OR (Sound amplifier) OR (Translation of voice to text)) AND ((Deaf people) OR (Deaf))
2	((Virtual avatar) OR (Sound amplifier) OR (Translation of voice to text)) OR (Deaf people)) AND (Deaf) AND (Inclusive education model)
3	((Virtual avatar) OR (Sound amplifier) OR (Translation of voice to text)) OR (Deaf people) OR (Deaf) OR (Inclusive education model) AND (augmented reality))

3.3 Source Selection

907 articles are obtained under the chain of search selected in the same data bases and the same criteria: Abstract / just in Journal / period 2013 - 2017 / Language English.

The numbers of articles that belong to each group is the following: a) Emerging technology ten articles b) Means of communication 889 articles, c) Possible means five articles c) Education three articles.

For the selection process, the MacDonnell and Shepperd [6] strategies are applied in which consists in reviewing the titles and summaries to articles found to classify

them. For those articles that lack information in their summary, their introduction and conclusions were read in order to identify the required information.

4 Study Selection

The titles were reviewed, summaries and conclusions of each article found under this pre-selection 350 of 907 articles are distinguished in which 156 articles are pre-selected, 94 articles were discarded, since technology is not mentioned.

4.1 Data Collection Process and Data Extraction

The relevance of the article is considered based on the study field, the technological tools used and the contributions to inclusion model. It is also considered if the article is a case study or a theoretical proposal. Besides, the number of quotes that contain the articles and the impact factor when it was published are taken into account. Figure 2 (In the last two cases they were found using Scholar Google).

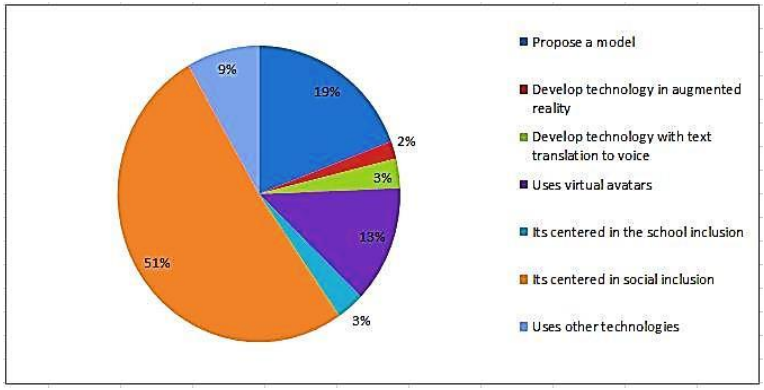


Fig. 2. Percentage of articles

4.2 Synthesis of Results

The 156 selected articles in the research themes were presented below: More articles are found centered on social inclusion, corresponding to the 51%. 60/80 were found doing it in case studies, 20/80 of theoretical method approaching the improvements in quality life of people with hearing loss through the existing technology or SL interpreter, only 19% talk about the creations of models to improve the quality of school education or mobility in the cities, 3/30 talk about models for smart cities and technologies that foster the life quality of their population with disabilities, 3% have developed technologies with translation from voice to text, 3/5 have made usage of it in pilot tests and 2/5 are already enabled (these include the subtitle in videos). 13% propose the usage of virtual avatars for the translation of SL, 2% have used augmented reality to enable usage of this technology for the learning of sign

language, 3% are centered on school inclusion through some technologies, 9% are centered on the usage of other technologies like antennas and cochlear implants.

Table 3. Most relevant technologies and characteristics for the study.

	Diffused in congress	Diffused in another medium	Paper	Description	Type	Country	Proposes model	Available	company or institution	Language
Dilo con señas	Yes	Yes	NO	LEARNING SIGN LANGUAGE	APP	México	YES	YES	JAGUAR LABS	SPANISH
Interpreter sign language for communication hearing impaired using image processing	Yes	NO	Yes	SIGN LANGUAGE TRANSLATOR	SOFTWARE	Colombia	NO	NO	INSTITUTION ANTONIO JOSE CAMACHO	SPANISH
Sign language recognition and translation with kinect	Yes	Yes	Yes	SIGN LANGUAGE TRANSLATOR	SOFTWARE	China	NO	NO	MICROSOFT RESEARCH ASIA	ENGLISH
Hand talk	NO	Yes	NO	SIGN LANGUAGE TRANSLATOR	APP, WEB PAGE	Brasil	YES	YES	HAND TALK	PORTUGUESE
Hablalo	Yes	Yes	NO	VOICE TO TEXT	APP	Argentina	YES	YES	MATEO SALVATO	SPANISH
LTCCS	NO	Yes	YES	SIGN LANGUAGE TRANSLATOR	PROPOSE	India	YES	NO		ENGLISH
Microsoft trasnlator	YES	Yes	YES	VOICE TO TEXT AND TRANSLATOR	APP. WEB PAGE	USA	NO	YES	MICROSOFT	66 LANGUAGES
Ava	NO	Yes	NO	VOICE TO TEXT	APP	USA	NO	YES	&AVA	ENGLISH

4.3 Risk of Bias within Studies

Furthermore, as a second search made in order to decrease the possibilities of publication bias, the bibliographic list of the selected articles in the main search were examined, with the objective of identifying studies not detected in the revision, it should be noted that it is possible that exist technology developed to access the information and communication designed for deaf people but research has not been done about the impact or viability in society which could cause a problem while trying to study the progress of the adapted technologies.

5 Results of Individual Studies

The existing technologies for the inclusion of deaf people in the different areas of an smart city are analyzed, the following are the most impact: sign language [7], "interpretive sign language for hearing impaired communication through processing images "[8], recognition and translation of sign language with Kinect [9], talk about it, hand talk [10], ava [11], signpost, whose specifications are shown below in Table 3.

5.1 Evidence Summary

The results obtained in this present job, allow demonstrate the existence of scientific literature related to the implementation of technologies that assist the access to communication and information for people with hearing loss disability, defined to promote the interest of research. The analysis of the quality of the articles that were included in this systematic revision had helped to sustain the scale of importance that all the scientific community of each of the studies selected. Although there are not many studies that directly linked the deaf people with the augmented reality to provide accessibility in smart cities specifically, the research work is highlighted done by Mizaei [12] and Nikita [13]. Being this a divided water to link the usage of technology with the simultaneous translation from voice to text.

5.2 Speech View Application

The purpose of developing a Speech View application is to have an intuitive UI for users with hearing disability, which provides them access to information in their environment; it is about augmented reality, simultaneous translation from voice to text / text to voice and QR codes, (for reproduction of videos in Mexican sign language LMS, both informative and for the basic learning of this); to be used in work environments, equipment and customer service, typical of an smart city and validated by the user.

Commonly, when a simultaneous translation is made, a person (stenotipista) writes what it is spoken. With help of computers and a projector screen, there is also the alternative of having a sign language interpreter. This is usually used in congresses or mass events. Due to the high costs and the small number of specialists in these issues, other alternatives should be analyzed, such as the one proposed in this research, to promote access to information without depending on an SL interpreter at all times. With the help of the VR application, information can be projected directly from the computer or cell phone.

Description of the Application Service

The user activates the voice-to-text service, having pre-configured the translation language. Using your phone camera, you can see the environment and expect a person's dialogue to be presented as a subtitle. At any time, the person can activate the text-to-speech option, where the user enters a dialogue via the keypad of the phone or tablet. By pressing the "talk" button, the text will be dictated by a male or female voice, depending on the configuration. These functions require connectivity. The data consumption must be measured in the usability tests to improve the user experience. In case the user does not have internet, an error message must be displayed in connection.

The processing capacity depends on the data and audio quality, so it is recommended to use a stable connection. This service is paid in packages with different costs (so you must attach the option to purchase characters in the menu).

The application is programmed in UNITY using C# and the APIS by IBM Blumix WATSON Figure 4, which have three main functions:

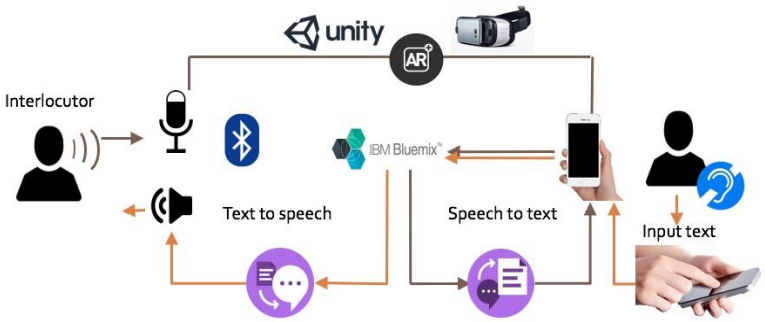


Fig. 3. The translation from voice to text using the APIS by IBM Blumix WATSON.

1) It plays videos in sign language, activated by QR codes loaded in the database to show relevant information:

Videos reproduced describe a process, instructions or relevant information from government agencies, hotels, restaurants, and they are unique. This service is for all those who request or agree to place a QR code in their establishments or offices. Your video is recorded by certified interpreters and is provided free of charge. The video is uploaded to the system database, so a QR code will be sent by mail so that your users can display the video in SL.

This service requires data, to enable access to the database and streaming access to the videos. This service is free for the application users.

2) Basic dictionary of sign language:

Through a printable format on the official page, the user can download a series of words that act as activation codes to access the content of each of these words; the user, when one of these words approaches the camera of his cell phone or tablet, can observe an image that clicks on a visual and auditory action, followed by a video that explains what the sign is in SL.

3) Configuration specifications:

Users can register and generate a password to access the applications, if they wish to acquire some of the packages proposed in the character acquisition plan (for the function Text to Speech and Voice to Text). Users can change the language of the voice to text / text to voice application from Spanish to English, but all other functions remain in Spanish. Users can change to AR lenses mode. This allows you to split the screen to use this option.

Prototype

The prototype of the interface is designed following the standards for IOS applications, taking into account the responsive design and colors. Icons are designed

considering two aspects: 1. The meaning of the application 2. The consideration of the target audience, since, according to Fajardo [14], icons can represent certain actions reflected in figures, since it is likely that one icon is easier for deaf people to understand than the other. That is, the icons are created from real elements, which are part of the culture of the selected users. This can improve the construction of icons that communicate what is intended: a good visualization of information to obtain a correct mental model.

According to Sánchez [15] the heuristics evaluation is an ideal evaluation method in the development of tools for people with disabilities as the usability is given by inspection, created by the expert judges from a previous principals established.

5.3 Limitations

The future limitations that can be presented when accessing, proving, using and validating the adapted tools, these are linked to three factors: the first one is the language for which it was built, second will be the availability from the developer to share it, it could be under development or the Project was dropped, third in case the usage of SL there is a risk that just is functional for the country that it was developed because the SL is different in every region. In addition, the majority of technologies developed do not seek the creation of models that can be replicated, since they were designed solely as a business or in some case were only in research prototypes and is not considered to hearing people as an important factor in the teaching/learning of the LS to make them partakers of the deaf culture.

On the other hand, the autonomy of the cellphone batteries play an important role in the augmented reality applications as the usage of the camera demands a high battery consumption, although this is an external problem from the application, this is a limitation of the battery type that we currently have today.

6 APP Usage Results

In the usability testing (in the first version), two focal groups were established where a qualitative semi-structure interview is performed with an inductive focus, in the first one the listener people isolated from any noise to interpret with the application what they were talking about, the second group of seven deaf people that had to interpret with sign language what they spoke about and written in text by the app.

The translation in real time is functional in a 65% in the Spanish language and 85 % in the English language (This is because of the characteristics of the Watson service), data that is extract from the lecture of a key text in the screen is display the probability that the phrase get close to the exact translation (This value is given by the IBM Watson service), this allows the user know if the translation was done in the correct way. Below is a table showing the main problems and advantages found in the tests:

Table 4. The codes that were more saturated in the interviews.

Control Group	Advantages	Disadvantages
Listener People	Fluent when reading and gets more accurate when describing the content of the test dictated and translated to text in the app.	The majority of them mentioned that they could not put attention in the text and the person at the same time. When using the augment reality glasses they fill dizziness in some occasions.
Deaf People	They could visualize the text and the person at the same time, this is an important factor in the communication for deaf people; they considered it as a good complementary tool see his/her interpreter and the text to fully understand the Spanish language and found the app very useful.	The velocity of reading is not the same as a listener person, in some cases the meaning of the word was unknown to describe it in the sign language, the wrong words transcribed by the app change the meaning of the context in some cases.

During some testing, a test was done in Xcode for IOS operation system, using a cellphone with Chip A9 with 64 bits of Architecture, rechargeable lithium-ion battery, this test shows that 79% of the resources are required of processing of this equipment when using the service voice to text, vuforia and the camera activation, refer to Figure 5 below. A problem origin with the battery performance, because of the camera usage and also with the continues data that is being sent and received, this application demands a high impact in the energy consumption of the equipment.

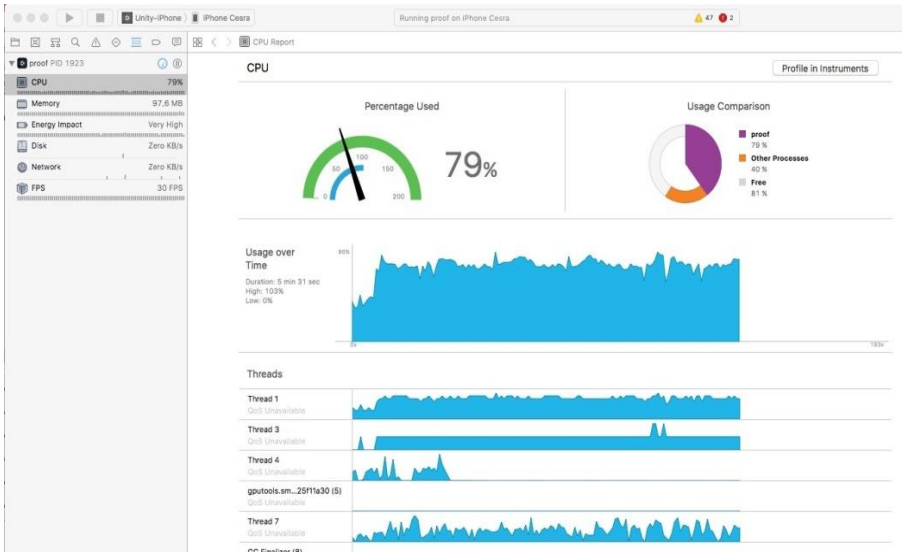


Fig. 4. CPU Test in XCODE when running the application.

Fig. 5. Battery test performance in XCODE when running the application.

7 Conclusion

In some cases universities had adopted a process of inclusion in their deaf students with support of sig language interpreters, technologies of communication, and others that are not contemplated or the budget is limited this is a problem of inclusion because it is supposed in smart cities every person can have access to the technological resources available to improve the communication, regardless of the condition of each person; however, this access currently today is limited to deaf people, because there is not sufficient technology that allows the simultaneous translation from voice to text or in signs language, the lack of services that get into interactive menus or interpreter, one of the circumstance is because of the lack of implementation and develop of technologies; achieve the accessibility in different environments without depending of an interpreter of SL will benefit all users and people that interact with services in the counter, schools, hotels, etc.

This revision shows the progress for the inclusion of deaf people has been slow but shows a favorable way, the new technologies provide a better outlook for future development, it is worth mentioning that in all moments that resources must be designed to incorporate the multiple visual strategies base on research and must be educational and with the best quality to obtain a positive impact in children. Golos & Moses [16]. Which takes us to question what it was questioned in the begging?

According to with was mentioned previously and from the point of view socio anthropological, it is important to understand the proper situations of the deaf community, recognized as a linguistic minority, the signs language user.

Therefore, development of tools for smart cities should be first priority, to respond to these needs of communication, the usage and interpretation of the information through the different methods and also usage of graphic information, rich in visual level, spatial and temporal, allowing the deaf person uses his/her cognitive domains to access the information and help in his/her learning process Rincón [17].

However, none of the studies found mention the importance of hearing people taking an open stance to learn sign language and access to deaf culture, in terms of a basic knowledge of SL to guarantee a society more inclusive.

At the same time investigate about the usage of technologies for the teaching / learning of Sign Language, re-define the way in which the reading and writing is transmitted to deaf people it turns in an area of opportunity to generate technological resources that warranty the access to information, education and better job positions for people with disabilities.

8 Future Work

The augmented reality, the virtual reality and the engines of voice recognition, currently are part of the new systems of accessing the information for education, geolocation, training, customer support or of amusement; provide a society alternatives to access the contents in an interactive way and experiential, improving with this the experience of learning, buying, managing, interact with virtual objects, and so much more applications, usage that can be also well use in the inclusion environment; to provide an alternative in the usage of technological tools like a support for deaf people inside the society, since it is possible to observe the interlocutor, an important factor in communication as human-to-human communication consists of words, non-verbal linguistic modifiers (paralanguage), and visual cues or gestures.

The growth in the research in this nature analyzed in this systematic revision are showing a positive trend to the augmented reality, virtual reality and the inclusion of people with any disability in the Smart cities.

Base on the systematic revision shown and the analysis of the existing technology (Looking for the optimization of it), undertakes from the Smart Cities Innovation Center at the University of Guadalajara the Project: “Technological-Social Model. Based in augment reality platforms for the inclusion of deaf people in the university class-room and the cities.” like a proposal for the innovation in the Smart Cities in social inclusion matter. In which has two main research:

1. Testing usability in an application developed with augment reality, translation from voice to text (using a sound filter since the recognition rate of a speech or speaker recognition system can decline to a lot by the influence of noise [18]) and incorporate videos activated with codes QR in SL as a means of access to information. (used Linguistic Knowledge for Machine Translation Evaluation as proposed by Samiksha [19]).
2. Explore the implementation of a contextualized off-line search engine that allows the users to continue using the application [20].
3. Measure the social impact of the Usage of serious game developed for the teaching of SL using user interfaces of gesture recognition and voice commands, As a means of access to the learning of basic SL.

4. Testing usability in an application developed with augment reality for the teaching of SL which displays objects associated with its meaning in Spanish (the word) and their sign language.

With the usage of augment reality to improve the level of visual communication allowing the users to access the gesture communication of its interlocutor, for instance, in the case of a classroom even if the teacher is writing when he is standing with the back turned the user will be receiving the translation and also see what the professor is saying, although this application can be used in any context of the city.

Searching that this technology be validated by the user; Because of this reason, statistics will be taken to analyze the percentage of translation errors, estimated time of usage, how easy is for the user create an activity, how fast can get used to its usage, how pleasant is using the interface (learning, efficiency, memorability, task error, satisfaction) and its utility (if the interface performs what is needed from the user) the objective is to analyze if the VR and the simultaneous translation allows the end user has access the information in an effective way.

Furthermore, take advantage of the technologies of virtual interfaces like Kinect (adapting the model of Speech to sign language of Caballero [21]) to create a serious game for the teaching/learning of SL and promote an inclusive social environment This can be used by people to learn basic signs, which may put in practice in front of a Kinect (as a way of validating if the sign that is being learned corresponds to the correct way to play it, analyzing regular correspondences between related languages [22]).

The serious game will be shared at the Language Center of the University Center of Administrative Economic Sciences of the University of Guadalajara and in some public transport of the city of Guadalajara (those with screens, in the area of operations of Digital Creative City), with the intention to measure the time that the user learns a sign and the interest in learning SL in a playful way by members of the society.

The model becomes important when trying to create a tripartite link between technology-society-institutional strategy, since it not only seeks to provide technology to people with disabilities but to consent to the society of the importance of the learning of SL to provide inclusive spaces; as well as the analysis of the institutional strategies that receive the deaf community to be replicated. Since the opportunity that provides access to the information opens the possibility of adding to the labor force to this social group, a force that has been wasted; for example, in Mexico, according to re-ports of the INEGI, 70% of the Mexican population with disabilities is economically inactive, The search for a replicable model also opens doors to provide opportunity to access an educational environment, without leaving behind the importance of having a sign language interpreter, but with technology, the user does always need to depend on him/her interpreter at all times (for the lack of certified interpreters in the institutions).

