Advances in Intelligent Systems and Computing 1249

Ernesto León-Castro
Fabio Blanco-Mesa
Anna Maria Gil-Lafuente
Jose M. Merigó
Janusz Kacprzyk *Editors*

Intelligent and Complex Systems in Economics and Business

Springer

Advances in Intelligent Systems and Computing

Volume 1249

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by SCOPUS, DBLP, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST), SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/11156

Ernesto León-Castro · Fabio Blanco-Mesa ·
Anna Maria Gil-Lafuente · Jose M. Merigó ·
Janusz Kacprzyk
Editors

Intelligent and Complex Systems in Economics and Business

Springer

Editors
Ernesto León-Castro
Universidad Catolica de la Santisima Concepcion
Concepcion, Chile

Fabio Blanco-Mesa
Universidad Pedagogica y Tecnologica de Colombia
Tunja, Colombia

Anna Maria Gil-Lafuente
Department of Business Administration
University of Barcelona
Barcelona, Spain

Jose M. Merigó
Department of Management Control and Information System
University of Chile
Santiago, Chile

Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
Warsaw, Poland

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-030-59190-8 ISBN 978-3-030-59191-5 (eBook)
https://doi.org/10.1007/978-3-030-59191-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Introduction

Nowadays, most of the problems within the economy and finance respond to realities of uncertainty; Kaufmann and Gil-Aluja (1986) point out that when a phenomenon cannot be specified it is necessary to use fuzzy studies to assign a degree of truth to variables and not a binomial logic of true and false.

The fuzzy sets began with the studies of Zadeh (1965), and from these bases different studies, models and methodologies have been generated within the economic and financial management, such as the case of this book.

This volume is a collection of 10 chapters that use different intelligent and complex systems within different cases. The book is based on the fields on which these methodologies are focused and is divided into three important aspects:

- Global and governmental analyses,
- Analyses within organizations, and
- Analysis of modern methods, and tools and techniques which are used.

Contents

Multicriteria Analysis Model for the Evaluation of the Competitiveness of the States in Mexico

Manuel Muñoz-Palma, Pavel Anselmo Alvarez-Carrillo, Eva Luz Miranda-Espinoza, Ezequiel Avilés-Ochoa, and Ernesto León-Castro

Abstract The objective of the article is to present the competitiveness of the federative entities in Mexico using a multicriteria analysis model based on composed indexes using ELECTRE III. The method considers a multicriteria decision support approach using expert preferences. The result is a model that allows the identification and ranking of the states in Mexico with the highest competitiveness rates. The limitations of the model are presented in the expert's cognitive effort to define the parameters of the ELECTRE III method. In addition, a multiobjective evolutionary algorithm for the generation of a total or partial pre-order is used. The findings are the ranking of the federative entities in Mexico compared with the competitiveness index by the Mexican Competitiveness Institute (IMCO for their acronym in Spanish).

Keywords Competitiveness · Composite indices · Multicriteria ranking · ELECTRE III

M. Muñoz-Palma
Department of Management, Department of Management Universidad de Sonora, Hermosillo, Mexico
e-mail: manuel.munoz@unison.mx

P. A. Alvarez-Carrillo (✉) · E. Avilés-Ochoa
Department of Economic and Management Science, Universidad Autónoma de Occidente, Culiacan, Mexico
e-mail: pavel.alvarez@uadeo.mx

E. Avilés-Ochoa
e-mail: ezequiel.aviles@udo.mx

E. L. Miranda-Espinoza
Decision-Making in Social Problems Research Group, Universidad Autónoma de Occidente, Culiacan, Mexico
e-mail: mirandaeeva@hotmail.com

E. León-Castro
Faculty of Economics and Administrative Sciences, Universidad Católica de la Santísima Concepción, Concepción, Chile
e-mail: eleon@ucsc.cl

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_1

JEL Classifications D72 · D70

1 Introduction

Territorial competitiveness is a significant variable to attract national and foreign private investment. This forces organizations to pay more attention to continuous improvement processes. The World Economic Forum (WEF), the body responsible for measuring the competitiveness of nations since 1979, defines it as the set of institutions, policies and factors that determine the level of productivity of a country (WEF 2016). The prosperity of a nation, says Porter (1990), depends on its competitiveness, which is determined by productivity. In this sense, competitiveness has a positive effect in creating value for the business and people in the country (Charles and Zegarra 2014).

The concept of competitiveness is important not only to know why some countries grow faster than others, but also to identify why other countries have a better income distribution (Charles and Zegarra 2014). Also, different dimensions of competitiveness must be analyzed to get a whole understanding of the concept; Lombana and Gutiérrez (2009) says that territorial competition considers political, social and economic mechanisms.

For the complexity associated with the competitive definition, a useful technique to evaluate it is the multiple criteria decision-making (MCDM) methods that are analytical methods to evaluate a set of alternatives on multiple criteria (Wang and Lee 2009; Huang and Peng 2012). For that, the analysis of the weights that are associated with each criterion is crucial to measure the importance of each criterion (Zhang et al. 2011).

The paper presents an evaluation and ranking of the competitiveness for the states in Mexico based on composite index from a MCDM support approach using ELECTRE III. The main idea is to compare the results obtained by the proposed method with the one published by IMCO in 2016. Among the main results is the presentation of a ranking based on 5 levels according to the results of the different competitive components.

The remainder of the paper is structured as follows: Sect. 2 presents an overview of competitiveness in Mexico and the methodology that is used to evaluate it, Sect. 3 explains step by step the way the methodology proposed in the paper is used, Sect. 4 shows some discussions of the results, and Sect. 5 summarizes the conclusions.

2 An Overview of Competitiveness in Mexico

Business competitiveness is defined as the company's ability to offer products and services that meet certain quality standards of local and global markets at competitive prices and that provide an adequate return for the resources used or consumed (Oyarce

2013). Therefore, the private capital organizations are oriented to global markets and in this sense to global competition, this leads the companies to innovate and improve their process to offer more quality and dynamic products and services. In the local perspective of a nation, the divergence in budgetary allocations generates a competitive disadvantage among the states/municipalities/towns and, consequently, concentration of investment in few regions (Crescenzi 2009).

To this situation, Labarca (2007) proposes an alternative vision to the explanations of competitiveness, specifically focused on the determinants that make an industry more competitive, that is, demand, strategy, the rivalry of companies, pressures and company capabilities.

In Mexico, the organization that measures the competitiveness of the states is the Mexican Institute for Competitiveness based on 10 criteria that are defined as follows (IMCO 2016).

$C1$ Legal system (LS) measures environment of public and legal security in regions. It closely linked to the quality of life of citizens through the prevention and elimination of what puts liberties, order and public peace at risk, safeguarding the physical integrity and the rights of people. A functional State of Law generates favorable conditions for the attraction and retention of investments in regions.

$C2$ Sustainable environmental management (SEM) measures the ability of regions to relate sustainably and responsibly to natural resources and their environment. It provides information on the availability and management of water, air and solid waste. It also outlines certain risks that could be incurred by companies that want to invest in the state. Both elements directly affect the quality of life of the inhabitants.

$C3$ Inclusive, prepared and healthy society (HIS) measures the quality of life of the inhabitants through three areas: inclusion, education and health. These give an indication of the opportunities that exist in a state to form, attract and take advantage of human capital. It includes indicators of academic performance, medical offer and health services, socioeconomic conditions, poverty and inequality. A state that offers high levels of quality of life for its entire population is much more attractive for talent and investments.

$C4$ Stable and functional political system (SPS) measures the potential of state political systems to be stable and functional. The good quality of the political system can encourage investment by creating an environment of healthy competition that leads to greater accountability. Indicators that give information on corruption, citizen participation in the political life of the state and civil liberties are incorporated. The good quality of the political system can encourage investment through the creation of a stable environment and public management accustomed to rendering accounts.

$C5$ Efficient and effective governments (EEGs) measure the way in which governments are able to positively influence the competitiveness of their states. Among the actions necessary to achieve this objective are public policies aimed at promoting local economic development. Therefore, this subindex includes indicators related to the promotion of economic development and the formality of the economy. In addition, it includes indicators on the capacity to generate own revenues, the quality of the information of its public finances and the approach with the citizenship by electronic means.