References

1. Bouzid, M.A., Khenissi, F., Essalmi, M., et al.: Using Educational Games for Sign Language Learning - A Sign Writing Learning Game: Case Study Learning Game. Educ. Technol. Soc., 19(1), pp. 129–141 (2016)
2. Ruiz, A.L.: ¿Educación inclusiva? Análisis del marco jurídico sobre el derecho de acceso y permanencia en la educación superior para las personas Sordas en la ciudad de Bogotá. Incl. Educ. Anal. Leg. Framew, 33, pp. 1–55 (2016)
3. Baker, P.M., Bellordre, C.: Adoption of information and communication technologies: key policy issues, barriers and opportunities for people with disabilities. System Sciences, Proceedings of the 37th Annual Hawaii International Conference on IEEE (2004)
4. OMS: http://www.who.int/disabilities/world_report/2017/es/index. html (2017)
5. Kitchenham, B., Charters, S.: Guidelines for Performing Systematic Literature Reviews in Software Engineering. Keele University and Durham University Joint Report (2007)
6. Shepperd, S.G., MacDonell, M.J.: Comparing Local and Global Software Effort Estimation Models Reflections on a Systematic Review. Empirical Software Engineering and Measurement (2007)
7. Dilo con señas: http://www.diloensenas.com/inicio (2016)
8. Greene, J.R., et al.: Electronic translator for assisting communications (2002)
9. Chen, X., Li, H., Pan, T., Tansley, S., Zhou, M.: Kinect Sign Language Translator expands communication possibilities (2013)
10. Hand Talk: Digital translation for Brazilian Sign Language. https://handtalk.me/ (2012)
11. AVA: https://www.ava.me/ (2017)
12. Mirzaei, M.R., Mortazavi, M. et al.: Combining augmented reality and speech technologies to help deaf and hard of hearing people. Virtual Augment. Reality, pp. 174–181 (2013)
13. Nagori, N.P., Malode, V.: Communication Interface for Deaf-Mute People using Microsoft Kinect, Int. Conf. Autom. Control Dyn. Optim. Tech., pp. 640–644 (2016)
14. Fajardo, I., Abascal, J., Cañas, J.J.: Bridging the digital divide for deaf signer users. In: 15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction, ECCE, pp. 1–37, ACM (2008)
15. Sánchez, J., Zapata, M., Jiménez, J.: Evaluación heurística de la usabilidad de software para facilitar el uso del computador a personas en situación de discapacidad motriz. Revista EIA (2017)
16. Golos, D.B., Moses, A.M.: Supplementing an Educational Video Series with Video (2015)
17. Rincón, M. L.: ¿Cómo la comprensión de lectura en estudiantes sordos se ve facilitada por el uso de tecnologías de la comunicación e información?, Rev. la Fac. Med., 63(3) Sup, pp. 83–91 (2008)
18. Lai, W.H., Yang, C.J., Wang, S.L.: Post-Processing for the Mask of Computational Auditory Scene Analysis in Monaural Speech Segregation. Computación y Sistemas, 21(4) doi: 10.13053 (2017)
19. Tripathi, S., Vineet, K.: Using Linguistic Knowledge for Machine Translation Evaluation with Hindi as a Target Language. Computación y Sistemas, 21(4) (2017)
20. Centelles, J., Costa-Jussà, M.R., Banchs, R.E., Gelbukh, A.: On-line and Off-line Chinese-Portuguese Translation Service for Mobile Applications. Computación y Sistemas, 18(3), pp. 603–610. DOI: 10.13053/CyS-18-3-2047 (2014)

21. Caballero, S.O., Trujillo, F.: 3D Modeling of the Mexican Sign Language for a Speech-to-Sign Language System. Computación y Sistemas, 17(4). DOI: 10.13053/CyS-17-4-2013-011(2013)
22. Fischer, A.K., Vreeken, J., Klakow, D.: Beyond Pairwise Similarity: Quantifying and Characterizing Linguistic Similarity between Groups of Languages by MDL. Computación y Sistemas, 21(4) (2017)

ISSN 1870-4069

Applying Blockchain to Supply Chain Operations at IBM Implementing Agile Practices in a Smart City Environment

Jesús M. Montes[1], Victor M. Larios[2], Manuel Avalos[1], Cecilia E. Ramírez[3]

[1] International Business Machine Corporation, Guadalajara, Mexico

[2] University of Guadalajara, Guadalajara, Smart Cities Innovation Center CUCEA, Mexico

[3] Division of Industrial Process, Technical University of Jalisco, Guadalajara, Mexico

{jesusmd, jmavalos}@ mx1.ibm.com, victor.m.lariosrosillo@ ieee.org, cecilia.ramirez@ utj.edu.mx

Abstract. Nowadays, performance and promptness in the production line are significantly relevant; solutions are being delivered each quarter lead by the evolving innovation of technology which is moving at a fast paste. Taking into consideration concepts from Smarter Cities, Smart Contracts, and Blockchain, we will include these tools in the daily operations of International Business Machines' supply chain to reduce these transaction times radically, implementing agile practices. Fundamentally using IBM Bluemix®, which is a fast-growing Platform that provides support for development organizations by defining various users and roles in the cloud environment. IBM Bluemix® provides an extensive inventory of runtime and services that enterprises and developers can use in cloud applications without requiring significant development effort. We as developers can quickly build applications with a choice of several runtimes as can be needed in Supply Chain operations. Furthermore, as the workloads increase, applications can promptly scale on-demand to add more memory or processing power to ensure an optimal block chain experience application. Supply Chain operations can now shift their client-side code into a cloud system environment to leverage service integration, internal confidential activities and services, cloud data and the infrastructure capabilities. It will allow operations to be more productive and highly available and flexible, achieving efficient and sustainable development based on the analysis of real-time information shared by stakeholders and connected devices in a Smart City Environment. IBM is seen as a leader in blockchain, from recent surveys it is revealed that IBM is leading the way of blockchain technology, with more than 40% of respondents raking it number one. International customers are working with IBM to introduce Blockchain to their businesses. These customers are currently testing out IBM's blockchain technology to help small/medium sized enterprises become more evident and attract investment. Another result of working with IBM is that the customers

can track and manage their paper trail from their global shipping process. Another technology that is thriving is artificial intelligence, and in one way or another can influence blockchain, taking into consideration scalability; blockchain grows at a steady pace of megabytes per minutes, here is where AI can introduce decentralized learning systems or new data sharing techniques to make a more efficient system. Another relationship between Blockchain and AI is their effectiveness with security, but AI makes an excellent ally for blockchain to guarantee secure deployments of applications, specifically with the given fixed structure of the systems.

Keywords: blockchain, supply chain, smarter cities, smarter contracts, IBM Bluemix®, agile practices, artificial intelligence.

1 Introduction

The accomplishment of Smart services by Supply Chain Operations which are the essential elements of a Smart City since they sustain the realization of urban "intelligence" regarding people, industry, economy, governance, environment, mobility, health and living. Smart services aim to improve the quality of life within a city and in this respect to improve "livability" [1]. The types and purposes of Smart Services cannot be predetermined since they are the outcome of innovation, which cannot be decided either without the management of sensitive information that must be kept in log files on an untrusted computer. Service, Supply Chain Operations and Manufacturing areas embraced in investigating manufacturing for design, planning, control, and improvement of services, including distribution operations. Inside the production line, topics include the preparation of production, process design, projects, and personnel, capacity planning, production control, resource planning, quality control, facility layout, materials handling, and new product and process development. Within services, topics include service system design and oversight, service quality, capacity planning, workforce planning, and demand management. Supply chain operations include studies of replenishment policies, inventory and logistics and distribution network design and administration, purchasing, supply management and channel coordination [2, 4]. If an attacker captures Supply Chain Operations data from a machine, it is imperative to guarantee that they will gain little to no information from the log files and to limit their ability to corrupt the log files.

For IBM, Supply Chain Operations are managed to service focusing on the margin-develop procedures and high-cost regions of our clients' value chains, including product development, Supply Chain Planning, Procurement, and Logistics. We deliver value-based, industry-oriented offerings that contribute to operational improvements and financial gains as depicted in Fig. 1.

Our Clients value chains in a Smart City are frequently merge to an entity named commerce that enables businesses to make Smarter-souring decisions in real time and give their customers the best buying experience possible. We provide Smart Solutions that help organizations deliver engaging customer experiences by intelligently processing and fulfilling products orders most efficiently and profitably possible. For instance, retailers, wholesalers, manufacturers and software developers today are

challenged by Smart Commerce to meet the demands of empowered consumers; who require advanced delivery methods such as "buy online and pick-up in store." E-Commerce is under increasing pressure to offer these flexible options across more channels to meet their customer's expectations and our client's expectations becoming Smart Business. Without Blockchain into the complete Supply Chain Operations-Management, businesses are forced to make snap decisions based on the sensitive information. Well known in the Informartion Technology field as granular data, resulting in inefficient fulfillment processes, shipping methods, and inventory positions far along as should be in a Smart City Environment being disadvantageous in the broadest commerce industry as seen in Fig. 2.

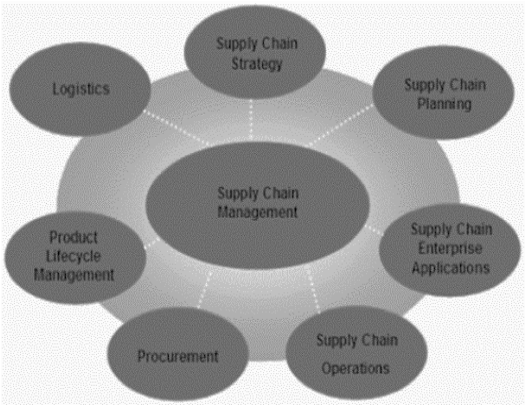


Fig. 1. IBM Supply Chain Management long established [5].

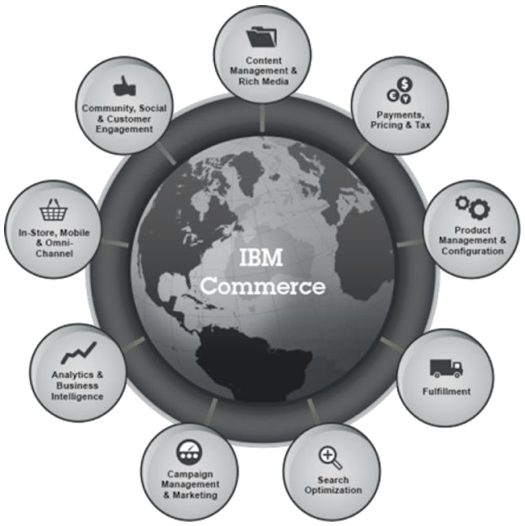


Fig. 2. IBM Commerce Supply Chain Operations in Smart Cities [6].

Today, involved across the world are smart cities with the contribution of smart solutions, developed by multidiscipline sciences, which vary from politics and government, to health and education, or to construction and city facilities (water, heat, energy, transportation) [7].

As a result, this compound smart city context has engaged almost all traditional industries that require applying Blockchain in their Supply Chain Operations; blockchain is mostly a database, an extensive network, known as a distributed ledger, which allows any person with access to view and participates in the records ownership and value. The system is updated and verified through consensus of all the parties involved. When adding something it cannot be altered and if it looks valid to everyone, the update is approved [8].

The Blockchain arises as an innovative tool which proves to be useful in some application scenarios. Some large industrial players, such as IBM, Microsoft, Intel, and NEC, are currently investing in exploiting the Blockchain to enrich their portfolio of products. Some researchers and practitioners speculate that the Blockchain technology can change the way we see some current online applications; though it is still early to tell for sure, the Blockchain will fuel substantial changes to a large number of products and will positively impact the digital experience of many individuals around the world [9].

Regarding governance, the growth of smart cities is increasing the use of Information and Communication Technologies (ICTs), that require ameliorating political participation implement public policies or providing public sector services. By doing so, traditional processes of governing do not seem to be appropriated to manage the cities, and therefore new and innovative forms of governance are needed. Because it appears to be clear that transforming public processes will only be achieved with better urban management applying Blockchain to Supply Chain Operations to establish a new governance model for smart cities named as "smart governance." Consequently, governments in smart cities are called to play a vital role in promoting and developing smart cities, using efficient Supply Chain Operations to improve a Smart City Environment [10].

2 Background

Throughout the centuries cryptosystems have been used by the military and by the consular services. The nowadays widespread use of computer controlled communication systems in industry or by civil services, often asks for special protection of the data using cryptographic techniques. From the time when storage, and later data recovery, is viewed as a transmission of this data in real-time, we shall always use the term transmission when discussing a situation when information is stored and transmitted [11].

Security systems that protect privacy, secure electronic commerce transactions, cryptography, do not directly prevent fraud. Instead, they detect attempts at fraud after the fact, provide evidence of that fraud to convict the guilty in a court of law, and assume that the legal system will provide a "back channel," currently renowned as blockchain to deter further attempts [12].

Satoshi Nakamoto conceptualized the first blockchain in 2008 and implemented in the following year as a critical element of the digital currency known as Bitcoin, where it serves as the public ledger for all transactions [13].

Over the use of a peer-to-peer network and a distributed timestamping server, a blockchain database can befall management autonomously.

The creation of the blockchain technology for Bitcoin made it the first digital currency to solve the double spending problem. The Bitcoin scheme has been the inspiration for other applications [14].

The blockchain is an essential topic in supply chain management research and practice. For manufacturing companies like IBM, one of the most pressing challenges associated with blockchain in supply chain management is that they typically do not possess sufficient information on what is occurring in the compound supply chain environment, as demonstrated by numerous incidents lacking blockchain. Using eight in-depth case studies across four industries and elaborating on information processing theory, we identify three forms of applying blockchain related uncertainty that each firm faces in the supply chain. Refer to as supply chain uncertainty, source uncertainty, and task uncertainty. The study displays that the extent to which these changes translate into information processing needs depends on a newly identified boundary condition labeled uncertainty intolerance. On the management of such information processing requirements, previous research has pointed primarily at matching information processing requirements with fitting information processing capacity and secondly at mitigating information processing needs with corrective measures. In doing so, this research exemplifies how applying blockchain in the supply chain management may ultimately turn from an adjustment to a firm's daily business to a decisive factor for shaping future supply chains.

Also, the study constitutes a nascent step to elevate information processing theory to the supply chain level. It may contribute to establishing Blockchain in Supply chain operations in Smart Services that a smart city can deliver: Smart Water, Smart Energy, Smart Transportation, Smart Healthcare, Safety/Emergency, Education and Tourism, Smart Waste Management, Smart Buildings, e-government and e-business supported by IBM Bluemix services and implementing agile practices. In the software business, the Agile movement has produced much discussion since its beginning at the 2001 meeting in Snowbird, Utah. Although Agile methods have been applied quite widely in various software development environments, the field of Blockchain in Supply Chain Operations has not been seen in agile projects [15].

3 Theoretical Framework

The protection of classified information against unauthorized access or fraudulent changes has been of prime concern throughout the times. Current communication techniques, using computers connected through networks, make all data even more vulnerable to these threats. Since its beginning in 2009, Bitcoin’s blockchain has fueled innovation and some new applications, such as smart contracts, have been intended to take advantage of the blockchain [16]. Further concerns have come up that were not relevant beforehand, e.g., how to append a digital signature to a digital document in such a manner that the signer cannot deny later that the paper was signed by him/her [17].

Cryptology addresses the above issues. It is at the base of all information security. The techniques engaged to this end have become increasingly mathematical of nature. The study of cryptosystems, cryptology, can be subdivided into two regimens. Cryptography deals with the design of cryptosystems, while cryptanalysis considers the breaching of cryptosystems. These two aspects are closely related; when setting up a cryptosystem the analysis of its security plays an important role [18].

A blockchain database contains two kinds of records: transactions and blocks. Blocks held batches of hashed logical operations and encoded into a Merkle tree as depicted in Fig. 3. Each block comprehends the hash of the previous block in the blockchain, linking the two. Variants of this format were used previously, for example in Git, and it is not by itself sufficient to qualify as a blockchain. The linked blocks form a chain; constant process ratifies the integrity of the previous block, all the way back to the source block. Division structures block into layers defined by protocols [19].

Some blockchains create a new block as rapidly as every five seconds. As blockchains phase, they are sometimes separate blocks that can be validated concurrently, creating a provisional fork. Now with a secure hash based history, all blockchains have a specified algorithm for storing different versions of the history so that one selected with a higher value over others. Unselect blocks for inclusion in the chain are called orphan blocks [20].

Peers supporting the database do not have the same version of the history. Instead, they keep the highest scoring version of the database that they currently know.

Each time a peer receives a higher scoring version, then generally the previous version with a single new block added, is extended or it overwrites their database and retransmits the enhancement to their peers. There is never an absolute assurance that any particular entry will remain in the best version of the history always.

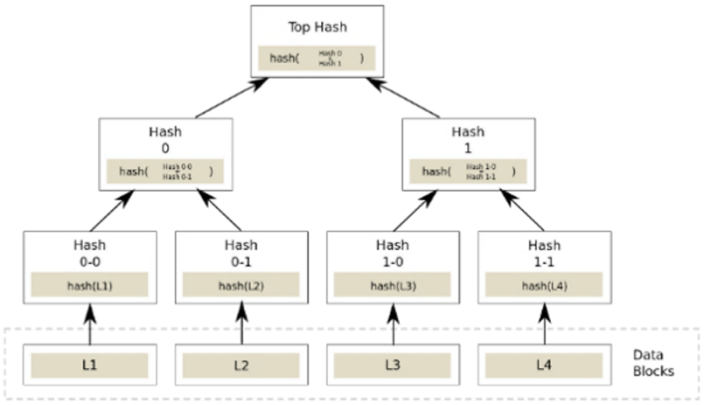


Fig. 3. An example of a Merkle tree [21].

But cause blockchains are characteristically built to sum the score of new blocks onto old blocks, and there are motivations to only work on extending on new blocks instead of overwriting old blocks. The possibility of an entry becoming superseded goes down exponentially, built on top of it are more blocks, eventually becoming very small.

For instance, in a blockchain using the proof-of-work system, the chain with the most common proof-of-work is considered the valid one by the network all the time.

In practice, some methods can demonstrate a sufficient level of computation. Within a blockchain, the calculation is carried out redundantly rather than in the traditional segregated and parallel manner [22].

4 Methodology

For this endeavor, the tasks will be broken down into five process groups; initiating, planning, executing, monitoring and controlling, and closing. In which we will apply knowledge, tools, skills, and techniques to launch activities to meet the project's monitoring & controlling and closing phases [23]. Diving the activities within these groups will give order to our project‘s research and will allow us to work on any algorithm needed one stage at a time. For this project to be successful, we come across the primary challenge of achieving all of the project goals within the given constraints.

Typically the critical restrictions are scope, time, quality and budget within the projects. These will be our primary focus in the initiating and planning phases, as long as we attend to these requirements in order and fashion, we should have a bright and smooth start in our journey to implement this project. The secondary, and probably more ambitious task is to optimize the assignments of all the essential contributions and apply them to meet predefined objectives, which are set as milestones to acquire a straightforward implementing of agile practices. The practices of continuous integration and test-first programming appear to be practices that can make measurable improvements in the quality and development time for software projects, focusing in [15]:

- Value individuals and interactions over processes and tools.
- Value is working software over comprehensive documentation.
- Value customer collaboration over contract negotiation.
- Value is answering to change over following a plan.

Once we have our path set out, using the project management discipline, we will then need to integrate the Blockchain technology in our supply chain system, for this endeavor, we will work with the IBM Bluemix® platform. Defining IBM Bluemix® as a cloud platform as a service (PaaS). It supports several programming languages and services along with integrated Development Operations to build, run, deploy and manage applications on the cloud. Bluemix® is founded on Cloud Foundry open technology and runs on SoftLayer infrastructure; with Bluemix® we are working with our ledger on the cloud platform, looking to offer solutions for this case study and drive business value with services, applications, and infrastructure. With the cloud resources at our disposition, it will be possible to bring together the data sources for this environment, scale systems, and incorporate customized services to drive the business results quickly and within the budget.

5 Experimental Results

Smart Cities require efficient management of a vast number of monitored data regarding processing, storing and analysis is an important matter to deal with for its large-scale adoption in Supply Chain Operations. In a Smart City Environment, any industry is entering a new era of safe and fast transactions with Blockchain technology.

Though they require a robust and scalable high-performance computing and massive storage organization for real-time processing and storing of the data as well as analysis (on-line and off-line), of the processed data under context using integrally complex models to extract information from specific supply chain operations. The adoption of blockchain can lead to performing more enterprise operations, letting the Supply Chain be the turning point. We can see blockchain applied to all departments, and being offered as a solution to the world.

A goal is to document and track all the activities of an average business day in the supply chain, and build our case study; and eventually, apply the improvements produced by the Blockchain technology. The results will be presented in different phases as depicted in Fig. 4.

Adapting planning to business needs, supply chain operations and market situation expectations change so rapidly that the preparation must be verified and modified on an ongoing basis as is shown in Fig. 4, where development needs to be very reactive where short-term goals change often.

Evolutionary development is something nobody can afford to spend months on, to deliver complete applications. Quick time to supply chain is critical even with limited functionality. Following the initial delivery, the Bluemix application goes through an ongoing refinement to meet all required features and to satisfy continuous flow of the supply chain operations requirements.

In the case of an early delivery, we need to advertise and deliver the new service very quickly before competition enters the game. There will be a delivery as soon as there is a Beta quality version. Even though it will not have the full set of functionalities, it will let customers experience some of the benefits of the application and give a feel for what is coming.

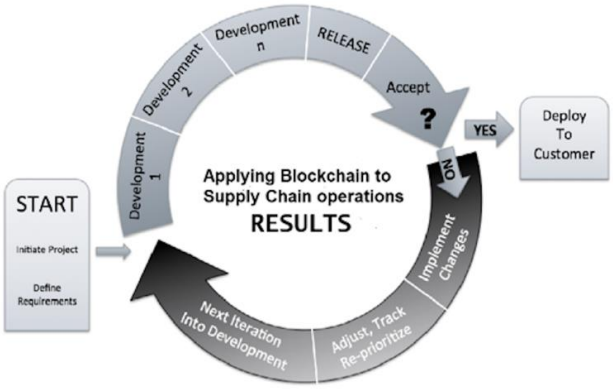


Fig. 4. Applying Blockchain to the supply chain operations [24].

Continuous improvement, starting with the Beta Bluemix application, we need to improve the use on an ongoing basis. Bi-weekly or monthly updates to the application will be the goal. Implementation of the updates requires working in non-downtime fashion for business continuity. It will allow for quick defect fixes and provides an ongoing commitment to the supply chain operations.

A. Impact

To quicken the times of distribution and take advantage of the internet, to have faster online services. These transactions would consist of orders and contracts, making them available to all interested parties instantly and with a legal validity.

Also, this project will aim to impact the process cost, by cutting time in order processing and reducing inventory space with better tracking and logistics that smart cities prioritize establishing local economic growth and taking care of communities with special needs (unemployment, youth, older adults, accessibility and young families). Environmental protection and climate change follow, even if we consider that smart water, smart energy, and smart buildings deal with the same issue. Cities smart ecological performance needs to increase to achieve an urban sustainability vision.

6 Conclusions and Future Work

IBM's teams will be interacting with one another, collaborating within the supply chain, Bluemix®, and Blockchain, ironing out details, obtaining agreements, asking questions, getting to know each other, and developing trust and bonds. To have face to face interaction, nonetheless, taking into consideration that the most challenging task is still when Supply Chain operations have a high dependence with overseas partners, adding distiance and time barriers to the routin operations. In this new agile world, it is essential to reduce transactions times, be more confidential than we used to have been in the past. From experience learned from how to improve Supply Chain processes applying Blockchain, we can too review and enhance Smart Cities services and procedures as the primary contribution to the field.

This project is ambitious and complex to tackle, there is a lot of theory out there, and very few companies that have successfully applied blockchain to their activities. The project will be a significant challenge and will require identifying entire departments within the Supply Chain process to start with the investigation and application of a case study with project management discipline.

To migrate the supply chain transactions onto Bluemix, our cloud platform, education, and skills will be an obligated requirement. Bluemix® is a fast-growing Platform as a Service (PaaS), in the cloud environment. We will need to get onboard to one of the most critical software development paradigm shifts in the subsequent eras. Online courses offered at IBM, and in-person classes of the subject matter are available nationally and internationally. By the end of the project's research, we have to demonstrate and understand why Bluemix is all we need to prototype, develop and deploy our applications to the supply chain production environment.

Which can be a natural focus on smart service standardization that will require competitive standards and suggest technical specifications and guidelines for similar solutions development.

Among the elements that these rules try to identify concern smart city services, which conveys the products/services and components that can generate a smart city ecosystem, such as smart transportation, smart water, smart energy, which are the most common prioritize smart city areas.

Acknowledgements. To the IBM experts for their advice. To the University of Guadalajara's academic researchers for tutoring, guidance, and their collaboration. To the Technical University of Jalisco for its knowledge support and availability.

References

1. Ben, A. et al.: Industrial and Business Systems for Smart Cities. Proceedings of the 1st International Workshop on Emerging Multimedia Applications and Services for Smart Cities (2014)
2. Duenyas, I., Jing-Sheng, S.: Manufacturing, Service, and Supply Chain Operations. Operations Research (2009)
3. Busse, C., Meinlschmidt, J., Foerstl, K.: Managing information processing needs in Global Supply Chains: A prerequisite to sustainable Supply Chain Management. Journal of Supply Chain Management (2017)
4. Adiseshu, H., Lakshman, T.V.: On the Security and Performance of Proof of Work Blockchains. Proceedings of the 15th ACM Workshop on Hot Topics in Networks (2016)
5. https :// www-07.ibm.com/services/hk/bcs/supply_chain_ops.html (2016)
6. http:// www.decidesoftware.com/integrated-Omni-channel-supply-chain-ibm-jda/ (2014)
7. Ghassan, K: On the Security and Scalability of Bitcoin's Blockchain. Conference on Computer and Communications Security (2016)
8. Adiseshu, H., Lakshman, T.V.: The Internet Blockchain: A Distributed, Tamper-Resistant Transaction Framework for the Internet November. Proceedings of the 15th ACM Workshop on Hot Topics in Networks (2016)
9. Maxwell, D., et al.: Effing' the ineffable: opening up understandings of the blockchain. Proceedings of British HCI Conference (2015)
10. Sakamoto, M., Nakajima, T.: Gamifying smart city services to flourish our society. International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers (2015)
11. Miller, A., et al.: The Honey Badger of BFT Protocols. Conference on Computer and Communications Security (2016)
12. Decker, C., et al.: Bitcoin Meets Strong Consistency. Proceedings of the 17th International Conference on Distributed Computing and Networking (2015)
13. Gervais, A., et al.: On the Security and Performance of Proof of Work Blockchains. Proceedings of the ACM SIGSAC, Conference on Computer and Communications Security (2016)
14. Bogner, A., et al.: A Decentralised Sharing App running a Smart Contract on the Ethereum Blockchain. Proceedings of the 6th International Conference on the Internet of Things (2016)
15. Matthews, C.E.: Agile practices in embedded systems. SPLASH '11, Workshops: Proceedings of the compilation of the co-located workshops on DSM'11 (2011)
16. Judmayer, A., Weippl, E.: Condensed Cryptographic Currencies Crash Course (C5), Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (2016)
17. Henk, C.A., van, T.: Fundamentals of Cryptology: A Professional Reference and Interactive Tutorial. Kluwer Academic Publishers (2000)
18. Underwood, S.: Blockchain beyond bitcoin. Communications of the ACM: 59(11) (2016)
19. Faísca, J.G., Rogado, J.Q.: Decentralized Semantic Identity. Proceedings of the 12th International Conference on Semantic Systems (2016)
20. http:// www .marcoyuen.com/articles/2014/02/01/merkle-tree.html (2014)

21. Bogner, A., et al.: A Decentralised Sharing App running a Smart Contract on the Ethereum Blockchain. Proceedings of the 6th International Conference on the Internet of Things (2016)
22. Norta, A.: Conflict-Resolution Lifecycles for Governed Decentralized Autonomous Organization Collaboration. Proceedings of the 2nd International Conference on Electronic Governance and Open Society: Challenges in Eurasia (2015)
23. Project Management Institute: A Guide to the Project Management Body of Knowledge (PMBOK Guide). Global Standard (2013)
24. http://www .code4nord.com/agile-methodology/ (2014)

ISSN 1870-4069

Multivariate Analysis of University Student Engagement through Visualization Techniques

Erika Yunuen Morales Mateos[1], María Arely López Garrido[1], José Alberto Hernández Aguilar[2], Carlos Alberto Ochoa Ortiz[3], Oscar Alberto González González[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, arely.lopez, alberto.ochoa}@uacj.mx, oscar.gonzalez@dais.ujat.mx, jose_hernandez@uaem.mx

Abstract. The objective of this work is to apply visualization techniques for the representation of multivariate analysis results. The techniques were used to publicize the results of student engagement at universities of careers in technology at a university in the south of México. To measure student engagement, the UWES-S instrument was used, composed of vigor, dedication and absorption dimensions. The graphs presented in this paper are: Graphics Matrices, GGE Biplot and Chernoff faces, which were developed with the R language. It concludes on the utility that visualization techniques have in the multivariate analysis, given that in this it is studio was able to identify how students interact (genotypes) with the environment (variables of student engagement), identifying that the majority presents a stable level of student engagement. With the Chernoff faces, it was possible to visualize through the facial features in detail the presence of student engagement as a result medium level.

Keywords: multivariate analysis, visualization techniques, biplot charts, Chernoff faces, student engagement.

1 Introduction

Populations are commonly characterized by a simultaneous manifestation of many processes, where results at best can be measured, which is why it is necessary to have methods capable of considering at the same time, several characteristics to define in an integral way to an object of study [1]. In this way, the need to apply multivariate data analysis arises.

Similarly, there are techniques of visuals for multivariate data that allow exposing what happens in a set of data that is represented by several dimensions. The presentation of data using visual techniques helps to understand the behavior of data with a look, even complex data, so that information analysts can save time when making judgments about patterns, trends, variability, among others, of data. One of the most important advantages of using visual techniques to represent data is that they can display features that otherwise would have been complex or impossible [2].

Visualization techniques for this type of data challenge the combination of statistics and computer science. For the development of this type of visualizations for multivariate data the R language is recommended, which is an integrated collection of software services for data handling, calculation and visualizations through graphics. Software R is free under terms of the GNU General Public License of the Free Software Foundation as source code [3].

There are research areas where it is practical to use multivariate analysis and above all the representation of results through visual techniques, as it facilitates the interpretation of results for decision making.

One of the areas of research where this type of representations is useful is the educational one, an example being the universities that are nowadays in charge of covering their quality indicators operating with educational models that require the integral formation of the students to increase the quality of their educational process, so it is necessary to know the information of their students, which can be interpreted in the best way, especially by the academic authorities. Taking into account this new model where the student is responsible for his own learning, where he is the one who builds the knowledge, who learns, it is necessary to know the different aspects such as that of psychological linkage with studies [4] or also called academic engagement, that is, how dedicated, engaging and vigorous students are.

There are some works where graphic representations have been developed to know the behavior of the data in educational areas at the license level, such as radar graphs where it is possible to visualize the comparison of the data in a fast way, facilitating the interpretation of the data through colored lines that form polygons [5, 6]. Likewise, research has been found related to academic performance and the elements that influence university students where data analysis techniques have been applied [7, 8] or techniques for evaluating the results obtained [9].

The objective of this work is to develop multivariate analysis of student engagement presenting the results through visualization techniques such as a graphical correlation matrix, GGE Biplots type graphs and Chernoff Face graphs. These graphs allow explicit visualization of the associations between variables, analysis where the main components can be observed, as well as the identification of the characteristics of the cases involved in this study.

2 Materials and the Proposed Method

2.1 Data Description

The purpose of this research is to present the results of a multivariate analysis through visualization techniques on the student engagement that students of two undergraduate

degrees and technologies present in a sample population belonging to a computer and systems faculty in a university in the south of Mexico. The bachelor's degrees considered for this study were the bachelor's degree in administrative informatics and bachelor's degree in computer systems. The sample was non-probabilistic, directed and for convenience [10], an Utrecht Work Engagement Scale for Students (UWES-S), survey was applied to a group of 32 students, 16 of each undergraduate degree, who agreed to answer the questionnaire on a voluntary basis, the period of application of the survey was February-August 2015.

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

The engagement at work is defined as [11]: "a positive psychological state characterized by high levels of energy and vigor, dedication and enthusiasm for work, as well as total absorption and concentration in work activity." Because the studies are considered as an activity in which responsibilities are acquired and one has an objective, the same concepts and instruments of engagement at work for the student engagement were used, with some slight variations. The Utretch Work Engagement Scale for Students (UWES-S), questionnaire was developed to measure student engagement, which consists of 17 items that indicate [12]: "vigor is evaluated by six items, which refer to high levels of energy and resilience, willingness to devote efforts, not fatigue easily, and persistence in the face of difficulties.

The dedication is evaluated through five items that refer to the meaning or meaning of the work, to feel enthusiastic and proud of its work, and to feel inspired and challenged by the work. Absorption is evaluated by six items that refer to being happily immersed in their work and have difficulty leaving it, in such a way that time passes quickly and one forgets everything around them.

The responses of the instrument are measured according to a Likert scale, where zero means "never" and six "always", so scores range from zero to six for each dimension that makes up student engagement. The original internal consistency obtained by the authors for the UWES-S version of 17 items in Dutch students was for vigor, dedication and absorption of 0.63, 0.81 and 0.72 respectively, thus meeting the criterion of superiority to 0.60 for an instrument of recent development. The content of the questionnaire can be consulted in the Handbook of UWES [12].

2.3 Data Analysis

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

Table 1. Descriptive statistical population study sample.

Variable	Minimum	Maximum	Mean	Standard Deviation
Vigor	1.8	5.7	4.141	1.0073
Dedication	1.8	6.0	4.906	1.0656
Absorption	1.2	5.8	4.125	1.1334
Engagement student	1.7	5.8	4.394	0.9990

To carry out the multivariate analysis and the representation of visual results, the R language was used, which has been considered as one of the most powerful and versatile software, especially in the construction of graphics. The R language has a variety of packages that allow programming at a higher level, complex analyzes and custom graphics design [13].

A correlational analysis was applied, presenting the results in a graphical way, in order to know how associated the variables of the student engagement are. Another way to know the characteristics of the cases presented was through the implementation of the GGE Biplot multivariate analysis, where two elements called genotypes (G) and another of genotype-environment interaction (IGA) are used, this technique facilitates the visual identification of genotypes and environments for their interpretation [13].

In the same way, an analysis was developed through the technique of Chernoff faces where the characteristics of the student engagement are represented. Chernoff proposed a representation method for multivariate data, the use of faces in the form of caricatures to represent data having several dimensions [14]. In Chernoff faces the values of the variables involved in the analysis correspond to facial features, this technique aims to improve the user's ability to detect and understand the importance of the phenomenon and serve as a mnemonic mechanism to remember the main conclusions [14].

3 Results

A descriptive analysis was carried out, which can be observed in Table 1, with a sample of 32 students of two careers in technologies, it is possible to observe the values obtained in minimum and maximum, average, and standard deviation for the student engagement and its dimensions. The minimum value obtained in the cases is observed in the absorption with 1.2, the maximum value is presented in the dedication with 6.0. The values shown in the mean of the dimensions from highest to lowest are for dedication 4,906, vigor 4,141 and absorption 4,125, finally the average of this population shows in student engagement is 4.394.

Table 2 presents a scale that was developed to qualify the student engagement, this table with established values allows to evaluate the observations of a case or group of cases 10. Other methods can also be used to calculate levels of student engagement with the UWES-S, consult Schaufeli and Bakker [12]. Using the scale can be interpreted for this population sample that the average vigor 4,141 and absorption 4,125 is average, that is, the students feel moderately vigorous and absorbed, on the other

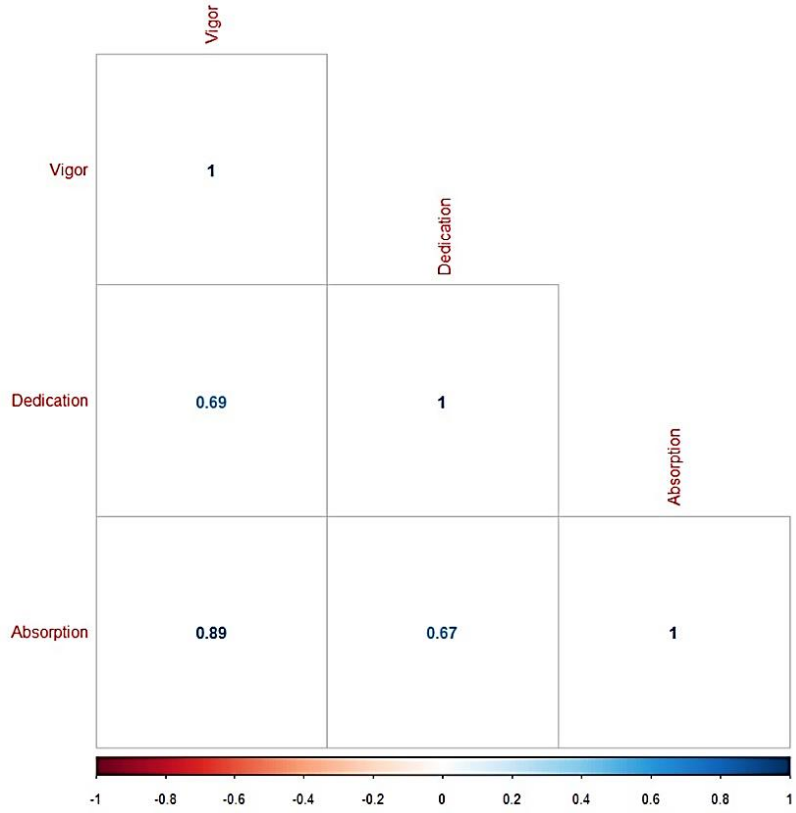


Fig. 1. Correlation matrix chart of the dimension's vigor, dedication and absorption of the student engagement. Source: (Self-realization, 2017).

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

hand, the dedication has a value of 4,906, considered high, so the students are enthusiastic, motivated and proud of their studies.

3.1 Visual Representation Technique of Correlational Analysis

A correlation analysis was carried out presenting its results through a matrix chart, which measures the degree of association that exists between the dimensions that make

up the student engagement of the sample population of this study. The method used is the Pearson correlation coefficient, which represents the linear relationship between two quantitative variables [15], the level of significance is 0.01. The magnitude of the coefficient indicates the degree of relation between the variables, so if this data is closer to one the correlation is greater, on the contrary, it is smaller if this data is closer to zero [16].

In Figure 1 shows that the resulting correlations between the variables force, dedication and absorption are close to one, so there is an important degree of association between these variables, it is concluded that the instrument that has been selected to measure student engagement is appropriate for the sample population.

3.2 Visual Representation Technique of GGE BIPLOTS

Genotype plus Genotype-by-Environment (GGE) Biplot is another visual technique that allows to represent the type of multivariate analysis [17]. The GGE biplot is constructed from the first two principal components. The first component, when highly correlated with the main genotype effect, represents the proportion of the yield that is due only to the characteristics of the genotype.

The second component represents the part of the yield due to the genotype-environment interaction. The genotypes close to each other in the GGE biplot present similar patterns of response across the environments. The environments close to each other, given by the acute angle between their vectors, indicate positive environmental association, that is, similar patterns of response in the relative behavior of a set of genotypes. The absence of association between environments is given by the right angle between vectors and the negative association by the obtuse angle [18].

In Figure 2, we can see the called Genotypes (G), which represent the students, each case, where they are identified by a label that is integrated by the degree program and a unique additional number, just as it is visualized to the interaction Genotype-by-Environment (GE), formed by the variables, in this case student engagement and its dimensions of vigor, dedication and absorption. The graph shows the first two main components, derived from the decomposition of the combined effects of G + GE [13]. The first main component explains 91.14% and the second main component is 5.64%, so the variability explained by the two axes is greater than 50% [13], so that at 96.78% the interpretations shown are considered valid.

The GGE Biplots are interpreted in terms of distances between elements and orientation of the axes, so from the graph one can know approximately the values of all variables for each case [13]. Degrees programs considered for this study were labeled in the graphics as LIA for the bachelor's degree in administrative informatics and LSC for computer systems.

In Figure 2, all genotypes (G) are observed, most of them agglomerate in a group, observing a significant relation of all these genotypes with their environment (GE), that is, with all the variables that integrate the engagement student. On the other hand, genotypes (G) that are far from the group, which are LIA2, LIA10, LIA6 and LSC9, are observed in the opposite direction to the variables of the student's engagement, reason why it is deduced that they present low levels of vigor, absorption, dedication and consequently also under student engagement.

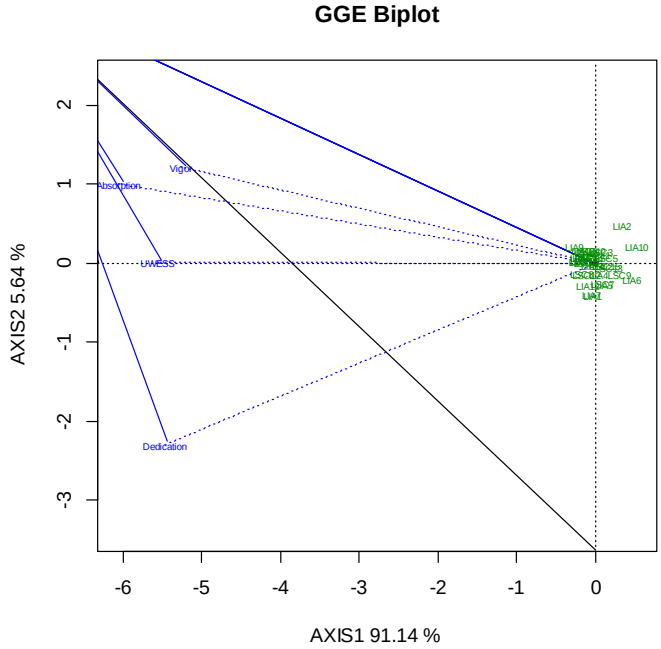


Fig. 2. Graph GGE Biplot of the dimension's vigor, dedication, absorption and UWES-S, obtained from the students. (Source: Self-realization, 2017).

Table 3. Effect of variables in R Language.

Modified Item	Variables
Height of face	Vigor
Width of face	Dedication
Structure of face	Absorption
Width of ear	UWES-S
Height of eyes	Bachelor's degree
Height of hair	Gender

3.3 Chernoff Faces Visual Representation Technique

In the representation across the faces of Chernoff each dimension of the data is determined by the size, location, or shape of some facial component of the caricature [14]. In this study, we developed this type of visualization technique to show the characteristics of student engagement taking into account the sample population.

Using the face technique of Chernoff, a facial feature was assigned to each of the variables, the variables corresponding to vigor, dedication, absorption, UWES-S student engagement were represented, and two variables were added that correspond to career and gender of the student, in total six variables.

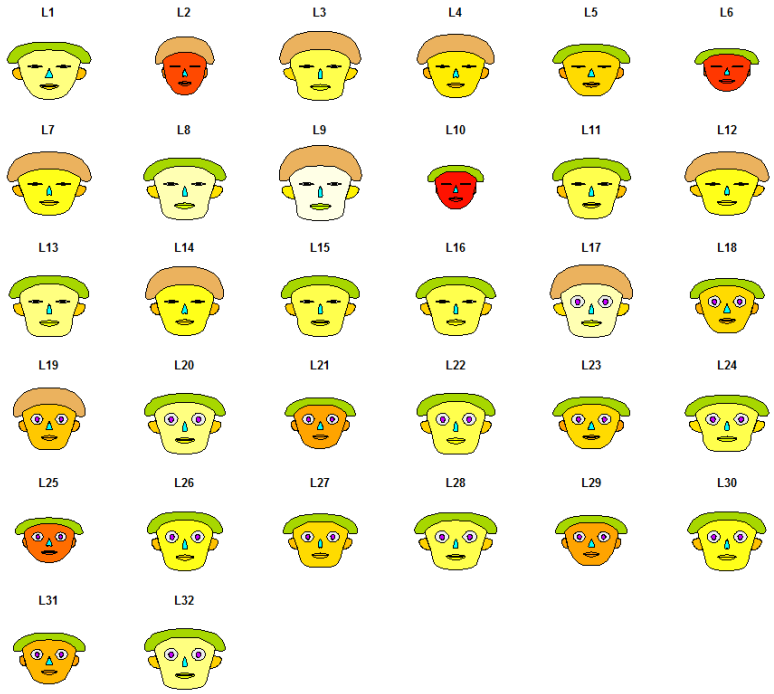


Fig. 3. Representation of Chernoff faces of the dimension's vigor, dedication, absorption and UWES-S, obtained from the students, Source: (Self-realization, 2017).

Table 3 shows the correspondence between the facial feature (Modified Item) and the variables that integrate student engagement (Variables). Examples of these representations are: the variable Vigor that is represented by the height of the face; the student engagement variable, identified as UWES-S, which is represented by the width of the ear; and the variable gender that is represented by the height of the hair.

In Figure 3, the results of the application of the Chernoff faces technique are presented, all cases, 32 figures of both bachelor's degrees: administrative informatics and computer systems. An identifier that begins with the letter L followed by a unique number was used.

One of the representative cases is the student of L3, where the vigor is represented by the height of the face, which is considered of the high ones in comparison with the others, reason why considers a student with a high level of vigor; the dedication that is identified by the width of the face, is remarkably wide, that is, has a high level of dedication; the absorption is determined by the structure of the face, which is very large, notably square, so that the student has a high level of absorption; the student engagement UWES-S, is represented by the ears, which are very wide in this case, so the student has a high level of student engagement.

For additional variables such as bachelor's degree, this is especially identified by the height of the eyes, for almost closed eyes corresponds to the bachelor's degree in

administrative informatics, for wide eyes corresponds bachelor's degree in computer systems; for the case of the variable gender, the height of the hair represents whether it is male or female, very high corresponds to female and short corresponds to male. In this case of the student of L3, it is identified as a part of the degree program in administrative informatics and is a woman.

Another of the cases in Figure 2, where it is observed that its characteristics show remarkable differences, is the L25, considering Table 3, it can be concluded that given its facial features, it has a low level of vigor, dedication, absorption and consequently a low student engagement, in the same way it is observed that it is a man student in computer systems.

4 Conclusions

The use of visual techniques to represent analysis of multivariate data presents advantages such as the rapid identification of the differences between the cases and the rapid interpretation of the same. In this paper, we have gone into the search for new forms of multivariate visualization, since a great amount of information needs to be analyzed from multiple dimensions, three forms of representation of results were selected for this type of studies, the development of correlation charts, GGE Biplot, and Chernoff faces. The R language was used for the implementation of these techniques, since it was found to be one of the most complete languages in the area of visual techniques deployment, since it allows the development of high level graphics.

The multivariate visualizations were developed to give information of the engagement in university students, conformed in three dimensions: vigor, dedication and absorption; variables such as undergraduate and gender were also considered. The correlation technique used allowed us to visualize how closely related the variables of the engagement are. Regarding the visualization through GGE Biplot, this contributed to visualize the relation between the cases according to the variables that make up the student's engagement, or it could also be said according to this technique: the positioning of genotypes in the environment.

When using the Chernoff faces method, a direct translation of the actual data into facial features is done, for each of the cases that integrated the sample. The techniques used for the visualization of multivariate analysis in this work, allow observing the information from different perspectives, contributing to the understanding of the data in a faster way, since it is possible to simplify the tedious and complex, reason why it is recommendable carry out this kind of visual transformations. It is proposed to continue the studies related to the development of visual techniques in the R language to represent data that facilitate the understanding of the results. The next step is to combine analysis of the type of main components with a discriminant analysis, both represented in the same graph.