$C6$ Factor market (FM) measures the productivity of workers and other essential characteristics of employment, since this human capital represents the most important production factor for the competitiveness of each federative entity. Those entities where workers are more qualified and salaries are higher become more attractive for talent and, therefore, attract investment.

$C7$ Stable economy (SE) measures the main characteristics of state economies, as well as the credit situation for companies and families. It includes indicators that describe the distribution of GDP, the dynamism of the economy, the level of debt, as well as economic dependence and diversification. The states that present a stable economy as well as large credit markets attract more talent and investment and are, therefore, prone to a greater generation of employment and wealth.

$C8$ Precursors (P) measure the financial, telecommunications and transport sectors. These sectors are of great importance because they are considered as necessary conditions to boost economic growth, investment and employment generation, by directly affecting many other sectors of the economy. Therefore, its development is fundamental to improve the competitiveness of the states. This subindex considers indicators related to access to and use of the Internet, physical means of communication, whether aerial or terrestrial, and the use and access to financial services.

$C9$ Exploitation of international relations (EIRs) measures the degree to which the states capitalize their relationship with the outside to increase their competitiveness. Therefore, the subindex considers indicators related to international tourism and the flow of capital. In an environment of globalization, the competitiveness of the states of our country depends more and more on their ability to exploit the links they have with the outside world.

$C10$ Innovation in sectors of the economy (ISE) measures the capacity of the states to compete successfully in the economy, particularly in sectors of high added value, intensive in knowledge and cutting-edge technology. The ability to generate and apply new knowledge is considered, which includes indicators related to the characteristics of the companies, the research context and the generation of patents. A state that has more innovative economic sectors is able to attract and retain more investment and talent.

Also, each criterion is divided into different elements to analyze and give a score. This information is presented in Annex 1. The main idea of using this methodology provided a whole overview based on that information when a different approach based on the expectations of the investors is applied. The purpose is to identify which are the Mexican States that have a greater capacity for competitiveness so they would be more attractive to the investor. The process to do that is explained in the next section.

3 Evaluating and Ranking Competitiveness in Mexico

To construct the MCDM model that will be used to evaluate and rank the states in Mexico, the approach proposed by Bartolini et al. (2005) will be used. This approach identifies two important phases in order to do a multicriteria analysis: the construction of the preferential model (aggregation phase) and the exploitation of the preferential model (exploitation phase). In this sense, the decision support procedure through the multicriteria analysis approach is executed through a series of steps that are identified as follows: problem definition; identification of alternatives; identification of the evaluation criteria; obtaining information and measuring the evaluation criteria; determination of weights, indifference and preference parameters; aggregation and calculation of evaluation parameters; exploitation of the aggregation model; and proposal of ordering as final solution. Each one of these steps is described and explained as follows.

Step 1 Definition of the problem

In this step, the identification of the main variables is done. In this case, the information provided and published by IMCO will be used and the objective is to evaluate and rank the competitiveness of the states in Mexico. The decision-making process is based on the group decision support system with genetic algorithms and ELECTRE III (SADGAGE for its acronym in Spanish) system (Leyva et al. 2016); the decision problem lies in assessing the competitiveness of states in Mexico by generating a multicriterion ranking. The multicriteria ranking problem consists in the generation of decreasing order of preferences of a set of alternatives $A = \{a_1, a_2, \ldots, a_m\}$, considering a set of decision criteria $F = \{g_1, g_2, \ldots, g_n\}$.

Step 2 Identification of alternatives

In the evaluation of alternatives, the decision is the one that mitigates the problem (Bardach 2001). To do that is important to identify those alternatives to consider in the course of the analysis, to later discard the unsatisfactory ones, combine others and reorganize a single basic alternative with one or more variants, to finally choose the ones that provides the best option according to established criteria. In this case of study, the possible alternatives that can be chosen are the 32 states in Mexico (see Table 1).

Step 3 Identification of the evaluation criteria

To the case presented in the paper, the evaluation of competitiveness in the states of Mexico is derived from a set of dimensions defined by the IMCO (2016). The measurement of competitiveness is achieved by evaluating and comparing each of the states against each other, considering each of the dimensions that correspond to the decision criteria.

IMCO develops a theoretical framework that includes concepts, which derive in a model, where state competitiveness is determined by 10 dimensions (IMCO 2016).

Table 1 States in Mexico (alternatives)

Label	State	Label	State
A1	Aguascalientes	A17	Morelos
A2	Baja California	A18	Nayarit
A3	Baja California Sur	A19	Nuevo León
A4	Campeche	A20	Oaxaca
A5	Coahuila	A21	Puebla
A6	Colima	A22	Querétaro
A7	Chiapas	A23	Quintana Roo
A8	Chihuahua	A24	San Luis Potosí
A9	Ciudad de México	A25	Sinaloa
A10	Durango	A26	Sonora
A11	Guanajuato	A27	Tabasco
A12	Guerrero	A28	Tamaulipas
A13	Hidalgo	A29	Tlaxcala
A14	Jalisco	A30	Veracruz
A15	México	A31	Yucatán
A16	Michoacán	A32	Zacatecas

Being $F = \{g_1, \ldots, g_n\}$ a finite set of criteria and $n = |F|$ the total number of criteria. These criteria are presented in Table 2.

The formulations that will be used to calculate each of the criteria are presented in Annex 1.

Table 2 IMCO competitive criteria

Label	Criteria
$C1$	Reliable and objective law system
$C2$	Sustainable management of the environment
$C3$	Inclusive, prepared and healthy society
$C4$	Stable and functional political system
$C5$	Efficient and effective governments
$C6$	Market factors
$C7$	Stable economy
$C8$	Precursors
$C9$	Use of international relations
$C10$	Innovation of the economic sectors

Table 3 Specified scale of importance

Importance value	Definition
1	Equally importance
2	Moderate importance
3	Strong importance
4	Very importance
5	Extremely importance

Step 4 Define the weights for each criterion

A substantial element in the construction of composite indicators (criteria) is to support the decision maker to assign the relative importance weighting for each of the individual indicators. However, there are several methods to generate them; in this case, it is resolved to use the comparison matrix, a technique that is addressed by Alireza et al. (2010).

The procedure consists in constructing a matrix of comparison by pairs ($n \times n$) of criteria, by using a scale of comparisons by pairs of attributes. For each comparison, it is decided which of the two attributes is more important and then assigned a score (Alireza et al. 2010).

In this way, the weights of the attributes are calculated using comparison matrix. Meanwhile, the data will be suggested by the decision maker in opinion of the instrument in each dimension considered to determine the competitiveness of the states in Mexico by using values of the scale from 1 to 5 as shown in Table 3. To understand better the process, the example for $C1$ is presented (see Table 4).

Step 5 Make the evaluation for each criterion

Once the methodology of obtaining the weights is defined, then based on the scores proposed by IMCO (2016) the score for each criterion and each Mexican State is

Table 4 Attribute comparison matrix for reliable and objective law system

Attribute	$a1$	$a2$	$a3$	$a4$	$a5$	$a6$	$a7$	$a8$	$a9$	Sum	Weight
$a1$	1	4	3	1	3	2	3	1	2	20.00	0.1684
$a2$	0.25	1	3	2	2	3	2	3	1	17.25	0.1452
$a3$	0.33	0.33	1	4	3	3	2	2	3	18.67	0.1571
$a4$	1.00	0.50	0.25	1	2	2	3	4	2	15.75	0.1326
$a5$	0.33	0.50	0.33	0.50	1	2	2	3	3	12.67	0.1066
$a6$	0.50	0.33	0.33	0.50	0.50	1	2	4	3	12.17	0.1024
$a7$	0.33	0.50	0.50	0.33	0.50	0.50	1	3	3	9.67	0.0814
$a8$	1.00	0.33	0.50	0.25	0.33	0.25	0.33	1	4	8.00	0.0673
$a9$	0.50	1.00	0.33	0.50	0.33	0.33	0.33	0.25	1	4.58	0.0385
Total	5.25	8.50	9.25	10.08	12.67	14.08	15.67	21.25	22.00	118.75	1.00