The analysis of main components allows transforming the set of original variables into another reduced set of variables, the components, which conserve most of the information about the variability of the data. Discriminant analysis is used to separate the elements into groups or known classes.

By combining both methods, a more complete description will be obtained for the analysis of data, suitable when there are many variables correlated with each other and a classification criterion to associate the variables.

References

1. Hazarmabeth, I.: Métodos estadísticos exploratorios y confirmatorios para análisis de datos. Facultad de Estudios Superiores Zaragoza, Universidad Nacional Autónoma de México (2013)
2. Correa, J. C.: Gráficos Estadísticos con R. Universidad Nacional-Sede Medellín (2002)
3. R Development Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing, http://www.R-project.org (2016)
4. 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)
5. Morales, E.Y., Hernández, J.A., Ochoa, C.A., Lopez, M.A.: Comparison Represented in the Form of Radar of University Student Engagement in Degrees in Technologies. In: Sidorov, G., Ritter, G., Serra, J. Cortes, U.(eds.), Research in Computing Science, 122 (2016)
6. López, M.A., Hernández, J.A., Ochoa, C.A., Morales, E.Y., González, C.: Comparative Study of Learning Strategies of Bachelor Students in Nursing. In: Sidorov, G., Ritter, G., Serra, J. Cortes, U.(eds.) Research in Computing Science, 122 (2016)
7. Quintana, M., Trinidad, J.C., Morales, S.J., Landassuri, V.M.: Análisis Comparativo de Algoritmos de Minería de Datos para Predecir la Deserción Escolar. In: Santos, E.A., Chimal, J.C., Cabrera, L., Castillo, E. (eds.). Research in Computing Science, 67, Special issue Advances in Computer Science, 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. In: Miranda, S. (ed.), Research in Computing Science, 95, Minería de textos, lógica y ontologías, pp. 135─146 (2015)
9. Cruz, C., Reyes, L.A., Hernández, G., Sánchez-Cervantes, J.L.: Diseño de una ontología para el proceso de evaluación de las asignaturas técnico científicas del Instituto Tecnológico de Orizaba. In: Sidorov, G., Ritter, G., Serra, J. Cortes, U.(eds.), Research in Computing Science, 128 (2016)
10. Hernández, R., Fernández, C., Baptista, M.: Metodología de la investigación. McGraw-Hill Interamericana (2010)
11. 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)
12. Schaufeli, W.B., Bakker, A.: Utrecht Work Engagement Scale (UWES). Escala de Engagement en el trabajo de Utrecht, Ocuppational Health Psychology Unit: Utrech University (2003)
13. Guisande, C., Vaamonde, A.: Gráficos estadísticos y mapas con R. España. Ediciones Díaz de Santos (2013)
14. Raciborski, R.: The Stata Journal. Graphical representation of multivariate data using Chernoff faces. 9(3) pp. 374─378 (2009)
15. Cuñer, N.: Escala de Inteligencia para Niños de Wechsler, WISC-IV. Diccionario de Psicometría (2013)
16. Aquije, R., Ramírez, R., Castillo, H.: Guía para la aplicación del análisis multivariado a las encuestas de hogares. Lima, Perú: INEI (2002)
17. Dumble, S., Frutos, E., Galindo, P.: Package'GGEBiplots (2017)

18. Ibañez, M.A., Cavanagh, M. M. Bonamico, N.C., Di Renzo, M.A.: Análisis gráfico mediante biplot del comportamiento de híbridos de maíz. RIA, 35(3), pp. 83─93, INTA (2006)

ISSN 1870-4069

Representation of Learning Strategies through Visual Techniques of Multivariate Analysis in University Students

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

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

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

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

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

Abstract. This research aims to identify the learning strategies employed by university students; and represent them through graphic techniques of multivariate analysis. The sample consisted of 30 students from the Multidisciplinary Academic Division of the Rivers; of the degrees in Administration and Administrative Computing of the generational cohorts 2010, 2011 and 2012. The instrument used is the Inventory of Strategies of Learning and Motivational Orientation (EDAOM), this instrument consists of a self-assessment that the student makes on the learning strategies that employs. For the implementation of the visual techniques was used the Language R, with these techniques were represented individually to the students with their respective learning strategies. The results show the interaction of the students with the learning strategies and the academic performance, visualization that the students of the Degree in Administrative Computing obtain the lowest values in the strategies of performance and academic performance.

Keywords: EDAOM, learning strategies, data analysis, visualization, graphs.

1 Introduction

Higher education institutions have a large amount of data that has been generated from academic and administrative activities, in many cases this information has not been explored, analyzed and visualized. In this paper we used the data obtained from the

application of the Inventory of Strategies of Learning and Motivational Orientation (EDAOM), to new students; with the purpose of identifying the learning strategies employed by students of the Bachelor's in Management and Administrative Computing and represent them through graphic techniques of the multivariate analysis.

For the analysis of a set of data with many variables it is convenient to use multivariate visualization techniques, which allows in this case in an individual and group way the interaction of the learning strategies with academic performance; In this sense, several studies have been found in the literature in which they employ multivariate data analysis techniques such as GGE Biplots, generally applied to agriculture with the objective of evaluating the variability in the yield of corn grain and determining the response patterns between hybrids and environments, with the analysis of the biplot they identified the hybrids of superior performance [1]; another study consisted in determining the magnitude of the genotype x environment interaction in native corn populations of Chiapas in three contrasting environments, obtaining significant differences between environments, populations and interaction environments by populations [2], other studies of this type can be consulted in [3,4].

1.1 Multivariate Analysis

The main objective of data analysis is to summarize and interpret data sets, detailing the contents and showing the important characteristics. In this sense the visualization of the multivariate data plays an important role in the fulfillment of the objectives [5].

Some studies in the educational area use graphical visualization techniques such as radar charts to make comparisons that are easy to interpret [6, 7]; In other studies, on university academic performance and the aspects that affect it, data analysis techniques have been applied [8, 9] or techniques of evaluation of results obtained [10].

Multivariate analysis is the part of statistics and data analysis that studies, analyzes, represents and interprets data that results from observing more than one variable of an object [11].

1.2 Visual Techniques of Multivariate Analysis

The problems of multivariate space are difficult to understand, due to the large amount of data that must be represented in a very limited space; the problem is to consider a large number of variables and their relationships simultaneously. Traditional methods only produce a partial view of information. Visualization methods are the most convenient way to present data and achieve a graphic illustration of all the information in which it shows the essential facts, make comparisons and identify trends [5].

Advances in the development of automated information processing tools have simultaneously provided the creation and use of multivariate analysis techniques [12], increasingly sophisticated, dynamic and interactive graphs to visualize data or model results statistics [13].

Through the visual representation of data, the relationships described by graphs are easily understood and are easy to remember, in addition to a graph are displayed large amounts of data [14].

The graphs used for the visual representation were: of stars and GGE Biplots representing each of the individuals of a population or sample.

Stars's graphs represent a method for representing multivariate data in which each star represents an element of the sample or population, and each axis represents a variable, its length is proportional to the value of the variable. The ends of the rays are connected with segments of straight lines to form a star [15]. Star charts allow the identification of dominant variables in an element of the population; visualizes if there are similarities between the elements of the population and therefore if there are groupings, also the atypical values can be identified if they exist [16].

The GGe Biplots graphics employ the combined effects of genotypes (G) and genotype-environment interaction (GE), which facilitate the visual identification of genotypes and evaluation environments. These graphs are constructed using the first two main components, derived from the decomposition of the combined effects of G + GE. The first component, when highly correlated with the main effect of the genotype, represents the proportion of the yield that is due only to the characteristics of the genotype. The second main component represents the part of the yield due to the genotype-environment interaction [17].

The importance of identifying the learning strategies used by the students consists of knowing the causes of low or high performance [18], that is, improving learning outcomes can be achieved by incorporating assessment procedures to identify risks and determining strategies should be encouraged in students [19].

2 Method and Tools

The scope of this research is descriptive because they will specify properties and characteristics describing trends, the design is non-experimental transactional [20].

The population is made up of the new students belonging to the cohorts 2010, 2011 and 2012 of the Bachelor's Degree in Administration and Administrative Computer Science, the sample is of 30 students who answered in their entirety the inventory of Strategies of Learning and Motivational Orientation (EDAOM).

2.1 Inventory of Learning Strategies and Motivational Orientation (EDAOM)

The instrument used to know student learning strategies is the Learning Styles and Motivational Orientation questionnaire (EDAOM), the result of this questionnaire is a self-assessment that the student makes about their learning strategies and motivational orientation to the study [19]. The self-report section measures self-assessments of students on: a) frequency, b) ease or difficulty, and c) the results of using a wide variety of learning strategies.

The EDAOM is composed of four scales and 13 subscales; the Self-regulation, Metacognitive and Metamotivational scale consists of three components: those of the person, those of the learning task and learning materials; Table 1 shows the structure of the EDAOM, the first column specifies the scales and the second the subscales that make it up.

Table 1. EDAOM scales and subescales.

Scales	Subscales
Acquisition of information	Selective Generative
Recover of information	Subjected tasks During exams
Processing information	Convergent Divergent
Self-regulation, Metacognitive and Metamotivational	Person: Efficiency perceived Internal Contingency Perceived autonomy Orientation to external approval
	Learning Task: Orientation to the task itself Orientation to the achievement of goal
	Materials

In this work the learning strategies were analyzed: selective, generative, task recovery and during exams and convergent and divergent processing.

2.2 Software Used for Multivariate Analysis

A very important element to implement the techniques of multivariate analysis are the tools, in this case we used the software SPSS and language R.

IBM SPSS Statistics 22 is a software for statistical analysis that contains the primary functions to perform a complete analysis process. It is easy to use and contains an extensive category of procedures and techniques that help direct research and make the best decisions. Its main characteristics are to have a wide range of statistical procedures to perform accurate analysis, includes techniques to prepare data for analysis quickly and easily, can create very efficient charts for reporting and supports all types of data [21].

R is a free software for statistical computing and graphics. Compile and run and compile on a variety of UNIX, Windows and Mac OS platforms. It is widely used for the development of statistical software and data analysis, provides a wide variety of statistical models: linear and nonlinear models, classical statistical tests, time series analysis, classification, grouping among others), and graphic techniques. R is an environment in which statistical techniques are implemented and which is extensible through packets, there are about eight packets that are supplied with the R distribution [22].

2.3 Descriptive Analysis

A descriptive analysis of the learning strategies was carried out: selective, generative, retrieval of information before tasks and during exams, processing of convergent and

divergent information, the results obtained are shown in Table 2, in this it is observed that the value minimum for the majority is 60 or more with the exception of the acquisition of selective information that obtained 54 and the processing of divergent information 51; in the maximum value values equal or greater than 90 were obtained; in terms of the average the strategy of acquiring selective information and retrieval of information before tasks obtained a value greater than 70 and the other strategies values greater than 80. The minimum average bachelor is 7.30, the highest average is 9.95 and the mean of the averages is 8.70.

Academic performance is classified as low, regular and high; was derived from the indicators: approval rating in ordinary, average and promotion index in the first year [23]. Table 3 shows that 36.7% of the sample had low academic performance; 46.7% high and 16.7% regular.

Table 2. Descriptive Statistics.

Subscale	**Min**	**Max**	**Half**	**Standard deviation**
Selective	54	94	73.03	8.818
Generative	69	100	82.90	10.018
Before Task	60	90	72.97	7.025
During Exams	60	100	85.77	9.957
Convergent	63	97	81.60	10.074
Divergent	51	100	81.27	10.041
Average High School	7.30	9.95	8.7037	.75926

Table 3. Frequencies of Academic Performance.

Performance	**Frequency**	**Percentage**
Low	**11**	**36.7**
Regular	**5**	**16.7**
High	**14**	**46.7**

3 Results

Pearson's correlation was made between learning strategies and academic performance; the correlation measures the linear association between learning strategies, the value of the association may be from -1 to 1, the closer the value is to 1 the greater the association between variables [24]. Figure 1 shows the correlation between learning strategies and baccalaureate average and academic achievement in which there is a strong association between the average baccalaureate and academic achievement; learning strategies are correlated with values greater than .5 and .8.

The correlations between the subscales of Acquisition, Recovery and Processing of EDAOM Information, it is observed that they are related to each other, so it is

determined that they are not mutually exclusive, that is, a student can have a good command of the different learning strategies [25, 7], in the same way a strong association of the high school average with the academic performance obtained in the first year of the bachelor's degree is observed. The subscales that present a higher relation considering the selection criteria equal to or greater than .70, according to Figure 1, identify the selective subscale with recovery during the exams and convergent processing; and the generative with during the exams and divergent.

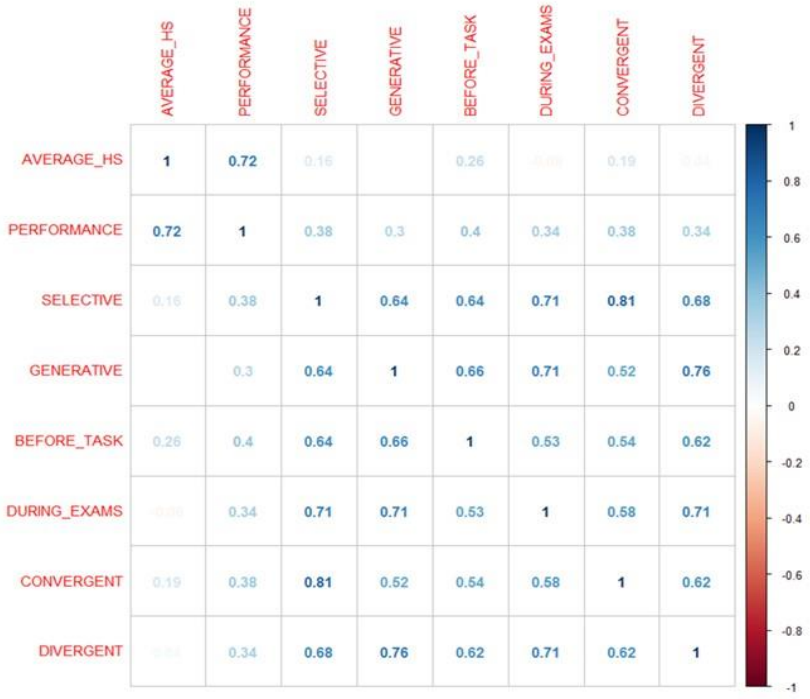


Fig. 1. Correlation of EDAOM learning strategies and academic performance, Source: (Self-realization, 2017).

3.1 Data Analysis with Star Graphics

The chart of Stars with the language R was made, in Table 4 the variables of each axis with their respective meaning are shown.

Table 4. Definition of the variables (axes or rays of each star).

Variable	Significado
SEL	Selective
GEN	Generative
SUB_T	Subjected_Task
DUR_E	During_Exams
CON	Convergent

DIVE	Divergent
AVG	Average_HS
PER	Performance

Graph 2 shows 6 stars where each one represents a student and the values obtained in each of the variables: Selective, Generative, Homework Recovery, during exams, convergent, divergent, baccalaureate average and academic performance, to identify the degree to which they belong were assigned a label composed by the initials of the degree and a consecutive number. This graph allows us to observe that students LIA7, LA6, LA10 and LIA1 have a high academic performance, the student LIA5 has very low results in almost all variables the highest value is shown in the variable of information retrieval during exams, the LIA7 student has a good academic performance, an average high school diploma and a high value in selective strategies. Students LA6 and LA10 observe a similar pattern obtaining high values in the averages and learning strategies.

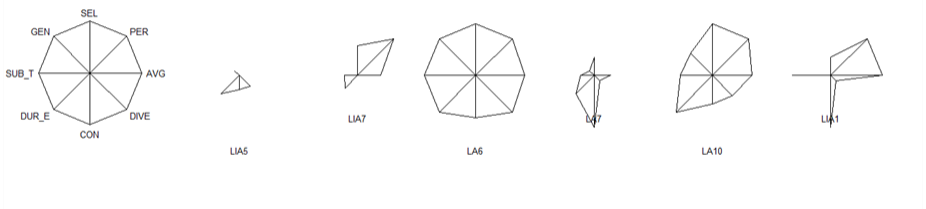


Fig. 2. Graph of stars of the learning strategies and the academic performance of the students of the Bachelor's degrees in Administration and Administrative Informatics, Source: (Self-realization, 2017).

3.2 Data Analysis with Graphs GGe Biplots

Another visualization technique applied to the data was the GGE Biplot graph for each case of the sample, in relation to the learning strategies, average of bachiller and academic performance. GGE Biplot is effective for the analysis of mega-environments, for example (which genotype (student), -got-where (in which locality (learning strategies), the evaluation of genotypes (from the point of view of stability and performance) and the evaluation of environments (discriminant power), there are several studies that have shown that this graph describes the IGA of populations in an effective way[26], due to this characteristic its use is adequate for the representation of student populations and their interaction with the strategies of learning used with academic performance obtained.

In Figure 3, the genotypes (G), which are the students, are identified where they identify by race and a unique additional number, also identifies genotype-environment interaction (GE), such as academic performance and strategies of learning: selective, generative, task recovery and during exams, convergent and divergent processing. The graph shows the first two main components, derived from the decomposition of the combined effects of G + GE [17].

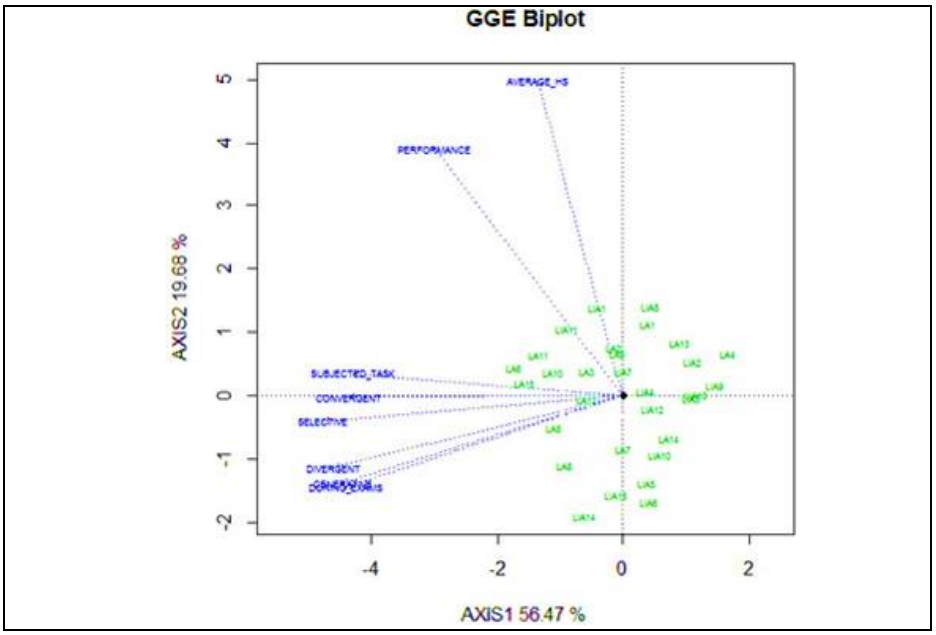


Fig. 3. Graph GGE Biplot of the learning strategies and the academic performance of the students of the degrees of Administration and Informatics Administrative, Source: (Self-realization, 2017).

The first main component explains 56.47% and the second main component 19.68%, obtaining a total 76.15% variability of the two axes and The GGE Biplots graphs are interpreted in terms of distances between elements and orientation of the axes, this allows to know roughly, from the graph, the values of all variables for each genotype [17]. For the description of the results, genotypes with more representative values (extremes), have been considered, and other cases with characteristics close to them can be observed in the graph. The academic performance environment takes high values for the majority of the genotypes and lower for the genotypes LIA14, LIA15, LIA6, LIA5, LIA10, LA14 and LA4 observing that the students of the Degree in Computer science have a lower academic performance. As for the learning strategies, it is observed that the genotypes LIA2, LIA4, LIA9 and LIA13 have the lowest values. The genotypes with low values in both academic performance and strategies are: LIA14, LIA5, LIA15, LIA6, LIA10 and LA14.d being greater than 50% [17], are considered valid interpretations presented.

4 Conclusions

The graphical techniques of multivariate analysis provide numerous advantages over any other type of representation, being more attractive, provoking interest and facilitating the transmission of information in an easy and fast way.

The visualization in the analysis of data allows showing information easy to assimilate and to understand. Also in these representations of information like the graphs of stars and the GGE Biplot one can identify groupings, atypical values and tendencies.

In this work the objective was to use visual techniques of the multivariate analysis to represent the learning strategies used by students, it is proposed the continuation of this work by implementing other visual graphs that describe characteristics of students and are easy to understand for the academic administration or tutors to help them in decision making.

References

1. Ibañez, M., Cavanagh, M., Bonamico, N., Renzo, M.: Análisis gráfico mediante biplot del comportamiento de híbridos de maíz. Revista de Investigaciones Agropecuarias, 35(3):83–93, 86435306 (2006)
2. Martinez-Sánchez, J., Espinosa, N., Villegas, Y.: Interacción genotipo-ambiente en poblaciones de maíz nativo de Chiapas. Revista Mexicana de Agroecosistemas, 3(1):38–48 (2016)
3. Ledesma-Ramírez, L., Solís-Moya, E., Suaste-Franco, M.P., Rodríguez-Caracheo, J.F., de la Cruz-Gonzalez, M.L.: Análisis GGE Biplot del rendimiento de trigo (triticum spp.) con riego normal y restringido en el bajío. Agrociencia, 46(2):119–131 (2012)
4. Lozano-Ramírez, A., Santacruz-Varela, A., San-Vicente-García, F., Crossa, J., Burgueño, J., Molina-Galán, J.: Modelación de la interacción genotipo x ambiente en rendimiento de híbridos de maíz blanco en ambientes múltiples. Revista Fitotecnia Mexicana, 38(4):337–347 (2015)
5. Alvarado, J.: Visualización multidimensional, nuevas técnicas de exploración de datos y representación de información. Ediciones Universidad Cooperativa de Colombia (2010)
6. Morales, E.Y., Hernández, J.A., Ochoa, C.A., Lopez, M.A.: Comparison Represented in the Form of Radar of University Student Engagement in Degrees in Technologies. Research in Computing Science, 122 (2016)
7. López, M.A., Hernández, J.A., Ochoa, C.A., Morales, E.Y., González, C.: Comparative Study of Learning Strategies of Bachelor Students in Nursing. Research in Computing Science, 122 (2016)
8. 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, Special issue Advances in Computer Science, pp. 51–61(2013)
9. 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)
10. Cruz, C., Reyes, L. A., Hernandez, G., Sanchez-Cervantes, J.L.: Diseño de una ontología para el proceso de evaluación de las asignaturas técnico científicas del Instituto Tecnológico de Orizaba. Research in Computing Science, 128 (2016)
11. Balzarini, M., Bruno, C., Córdoba, M., Teich, I.: Herramientas en el Análisis Estadístico Multivariado. Escuela Virtual Internacional CAVILA. Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba. Córdoba, Argentina (2015)
12. Pérez, M.: Minería de Datos a través de ejemplos. AlfaOmega (2014)

13. Ledesma, R., Molina, J., Forrest, W., Valero-Mora, P.: La visualización múltiple en el análisis de datos: una aplicación en Vista para el análisis de componentes principales. Psicothema, 19(3), pp. 497–505 (2007)
14. Sáez, M.: Diseño e implementación de una aplicación en Processing para la representación de datos multidimensionales (2015)
15. Schiattino, I., Silva, C.: Representación gráfica de información multivariante. Aplicación al sistema de salud de Chile (2010)
16. Cabas, C.: Análisis de datos multivariantes (2017)
17. Guisande, C., Vaamonde, A.: Gráficos estadísticos y mapas con R. España. Ediciones Díaz de Santos (2013)
18. Beltrán, J.A.: Estrategias de aprendizaje. Revista de Educación. Madrid, 332, pp. 55–73 (2003)
19. Castañeda, S.: Educación, aprendizaje y cognición. Teoría en la práctica (2004)
20. Hernández, R., Fernández, C., Baptista, M.: Metodología de la investigación. McGraw-Hill Interamericana (2010)
21. IBM SPSS Statics: SPSS Statics Base. Recuperado de: http://www-03.ibm.com/software/products/es/spss-stats-base (2015)
22. R-Project: R-Project. https://www.r-project.org/about.html (2017)
23. Chain, R., Cruz, N., Martínez, M., Jacomé, N.: Examen de selección y probabilidad de éxito escolar en estudios superiores. Estudio en una universidad pública estatal mexicana. Revista Electrónica de Investigación Educativa, 5(1) (2003)
24. MacRae, S.: Modelos y métodos para las ciencias del comportamiento (1995)
25. López, M.A.: Predicción del rendimiento académico de estudiantes universitarios aplicado minería de datos (2017)
26. Frutos, M.: Interacción genotipo – ambiente: GGE Biplot y modelos AMMI, https://gredos.usal.es/jspui/handle/10366/121900 (2011)

ISSN 1870-4069

Intelligent Internetworking Tutoring System by Using Adaline Neuron and Mel Cepstral Coefficients

Miguel A. Ruiz-Jaimes[1], Jorge A. Ruiz-Vanoye[2], Yadira Toledo-Navarro[1], Diego Márquez-Salazar[1], Rodrigo Beltran-Labra[1]

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

[2] Autonomous University of Hidalgo, Hidalgo, Mexico

{mruiz, ytnavarro, mado140055, blro130524}@upemor.edu.mx, jorge@ruizvanoye.com

Abstract. We present a speech recognition prototype system in Spanish language to contribute to the study of the communication networks branch, alternative and innovative techniques of human computer interaction and the recognition with learning of the use and details about internetworking terms, in addition to the instruction on remote networks to carry out the search of all the possible routes, so subsequently choose the most efficient routes for data exchange among themselves, using dynamic routing by using one or more routing protocols such as: RIP, IPV 2, IGRP, EIGRP, OSPF, to later identify your metric either by jump count, bandwidth, delay and costs which the algorithm tells the protocol values to determine its metric value according to the lowest result you get a best route. The recognition of voice and taking of actions through the same is done using the artificial neuron (ADAptive LInear NEuron) and cepstrales coefficients of Mel as a method of extracting information from the audio files. This prototype, due to its basic neuron nature, was implemented and proved only in an controlled environment, the word's recognition rate it's between the 40% and 80% percent. The recognition eperimentation results are exposed in addition to the design of the intelligent tutor.

Keywords: artificial intelligence, intelligent tutor, speech recognition, adaline, cepstral coefficients, internetworking.

1 Introduction

The problem of learning subjects with high theoretical content is of great impact in various sectors of the population, for example, in energy sector people in charge of maintaining equipment must read a great manual to understand the use and management of each of the devices, this makes people as they read lose interest in not having interactive content that arouses their interest to continue learning. The same thing

happens in the educational sector, for this work we focus on the study of the subject of internetworking where concepts are used that, if they do not know how to use them correctly would endanger very important information of a complete network.

To maintain constant communication from one network to another requires routing protocols, which are responsible for finding the best route between the nodes (routers), to facilitate communication across the network [1]. The most commonly used dynamic routing protocols are: RIP (routing Information Protocol), EIGRP (Enhanced Interior Gateway Routing protocol), OSPF (Open Shortest Path First), is-is (intermediate system to intermediate System), and BGP (Border Gateway Protocol) [2].

The development of an intelligent tutor that provides the student with information about the aforementioned protocols through interactive content such as augmented Reality (RA), and voice recognition, would provide comfort and versatility for the process of Learning, in addition, speech recognition offers the possibility for students to interact with an electronic team [3].

Within the various applications that Artificial intelligence has we can find the recognition of voice of which there are innumerable uses, the functions and characteristics of the speech recognition technology are very wide, for this project we conceptualize the process of interpreting a spoken word to a pre-programmed instruction within a system, this is achieved through the training of an Adaline neuron using five samples of each word to recognize by the system.

2 Related Work

There are several works on the learning of different academic subjects such as STAAM [4] the system of mathematics that focuses only on solving mathematical problems of primary education, the system is based on the Web designed to help elementary students. Barron et al. make use of gamification elements as main components of the EMATIC system (Mathematics Education through ICT) [5], the system is oriented to mobile devices for the teaching of basic mathematical operations. On the other hand, Zatarain et al. present an intelligent learning system that integrates a social network for the learning of mathematics and where it also presents an architecture for the design of the system [6], 3 emphasizing the recognition of the user's emotions.

Thinking about the implementation of speech recognition, Perez and Caballero developed a voice-based emotion recognition system [7], the work considers 640 sentences of eight users, and the following basic emotions were considered: Anger, Happiness, Neutral and Sadness. The percentage of overall effectiveness of said system is 87%. Hernández and Lemus developed an application for training in TOEFL certification through voice recognition [8], the application has a database where it protects synthesized voices for recognition through SAPI (Set of tools for voice recognition created by Microsoft).

Using the Adeline Neuron, Medina et al. developed a biomedical signal processing system [9], the system supports the doctor in the diagnosis of mental illnesses, presents the structure of the neuron and the hardware used to detect eye movements.

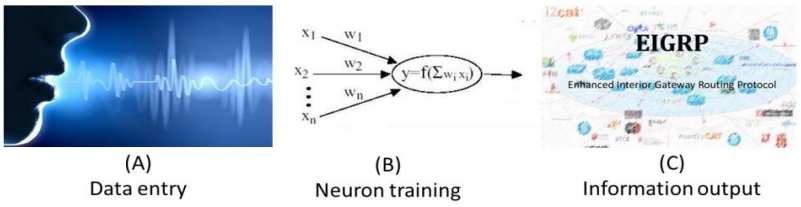


Fig. 1. System design.

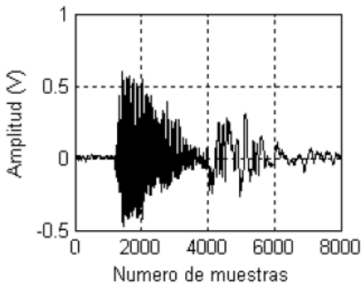


Fig. 2. Voice signal generated by Matlab (Taken from [10]).

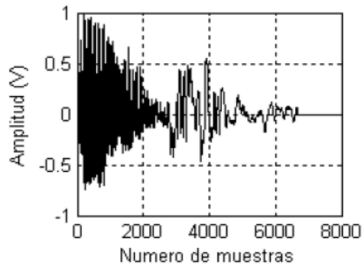


Fig. 3. Filtered and normalized voice signal (Taken from [10]).

3 Materials and Methods

For the development of this research was used a computer with a Operating System Windows 10 with an Intel Core i5 processor (6MB in Caché and 3.20 GHz.), a graphics card NVIDIA GeForce GTX 1070 with 8GB of RAM, integrated microphone and Matlab software. The design of the system developed in this work is shown in Figure:

A. Data entry: In this first stage the word of the Protocol is recorded to be recognized, then it is codified and finally normalized the voice, it is noteworthy that the recording of the voice must be under a controlled environment without ambient noise.

```
function newRecord = norm(record)
    maxi = max(abs(record));
    n = length(record);
    newRecord = zeros(1,n);
     for i=1:1:n
       newRecord(i)=record(i)/maxi;
     end
end
```

Fig. 4. Signal normalization function.

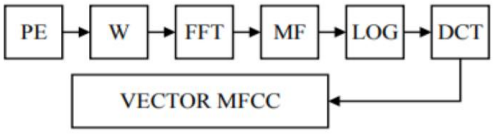


Fig. 5. Scheme for obtaining the MFCC (Take from [13]).

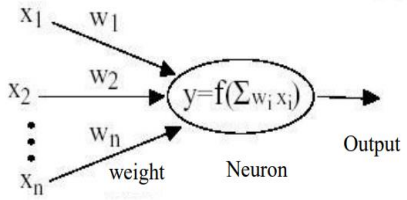


Fig. 6. Neuron Adaline model (Take from [14]).

B. Neuron training: Once the voice is properly processed, the neuron is trained to recognize the previously recorded word.

C. Information output: Finally when the neuron recognized the type of voicerecorded protocol, the system displays the protocol information on the screen.

The type of audio files that are handled in the system is WAV, in mono system. When the system records the voice, you get a signal very similar to Figure 2, however, when the voice is filtered and normalized there is a signal with a little more amplitude as seen in Figure 3.

Voice normalization is performed through the wave-amplitude standardization function; Figure 4 shows the function that normalizes the audio signal:

MFCC (Cepstrales coefficients in the frequency of Mel), are coefficients for speech representation based on human auditory perception, their frequency bands are located logarithmically [11], this makes the human auditory response modeled more appropriately than linearly spaced bands [12]. Figure 5, contains a diagram for obtaining the FC, where:

- PE: pre emphasis filtering,
- W: Windows,

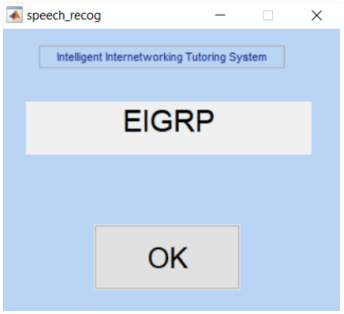


Fig. 7. Developed software user interface.

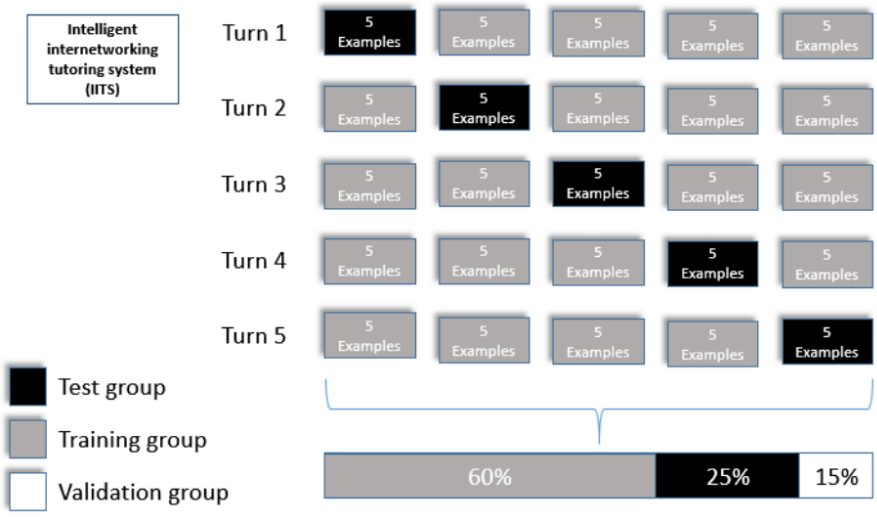


Fig. 8. Proportion of used audio scheme.

- FFT: Fast Fourier transformed,
- LOG: Logarithmic scale,
- DCT: Discrete cosine transformed [13].

The training of the system is carried out using the samples corresponding to each word, they establish random weights in a beginning, for later, having the neuron with the corresponding training, they obtain the appropriate values for Classify the words obtained by means of the recording of signal in mono frequency to 44100Khz. Figure 6 refers to the neuron model used in this research work.

Characteristics extraction and neuron trainment was made by using 12 different values taken from the MFCC model, so as it is showed in Figure 6, neuron has X_{12} entries that will define word's singularities to identify each word used in the system trainment.

Table 1. Words used in the recognition prototype implementation and its recognition rate at final test, words are routing protocols.

Words	Recognition rate				
Test	**1**	**2**	**3**	**4**	**5**
EIGRP	30%	42%	60.5%	74%	88%
RIP	26%	36%	48%	65%	79%
OSPF	20%	26%	49%	59%	70%
BGP	19%	23%	32%	52%	68%
IS-IS	10%	20%	33%	49%	57%

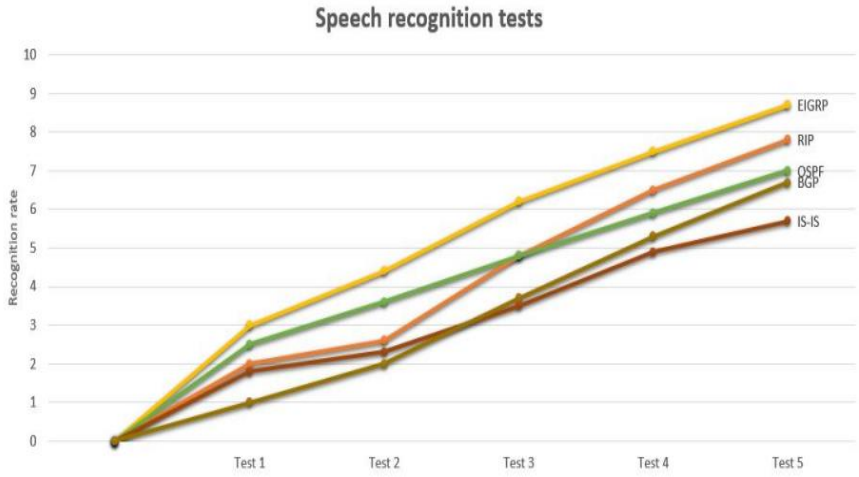


Fig. 9. Graph of speech recognition rate.

4 Problem Statement

Voice recognition plays an important role in networks and telecommunications, which is why it was thought to develop this system that will allow a better understanding and interaction of the cisco network operating systems which will allow the students who have the curricula of CCNA of Cisco [15], to have a better understanding and understanding of each of the network concepts (technologies, standards and protocols) and commands that make up each of the configuration modules of the network devices (routers and switches).

This system will allow through voice recognition that the student has a more dynamic, flexible and understandable knowledge that allows them to have a clearer knowledge of the concepts with the intention of motivating students to accredit each of the modules so that later present your certification exam.

Selected words are the more relevant protocols that CCNA of cisco [15], currently teach in bachelor's degree, due to it is a project prototype, methods that are implemented are just in order to discover and test the system functionality in a real context.

5 Analysis of Results

As a result of the tutored system, the software prototype shown in Figure 7 was obtained, which is able to perform the voice recognition of the five different dynamic routing protocols with which the project is being worked on, the process software operation is to start the software, then record the word or in this case the desired protocol, then the system will make the selection of the appropriate term based in its previous training, after that the system shows on screen information of the selected protocol in addition to the recognition of the same.

Figure 8, shows the selection and proportion of words or recordings used for the training and subsequent operation of the system, in addition to the validation group of the same, in the system were used 5 groups of recordings per word, four are for training and one test, validation was performed later when the system was put into practice with different users.

Figure 9, shows the graph of the training process and the rate of recognition of the words (protocols) used as samples in the project, it is remarkable how the improvement through the training process is notorious and can be concluded from this means that a greater number of training cycles, the better the recognition of the words in the system.

The words used for the recognition and operation of the system were the five dynamic routing protocols previously named (eigrp, ospf, rip, bgp and is-is), based on the recognition of each one of them is displayed relevant information and it is sought that the user out in the same, as well as in related topics where the main objective is to describe the operation of the protocols of dynamic routing as base in the internetworking.

For now, the interaction with the system by the user is through voice commands and selective information elements, which are accessed through the use of the mouse.

Word's recognition rate and words used into system training and implementation divided by tests, are shown in table 1.

Results in table 1 were graphed in figure 9, all tests including final result in test 5. We tried to continue testing after test 5 but results were not improved anymore.

Due to it is a controlled environment, all testings including final test were made with only one student and in the same environment sound conditions.

Selected words (protocols), were used due to are the most important protocols currently teached in Cisco Networking Academy course in informatics bachelors degree in the Politechnical University of State of Morelos and it actually solve the main fact of lack of networking hardware devices and because that is the main porpouse of this research, as a future work it is planning to make a more complex system, using a bigger artificial neural network and implementing the multi-user recognition function, even in an uncontrolled sound environment.

6 Conclusions and Future Research

A system was obtained with a trained neuron that is able to recognize with 70% effectiveness the words established in the program in a controlled environment. The system is able to interpret the voice message recorded by the user and then show on the screen the response obtained as well as information about each protocol.

As future work is intended to implement the use of augmented reality in the intelligent tutoring system, with which the user obtains a greater interactivity with the information, as well as with the physical objects in the system, situation that represents relevance because the direct interaction with the physical components of a training network always improves performance at the time of taking the practice to the implementation of the case in real life.

In addition, you want to expand the catalog of available terms for your search within the system, thereby expanding the possibilities of information and add interaction between related terms.

Also, as part of the improvement and scalability of the project, is intended to make an improvement of the network of neurons for the recognition, thereby expanding also the possibilities of use in different controlled environments or semi controlled, where the user can perform searches with greater freedom and in different sound environmental circumstances.

References

1. Enciso, L., Quezada, P., Fernández, J., Figueroa, B.: Análisis deparámetros de desempeño de los protocolos de enrutamiento Ad Hoc aplicado a un entorno urbano. TELECOM UNI, pp. 1–7 (2015)
2. Paciello, G.: Routing, Fundamental concepts. Journal Cisco Support, pp. 03–47 (2017)
3. Torres, C., Perpiñan, G., Mattos, L., Castro, J.: Automatic speech recognition using fourier transform and neural network. Colombian Journal of Avanced Tecnologies, pp. 116–120 (2005)
4. Zatarain, R., Barrón, M., García, J.: Sistema tutor afectivo para el aprendizaje de las matemáticas usando técnicas de gamificación. Research in Computing Science 111, pp. 83–96 (2016)
5. Barrón, M., Zatarain, R., Aranda, M., Gómez, H., Mejía, J.: Un tutor inteligente, afectivo y configurable para el aprendizaje de números naturales de 3er grado. Research in Computing Science 77, pp. 45–54 (2014)
6. Zatarain, R., Barrón, M., González, F., Oramas, R.: Diseño e implementación de un sistema de aprendizaje inteligente y afectivo para la enseñanza de las matemáticas del segundo año de primaria. Research in Computing Science 77, pp. 31–44 (2014)
7. Pérez, L., Caballero, S., Trujillo, F.: Integración de optimización evolutiva para el reconocimiento de emociones en voz. Research in Computing Science 93, pp. 9–21 (2015)
8. Hernández, M., Lemuz, R.: Multimedia Application for Training in TOEFL Certification Using Speech Recognition. Research in Computing Science 128, pp. 67–75 (2016)
9. Medina, A., Gómez, L.G., Cisneros, A., Melgar, E., Nango, G., Morales, N., Villegas, J.M., García, N.R., Bermudez, J.I.: Processing of Biomedical Signal with Neural Network Adaline. In: Proceedings of the 12th Iberian Conference on Information Systems and Technologies, pp. 2256–2261 (2017)
10. Barrero, V., León, J.P., Sánchez, C., Rosas, F.: Determinación de la arquitectura óptima de una red neuronal tipo perceptrón multicapa para la extracción de características relevantes y

reconocimiento de señales mioeléctricas. IEEE Transactions on Biomedical Engineering, pp. 11–16 (2014)
11. Mantilla-Caeiros, A.V.: Análisis, reconocimiento y síntesis de voz esofática: Instituto Politécnico Nacional (2007)
12. Del Pino, P., Granadillo, I., Miranda, M., Jimenes, C., Díaz, J.A.: Diseño de un sistema de medición de parámetros característicos y de calidad de señales de voz. Journal Ingeniería UC, pp.13–20 (2008)
13. Matínez-Mascorro, G.A., Aguilar-Torres, G.: Sistema para identificación de hablantes robusto a cambios en la voz. INGENIUS Journal of Science and Technology, pp. 45–53 Judith Shamian (2014). The power of unity. In: International Council of Nurses, pp. 301, International Nursing Review (2012)
14. Hong-wu, Y., Ya-Li, L., De-Zhi, H.: Speaker Recognition Based on Weighted Mel-cepstrum. Journal Computer Sciences and Convergence Information Technology, pp. 124–1136 (2009)
15. Ownn, D., Larson, W.: Development of a Genetic Linkage Map for Cisco (Coregonus artedi), to Facilitate Integrated Studies of Adaptive Diversity. Research Gate (2017)

ISSN 1870-4069

Optimization of a Production Process from Demand and Number of Defective Units through Diffused Logic and Genetic Algorithms

Christian A. Mejía Ramírez, Martin Montes Rivera

Universidad Politécnica de Aguascalientes, Aguascalientes, Mexico

mc170003@alumnos.upa.edu.mx, martin.montes@upa.edu.mx

Abstract. Production planning and scheduling are important activities in enterprises since they allow maximizing productivity by efficiently managing financial, material and human resources for achieving production objectives, avoiding excessive or insufficient stock material. Production time and faulty pieces are common analyzed variables with different mathematical techniques applied to improve production processes, nevertheless, these methods can obtain solutions with variable quality depending on the given input data. An alternative for classical methodologies that does not require complex mathematical models or high precision data in an environment of imprecision and uncertainty is fuzzy logic, but even with this technique, tuning inference systems demand long time for achieving good results in a production process, an alternative is to hybridize this algorithm together with optimizing algorithms for adjusting the required parameters. Genetic Algorithms (GA) are metaheuristic optimizing systems for exploring possible solutions to well-structured problems, which have shown good results in the parametrizing of complex problems, using natural selection principles. In this paper is proposed a fuzzy inference system that allows planning and scheduling production using a Mamdani inference system tuned with a GA.