Table 5 Computation of the Criterion 1 ($C1$) with eight attributes to Aguascalientes

	$a1$	$a2$	$a3$	$a4$	$a5$	$a6$	$a7$	$a8$	$a9$	$C1$
Aguascalientes	98	99	76	51	48	45	66	32	97	
Peso	0.1684	0.1452	0.1571	0.1326	0.1066	0.1024	0.0814	0.0673	0.0385	1
Total	16.50	14.37	11.94	6.76	5.12	4.61	5.37	2.15	3.73	70.60

obtained (see Table 6). The evaluation score for each alternative a_k in each criterion C_j ($C_j(a_k)$) is calculated by multiplying the x_{ki} value given to each attribute i for their corresponding relative importance weights w_i assigned by the decision maker, followed by the sum of the products for all attributes, such as:

$$C_j(a_k) = \sum w_i x_{ki} \tag{1}$$

To provide a better understanding of how the scores were obtained, and explanation for Aguascalientes $C1$ is detailed. The row Aguascalientes in Table 5 shows the value of each attribute for the criteria reliable and objective law system. The obtained weight by Step 4 (see Table 4) is used to compute the final performance for Criterion 1. Then, Eq. (2) is used to compute the final performance of Criterion 1 for the Aguascalientes and the results are presented in Table 5.

$$C_1(\text{Aguascalientes}) = \sum_{i=1}^{8} a_i * w_i \tag{2}$$

where a_i is the value of attribute i and w_i is the weight of the attribute i (Table 6).

Step 6 Ranking of the alternative through ELECTRE III method

The ELECTRE III method is an outranking approach developed by Roy (1978), which compares each pair of alternatives in the set $(a, b) \in A \times A$ to assess the credibility of the statement "a is at least as good as b" denoted by aSb. The method uses extra-criterion and intra-criterion information as input and preferential information for the construction of a partial and comprehensive concordance index (Roy 1978). The partial concordance index is defined by $C_j(a,b)$, evaluating the difference between two alternatives by the criterion g_j, using the DM's preference defined in the indifference q_j and preference p_j thresholds.

$$C_j(a,b) = \frac{p_j\big(g_j(a)\big) - \big[g_j(b) - g_j(a)\big]}{p_j\big(g_j(a)\big) - q_j\big(g_j(a)\big)} \tag{3}$$

The comprehensive index $C(a,b)$ is calculated by Eq. (4). This corresponds to the sum of the partial concordance indices $Cj(a,b)$ on each criterion considering the weights of each criterion w_j. The value of $C(a,b)$ expresses to what extent the performance of all the criteria is in accordance with the statement "a outranks b".

Table 6 Performance of the federative entities with respect to the decision criteria

	$C1$	$C2$	$C3$	$C4$	$C5$	$C6$	$C7$	$C8$	$C9$	$C10$
Aguascalientes	70.60	69.34	65.20	54.79	69.95	53.03	70.71	44.93	13.92	58.10
Baja California	21.24	48.29	59.84	47.04	40.48	51.05	55.62	46.79	15.24	45.46
Baja California Sur	2.23	54.22	60.59	49.43	41.30	72.13	62.64	54.04	21.89	25.73
Campeche	82.32	44.33	49.72	72.25	56.52	52.28	50.91	32.77	5.08	37.47
Coahuila	76.39	65.73	57.84	68.56	50.67	54.06	57.40	34.72	12.32	51.84
Colima	10.00	65.79	64.29	63.14	78.81	50.49	68.35	40.93	9.72	31.20
Chiapas	1.40	29.10	26.60	47.00	40.55	16.83	44.89	18.86	2.52	20.28
Chihuahua	0.21	63.48	46.93	52.72	55.60	50.17	51.97	38.57	37.34	42.40
Ciudad de México	51.53	56.22	81.31	42.24	57.48	64.99	70.55	79.08	64.70	51.01
Durango	7.34	55.33	61.69	46.30	34.99	36.67	55.18	28.56	2.61	23.26
Guanajuato	70.64	51.18	47.26	52.17	51.75	42.58	60.24	29.33	9.51	37.96
Guerrero	45.56	25.30	30.04	43.87	17.09	20.58	52.35	23.86	10.92	16.13
Hidalgo	77.49	47.46	49.93	58.48	46.07	42.36	54.61	24.34	1.35	33.44
Jalisco	64.60	62.98	58.19	47.61	63.15	47.14	61.11	47.32	30.01	36.23
México	46.89	50.45	51.19	41.47	51.39	45.14	61.78	40.28	8.47	35.16
Michoacán	60.05	44.09	42.23	45.14	35.71	37.95	58.33	26.80	3.51	22.62
Morelos	39.10	54.86	54.38	41.04	45.80	38.40	50.33	41.92	8.37	62.99
Nayarit	81.47	63.95	55.22	55.56	40.64	61.49	55.82	34.32	13.17	21.14
Nuevo León	71.86	70.88	64.77	36.88	64.02	59.52	59.29	49.95	38.24	50.00
Oaxaca	69.50	36.70	32.57	39.31	47.13	16.93	44.19	16.93	8.57	28.22
Puebla	72.99	45.03	47.90	48.44	62.19	34.55	60.62	37.93	9.08	48.85
Querétaro	69.02	54.97	58.98	56.93	63.35	50.22	68.50	37.42	11.47	57.46
Quintana Roo	67.82	46.77	60.03	59.50	38.07	70.49	51.22	62.57	61.94	13.05
San Luis Potosí	70.91	52.04	49.69	54.65	55.52	39.73	58.04	30.70	11.85	39.40
Sinaloa	64.22	58.23	64.97	56.82	57.32	54.60	65.36	38.30	5.91	25.92
Sonora	70.90	59.88	60.92	47.43	57.96	57.41	63.60	50.62	19.22	54.09
Tabasco	58.42	38.67	44.99	48.91	47.03	47.36	50.61	32.51	3.86	27.09
Tamaulipas	53.46	63.57	57.48	57.49	51.13	48.95	53.05	39.94	8.94	38.46
Tlaxcala	73.83	56.30	47.41	46.22	56.86	33.87	55.48	32.55	4.87	38.57
Veracruz	70.16	44.41	43.40	51.10	45.87	40.54	48.63	23.31	5.93	31.13
Yucatán	78.78	39.94	48.38	58.85	57.12	41.48	65.32	37.15	3.65	38.04
Zacatecas	65.35	57.25	53.76	52.74	49.93	32.67	37.11	27.05	14.70	25.38

$$C(a,b) = \frac{\sum_{j=1}^{n} w_j C_j(a,b)}{\sum_{j=1}^{n} w_j} \tag{4}$$

In case of disagreement because the decision maker expresses veto values, some criterion calculates the index of discordance index $D_j(a,b)$. The discrepancy of criterion g_j indicates the extent to which this criterion disagrees with the statement "a outranks b". The discordance index D_j reaches its maximum value when the criterion g_j adds its veto to the relation of outranking. The index obtains its minimum value when the criterion g_j is not discordant with the relationship. The formulation of this idea is as follows

$$D_j(a,b) = \min\left\{1, \max\left\{0, \frac{\left[g_j(b) - g_j(a)\right] - p_j(g_j(a))}{v_j(g_j(a)) - p_j(g_j(a))}\right\}\right\}, \tag{5}$$

The relationship about fuzzy classification is defined for each pair of alternatives (a, b) as credibility index. The relationship $\sigma(a,b)$ fully expresses to what extent "a outranks b" considering the comprehensive concordance index and the discordance index for each criterion g_j. The credibility index is simply the exhaustive concordance index weakened by the discordance index. Credibility is reduced in the presence of one or more discordant criteria, when $D_j(a,b) > C(a,b)$. In accordance with the veto effect, $\sigma(a,b) = 0$ if $\exists j | D_j(a,b) = 1$ regardless of the relative importance w_j of the criterion. With (5) and (6) calculate (7).

$$\sigma(a,b) = \begin{cases} C(a,b) & \text{if } \overline{F}(a,b) = 0 \\ C(a,b) \times \prod_{J \in \overline{F}(a,b)} \frac{1-D_j(a,b)}{1-C(a,b)} & \text{if } \overline{F}(a,b) \neq 0 \end{cases} \tag{6}$$

where $\overline{F}(a,b) = \{j \in F / D_j(a,b) > C(a,b)\}$.

The value of $\sigma(a,b)$ is an interval [0,1], and it is constructed in such a way that it complies with certain qualitative principles. Mainly, the index excludes the possibility that a large loss in a criterion can be compensated by a small number of gains from the remaining criteria. In Table 7, the parameters defined by the decision maker are presented.

Table 7 Parameters: weights, indifference thresholds and preference for each criterion

	$C1$	$C2$	$C3$	$C4$	$C5$	$C6$	$C7$	$C8$	$C9$	$C10$
Dir	Max	Max	Max	Max	Max	Max	Max	Max	Max	Max
w_j	0.15	0.1	0.05	0.06	0.1	0.08	0.2	0.08	0.03	0.15
q_j	5	5	5	5	5	5	5	5	5	5
p_j	10	10	5	10	10	10	15	5	5	15
v_j	–	–	–	–	–	–	–	–	–	–

Note The *vj* threshold is not defined for the competitiveness problem

Table 8 Categorization of entities in relation to their level of competitiveness

Position	Ranking	Level of competitiveness
1	$R1$	High
2	$R2$	Considerable
3	$R3$	Medium
4	$R4$	Low
5	$R5$	Very low

The parameters defined in Table 7 allow the decision maker to express their preferences in relation to the performance of the alternatives in each of the criteria. The weight w_j corresponds to the relative importance of the criterion g_j. The indifference (q_j), preference (p_j) and veto (v_j) thresholds defined by the decision maker in Table 7 allow us to identify preference situations between alternatives for the same criteria. For example, the criterion labeled $C1$ has values 5 and 10 of indifference and preference, respectively. These thresholds indicate that an indifference is established between states in criterion $C1$ when the value of the entity is less than or equal to 5 in relation to another entity, and a strict preference when the value of the entity is greater than 10 in the criterion with regarding another state. The veto threshold is a condition used the decision maker to say that it is not possible to support the idea that the worst of the two actions under consideration on a certain criterion may be comprehensively considered as good as the better one, even if its performances on all the other criteria are better. For the competitiveness problem, the veto threshold was not defined. Thus, in the decision model the veto condition is not used.

To construct the decision model, Eq. (6) uses the data of Table 6 (it is the input data generated by composed index) and the parameters of Table 7 (DM's preference information). The decision model contains the evaluation of regions based on the DMs' preferences. By the use of the methodology explained before, a categorization of 5 levels have been done to measure the level of competitiveness of the Mexican States and with that the ranking of the most suitable states to invest is obtained (see Tables 8 and 9).

4 Discussion of the Results

The order generated by the current proposal is compared to that generated by the IMCO (2016), which corresponds to the ranking of the entities considering the highest competitiveness index for the first positions and the worst index for the lower positions (see Table 10).

The current proposal presents an arrangement with some variations in relation to the IMCO. These deviations are presented as investments between systems. That is, one element is better positioned than another element in an order, but it is worse positioned than the other in another order. Investments indicate these discrepancies

Table 9 Level of competitiveness generated with composite indices and ELECTRE III

Ranking	State
$R1$	Aguascalientes
	Ciudad de México
$R2$	Sinaloa
	Nuevo Leon
	Colima
	Coahuila
	Querétaro
	Puebla
	Sonora
	Jalisco
$R3$	Hidalgo
	San Luis Potosí
	Guanajuato
	Quintana Roo
	Morelos
	Nayarit
	México
	Tamaulipas
	Yucatán
	Campeche
	Tlaxcala
	Baja California
	Chihuahua
$R4$	Tabasco
	Zacatecas
	Veracruz
	Baja California Sur
	Durango
	Michoacán
	Chiapas
$R5$	Guerrero
	Oaxaca

between the methods used to generate systems by the way they exploit the information and generate the ranking proposal.

With the information provided by Table 10, it is possible to see some changes in the ranking of the states. The analysis will focus on the states that increase or downgrade their ranking not specifically in the position that they have. In this sense,

Table 10 Ranking comparison between ELECTRE III and IMCO

Ranking	ELECTRE III	IMCO
*R*1	Aguascalientes	Ciudad de México
	Ciudad de México	Aguascalientes
*R*2	Sinaloa	Nuevo León
	Nuevo León	Colima
	Colima	Querétaro
	Coahuila	Sonora
	Querétaro	Coahuila
	Puebla	Jalisco
	Sonora	Sinaloa
	Jalisco	Yucatán
*R*3	Hidalgo	Campeche
	San Luis Potosí	Baja California Sur
	Guanajuato	Quintana Roo
	Quintana Roo	Puebla
	Morelos	Tamaulipas
	Nayarit	Chihuahua
	México	Nayarit
	Tamaulipas	Guanajuato
	Yucatán	San Luis Potosí
	Campeche	Hidalgo
	Tlaxcala	México
	Baja California	Durango
	Chihuahua	Morelos
*R*4	Tabasco	Tlaxcala
	Zacatecas	Baja California
	Veracruz	Zacatecas
	Baja California Sur	Tabasco
	Durango	Veracruz
	Michoacán	Michoacán
*R*5	Chiapas	Chiapas
	Guerrero	Oaxaca
	Oaxaca	Guerrero

it is possible to see that in *R*1 and *R*2 there is not a change at all; this means that the best and the worst states to invest are the same between both rankings. But in the case of *R*2, *R*3 and *R*4, it is possible to see important changes, for example, in Puebla that in IMCO it is found in *R*3 and in ELECTRE III method it goes up to *R*2. Another example is Yucatan that downgrades from *R*2 to *R*3, and Baja California

Sur that downgrades from $R3$ to $R4$ and is the same case for Durango that goes from $R3$ to $R4$. These variations are important to consider because it is possible to identify how much the ranking can change if we use different parameters with the same information. In this sense, the rankings are not absolute, but they can change to the preference, importance, weight and many other elements related to different quantitative parameters. That is why, it is important to use methodologies that can adapt to the reality of the decision maker, country or company.

5 Conclusions

The analysis of the competitiveness of the state of a country is not a new topic. This type of information is used to a different public policy making and by the investor in order to know where to put their money. The objective of the paper is to present a procedure for ordering the federative entities of the Mexican Republic, considering its level of competitiveness, using multicriteria analysis techniques to aid decision making, particularly ELECTRE III method.

This method provides support for decision making for complex real-world problems. With the decision process described, it was possible to observe the changes that the ranking of competitiveness can have by including more information and parameters obtained by the decision maker, in this case, investors. The comparative analysis between the IMCO ranking and ELECTRE III is presented. The reason of this differences is because of the methodology used, in the case of IMCO they use a compensatory method, in contrast to the multicriteria analysis method that is proposed.

For future research, we expect to analyze the levels of competitiveness considering criteria hierarchies to analyze the federative entities interactively including criteria at each level of the hierarchy and expand the ELECTRE III method by the incorporation of other aggregation operator (AGOP) techniques (Blanco-Mesa et al. 2019).

Annex 1: Formulations for Each Criterion

C1 Reliable and Objective Law System

$$\begin{aligned} C1 = &(w_1 * \mathrm{Ho}(a_i)) + (w_2 * \mathrm{Se}(a_i)) + (w_3 * \mathrm{Rv}(a_i)) \\ &+ (w_4 * \mathrm{Cd}(a_i)) + (w_5 * \mathrm{Id}(a_i)) + (w_6 * \mathrm{Dd}(a_i)) \\ &+ (w_7 * \mathrm{Ps}(a_i)) + (w_8 * \mathrm{Cs}(a_i)) + (w_9 * \mathrm{Cc}(a_i)) \end{aligned} \tag{7}$$

where

Ho Number of homicides.
Se Number of kidnappings.

Rv Vehicle theft number.
Cd Number of pesos in crime costs.
Id Crime incidence number.
Dd Number of crimes reported.
Ps Percentage of the population with security perception.
Cs Number of notaries.
Cc Average number of contract compliance.