Keywords: planning of production, fuzzy, genetic algorithms, optimization.

1 Introduction

1.1 Description of Problem

Nowadays, the planning of production is one of the major problems faced by the managers and administrators of this process; among the variables that can be found in the problem are those corresponding to production levels, the supply chain, demand,

and capacity in stock, defective units, among others. The main objective of the production is the reduction of the total costs; however, in many situations there may be others, such as minimizing defective units in the process; or, to maximize production levels.

For example, one of the main objectives regarding the problem of the supply chain is the minimization of the costs of controlling the flow of materials among suppliers, industries, warehouses and users. With regard to the economic quantity of order in a company, it is primarily sought to minimize the inventory costs from the demand that is had, usually in regular periods of time. Another important problem is the aggregate planning, which consists in the determination of the quantity of production and its development of time to certain term.

In production planning there are several factors that are not foreseen when they are carried out, among which are the defective units that can be had, this problem can delay the delivery of products to meet the demand; or, not anticipate in time to fulfill the demand; that is why, by programming a diffuse inference machine, the best decision to be made is sought with respect to the level of production that one wants to reach, when there is a certain demand and a quantity of defective units.

1.2 Related Works

From its very beginnings fuzzy logic have been implemented for the solution of several industrial problems, where in diverse situations there are scenarios of imprecision and uncertainty; however, to such problems a decision must be made in the most appropriate way possible. In the article by Kaharam, Gülbay and Kahraman [11] several problems that have arisen in the industry are mentioned, among them the planning of the production; However, no mention is made of the defective units that may be reached in the production process, this is a variable that cannot be overlooked in planning, since this can generate delays with compliance with the satisfaction of the demand.

In [12], a diffuse model is applied to the planning and programming of steel production; This model had great benefits in terms of reduction of stock and storage costs, increases safety when making decisions and improves quality; however, the time factor to be used in the process must be taken into account, since in many situations it does not depend directly on the manager; the malfunction of a plant or machinery that causes defective units in most cases, is a mechanical problem. In the present article this variable is not omitted, in fact this can be implicit in the demanding and defective units, since from these variables we can determine the level of production (greater or less than stable), that is required, and with it, to know the operation time of the plant.

On the other hand, the investments that handled the planning of the production are important since it is intended that the costs be the minimal possible, in the article by Kahraman and Tolga [13], diffuse logic techniques are proposed to choose the most appropriate way to carry out investments in planning processes; however, not always the failures in the process are directly attributed to the manager of the planning, in different situations it depends on the capacity and effectiveness of the production plant. In this article, costs are not directly minimized, but if a strategy for cost minimization

is given in the action of production based on the demand and the defective units, then costs would be minimized.

Another contribution in this research is the genetic algorithms application, since they were used in the tuning of membership functions. These can carry out a better proposal of the belonging functions. In the article Gladkov, Gladkova and Gromov [14] it is proposed a solution to the planning of production processes by means of fuzzy logic and genetic algorithms, very similar to what is proposed here, but Gladkov article does not define variables such as defective units and the demand for make the decisions.

1.3 Theoretical Framework

The optimization applications are countless. Each process is feasible and optimized. There is no company that does not involve optimization problems within its activities. Many of the industrial processes can be formulated as optimization processes, such operations are presented in minimizing the time invested for the accomplishment of a task, the cost of elaboration of a product and the risk in an investment, or they can also be presented in the maximization of profits, product quality and efficiency of a device, among others.

The great part of the problems of optimization with practical implications in the industrial engineering are very complex and difficult to solve, among them we can mention the maximization of the production in a production line in such a way that the demand is satisfied without defective units. Such problems cannot be solved in an exact manner using classical optimization methods; however, under these circumstances the fuzzy logic methods and genetic algorithms have been consolidated as an alternative solution, because they do not require high precision information, only very general information given by an expert.

Production planning policies must balance production costs and losses for defective units, with the costs of not being able to meet an order on time. Below there is a model of production planning based on the demand that is held in a certain period and the defective units that can be present.

A diffuse set $\mathcal{A}$ (1.1), is defined as an assembly that can contain elements in a partial way; that is, the property that an element belongs to the set can be true with a certain degree of truth, said degree of truth is given by a function $\mu_{\mathcal{A}}: U \to [0,1]$, where U is a universe and $[0,1]$ is the closed interval with ends 0 and 1.

This definition associates each element $x \in A \subseteq U$ with a real number $\mu_{\mathcal{A}}(x)$ in ranges $[0,1]$, which is assigned to x. Then the diffuse set A can be written as in (1):

$$\mathcal{A} = \{(x, \mu_{\mathcal{A}}(x)) : x \in A, \mu_{\mathcal{A}}(x) \in [0,1]\}. \quad (1)$$

Let A and B two diffuse sets with $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{B}}$ membership functions, respectively: The intersection operation (2) of $\mathcal{A}$ and $\mathcal{B}$ denoted by $\mathcal{A} \cap \mathcal{B}$ is defined by:

$$\mu_{\mathcal{A}\cap\mathcal{B}} = \min(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)), \; x \in U. \quad (2)$$

The union operation (3) of $\mathcal{A}$ and $\mathcal{B}$ denoted by $\mathcal{A} \cup \mathcal{B}$ is defined by:

$$\mu_{\mathcal{A}\cup\mathcal{B}} = \max\big(\mu_{\mathcal{A}}(x),\ \mu_{\mathcal{B}}(x)\big),\ x \in U\ . \tag{3}$$

Membership function bell (4) is defined as:

$$bell(x; a, b, c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}}\ . \tag{4}$$

Fuzzy logic models use fuzzy sets to handle and describe vague and complex phenomena and use logical operations to reach the conclusion. Fuzzy arrays and fuzzy logic applied to control problems is related to a field of knowledge called Fuzzy Logic Control (FLC). It deals with control problems in an environment of uncertainty and precision. It is very effective when high accuracy is not required and the control object has variables available for measurement or estimation.

The diffuse control always involves the fuzzification process, which is performed always, it is the gateway to the diffuse inference system. It is a mathematical procedure in which an element of the universe of discourse (variable measured from the process) is converted into fuzzy set with a membership function.

Fuzzy controllers use rules, which combine one or more fuzzy sets of inputs called antecedents or premises and associate a fuzzy set of output called consequent or consequence. They involve fuzzy sets, fuzzy logic and fuzzy inference. These rules are called fuzzy rules, which are affirmations of the type IF-THEN. Fuzzy antecedent arrays are associated by blurred logical operations AND, OR, etc.

The rule used in this controller is Mamdami's (1975) fuzzy rule (5), which is a conjunction rule expressed by the operation ∧ (minime); If r_k is the conclusion or consequent and p_i and q_j are the premises or antecedents, then:

$$,\ (p_i \wedge q_j) \rightarrow r_k = \min(\mu_{\mathcal{A}_i}(x),\ \mu_{\mathcal{B}_j}(y), \mu_{\mathcal{C}_k}(z))\ \ r_k = r_{ij}\ , \tag{5}$$

with $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$; $k = 1, 2, \ldots, l$; y $(x, y, z) \in U_1 \times U_2 \times U_3$.

When evaluating the rules, you get as many fuzzy sets as there are rules, to defuse it is necessary to group these sets, this stage is called aggregation and there are several criteria to carry out this step. One of the most frequently used criteria is to group the inferred sets by means of the operation ∨ (maxime).

Defuzzification is a mathematical process used to convert a fuzzy set into a real number. The fuzzy inference system obtains a conclusion from the input information, but it is in fuzzy terms. This conclusion or diffuse output is obtained by the stage fuzzy inference, it generates a fuzzy set but the output data of the system must be a real number and must be representative of the whole set obtained in the aggregation stage, that is why there are different methods of defuzzification and give different results, the "most common and widely used" is the gravity centroid. With the centroid defuzzification method, the fuzzy output is transformed into a real number which is the coordinate of the center of gravity (1.6) of a fuzzy set output and is given by equation (6):

$$z_{CG} = \frac{\sum_{k=1}^{l} z_k \mu(z_k)}{\sum_{k=1}^{l} \mu(z_k)}. \tag{6}$$

A genetic algorithm (GA) simulates some aspects of Darwin's evolutionary theory of species and the laws of Gregory Mendel's inheritance. The first genetic algorithm was developed by John Holland in the 1960s at the University of Michigan; the pseudocode of his work is listed below:

- Define an objective function.
- Random Generation of possible solutions (population).
- Encoding population.
- Evaluate the population, thus starting the i-generation.
- Select solutions to be reproduced.
- Crossover operation.
- Offspring mutation.
- Replace elements of the population of the i-generation with the best element-elements of 6 and 7.
- Stop if the stop criterion is met, otherwise go to 4.

Fuzzy logic and genetic algorithms are part of the techniques that fall under the concept of soft computing, a term coined by Lofti Zadeh in 1992. When there is synergy between the above techniques, it is called computational intelligence.

2 Methodology

To make the diffuse inference machine, information was collected from an expert in the administration of the process of interest, with respect to the levels of demand, defective units and increase and decrease of units in the production process that are taken into planning, for example for the set increased demand was considered the range of 20 to 30 units with value of belonging to 1, from 10 to 20 units a linear function with value of membership 0 in 10 units and 1 in 20 units and from 30 to 40 units another linear function with membership value 1 in 30 units and 0 in 40 units. It is clearly seen that the belonging function of said set is trapezoidal, but to reduce the number of variables in the objective function of the genetic algorithm, a bell function was adjusted with domain equal to the interval [-50.50], which is the universe corresponding to the demand, and the function of belonged was defined by $\mu_A = bell(x; 10,4,25)$. Subsequently the same was done for each of the sets defined in the section. Then once having the defined sets, we continue to formulate the implications based on the experience of the manager or administrator of the process. The rule used in the controller is that of Mamdami, the aggregate was made through the maximum operation and the defuzzification by centroid.

In this work, the aim is not to minimize time directly, but to maintain an adequate level of production that will restore demand at a given time. The experience and knowledge of industrial engineers is of great importance in the construction of the model, since through them the possible alternatives that can be taken are considered.

The fuzzy models presented here have two input variables: The value of demand $\mathcal{D}$ for a product and the number of defective units $\mathcal{U}$ that are taken at the end of the

production process. The output variable, the production action ($\mathcal{P}$), suggests elaborating more units or no action at some point.

The entries are modelled by set $\mathcal{D}$ with five terms:

$$\mathcal{D} = \{S, D, E, A, C\},$$

where $S \coloneqq$ decreasing, $D \coloneqq$ diminished, $E \coloneqq$ stable, $A \coloneqq$augmented and $C \coloneqq$ increasing.

And the set $\mathcal{U}$ with four parameters:

$$\mathcal{U} = \{N, P, R, B, X\},$$

where $N \coloneqq$low $P \coloneqq$ few, $R \coloneqq$ average, $B \coloneqq$ several and $X \coloneqq$ exaggerated.

The output is given by $\mathcal{P}$ with four parameters:

$$\mathcal{P} = \{O, AP, AM, AG\},$$

where $O \coloneqq$ zero, $AP \coloneqq$ little augmented, $AM \coloneqq$average augmented and $AG \coloneqq$high augmented.

The universe considered in demand is $U_{\mathcal{D}} = \{a \in \mathbb{Z}: -50 \leq a \leq 50\}$, for the defective units is $U_{\mathcal{U}} = \{b \in \mathbb{Z}: 0 \leq b \leq 25\}$ and for the production is $U_{\mathcal{P}} = \{c \in \mathbb{Z}: 0 \leq b \leq 40\}$. The membership functions for each set corresponding to $\mathcal{D}$ (Demand) are:

$$\mu_S = bell(x; 25,1,-50), \quad \mu_D = bell(x; 10,10,-15), \quad \mu_E = bell(x; 7.5,3,0),$$

$$\mu_A = bell(x; 10,4,25), \quad \mu_C = bell(x; 20,1,50).$$

The membership functions for each set corresponding to $\mathcal{U}$ (Defective units) are:

$$\mu_N = bell(x; 0.9,1,0), \quad \mu_P = bell(x; 4,11,7.5), \quad \mu_R = bell(x; 5,5,9),$$

$$\mu_B = bell(x; 4,16,13.5), \quad \mu_X = bell(x; 1,1,25).$$

The membership functions for each set corresponding to $\mathcal{P}$ (Production) are:

$$\mu_O = bell(x; 1,1,0), \quad \mu_{AP} = bell(x; 4,9,5), \quad \mu_{AM} = bell(x; 8,8,16),$$

$$\mu_{AG} = bell(x; 8,14,35).$$

The number of rules obtained from the expert information is 25 (table 1), these rules were developed from information collected during one year on a production line and these conclude in the terms of departure.

Table 1. Rules given by expert.

	N	***P***	***R***	***B***	***X***
S	*O*	*O*	*O*	*O*	*AP*
D	*O*	*O*	*O*	*AP*	*AM*
E	*O*	*O*	*O*	*AM*	*AM*
A	*O*	*AP*	*AM*	*AM*	*AG*
C	*AP*	*AP*	*AM*	*AM*	*AG*

The following are some rules that lead to the production action according to table 1. Rule given by row 1 and column 1 of table (1): If $\mathcal{D}$ is decreasing (***S***) and $\mathcal{U}$ is low (***N***), then nothing must be done in production (***O***).

Rule given by row 4 and column 3 of table (1): If $\mathcal{D}$ is augmented (***A***) and $\mathcal{U}$ is average (***R***), then a medium increase in production must be made (***AM***).

For the objective function (fitness) of the genetic algorithm tuned by the fuzzy inference machine, the vector with the initial parameters equation (7) of the bell membership functions is given by an expert:

A = [25 1 -50 10 10 -15 7.5 3 0 10 4 25 20 1 50 0.9 1 0 4 11 7.5 5 5 9 4 16 13.5 1 1 25 1 1 0 4 9 5 8 8 16 8 14 35] (7)

The objective function (equation (8)) is:

$$y = \frac{1}{42}\sum_{i=1}^{42}|x_i - A_i| \tag{8}$$

Where A_i is the i-element of vector A. The objective is to minimize the function given in equation (8).

Since the problem to be optimized is of dimension 42, an initial population of 5000 individuals is used, each with 672 alleles, in which every 16 alleles form a chain of bits that represent a real number and a resolution of 1000 is used.

The selection method is by tournament, the tournament size applied here is 10, the number of crossover points is 5, the mutation percentage is 4% and 5000 generations are used.

3 Results

The comparative between the results obtained with genetic algorithm tuned values and the expert recommended system is shown below.

The table that shows the parameters given by the expert and the genetic algorithm of the membership functions corresponding to the demand are shown in table 2.

Table 2. Demand table.

Demand	Expert	Genetic Algorithm
Decreasing	$\mu_S = bell(x; 25{,}1, -50)$	$\mu_S = bell(x; 24.00{,}1.03, -32.76)$

Diminished	$\mu_D = bell(x; 10,10,-15)$	$\mu_D = bell(x; 10.01,9.98,-16.38)$
Stable	$\mu_E = bell(x; 7.5,3,0)$	$\mu_E = bell(x; 7.17,-0.004,-0.008)$
Aumented	$\mu_A = bell(x; 10,4,25)$	$\mu_A = bell(x; 10.00,4.00,24.57)$
Increasing	$\mu_S = bell(x; 20,1,50)$	$\mu_S = bell(x; 16.38,0.98,32.77)$

The table that shows the parameters given by the expert and the genetic algorithm of the belonging functions corresponding to the defective units is show in table 3.

Table 3. Defective units.

Defective Units	Expert	Genetic Algorithm
Low	$\mu_N = bell(x; 0.9,1,0)$	$\mu_N = bell(x; -0.0005,1.0085,0.0045)$
Few	$\mu_P = bell(x; 4,11,7.5)$	$\mu_P = bell(x; 4.10,11.00,7.5)$
Average	$\mu_R = bell(x; 5,5,9)$	$\mu_R = bell(x; 4.09,8.20,8.96)$
Several	$\mu_B = bell(x; 4,16,13.5)$	$\mu_B = bell(x; -0.0015,16.3925,13.5)$
Exaggerated	$\mu_X = bell(x; 1,1,25)$	$\mu_X = bell(x; -0.0045, 0.9995,24.58)$

The table that shows the parameters given by the expert and by the genetic algorithm of the belonging functions corresponding to production is shown in table 4.

Table 4. Production table.

Production	Expert	Genetic Algorithm
Zero	$\mu_O = bell(x; 1,1,0)$	$\mu_O = bell(x; 1.03,-0.0005,0.0005)$
Little Augmented	$\mu_{AP} = bell(x; 4,9,5)$	$\mu_{AP} = bell(x; 3.96,8.19,5.12)$
Average Augmented	$\mu_{AM} = bell(x; 8,8,16)$	$\mu_{AM} = bell(x; 8.20,8.19,16.39)$
High Augmented	$\mu_{AG} = bell(x; 8,14,35)$	$\mu_{AG} = bell(x; 8.00,14.02,32.76)$

The minimum value of the target function released by the genetic algorithm is 1.4136, which is the average error between the obtained value and the expert vector.

The functions tuned by the genetic algorithm are shown in the figure 1, this graph shows the tuning (red dotted lines), of the functions of belonging (blue continuous lines), of the diffuse sets corresponding to demand, defective units and action in production.

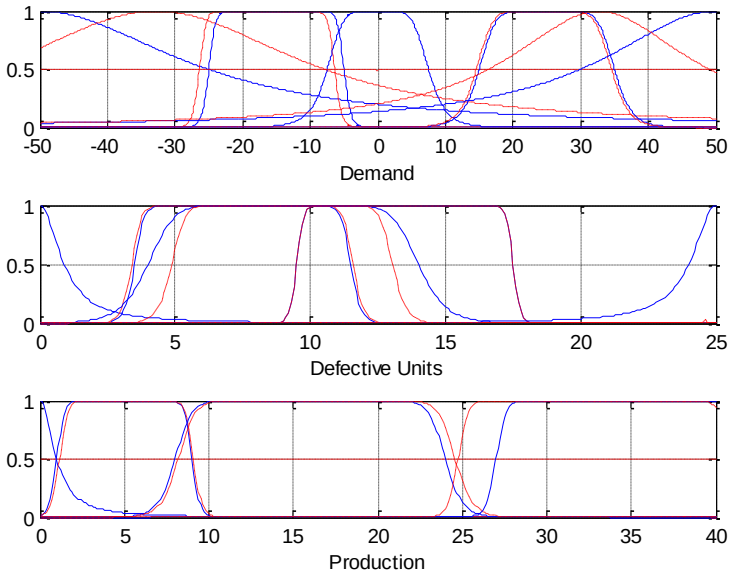


Fig. 1. Graph with comparative for demand, defective units and production sets with expert suggestion and Mamdani fuzzy logic system tuned with GA.

4 Conclusion

In this work is presented a self-tuned Mamdani fuzzy logic system by using GA, this fuzzy logic implementation it is applied for optimizing a production process based on demand required and the number of defective units, the proposed work achieves similar results to those suggested by a production expert.

One of the objectives of production planning is to anticipate the future for the efficient and effective completion of pre-established goals in planning. The objective of production planning is to make optimal use of material and financial resources to: a) meet demand, b) seize opportunities that may arise in the market and c) avoid the undesirable. However, in this process we cannot have a fixed demand nor avoid the undesirable, as in the case of having defective units, that is why experience in the processes can generate basic and efficient rules to have a control in the process and with them satisfy the demand and maximize the resources efficiency, in order to avoid the undesirable.

By means of the fuzzy control it is possible to make a production planning tuned with a GA, since when entering a value in the level of demand and in the level of defective units, the action in the production is generated. The parameters of the membership functions to the fuzzy sets corresponding to the action in the production of the table (4) shows that the GA is capable of tuning four of the output functions very similarly to those given by the expert; however, it cannot be said that the non-similar

functions thrown by the algorithm (Fig. 1), are not adequate, since expert system and genetic algorithm data differ only by an average 1.4136 units.

With regard to future work, we will seek to implement another objective function of the genetic algorithms to tune belongings function in such a way that the production time of the process is estimated. It will also be sought to implement these techniques to the inventory control systems of a pharmacy of public hospital.

References

1. Bojadziev, G., Bojadziev, M.: Fuzzy Logic for Business, Finance and Management. World Scientific, Singapore (2007)
2. Ross, T.J.: Fuzzy logic with engineering applications. Wiley (2004)
3. Sivanandam, S.N., Sumathi, S., Deepa S.N.: Introduction to Fuzzy Logic using MATLAB. Springer (2007)
4. Nguyen-Hung, T., Walker-Elbert, A.: A First Course in Fuzzy Logic. Champan & Hall (2006)
5. Bramlette, M.F.: Initialization, Mutation and Selection Methods in Genetic Algorithms for Function Optimization. Proceedings of Fourth International Conference on Genetic Algorithms, pp. 100–107 (1991)
6. Holland, J.H. Adaptation in Natural and Artificial Systems. Cambridge (1992)
7. Peña-Reyes, C.A.: Coevolutionary Fuzzy Modeling. Lecture Notes in Computer Science. 3204, Springer (2004)
8. Rahnamayan, S., Tizhoosh, H.R., Salama, M.M.A.: A Novel Population Initialization Method for Accelerating Evolutionary Algorithms. Computers and Mathematics with Applications, 53(10), pp: 1605–1614 (2007)
9. Wang, L.X.: A Course in Fuzzy Systems and Control. Prentice-Hall (1997)
10. Yager, R.R., Zadeh, L.A.: Fuzzy Sets, Neural Networks, and Soft Computing. Van Nostrand Reinhold (1994)
11. Cengiz, K., Murat, G., Özgür, K.: Applications of Fuzzy Sets in Industrial Engineering: A Topical Classification. StudFuzz, 201, pp. 1–55 (2006)
12. Skalna, I., Rebiasz, B., et al.: Advances in Fuzzy Decision Making. Studies in Fuzziness and Soft Computing, Springer (2015)
13. Kahraman, C., Tolga, A.: Fuzzy Investment Planning and Analyses in Production Systems. STUDFUZZ, 252, pp. 279–298 (2010)
14. Gladkov, L.A. et al.: Hybrid Fuzzy Algorithm for Solving Operational Production Planning Problems. Advances in Intelligent Systems and Computing 573 (2017)
15. Verma, R. et al.: Limits, Compression, Generalized Model and Heuristics. Computación y Sistemas, 21(4), pp. 787–798 (2017)
16. Smailovic, J. et al.: Automatic Analysis of Annual Financial Reports: A Case of Study. Computación y Sistemas, 21(4), pp. 809–818 (2007)
17. Cruz, E. et al.: Simulation and Application of Algorithms CVRP to Optimize the Transport of Mineral Metallic and Nonmetallic by Rail to Export Level, Research in Computing Science 122, pp. 77–88 (2010)
18. Briones, G.E. et al.: Design of Technology Management Infrastructure for Large Volumes of Data in Intelligent Power Network. Research in Computing Science 122, pp. 113–126 (2016)
19. Souverville, S. et al.: Fuzzy Logic Applied to Improvement of Image Resolution using Gaussian Membership Functions, Research in Computing Science 102, pp. 77–88 (2015)

ISSN 1870-4069

Heuristic Mechanism for Drone Swarm Auto-Organization

Juan Muñoz, Alberto Ochoa, Israel Hernández, Gilberto Rivera, Felipe Vázquez

Universidad Autónoma de Ciudad Juárez, Chihuahua, Mexico

alberto.ochoa@uacj.mx

Abstract. Current technology used for fires are: meteorological stations and satellite image and satellites. This last is a great option, but, from space, the fire detection is until it is large enough to be seen from orbital altitude. A drone can be used to monitor a forest looking for fire signs before the satellite observes it, but a single drone to cover a large acre surface is not optimal. A drone swarm with auto-organization capacity, equipped with atmospheric sensors that detect fire hazard conditions or even a fire in an early stage, needs to be used to optimize the area coverage. Implement a heuristic algorithm for drone swarm auto-organization applicable for wildfire alert and detection. Forest fires are a big environmental problem due they are mainly detected until they have burned some square kilometers. When these are detected at the developed stage, the fire will be difficult to contain. Some wildfires affect agricultural along as residential areas causing significant economic loses.