C2 Sustainable Management of the Environment

$$\begin{aligned} C2 = {} & (w_1 * \mathrm{Ea}(a_i)) + (w_2 * \mathrm{Ar}(a_i)) + (w_3 * \mathrm{Ua}(a_i)) \\ & + (w_4 * \mathrm{Mi}(a_i)) + (w_5 * \mathrm{Cf}(a_i)) + (w_6 * \mathrm{Pa}(a_i)) \\ & + (w_7 * \mathrm{Ap}(a_i)) + (w_8 * \mathrm{Rg}(a_i)) + (w_9 * \mathrm{Dr}(a_i)) \\ & + (w_{10} * \mathrm{Ie}(a_i)) + (w_{11} * \mathrm{Ec}(a_i)) + (w_{12} * \mathrm{Gf}(a_i)) \end{aligned} \tag{8}$$

where

Ea Percentage of the aquifer exploitation index.
Ar Number of liters of treated wastewater.
Ua Thousands of pesos per cubic hectometer.
Mi Number of infant deaths.
Cf Forest competitiveness index.
Pa Percentage of total area covered by trees.
Ap Number of protected natural areas.
Rg Number of kilograms per person.
Dr Percentage of households that dispose of their solid waste.
Ie Megawatts hour per million of GDP.
Ec Number of clean industry certificates issued.
Gf Number of pesos per inhabitant.

C3 Inclusive, Prepared and Healthy Society

$$\begin{aligned} C3 = {} & (w_1 * \mathrm{Po}(a_i)) + (w_2 * \mathrm{Ac}(a_i)) + (w_3 * \mathrm{Ai}(a_i)) \\ & + (w_4 * \mathrm{Ma}(a_i)) + (w_5 * \mathrm{Es}(a_i)) + (w_6 * \mathrm{Il}(a_i)) \\ & + (w_7 * \mathrm{Af}(a_i)) + (w_8 * \mathrm{Ep}(a_i)) + (w_9 * \mathrm{Ec}(a_i)) \\ & + (w_{10} * \mathrm{Ra}(a_i)) + (w_{11} * \mathrm{Ev}(a_i)) + (w_{12} * \mathrm{Mi}(a_i)) \\ & + (w_{13} * \mathrm{Md}(a_i)) + (w_{14} * \mathrm{Su}(a_i)) + (w_{15} * \mathrm{Ea}(a_i)) \\ & + (w_{16} * \mathrm{Ch}(a_i)) + (w_{17} * \mathrm{Me}(a_i)) + (w_{18} * \mathrm{Mp}(a_i)) \\ & + (w_{19} * \mathrm{Mn}(a_i)) \end{aligned} \tag{9}$$

where

Po Percentage of poverty index (%).
Ac Sewer access of the population (%).
Ai Access to health institutions (%).
Ma Economically active women.
Es Salary equity.
Il Labor informality index between women and men.
Af Percentage of population that can read (%).
Ep Average schooling.
Ec Examination quality schools.
Ra Academic performance.
Ev Life expectancy in years.
Mi Infant mortality of children under 1 year.
Md Mortality from diabetes and hypertensive diseases.
Su Suicides.
Ea Teen pregnancies.
Ch Hospital beds.
Me Doctors and nurses.
Mp Doctors with specialty.

C4 Stable and Functional Political System

$$\begin{aligned} C4 = &(w_1 * \mathrm{Pe}(a_i)) + (w_2 * \mathrm{Pp}(a_i)) + (w_3 * \mathrm{Pi}(a_i)) \\ &+ (w_4 * \mathrm{Pc}(a_i)) + (w_5 * \mathrm{Ce}(a_i)) + (w_6 * \mathrm{Bc}(a_i)) \\ &+ (w_7 * \mathrm{Ec}(a_i)) + (w_8 * \mathrm{Mi}(a_i)) + (w_9 * \mathrm{Ap}(a_i)) \end{aligned} \tag{10}$$

where

Pe Percentage of urban population (%).
Pp Perception of corruption in political parties (%).
Pi Perception about the availability of public information.
Pc Citizen participation.
Ce Electoral competition.
Bc Barriers to independent candidates.
Ec Equity in Congress in percent.
Mi Number of equal marriages.
Ap Number of attacks on the press.

C5 Efficient and Effective Governments

$$C5 = (w_1 * \mathrm{Ie}(a_i)) + (w_2 * \mathrm{Ip}(a_i)) + (w_3 * \mathrm{In}(a_i))$$

$$+ (w_4 * \mathrm{Rp}(a_i)) + (w_5 * \mathrm{Ae}(a_i)) + (w_6 * \mathrm{Em}(a_i)) + (w_7 * \mathrm{Hp}(a_i)) + (w_8 * \mathrm{Il}(a_i)) + (w_9 * \mathrm{Dm}(a_i)) \quad (11)$$

where

Ie Interaction with the government by electronic means.
Ip State budget information index.
In Own income.
Rp Registration of a property.
Ae Opening a company.
Em Regulatory improvement efforts.
Hp Households that are filled with water by pipe.
Il Labor informality.
Dm Distribution of women in the state administration.

C6 Market Factors

$$C6 = (w_1 * \mathrm{It}(a_i)) + (w_2 * \mathrm{Ic}(a_i)) + (w_3 * \mathrm{Ds}(a_i)) + (w_4 * \mathrm{Po}(a_i)) + (w_5 * \mathrm{Et}(a_i)) + (w_6 * \mathrm{Ep}(a_i)) + (w_7 * \mathrm{As}(a_i)) + (w_8 * \mathrm{Ef}(a_i)) + (w_9 * \mathrm{Cl}(a_i)) \quad (12)$$

It Percentage of job growth (%).
Ic Monthly income full time.
Ds Income ratio.
Po Percentage of employed population.
Et Terminal efficiency at secondary level.
Ep Terminal efficiency at the preparatory level.
As Higher education.
Ef Percentage of foreign education (%).
Cl Job training.

C7 Stable Economy

$$C7 = (w_1 * \mathrm{Ip}(a_i)) + (w_2 * \mathrm{Ie}(a_i)) + (w_3 * \mathrm{Cr}(a_i)) + (w_4 * \mathrm{De}(a_i)) + (w_5 * \mathrm{Df}(a_i)) + (w_6 * \mathrm{Pd}(a_i)) + (w_7 * \mathrm{Cp}(a_i)) + (w_8 * \mathrm{Im}(a_i)) + (w_9 * \mathrm{Pl}(a_i)) + (w_{10} * \mathrm{Dp}(a_i)) + (w_{11} * \mathrm{Ds}(a_i)) \quad (13)$$

where

Ip Income per capita per person.

Ie State per capita income.
Cr GDP growth.
De Percentage of state GDP.
Df Federal participation debt.
Pd Average term of debts.
Cp Percentage of the average cost of debt (%).
Im Percentage of people busy PEA (%).
Pl Economically active population of the total population.
Dp Economic dependence.
Ds Economic diversification.

C8 Precursors

$$\begin{aligned} C8 = &(w_1 * \mathrm{Sm}(a_i)) + (w_2 * \mathrm{Hi}(a_i)) + (w_3 * \mathrm{Tp}(a_i)) \\ &+ (w_4 * \mathrm{Ct}(a_i)) + (w_5 * \mathrm{Sc}(a_i)) + (w_6 * \mathrm{Ha}(a_i)) \\ &+ (w_7 * \mathrm{Ac}(a_i)) + (w_8 * \mathrm{Fp}(a_i)) + (w_9 * \mathrm{Ca}(a_i)) \end{aligned} \tag{14}$$

where

Sm Subscribers with mobile.
Hi Percent of homes with Internet.
Tp Points of sale for every 10 thousand.
Ct Number of ATMs.
Sc Road system.
Ha Traffic accidents.
Ac Accidents due to poor road conditions.
Fp Number of passengers per year.
Ca Weight of air cargo per person.

C9 Use of International Relations

$$\begin{aligned} C9 = &(w_1 * \mathrm{Tm}(a_i)) + (w_2 * \mathrm{Pa}(a_i)) + (w_3 * \mathrm{Pt}(a_i)) \\ &+ (w_4 * \mathrm{Ie}(a_i)) + (w_5 * \mathrm{Xp}(a_i)) \end{aligned} \tag{15}$$

where

Tm Internationally certified hospitals.
Pa Passenger to international flights.
Pt Percents of the tourism sector of GDP.
Ie Foreign investment in thousands of GDP.
Xp Exports of the country.

C10 Innovation of the Economic Sectors

$$C10 = (w_1 * \mathrm{Si}(a_i)) + (w_2 * \mathrm{Fi}(a_i)) + (w_3 * \mathrm{In}(a_i)) + (w_4 * \mathrm{Pa}(a_i)) + (w_5 * \mathrm{Et}(a_i)) + (w_6 * \mathrm{Ec}(a_i)) \quad (16)$$

where

Si Innovation percent index.
Fi Productivity percent index.
In Researchers.
Pa Number of patents.
Et Research center for every 100 thousand.
Ec Companies certified from ISO 9001 to 14001.

References

Alireza, A., Majid, M., & Rosnah, M. (2010). Simple additive weighting approach to personnel selection problem. *International Journal of Innovation, Management and Technology, 1*(5), 511–515.

Bardach, E. (2001). *Los ocho pasos para el análisis de políticas públicas. Un manual para la práctica*. México: Miguel Angel Parrúa.

Bartolini, F., Gallerani, V., Samoggia, A., & Viagg, D. (2005). Methodology for multicriteria analysis of agri-environmental schemes. *ITAES project, document number: WP10-P6-D11*

Blanco-Mesa, F., León-Castro, E., & Merigó, J. M. (2019). A bibliometric analysis of aggregation operators. *Applied Soft Computing, 81*(4), 105488.

Charles, V., & Zegarra, L. (2014). Measuring regional competitiveness through data envelopment analysis: A peruvian case. *Expert Systems with Applications, 41*(11), 5371–5381.

Crescenzi, R. (2009). Undermining the principle of concentration? European Union regional policy and the socio-economic disadvantage of European regions. *Regional Studies, 43*(1), 111–133.

Huang, J. H., & Peng, K. H. (2012). Fuzzy Rasch model in TOPSIS: A new approach for generating fuzzy numbers to assess the competitiveness of the tourism industries in Asian countries. *Tourism Management, 33*(2), 456–465.

IMCO. (2016). *Índice de Competitividad Estatal 2014*. México: Impresos Villaflorito S.A. de C.V.

Labarca, N. (2007). *Consideraciones Teóricas De La Competitividad. Omnia, 13*(2), 158–184.

Leyva, J., Alvarez, P., Gatelum, D., & Solano, J. (2016). A web based group decision support system for multicriteria ranking problems. *Operational Research, 17*(2), 499–539.

Lombana, J., & Gutiérrez, R. (2009). Marco analitico de la competitividad: fundamentos para el estudio de la competitividad regional. *Revista Cientifica Pensamiento y Gestión, 26*, 1–38

Oyarce, J. (2013). Excelencia empresarial y competitividad: ¿una relación fructifera? *Panorama Socioeconómico, 31*(46), 58–63.

Porter, M. (1990). The competitive advantage of nations. *Harvard Business Review*, 70–91

Roy, B. (1978). Un algorithme de classements fond sur une reprsentation floue des prfrences en prsence de critres multiples. *Cahiers du CERO (Centre dEtudes de Recherche Oprationnelle), 20*, 3–24.

Wang, T. C., & Lee, H. D. (2009). Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. *Expert Systems with Applications, 36*(5), 8980–8985.

WEF World Economic Forum (2016). Obtenido de World Economic Forum. https://es.weforum.org/agenda/2016/10/

Zhang, H., Gu, C. L., Gu, L. W., & Zhang, Y. (2011). The evaluation of tourism destination competitiveness by TOPSIS and information entropy—A case in the Yangtze River Delta of China. *Tourism Management, 32*(2), 443–451.

Degree of Global Covering and Global Overlapping in Solvency Fuzzy Classification

Fabián Castiblanco, Camilo Franco, Javier Montero, and J. Tinguaro Rodriguez

Abstract In this paper, the notions of degree of global covering and degree of global redundancy on a fuzzy classification system are established and applied. Such notions are implemented in the evaluation of a fuzzy partition. In particular, the characterization of the financial solvency of an economic sector and the possible estimation of financial insolvency are considered as problems of fuzzy classification. The results of implementing fuzzy classification systems and global indices are illustrated on this application.

1 Introduction

The detection of company financial difficulties is a subject which has been particularly study from financial ratio analysis. From the seminal papers of Altman (1968), Beaver (1966), and Sinkey (1975), the financial ratios together with the multivariate statistical techniques have presented wide relevance in the study of financial solvency or bankruptcy prediction. At present, due to computational advances, techniques based on artificial intelligence and machine learning have been explored (Barboza et al. 2017; Ravi and Ravi 2007; Chen 2011). Therefore, one of the most predominant ideas is to formulate the detection of financial insolvency or bankruptcy prediction as

F. Castiblanco (✉)
Gran Colombia University, Bogotá, Colombia
e-mail: fabianalberto.castiblanco@ugc.edu.co

C. Franco
Andes University, Bogotá, Colombia
e-mail: c.franco31@uniandes.edu.co

J. Montero · J. T. Rodriguez
Complutense University of Madrid, Madrid, Spain
e-mail: monty@mat.ucm.es

J. T. Rodriguez
e-mail: jtrodrig@mat.ucm.es

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_2

a classification problem. In particular, several authors have considered an unsupervised classification problem using fuzzy theory, i.e., algorithms that establish a fuzzy partition of the analyzed data set (Andrés et al. 2011; Alam et al. 2000; Boyacioglu et al. 2009).

The concept of a partition is, without any doubt, one of the pillars of classification problems. In the fuzzy case, a standard notion for a partition has been the one proposed by Ruspini (1969), where the total degree of membership is completely distributed among all classes, which was later generalized by Krishnapuram and Keller (1993), only requiring that the membership of an object to any class is positive. Therefore, under the notion of a Ruspini's partition, the total sum of the membership degrees of an object to all classes must be equal to 1, while in the perspective of Krishnapuram and Keller, it is sufficient that for each element, its membership degree is greater than zero.

Whether a partition meets Ruspini's condition or not, one of the biggest challenges in unsupervised classification problems is to establish the most appropriate cluster number, i.e., to evaluate the accuracy of the classification obtained. About this problem, several cluster validity indices have been proposed in the literature: Bezdek's partition coefficient (PC) (Bezdek 1974), partition entropy (PE) (Bezdek 1981), modified partition coefficient (Dave 1996), average silhouette width criterion (Kaufman and Rousseeuw 1990), and Fuzzy Silhouette (Campello and Hruschka 2006) have been some of the most used.

However, the variety of indexes complicates the problem of evaluating the quality of a fuzzy partition. Given a fuzzy partition and a set of indexes to evaluate such partition, rarely all indexes converges to select the best partition with the same number of classes, i.e., for some index the best partition can be with three clusters while for another index the best partition can be with four clusters. Only under a learning process and considering a calibration process, the decision maker can establish the most appropriate partition for the set of data studied.

Based on these ideas and considering that in most cases, fuzzy partitions do not verify the Ruspini condition, and fuzzy classification systems have been proposed by means of families of aggregation functions, analyzing the obtained classification (Amo et al. 2004). Therefore, a classification problem can be evaluated through a fuzzy classification system regardless of the notion of partition used.

The key idea of fuzzy classification systems is to obtain an aggregated value of the membership degree of each item to the classes of the fuzzy partition, i.e., for each element of the data set studied, the degree of satisfaction of a specific property is obtained. The property evaluated depends on the kind of operator employed. Thus, under the fuzzy classification systems, it is proposed to evaluate the coverage property of a fuzzy partition with disjunctive aggregation operators and redundancy with conjunctive aggregation operators.

According to the above, it is necessary to establish in a next phase, a global index of the property evaluated on each item or element. Such a need arises because under a calibration process, it is relevant to compare the properties studied on each item with the properties evaluated by the existing indexes in the literature. Similarly, because

with a set of global index established jointly, a broader vision and evaluation of the partition is obtained, in contrast to indices generated independently.