Keywords: wildfires prevention, heuristic algorithms.

1 Introduction

1.1 Background

According to Bala et al [1], the prevention of deforestation and promotion of afforestation have often been cited as strategies to slow global warming. Deforestation releases CO2 to the atmosphere, which exerts a warming influence on Earth's climate. However, biophysical effects of deforestation, which include changes in land surface albedo, evapotranspiration, and cloud cover also affect climate.

Deforestation has several causes, but wildfires and illegal tree cut, which are man-made, are the mayor. Current technologies used to detect wildfires are: meteorological stations and satellite image and satellites. This last is a great option, but, from space,

the fire detection is until it is large enough to be seen from orbital altitude. A drone can be used to monitor a forest looking for fire signs before the satellite observes it, but a single drone to cover a large acre surface is not optimal.

A drone swarm with auto-organization capacity, equipped with atmospheric sensors that detect fire hazard conditions or even a fire in an early stage, needs to be used to optimize the area coverage. Implement a heuristic algorithm for drone swarm auto-organization applicable for wildfire alert and detection. Forest fires are a big environmental problem due they are mainly detected until they have burned some square kilometers. When these are detected at the developed stage, the fire will be difficult to contain. Some wildfires affect agricultural along as residential areas causing significant economic loses. Unmanned Aerial Vehicle (UAV) is mainly known as drone. These devices have been flying for some decades now, mainly for military purposes. From 2010 to date, these have been more accessible for non-military purposes. Photography, real estate, utilities and construction are the main fields that have adapted these technologies.

2 Forest Fire Prevention Using Data-Logging/Transmitting Drone Swarm Triad

2.1 Forest Wild Fires in Chihuahua State

Forest fires are one of the main causes of forest loss. Vegetation covers in forest ecosystems and as a consequence erosion and soil degradation. Historically, our State has seen affected by this type of casualties, the years of greatest occurrence were in the 2011 and 2012, which showed a significant decrease to the year 2015, returning to present a rebound during the 2016, with a historical average of 870 annual fires between 1995 and 2016.

The main causes of fires are still those related to Agricultural activities, slash and burn and the Crops, with a percentage that fluctuates between 25 and 60%. Regarding the type of ecosystem (CONAFOR 2015), the most affected is the

Cold temperate climate, followed by natural pasture. The smaller area

Report is the arid and semi-arid, as for forests there were no fires. The most common type of fire is the superficial fire.

The year 2015 presented totally atypical conditions, since only Presented 252 forest fires, affecting a total area of 1,974.05 ha. The municipalities that historically have the largest affected area are Guadeloupe and Bald, Wood, Bocoyna, Guachochi and Balleza. The municipalities less affected were Janos, Casas Grandes, Ocampo and Urique [2].

Table 1. Historical data of wild fires in Chihuahua, Mexico.

YEAR	QUANTITY OF FIRES	BURNED SURFACE	AVERAGE SURFACE
2006	1,057	18,505	17.51

YEAR	QUANTITY OF FIRES	BURNED SURFACE	AVERAGE SURFACE
2007	625	10,560	16.9
2008	1,153	17,216	14.93
2009	842	10,704	12.71
2010	697	29,316	42.06
2011	1,687	87,920	52.12
2012	1,473	55,979	37.47
2013	1,137	30,554	26.87
2014	817	17,600	21.52
2015	251	1,974	7.83
2016	701	13,353	19.04
AVERAGE	**949**	**26,698**	**24**

3 Problem Formulation

As stated above, forest fires are a problem due the devastation it leaves, environmental and economic. Big efforts are applied to solve the wildfire, but due the large extensions of the forests, most are detected at a late state that commonly leads to large spread with high difficulty to contain.

Surface recognition is a very common need like for agriculture, disaster relief operations, goods delivery, etc. Depending on the area size, this could take a significant amount of time.

For this, there is a need to use tools like a drone, but, also, this need so to have a mechanism to efficiently perform these activities and, mainly, surface recognition to optimize time and resources. Figure 2 depicts need to coordinate the drone swarm.

3.1 Formalization of the Problem

We consider the problem of monitoring a large geographical area using a drone swarm to prevent forest fires. The area to be monitored is divided into well identified sub-areas. A drone swarm is composed by a set of heterogeneous drones, which are located at a starting point.

Thus, the problem that we tackle is to create a schedule containing the assignment of drones to geographical sub-areas for monitoring and detecting forest fires, such that the schedule completion time is minimized. In figure 1, the geographical area is represented in quadrants to define the tasks surface.

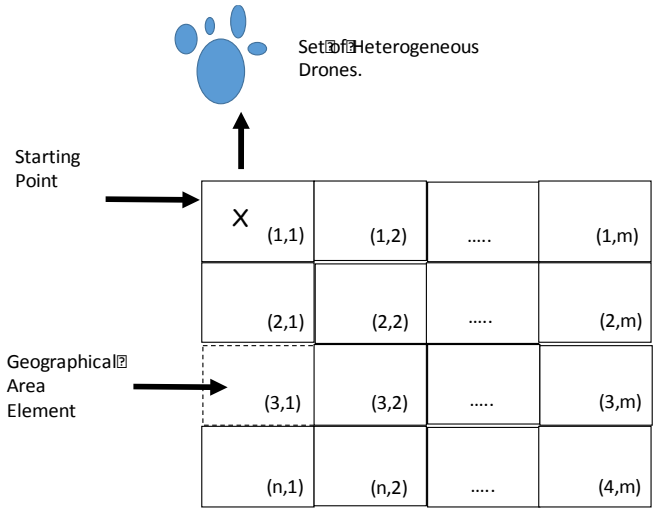


Fig. 1. Representation of the surface.

A scheduling system model for planning the visit of drones to geographical sub-areas consists of the following elements: the geographical area, drones swarm and an objective function for scheduling.

3.2 Large Geographical Area

The geographical area is denoted by *A*. Without a loss of generality, we assume that A has a square shaped area and does not contain any obstacle. The square shape was chosen for simplicity in the model. The area *A* can be divided into finite sub-areas forming a vector $A=\{a_0, a_1, a_2, .. a_{nm}\}$ of dimension n-by-m. For convenience, we consider a_0 as the base from which drones depart and return after complete their mission. We use $Geo(a_i)$ to denote the geographical position of a_i on A.

3.3 Heterogeneous Drone Swarm

The Heterogeneous Drone Swarm (HDS) can be represented by a DAG *HDS::(D, E).* *D* represents the set of heterogeneous drones that compose the swarm. *E* is the set of directed arcs connecting different pairs of drones, so $e(d_i, d_j)$ denotes a precedence that indicates that dron d_j cannot start its mission until d_i finishes its mission. For convenience, $Pred\ (d_i)$ denotes the subset of drones that directly precede d_i and $Succ\ (d_i)$ denotes the subset of drones that directly follow d_i.

The entry drone are those with $|Pred\ (d_i)| = 0$ and the output drone are those with $|Succ\ (d_i)| = 0$. For simplicity, in these cases we consider the use of dummy tasks such that the dag contains only one entry and output dron. Remembering that the drones are heterogeneous, we represent the estimated flying time from the base at a_0 with $EFT : D \times A \rightarrow Int$, where $EFT(d_i, a_j)$ denotes the time for a dron d_i to reach a geographical sub-

area a_j. For simplicity, we consider that the flying time to return to the base at a_0 is the same than the time to reach a particular area from a_0. A dron can be assigned to different missions, but it can only perform one mission at time. Thus, at time *t* we consider *avail* $:D \rightarrow [0..1]$, which captures the availability of each dron at time *t*. Note that the time of the mission of a particular dron is given when it is working at full availability. $W(d_i)$ denotes the time for a dron d_i to execute certain work once it reaches a geographical sub-area. $Setup(d_i)$ denotes the setup time for a dron to start a new mission. We assume that information about the flying and setup time are provided in standard time units, compatible with our drone performance measures.

3.4 Scheduling Problem

Scheduling drones to geographical areas requires the consideration of four events: (a) the time at which the dron starts its mission. (b) The time for a dron to reach a particular geographical area. (c) The time for a dron to perform certain work once it reaches its geographical area and (d) the time for a dron to return to the base.

Thus, we first need to predict the time at which a particular dron departs from a_0 to perform its mission to a particular sub-area and the time in which the dron returns to the base. We must first define two mutually referential quantities. $EDT(d_i, a_m)$ is the *Estimated Departing Time* of dron d_i to a_m, it is calculated by:

$$EDT(d_i, a_m) = Setup(d_i) + \max d_j \in Pred(di)\{ERT(d_j, a_0)\}. \quad (1)$$

$Setup(d_i)$ is preparation time for a dron to start a new mission. It is added to the result of the max block in Equation (3), which returns the maximum estimated returning time in which each drones in $Pred(d_i)$ return to the base. This is calculated by $ERT(d_i, a_0)$, which denotes the *Estimated Returning Time* of dron d_i to the base located at a_0 and it is calculated by:

$$ERT(d_j, a_0) = EDT(d_j, a_m) + (2 * EFT(d_j, a_m)) + W(d_j). \quad (2)$$

Once that all the drones have been scheduled, the estimated completion time of the schedule is determined by the estimated return time of the output dron. The estimated completion time is also known as the schedule make span:

$$ERT(d_{output}, a_{m.n}). \quad (3)$$

The objective function for drone scheduling aims to create a schedule containing the assignment of drones to geographical sub-areas such that its make span is minimized.

3.5 DERT Algorithm

The DERT algorithm is based on the well-known list scheduling approach. Our interest in this approach is to explore low computational complexity strategies and apply them to prevent and combat forest fires with the use of drones. Thus, the DERT algorithm basically consists of two phases: The *drone prioritization phase* in which a priority rank assignment is set to each dron. The *geographical sub-area assignation phase* where

each dron is assigned to that geographical sub-area which optimizes a predefined cost function. The DERT algorithm is shown in Figure 2.

I. Drone Prioritization Phase

We use $DRu(d_i)$, an upward rank defined as the length of the critical path from dron d_i to the output dron. $DRu(d_i)$ is calculated recursively as:

$$DRu(d_i) = \text{avg}(FT_i) + max\ v_j \in Succ(vi)\ (DRu(v_j)), \quad (4)$$

where avg(FT_i) is the average of the visit time for a dron d_i across all sub-areas:

$$avg(FT(di)) = \sum_{k=0}^{nm} \frac{(di,ak)}{n.m}. \quad (5)$$

1. Set the drone flying time.
2. Set the drone setup time.
3. Set the drone work time.
4. Calculate DR_u for each dron by traversing the graph from the exit node to the entry node and keep the values in *L*.
5. Sort the drones in *L* in descending order of DR_u values.
6. Create a list *LSA* with the sub-areas composing A.
7. **while** there are unvisited areas in ***LSA*** do
8. Select the first sub-area a_m from *LSA*
9. **for** each available dron d_i ($avail(d_i)=1$) in L **do**
10. Compute *EDT* (d_i, a_m) value.
11. Compute *ERT* (d_i, a_0) value.
12. Assign dron d_i to the sub-area a_m that minimizes *ERT* of d_i.
13. Set $avail(d_i) = 0$ from the time between *EDT* and *ERT*.
14. **end while**

Fig. 2. DERT algorithm.

II. Sub-area Assignation Phase

The DERT algorithm considers that a dron can be assigned to several missions, but it only can perform once at time. A mission involves to depart from the base a_0 to an assigned area a_m, perform a work once it reaches a_m and return to a_0. In our case, the work that a dron performs at a particular area is to monitor. The assignation phase where a dron is assigned to a geographical sub-area offering the minimum estimated returning time, takes $O(d\ x\ e)$ time complexity for *d* drones a *e* precedencies.

3.6 Example

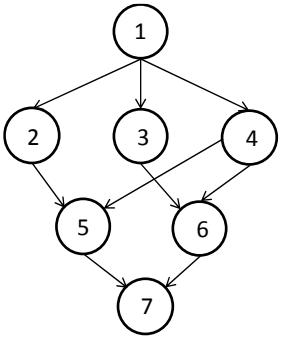
Table 2. Sub-area assignation matrix.

Dron	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	12
1	7	3	9	4	9	4	3	3	1	2	9	9
2	4	7	18	3	2	7	7	2	7	8	7	5
3	8	2	19	4	7	1	2	8	4	4	4	1
4	9	9	17	3	2	7	9	1	2	2	3	5
5	4	8	10	5	9	9	8	8	9	1	4	9
6	5	5	9	8	2	3	5	4	5	8	8	4
7	2	3	11	2	3	3	3	7	3	2	7	3
8	7	5	14	4	7	5	5	5	5	3	3	4
9	3	4	20	3	4	4	4	4	2	4	1	4
10	1	5	2	7	2	5	5	5	10	7	3	2

4 Future work

Overview. In this section, the future work for the drone swarm algorithm is described, the next challenges to be taken and opportunities.

4.1 Fire Propagation Simulation Using Support Vector Machine



(a) Drone Swarm

a1	a2	a3	a4
a5	a6	a7	a8
a9	a10	a11	a12

(b) Area to explore

Support vector machine (svm) will be used to analyze a given dataset from the forest fires described in this work. The objective of svm is to have a tool so forest fires could be characterized and serve as base for propagation prediction, that lead to adapting the drone swarm algorithm to search for fire spots in the early stages of wild fires so its vicious propagation could be prevented.

4.2 Illegal Tree-cutting Detection

Another opportunity detected is that, all around the world, illegal tree cutting is a big problem with several causes and big consequences to the global ecosystem as mentioned in the introduction of this chapter. This comes to account due one of the next work involving the presented heuristic algorithm for drone swarm auto-organization is to monitor forests to detect illegal tree cutting. Not only surveying forests with the flying unmanned aerial vehicles (UAVs), but equip this swarm agents with image recognition based on a machine learning model to be developed. The authors of this chapter are now starting to get involved on this endeavor, as depicted in figure 3.

Fig. 3. A wood portion with illegal tree cutting identified by a machine learning model.

5 Conclusions and Future Research

The investigation has proven to be functional to adequately detect the beginning of a fire and how to give notice to the corresponding authorities. The relevance of our study lies in being able to identify the dimensionality (size of the fire). In our future work, the propagation of a fire will be analyzed by Support Vector Machine, analyzing the factors of the climate as in [3, 4, 5].

Additionally, using the artificial deep learning intelligence technique, we are looking to identify recurring patterns in the beginning of the fire on the forest in the southwest

of Chihuahua and how to determine the amount of trees that must be had to properly reforest after a large magnitude fire has occurred.

References

1. Malhi, Y., Roberts, J.T., Betts, R.A., Killeen, T.J., Li, W., Nobre, C.A.: Climate change, deforestation, and the fate of the Amazon. Science, 319(5860), pp. 169–172 (2008)
2. CNF: http://www.cnf.gob.mx:8090/snif/seif_chihuahua/programas/combatedeincendios-prevencion (2017)
3. Yanmaz, E., Yahyanejad, S., Rinner, B., Hellwagner, H., Bettstetter, C.: Drone networks: Communications, coordination, and sensing. Ad Hoc Networks, 68, pp. 1–15 (2018)
4. Colorado, J., Pérez, M.G., Mondragón, I., Méndez, D., Parra, C., Devia, C., Martinez, J.J., Molina, L. N.: An integrated aerial system for landmine detection: SDR-based Ground Penetrating Radar onboard an autonomous drone. Advanced Robotics, 31(15), pp. 791–808 (2017)
5. X. Wang, S. Poikonen, B.L.: Golden: The vehicle routing problem with drones: several worst-case results. Optimization Letters ,11(4), pp. 679–697 (2017)

ISSN 1870-4069

Solution of the Assignment of Schedules Problem with the Gray Wolf Optimizer (GWO) Applying Evolutionary Parameters

Humberto Muñoz Bautista[1], Juana Canul-Reich[1], Alejandro Padilla Díaz[2], Julio Cesar Ponce Gallegos[2], Jaime Muñoz Arteaga[2], Pedro Daniel García González[3]

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

[2] Universidad Autónoma de Aguascalientes, Aguascalientes, Mexico

[3] Triple3 Aguascalientes, Aguascalientes, Mexico

{hmuntista, jcanulreich, apadillarobot, jmauaa} @gmail.com, jcponce@correo.uaa.mx danielgarcia@tripletres.com

Abstract. The educational institutions are faced with the problem of assignment their educational resources (students, teachers, classrooms, material) each scholar period, so there is a need to look for new software tools to satisfy this goals, due to the great amount of possible restrictions that has the assignment schedule problem, it is convenient to use metaheuristics algorithms to obtain acceptable solutions in reasonable times, for this reason, the present article shows the proposal for the resolution of the assignment schedule problem in the educational institutions using the metaheuristic GWO, taking as a use case a public university, a private university and a higher education institute. The test was carried out in a private university, obtaining better results in comparison with a genetic algorithm, demonstrating the potential of the metaheuristic GWO with evolutionary parameters.

Keywords: GWO, metaheuristics, assignment of schedules, optimization, artificial intelligence.

1 Introduction

The assignment of schedules problem, is a recurrent situation in all educational institutions [1], where in each scholar period the schedule of the students is planned; assigning signatures, classrooms and teachers, with the objective of avoiding an overlap of hours [2]. Therefore it can be defined as a decision-making problem for the assignment of educational resources (students, teachers, classrooms, etc.), subject to restrictions. As such, it corresponds to optimization problems in computational

complexity theory classified as NP-complete problems or NP-hard problems [3], due to its complexity, resource limitations and number of restrictions.

These problems do not have an algorithm that solves them in a polynomial time [4], that is to say, it is not possible to find its optimal solution with acceptable computational efforts, although you can have high-speed computers working in parallel. A major problem of optimization is the phenomenon called combinatorial explosion that means, when the number of decision variables of the problem grows, the number of feasible decisions and the computational effort grow exponentially, requiring the entry of applications such as intelligent algorithms [5]. However, not all combinatorial problems are so complex to solve; there are some for which there are algorithms that solve these problems with a computational effort that grows polynomial with the size of the problem [6].

There are several techniques for the resolution of the NP-complete problems, one of them are the metaheuristics, which deliver an acceptable solution within a reasonable time or in others words "Meta-heuristic algorithms are black box procedures that, provided a set of candidate solutions, solve a problem or a set of problems instances" [7]. A new metaheuristic algorithm is the Gray Wolf Optimizer, which is based on the behavior of the gray wolf, where the hierarchy of leadership and hunting mechanism of the gray wolves in nature are imitated. This metaheuristic corresponds to the so-called population metaheuristics, specifically considered within the branch of the Swarm Intelligence o SI, this concept was proposed in 1993 [8], and at the same time allows to use evolutionary operators, so when using such operators, can be considered an evolutionary metaheuristic and bioinspired algorithms, which are used in the optimization [9].

Taking the concept of Bonabeau, Dorigo and Theraulaz, SI is "The Intelligent collective of groups of simple agents" [10]. The most popular SI techniques are: Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO) and the Artificial Bee Colony algorithm (ABC) [11]. Such techniques have the following advantages:

- Save the search space information during the iterations of the algorithm.
- Usually use memory to save the best solution obtained so far.
- Usually have few parameters to adjust.
- Have fewer comparison operators than evolutionary methods.
- SI algorithms are easy to implement.

Considering the above is intended to carry out a parameter adjustment to the Gray Wolf Optimizer and implement it in the solution of the assignment schedule problem through the application of evolutionary parameters, to try to obtain a tool that equals or preferably obtains better results than those thrown by the processes currently used (many institutions perform this task manually) or by those systems that apply traditional metaheuristics such as the genetic algorithm.

2 State of the Art

2.1 Metaheuristics

A metaheuristic is a heuristic method to solve one type of general computational problem, using the parameters given by the user about generic and abstract procedures

in a way that is expected to be efficient. The name of metaheuristic comes from putting the prefix meta ("above" or of "higher level") at term "heuristic" (heuriskein, "find") [12]. A heuristic refers to methods based on experience that help us solve problems [13], this means that metaheuristic algorithms perform a process at a higher level to find the solution to a problem. Such algorithms can be conceived as general strategies of design of heuristic procedures for the resolution of problems with a high performance, using the parameters given by the user on generic procedures and abstract in a way that is expected to be efficient, so they can be applied to different optimization problems with minimal changes to be adapted to a specific problem.

The metaheuristics are not the solution to everything and many times are often less efficient tan specific heuristics, in several orders of magnitude, in problems that accept this type of pure heuristics; however, the advantage of metaheuristics is exploited, when are applied in problems that do not have a specific algorithm or heuristic to obtain a satisfactory solution, or when it is not feasible to implement an optimal method. Metaheuristics are widely recognized as one of the best approaches to tackle combinatorial optimization problems [14], so they have become popular in the last couple of decades for its characteristics of simplicity, flexibility, by having free bypass mechanism and that can avoid falling into local optima.

Amazingly some of them as the Genetic Algorithm(GA) [10], Ant Colony Optimization(ACO) [15] and Particle Swarm Optimization (PSO) [16] are well known not only in the area of computational sciences but also in other areas of science as: biology, economy, structural engineering, among many others. In addition a large number of theoretical works, such as optimization techniques, have been applied in various fields of study, because metaheuristics are quite simple, most have been inspired by very simple concepts, typically related to natural phenomena such as: the behavior of animals, evolutionary concepts or physical phenomena.