Therefore, in this paper, the definition of global indices given in (Castiblanco et al. 2017) is used to evaluate fuzzy partitions with different class numbers established on a set of financial ratios of an economic sector in Colombia. Global indexes are established on each partition and compared with indexes frequently used in the literature. Partition coefficient (PC), partition entropy (PE), modified partition coefficient (MPC), average silhouette width criterion (SIL), and fuzzy silhouette (SIL.F) are used.

The best partition will provide a characterization of financial solvency in the specific economic sector. It will also provide a first classification on possible companies in a state of financial insolvency. In order to do so, in the first section, some necessary definitions and concepts we introduce, in particular, the notion of global degree is presented. In Sect. 3, the proposed methodology and the analyzed data set are established. In the last section, we present the results, final comments and future works.

2 Preliminaries

In this section, some definitions and necessary results for the development of this paper are presented.

Definition 1 Amo et al. (2004) A recursive rule ϕ is a family of aggregation functions $\left\{\boldsymbol{\phi_n}: [\mathbf{0}, \mathbf{1}]^{\boldsymbol{n}} \to [\mathbf{0}, \mathbf{1}]\right\}_{\boldsymbol{n}>1}$ such that there exist an ordering rule π and two sequences of binary operators.

$$\begin{aligned} &\left\{L_n: [0, 1]^2 \to [0, 1]\right\}_{n>1} \text{ and} \\ &\left\{R_n: [0, 1]^2 \to [0, 1]\right\}_{n>1} \text{such that} \\ \boldsymbol{\phi_n}\left(a_{\pi(1)}, \ldots, a_{\pi(n)}\right) &= L_n\left(\boldsymbol{\phi_{n-1}}\left(a_{\pi(1)}, \ldots, a_{\pi(n-1)}\right), a_{\pi(n)}\right) \\ &= R_n\left(a_{\pi(1)}, \boldsymbol{\phi_{n-1}}\left(a_{\pi(2)}, \ldots, a_{\pi(n)}\right)\right) \end{aligned}$$

Recursiveness is a property of a sequence of operators $\left\{\boldsymbol{\phi_n}\right\}_{\boldsymbol{n}>2}$ allowing the aggregation of any number of items: $\boldsymbol{\phi}_2$ tells us how to aggregate two items, $\boldsymbol{\phi}_3$ tells how to aggregate three items and so on, in such a way that being faced with n items we shall aggregate them by means of $\boldsymbol{\phi_n}$. A standard recursive rule will be one based upon the identity ordering rule (i.e., the ordering rule that keeps the order of data as they are given to us). Obviously, as noted in (Amo et al. 2004), in case such a sequence of binary operators is given by a unique binary operator ($L_i = R_j$, for all i, j), we should be talking about a rule based upon an associative binary operator. Hence, recursiveness indeed allows the generalization of associativity (from this perspective, an associative rule is a recursive rule based upon a unique binary operator) [see again Amo et al. (2004)].

Recursive rules constitute an elegant formal mechanism to deal with aggregations of any arbitrary number of elements. Moreover, the recursive approach allows this mechanism to be robust under changes in the cardinality of the data, as it guarantees that all the operators in a family of aggregation functions are tightly related to a unique aggregation procedure: the binary operators that build up the recursive rule [see also Rojas et al. (2013)]. This is useful in the context of unsupervised classification, where the number of classes in which the data has to be segmented is typically unknown a priori. In this context, recursive rules provide a robust mechanism to deal with the aggregation of diverse class information for different number of classes (Amo et al. 2004), automatically assessing the classification performance. We recall next the definition of fuzzy classification systems.

Definition 2 Amo et al. (2004) Let us assume a finite set of objects X. A fuzzy classification system is a finite family C of n fuzzy classes, where each $c \in C$ has an associated membership function $\mu_c(x)$: $X \to [0, 1]$, together with a recursive triplet (φ, ϕ, N) such that

1. ϕ is a standard recursive rule such that $\phi_2(0, 1) = \phi_2(1, 0) = 0$
2. N: $[0, 1] \to [0, 1]$ is a strong negation function,[1] i.e., a continuos strictly decreasing function such that $N \circ N(x) = x$ for all $\mu(x) \in [0, 1]$
3. φ is a standard recursive rule such that

$$\varphi_n(x_1, \ldots, x_n) = N^{-1}[\phi_n(N(x_1), \ldots, N(x_n))] \forall n > 1$$

According to Definition 2, a fuzzy classification system can be denoted by (C, ϕ, φ, N). Notice that in a fuzzy classification system, each $x \in X$ has a membership degree $\mu_c(x)$ associated with each class $c \in C$. As our purpose is to analyze the fuzzy classes, from now on, we consider the aggregation of each standard recursive rule on such membership degrees, that is, $\phi\{\mu_c(x) | c \in C\}$ or $\varphi\{\mu_c(x) | c \in C\}$.

Thus, for instance, we have that

$$\varphi_n(\mu_1(x), \ldots, \mu_n(x)) = N^{-1}[\phi_n(N(\mu_1(x)), \ldots, N(\mu_n(x)))]$$

where $\mu_i(x)$ denotes the membership degree of the element x to the i-th class of C, $i = 1, ..., n$.

In the same line, notice that, φ_n is a disjunctive recursive rule, in the sense that $\varphi_n(\mu(x_1), \ldots, \mu(x_n)) = 1$ whenever there is j such that $\mu(x_j) = 1$, while ϕ_n is a conjunctive recursive rule in the sense that $\phi_n(\mu(x_1), \ldots, \mu(x_n)) = 0$ whenever there is j such that $\mu(x_j) = 0$.

Fuzzy classification systems, as proposed in (Amo et al. 2004) [see also Amo et al. (2001)], were conceived as a structure allowing the treatment of complex classification problems following two key ideas. On the one hand, under the theoretical framework proposed by Dombi (1982a, b) about aggregation operators, fuzzy

[1] Note that in (Amo et al. 2004), this definition considers a strict negation; however, the condition that such negation is bijective is imposed, and therefore, in general, it is considered a strong negation.

classification systems were proposed as an alternative approach for non-associative connectives through the use of the concept of recursiveness. On the other hand, fuzzy classification systems were proposed as a structure allowing the evaluation of the established classification from De Morgan triples, i.e., by using recursive rules (satisfying the De Morgan laws) to evaluate three key characteristics of the family of classes that are obtained in a fuzzy partition: redundancy, coverage, and relevance. Thus, fuzzy classification systems obtained a set of indices for the evaluation of partition, even better, for the knowledge of the partition.

Redundancy refers to a certain orthogonality of the family of classes, which is then viewed as a particular representation system of the set of objects (each family should not fully overlap with any family of classes). Coverage refers to how different aspects of reality are fully verified by a family of classes. Finally, the third characteristic of relevance is understood as the necessity of including, or not excluding, a class or family of classes from the classification. In this way, decision makers can have a hint on how to improve the classifier performance, e.g., searching for some missing classes, proposing greater or smaller classes, or deleting some classes.

Then, given a fuzzy partition, its coverage is analyzed by means of the disjunctive rule $\varphi\{\mu_c(x)|c \in C\}$, and its redundancy is analyzed by means of the conjunctive rule $\phi\{\mu_c(x)|c \in C\}$. Under the structure of a fuzzy classification system, it is desirable that the partition has high degrees of coverage (disjunctive operator) and low degrees of redundancy (conjunctive operator).

The above allows the aggregation of useful information for each item x on a particular property or characteristic. In this sense, it is desirable to analyze the aggregate information of such characteristic for all items x. Therefore, we propose the following definition, using an aggregation function ρ: [0, 1]→[0, 1], for all membership degrees of $x \in X$ to $c_1, \ldots, c_n \in C$, represent the degree to which the family of fuzzy classes satisfy a specific characteristic for an object $x \in X$;

Definition 3 Given a finite set of objects X, a family of fuzzy classes C over this set and an aggregation function ρ: $[0, 1]^n \rightarrow [0, 1]$. Then, the global degree of ρ denoted ρ^T on X is defined as the aggregation of the degrees of ρ for all items $x \in X$.