2.2 Metaheuristic GWO

Is a recent metaheuristic proposed by Mirjalili, Mirjalili y Lewis in the year of 2014 [17], which is based on the behavior of the gray wolf (Canis lupus). The gray wolves are in the top of the food chain, because they are considered one of the best predators. One of the reasons why this happens is by their lifestyle in herds. This lifestyle entails a very strict social dominant hierarchy. The herds are composed by five to twelve wolves on average, with a male and a female leaders (alphas), which make the decision to hunt, place to sleep, time to wake up; that means they make the most important decisions in the herd and their orders are followed by the herd [18]. In the second step of the hierarchy are the beta wolves, this wolves are subordinate that helps the alpha in decision-making or in others activities (alpha counselor, discipline the rest of the herds), for this reason they are the likely candidates to be the following alpha. The lowest ranking in the herds is omega, their plays the role of scapegoat, they always have to summit to all the other wolves in steps above of them and they are the last wolves to eat. In this complex social hierarchy each individual plays an important role, however small it may be, which helps maintain balance and strength of the herd.

If a wolf is not an alpha, beta or omega, is called subordinate or delta, this wolves have submit to alphas and betas, but they dominate the omega, their roles are scouts, sentinels, elders, hunters and caretakers, this wolves watching the boundaries of the

territory, warning the herd in case of any danger, protect and guarantee the safety of the herd, help in the hunting prey to provide food and are responsible for caring for the weak, ill, and wounded wolves in the herd. Elders are the experienced wolves who used to be alpha or beta. However, in some cases the alpha follows the other wolves in the Heard, demonstrating a democratic behavior, another example of such behavior occurs at meetings, where, the entire herd acknowledges the alpha by holding their tails down. The alpha wolves are only allowed to mate in the herd, but interestingly, they are not necessarily the strongest member, but they are the best managing the herd, demonstrating that the discipline and organization in the herd is more important than the individual strength.

Muro [19] describes the main phases of the hunting techniques of the gray wolves as the result of the perfect combination between the social hierarchy and group hunting, the main phases are as follows:

- Tracking, chasing and approaching to the prey.
- Pursuing, encircling, and press the prey until it stops moving.
- Attack towards the prey.

This hunting technique and the social hierarchy of grey wolves are mathematically modeled to design the GWO through of the main use of vectors for the representation of the positions of hunter and prey.

The algorithm presents the following processes, taken from the natural behavior of the gray wolf hunt:

- Encircling prey.
- Hunting.
- Attacking prey (exploitation)
- Search for prey (exploration)

2.3 Assignment of Schedule Problem

Describing in a general way the variables and the resources that the universities have for the assignment of schedule that they realize each scholar period, we have the following points [20]:

- Classrooms,
- Subjects,
- Teachers,
- Days of week,
- Class time,
- Assignments by period,
- Attendees,
- Room capacity,
- Maximum of attendees for subject,
- Department or faculty,
- Teacher availability.

There are many approaches to the assignment of schedule problem but in the present work will be addressed for a general case in the institutions. In that model starts from

the assignment of classes by department, where it is considered the classrooms, teachers, periods and assignments by period, as we can see in the next example.

Semester: 1
Department: A
Groups:

- Group: 1
    - Subjects:
        - Subject: S1
            - Imparted by: T1.
            - Classroom: 1A
            - Periods: 3
                - Monday – 9:00 to 10:00
                - Tuesday – 9:00 to 10:00
                - Friday – 10:00 to 11:00
        - Subject: S2
        - …
        - Subject: SN
- Group: 2
    - Subjects:
        - Subject 1
        - Subject 2
        - Subject N
- Group: 3
    - Subjects:
        - Subject 1
        - Subject 2
        - Subject N

Group: N

As we can see the assignment of schedules entails a big number of variables to be considered, becoming a complex problem. "On a daily basis, when we are faced with a complex problem, we turn to different people with more experience or experts in the area of the problem, to know and analyze their ideas and points of view. What leads us to perform an analysis of this information and group it by a classification to make a decision towards the problem we want to solve" [21]; however by extending the assignment of schedule problem to larger dimensions, as it is the programming of schedule in universities or higher education institute, where there is a large number of students, teachers, subjects and more resources to assign in each school period , taking an appointment to an expert would take a long time, so they have been proposed several techniques to try solve that problem, an example of this techniques are: linear programming [22] and the metaheuristics, like the evolutionary algorithms [23], ACO[24] y GRASP[25]. In such a way that it is proposed to use the algorithm GWO for the resolution of the assignment of schedule problem.

3 Methodology

For the realization of the research we make use of the quantitative methodology, taking into account the fulfillment of the following points:

- Splices of groups,
- Teacher splices,
- Free hours.

The objective is to minimize each point, taking as an optimal result to obtain 0 in each item. For the study case, the available schedules of teachers according to their regulatory academic burden, the subject (s) they can teach according to their academic profile, the subjects per group and the number of students that can be enroll by subject, are taken into consideration. In such a way that the teacher-matter-group relationship can be generated, obtaining the schedules assigned to each group, avoiding the points mentioned at the beginning of the section.

The tests are done using a laptop with a processor Intel Core i7 de 2.6 GHZ, 16 GB in RAM and with a SSD PCI-E of 512 GB and the development of metaheuristic is carried out with Matlab, Reading the information of teachers, subjects and groups from a file in Excel.

The expected results are to minimize the splices of groups, teacher splices and the free hours for the students, using the GWO metaheuristic adapted to solve the problem of assignment schedules as described in the next section.

4 Proposed Model to Resolve the Assignment of Schedule Problem by the GWO

Taking as a start the example of the previous model, the following parameters and restrictions are defined, to obtain a mathematical model that allows us to propose the assignment of schedule problem to be solved by the GWO.

4.1 Parameters

- r = Class rooms,
- s = Subjects or assignment,
- t = Teacher,
- wp = Week periods, (ex.4 times a week)
- sh = Schedule shifts,
- g = Group.

where, all of above integers:

$$\forall \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}, \boldsymbol{wp}, \boldsymbol{sh}, \boldsymbol{g} \in Z. \tag{1}$$

Cr= Room capacities. Where each element **r** of set **C** is an Integer that represents the capacity for r' room.

$$\forall \boldsymbol{r} \in [1, \ldots, N]. \quad (2)$$

Sa= Max Attendees by Subject. Where each element **a** of set **S** is an Integer that represents the Maximum of attendees that can be signed up to s' subject:

$$\forall \boldsymbol{a} \in [1, \ldots, N]. \quad (3)$$

Ph= Hours by Period. Where each element **h** of set **P** is an Integer that represents the Maximum of hours that a period p' is scheduled for:

$$\forall \boldsymbol{h} \in [1, \ldots, N]. \quad (4)$$

Th= Hours by Teacher. Where each element h of set T is an Integer that represents the Maximum of hours that a teacher t' can work:

$$\forall \boldsymbol{h} \in [1, \ldots, N]. \quad (5)$$

4.2 Variables

To build the viability V of schedules assignment, we analyze that there must be a classroom assigned to the subject with a single teacher in a specific day-hour period. We consider by default that the variable **V** is the viability for a single group **g** of a set of **N** groups. We obtain the next expression.

We can easily imagine this scenario as a Schedule sheet (figure 1) as you have in university; for example: Intelligent computing A that belongs to Department of Computer Science.

$$V\boldsymbol{r}'\boldsymbol{s}'\boldsymbol{t}'\boldsymbol{h}'\boldsymbol{p}'\boldsymbol{s}\boldsymbol{h}' = \begin{cases} \mathbf{1}, & if\ there\ a\ classrom\ \boldsymbol{r}'\ not\ assigned\ for\ a\ subject\ \boldsymbol{s}'with\ a\ teacher\ \boldsymbol{t}' \\ & on\ a\ specific\ hour\ \boldsymbol{h}', period\ \boldsymbol{p}'\ and\ schedule\ shift. \\ \mathbf{0}, & if\ not\ avaible. \end{cases}$$

As another point of view, we can see the same grid of schedule by subjects for groups by semester and by department (figure 2).

Hour\Week	Monday	Tuesday	Wednesday	Thursday	Friday
8:00-9:00	Classroom: C1 Teacher: T1. Subject: S1	...	...	...	...
9:00-10:00	...	...	...	...	...
10:00-11:00	...	...	...	...	Classroom CN Teacher TN Subject: SN

Fig. 1. Simple example of class schedule viewed as a calendar.

Assignments\ Days	Monday	Tuesday	Wednesday	Thursday	Friday
S1	Classroom: C1 Teacher: T1 Period: 9:00-10:00				
S2					
S3					
SN					Classroom: CN Teacher: TN Period: PN

Fig. 2. Example of schedule for group A, in Semester 1 by a Specific Department.

4.3 Restrictions

At this moment we have the main structure of the class schedule problem, but we need to define the hard and soft restrictions.

Hard restrictions

For the general model we consider some mandatory restrictions taking the parameters described above:

1. One Subject must be imparted by only 1 Teacher in 1 Classroom. For example: Biology – Mr. Lee – Classroom A:

$$V(\boldsymbol{s't'r'}) \leq \mathbf{1}\ ;\ \forall \boldsymbol{t, r.} \tag{6}$$

2. One Subject must be imparted in only 1 Classroom in a Single period. For example: Biology – Classroom A – 9:00 to 10:00:

$$V(\boldsymbol{s'r'h'p'}) \leq \mathbf{1}\ ;\ \forall \boldsymbol{r, h, p}. \tag{7}$$

3. Each Subject must complete the all Weekly periods. For example, Biology – 5 periods by week:

$$\sum_{i}^{s} SA[i] = \boldsymbol{wp}\ ; where\ SA = \{s_1, s_2, .., s_s\}. \tag{8}$$

There must not be more than 1 Subject assigned to the same period. For example: Monday- 10:00 - Biology, Monday -11:00- Artificial Intelligence:

$$V(\boldsymbol{s'h'p'}) \leq \mathbf{1}\ ;\ \forall \boldsymbol{h,p}. \tag{9}$$

4. Not exceed the maximum capacity of classroom. Example: 50 attendees are impossible to assign in the Classroom A, due it must be less or equals than 40 for Classroom A:

$$X \leq C[r]; where\ C = \{c_1, c_2, \dots, c_r\}\ capacities. \tag{10}$$

5. Not exceed the maximum teacher period assignment (T). For example, the Teacher T1 is only available for 4 weekly periods, so their assigned classes (X) should not be greater than 4:

$$X \leq T[h]; where\ T = \{h_1, h_2, \dots, h_t\}. \tag{11}$$

Soft restrictions

For this general model, we do not consider a specific set of soft restrictions due these restrictions are particularly not needed. But inside of this restrictions you can consider any specific situation like: A Subject cannot be imparted in a specific classroom, or a set of teachers prefer a specific turn, and so on.

4.4 Objective Function

For our main fitness function, we have to consider the set of hard restrictions (high cost) are accomplished at the same time that we have a positive result of our Viability result, adding as an optional soft restrictions (low cost).

The equations establish that the calculus of value of goodness an individual i at an instant t is given by:

$$r(i,t) = \sum_{j=1}^{Nc} |s(i,j) - c(i,j)|, \tag{12}$$

where:

- s (i, j) = Desired value for individual i in case j,
- c (i, j) = Obtained value for individual i in case j,
- Nc= Number of cases,
- i= individuals,
- j= Cases.

As we need to minimize “conflicts”, these occurrences that would be present for class schedule solution should be calculated:

1. Sum of conflicts for soft restrictions:

$$\boldsymbol{hrcs} = \sum_{i}^{w} confl(Ws[i], HR), \tag{13}$$

where:

- $HR = \{hr_1, hr_2, \dots, hr_n\}$ Hard restrictions,
- $Ws = \{w_1, w_2, \dots, w_n\}$ Wolf solutions,

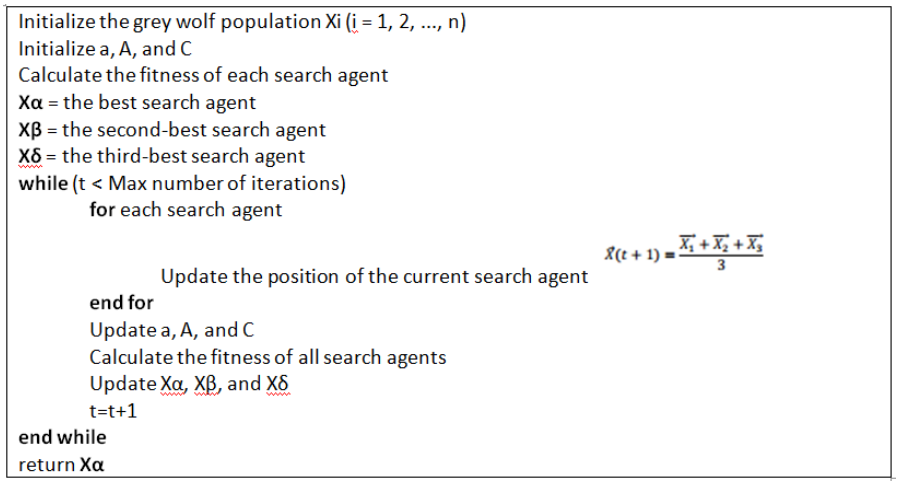


Fig. 3. Algorithm GWO proposed by Mirjalili in 2014.

2. Sum of conflicts for soft restrictions:

$$\boldsymbol{srcs} = \sum_{i}^{w} confl(Ws[i], SR), \tag{14}$$

where:

$\mathrm{SR} = \{\mathrm{hr}_1, \mathrm{hr}_2, \dots, \mathrm{hr}_n\}$ are soft restrictions.

Confl () is a function that evaluates a generated solution against a Soft or Hard set of restrictions.

Taking as reference the above equations we have the next objective function. Minimize the sum of conflicts for hard restrictions and soft restrictions:

$$f = MIN(\, hrcs + srcs). \tag{15}$$

In the literature we can see many examples of metaheuristics applied for general problems of scheduling, such as genetic algorithms, ant colony optimizations, swarm particle optimization, and so on.

In this article we are considering a new adaptation for this problem to be solved by an adaptation of general metaheuristic of the Gray Wolf Optimization. (GWO, figure 3)

In that adaptation of the metaheuristic we consider establish the next set of parameters and variables, which we described in the section III of the present article:

```
Initialize the grey wolf population Xi (i = 1, 2, ..., n)
Initialize a, A, and C
Initialize r, s, t, wp, Rc, and Sa
Initialize Hr, Sr restriction sets.
Calculate the fitness of each search agent
Xα = the best search agent
Xβ = the second-best search agent
Xδ = the third-best search agent
while (t < Max number of iterations)
        for each search agent
                Update the position of the current search agent unless V is feasible
                end unless
        end for
        Update a, A, and C
        Calculate the fitness of all search agents
        Update Xα, Xβ, and Xδ
        t=t+1
end while
return Xα
```

Fig. 4. Adapted algorithm GWO.

Table 1. Quantitative comparison of results between GWO, SCAIES (ACO) and manual method.

Method	Splices of groups	Teacher splices	Free hours
GWO	0	10	11
SACAIES	4	0	93
Manual method	0		
The smallest value is better	Avoid splicing groups at the same time and place	The manual method prevents the joining of teachers; however, he does it in a much longer time	Free hours is an important part of the solution to avoid having students with cut-off hours

- Parameters r, s, t, wp, Rc y Sa,
- Hard restrictions Hr [],
- Soft restrictions Sr [],
- Viability (Search agent positioning) V,
- Fitness function.

Resulting in the algorithm presented in the figure 4.

Fort the proposed algorithm the update of position of the current search agent occurs when the viability of solution is feasible.

5 Results

This section show, a comparison of the results obtained between GWO metaheuristic with evolutionary parameters, the SACAIES system [26], which uses a genetic

algorithm and a manual method for the assignment of schedules. Table 1 shows the results obtained by each of the methods. Taking into account the main objectives of the assignment of schedules (teacher splices, splices of groups and free hours).

As can be seen, the proposed metaheuristics improves the results obtained by a manual method and that obtained by an ACO metaheuristic, avoiding the splices of groups and reducing both free hours and the teacher splices.

Although it does not avoid 100% of teacher splices and free hours, the result obtained allows have the satisfaction of the students because they do not have cut-off schedules and make the most of their hours per day, while the teacher splices, being a single case, allows the intervention of the administrative responsible for the planning of schedules to solve a single case, instead of assigning all hours from 0. With these solutions, the metaheuristic adapted from the GWO demonstrates good results for the assignment of schedule problem, fulfilling the objective of avoiding splicing of groups and minimizing the free hours and the teacher splices.

6 Conclusions and Future Work

In the present work we established the necessary criteria for proposed the solution to the assignment of schedule problem with the GWO, resulting in an adaptation of the metaheuristic and obtained a general model to the solution of the problem. For future works we are in the approach of the problem to a higher education institute, a private university and a public university of the state of Aguascalientes, because each one of them has a different organization, which causes the problem approach to be modified.

For future works we proposed, the implementation of the GWO adapted to make an allocation of menus per day in a balanced diet, test the metaheuristic in public institutions of higher and middle higher level; as well as attacking the problem of assignment task by processes, recurrent in the industry.

References

1. Baldonado, M., Chang, C. CK., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1, pp.108–121 (1997)
2. Aguilar, L.D.M.O., Valadez, J.M.C., Soberanes, H.J.P., González, C.L.D., Ramírez, C. L., Soria-Alcaraz, J.: Comparativa de algoritmos bioinspirados aplicados al problema de calendarización de horarios. Research in Computing Science, (94), pp. 33–43 (2015)
3. Bruce, K.B., Cardelli, L., Pierce, B.C.: Comparing Object Encodings. In: Abadi, M., Ito, T. (eds.): Theoretical Aspects of Computer Software. Lecture Notes in Computer Science, 1281, Springer-Verlag, pp. 415–438 (1997)
4. Santoyo, A.M., Ortega, J.P., Salgado, G.R., Ortega N.N.A.: Caracterización de instancias difíciles del problema de Bin Packing orientada a la mejora de algoritmos metaheurísticos. Research in Computing Science, (19), pp. 295–308 (2015)
5. van Leeuwen, J. (ed.): Computer Science Today. Recent Trends and Developments. Lecture Notes in Computer Science, 1000. Springer-Verlag (1995)
6. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn. Springer-Verlag (1996)

7. Cosıo-León, M., Martınez-Vargas, A., Gutierrez, E.: An Experimental Study of Parameter Selection in Particle Swarm Optimization Using an Automated Methodology. Research in Computing Science, 82, pp. 9–20 (2014)
8. Ahumada, A., et al.: Generación de horarios académicos en INACAP utilizando algoritmos genéticos (2014)
9. Contreras-Varela, L.E., Mezura-Montes, E.: Comparativo empírico de medidas de diversidad en problemas de optimización evolutiva con restricciones. Research in Computing Science 134, pp. 69–83 (2017)
10. Caballero, J.M.M.: Asignación de horarios de clases universitarias mediante algoritmos. Universidad del norte división de postgrados e investigaciones en ingeniería, maestría en ingeniería industrial, Barranquilla atlántico (2008)
11. Baltazar, R., Vázquez, J.E., Rada, A., Díaz, C.: Desarrollo de un sistema capaz de optimizar rutas de entrega utilizando algoritmos genéticos, SINNCO (2010)
12. Isaza, R.A.H., Porras, C.A.R., Rendón, R.A.G.: Técnicas Heurísticas Aplicadas Al Problema Del Cartero Viajante (TSP), Grupo de Investigación en Planeamiento de Sistemas Eléctricos, Universidad Tecnológica de Pereira (2004)
13. Beni, G., Wang, J.: Swarm intelligence in cellular robotic systems. In: Robots and biological systems: towards a new bionics?, Springer, pp. 703–12 (1993)
14. Prospero, L.A.C., Mejía-Lavalle, M., Ascencio, J.R., Rincón, V.V.V.: Aprendizaje incremental basado en población como buena alternativa al uso de algoritmos genéticos. Research in Computing Science, 116, pp. 51–64 (2016)
15. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm intelligence: from natural to artificial systems. OUP USA (1999)
16. Basturk, B, Karaboga, D.: An artificial bee colony (ABC) algorithm for numeric function optimization. In: IEEE swarm intelligence symposium, pp. 12–4 (2006)
17. Rodríguez, A. G.: Algoritmo del Lobo Gris para la solución de un problema de asignación cuadrática, Universidad Autónoma de Aguascalientes, Ingeniería en Computación Inteligente (2016)
18. Dorigo, M., Birattari, M., Stutzle, T.: Ant colony optimization. Comput Intell Magaz, IEEE 1, pp. 28–39 (2006)
19. Cardona, P., Muñoz, Á., Muñoz, J., Álvarez, F., Velázquez, C., Hernández, Y., Collazos, C.: Determinación de Heurísticas en Interfaces de Objetos de Aprendizaje Multiculturales mediante Cadenas de Markov. Conferencias LACLO, 4(1) (2013)
20. Kennedy, J., Eberhart, R.: Particle swarm optimization, in Neural Networks. In: Proceedings, IEEE international conference on, pp. 1942–1948 (1995)
21. Mirjalili, S., Mirjalili, M., Lewis, A.: Grey Wolf Optimizer. Advances in Engineering Software, 69, pp. 46–61 (2014)
22. Escobedo, C., Spector, R., Coppinger, R.: Wolf-pack (Canis lupus) hunting strategies emerge from simple rules in computational simulations. Behav Process, 88, pp.192–197 (2011)
23. Bautista, M. H., Díaz, A. P., Gallegos, J. C. P., Suárez, H. M., Aguilar, J. A. H., Canul-Reich, J.: Accomplishment of generic skills of the middle higher education with the use of system SiCRE, XI Latin American Conference on Learning Objects and Technology (LACLO), San Carlos, pp. 1–8 (2016)
24. Sánchez, D., Cano, P., Reyes, M.: Caso de un Modelo de Optimización para un Problema de Calendarización de Horarios. Universidad Popular Autónoma del Estado de Puebla, Puebla (2009)
25. Hernández, R., Miranda, J., Rey, P. A.: Programación de Horarios de Clases y Asignación de Salas para la Facultad de Ingeniería de la Universidad Diego Portales Mediante un Enfoque de programación Entera. Revista Ingeniería de Sistemas, XXII (2008)
26. Barrientos, J. D. J.: Aplicación de algoritmos genéticos para la asignación de carga académicos en instituciones de educación superior (2000)

Impreso en los Talleres Gráficos
de la Dirección de Publicaciones
del Instituto Politécnico Nacional
Tresguerras 27, Centro Histórico, México, D.F.
febrero de 2018
Printing 500 / Edición 500 ejemplares