Such aggregation function can be of very different nature (e.g., conjunctive, disjunctive or averaging). According to the above, we use the following definitions to establish the notion of degree of global covering;

Definition 4 Castiblanco et al. (2017) Given a finite set of objects X and a family of fuzzy classes C over this universe, the degree of global covering of X is defined $\forall x \in X$, as the aggregation of the degrees of covering.

If we analyze the covering of each item x through the standard recursive rule φ, then we denote the degree of global covering as φ^T. Similarly, the notion of degree of global redundancy is presented through the following definition;

Definition 5 Given a finite set of objects X and a family of fuzzy classes C over this universe, the degree of global redundancy of X is defined $\forall x \in X$, as the aggregation of the degrees of redundancy.

If we analyze the redundancy of each item x through the standard recursive rule ϕ, then we denote the degree of global redundancy as ϕ^T. Remember that the aggregation operator corresponding to the global index can be of any nature: conjunctive, disjunctive, or average.

In this paper, five indices will be used to compare the results obtained with the degree of global covering and the degree of global redundancy. The criteria are: partition coefficient (PC), measures the amount of overlap between clusters. The closer this value is to the unit, the better. Entropy coefficient (EC). The closer this value to zero, the better. Modified partition coefficient (MPC), partition coefficient variation through a linear transformation, modify range of the value of $[1/c, 1]$ to [0, 1], the closer to the unit, the better. Average silhouette width criterion (SIL) allows distinguishing "clear-cut" from "weak" clusters in the same plot: clusters with a larger average silhouette width are more pronounced, the higher the better. fuzzy silhouette (SIL.F), the higher the better.

In this paper, the fuzzy c-means algorithm is used. Such algorithm is a method of clustering which allows one piece of data to belong to two or more clusters.

3 Data and Methodology

In order to apply the defined degree of global covering and the degree of global redundancy and establish its advantages in the characterization of the financial solvency of economic sectors, the following key aspects are established.

3.1 Data

A set of 54 companies from the commercial sector of the Colombian economy are selected. The commercial sector is selected due to its high participation in the generation of employment in Colombia and because it presents a high percentage of companies in insolvency process in recent years. The financial statements for 2017 of such company have been analyzed to compute the selected financial ratios.

The balance sheet and cash flow statement as of December 2017 have been considered. The year 2017 is selected for the wide availability of the information required in the Superintendence of Corporations Web site. The financial ratios corresponding to the static versus dynamic model proposed by Pessoa de Oliveira (2016) have been established;

1. E_1: **Short-term solvency**

$$E_1 := \frac{\text{Current Assets}}{\text{Current Liabilities}};\ \text{Solvency criterion: } E_1 > 1.$$

2. E_2: **Leverage**

$$E_2 := \frac{\text{Total Liabilities}}{\text{Equity}};\ \text{Solvency criterion: } 0 < E_2 \leq 1.$$

3. E_3: **Indebtedness of property, plant, and equipment**

$$E_3 := \frac{\text{Non current Liabilities}}{\text{Non Current Assets}};\ \text{Solvency criterion: } 0 < E_3 \leq 0.5$$

4. D_1: **Cash generated from operations**

$$D_1 := \text{PBT} + \text{amortization:Solvency criterion: } D_1 > 0$$

where PBT = profit before taxation.

5. D_2: **Short-term financial capacity**

$$D_2 \frac{\text{COA}}{D_1};\ \text{Solvency criterion: If } D_1 > 0,\ D_2 > 0.95.\ \text{If } D_1 < 0,\ D_2 > 1.05$$

where COA = Cash from operating activities.

6. D_3: **Management of long-term financial policy**

$$D_3 := \frac{\text{CFA}}{(\text{CIA} + \text{Dividend}) - D_1};\ \text{Solvency criterion: } D_3 \geq 1$$

where CFA = Cash flow from financing activities and CIA = Cash flow from investment activities.

The selection of these ratios corresponds to a perspective that considers; (1) cash flow statement as a fundamental element in determining the availability of financial resources and (2) the company's ability to generate cash in the long term. These are two key aspects in the financial solvency analysis process. Additionally, balance sheet ratios are selected to analyze the part corresponding to both long-term and short-term financial liabilities (Castiblanco et al. 2017).

3.2 Methodology

The first stage of the process is to compute the indexes selected in the literature for each of the established partitions. Table 1 summarizes the results.

Table 1 Indexes

Indexes\cluster	$c = 2$	$c = 3$	$c = 4$	$c = 5$	$c = 6$	Criterion
PC	0.25	0.56	0.52	0.55	0.59	Highest value
PE	0.55	0.75	0.89	0.91	0.87	Lowest value
MPC	0.25	0.34	0.36	0.44	0.51	Highest value
SIL	0.43	0.499	0.498	0.51	0.55	Highest value
SIL.F	0.54	0.635	0.631	0.66	0.67	Highest value

Source own elaboration

The values for which each criterion establishes the best partition have been highlighted. For instance, if we use the partition coefficient PC, the criterion is to find the partition with the highest value. In our case, the best partition would be with six clusters (PC = 0.59); however, we observe that after the partition with three clusters, the value for the partition coefficient decreases and grows again, and finally, it obtains the maximum value in six clusters. A similar behavior is observed with the indices SIL and SIL.F. On the contrary, if the partition entropy index is observed, the best partition is obtained with two classes.

According to the above, we are faced with the following situations; (1) considers the highest value for four indices PC, MPC, SIL, and SIL.F and omits the partition entropy index. In this case, the optimal number of classes is six, or (2) establishing a stop criterion according to the particular needs and interests of the problem addressed in order to satisfy the greatest number of indices and not lose interpretability of the results. For instance, if there is a partition for which the index decreases, the immediately preceding partition is selected. Thus, we could establish that the best partition is the one that generates three clusters, because it is the one that verifies the greatest number of indexes (PC, SIL, SIL.F, and PE. MPC is omitted), or (3) implement the algorithm again increasing the number of classes until obtaining higher values for the four indices. The behavior of the four indices reveals this possibility. In the latter case, the following results have been obtained; PC = 0.9 with 17 clusters, PE = 0.26 with 16 clusters, MPC = 0.9 with 17 clusters, SIL = 0.85 with 18 clusters, and SIL.F = 0.98 with 20 clusters. Clearly the interpretability of the clusters and selecting the best partition is more complicated under such situation.

Now, our purpose is to compute the degrees of global coverage and redundancy to establish possible advantages of our proposal. Keeping this in mind, we compute the six financial ratios for each company, and according to the solvency criteria for each ratio, a matrix $M_{54\times6}$ with dummy variables has been established. The rows correspond to the 54 companies studied, and the columns correspond to each of the solvency criteria (1 for solvent and 0 for insolvent). Such a matrix has been introduced in the fuzzy c-means algorithm. For the initiation of this algorithm, we have tested with the set $c = \{2, 3, 4, 5, 6\}$ where c is the number of clusters to estimate.

The output of the algorithm for each number of clusters is a matrix $U_{54\times c}$. On each row of such a matrix, we can consider a recursive triplet (ϕ, φ, N) such that,

Table 2 Degree of global covering and redundancy

Property\clusters	$c = 2$	$c = 3$	$c = 4$	$c = 5$	$c = 6$
Degree of global covering	0.59	0.60	0.57	0.61	0.65
Degree of global redundancy	0.40	0.04	0.04	0.0002	0.0001

Source Own elaboration

$$1.\phi_n(\mu_1(x), \ldots, \mu_n(x)) = \frac{3\prod_{k=1}^{n}\mu_k(x)}{1+2\prod_{k=1}^{n}\mu_k(x)}$$

$$2.\varphi_n(\mu_1(x), \ldots, \mu_c(x)) = \frac{1-\prod_{k=1}^{n}(1-\mu_k(x))}{1+2\prod_{k=1}^{n}(1-\mu_k(x))}$$

$$3.N(\mu(x)) = 1-\mu(x)$$

Such aggregation operators allow obtaining a degree of redundancy and a degree of coverage of the partition for each company. Disjunctive operator φ measures the coverage, and conjunctive operator ϕ measures the redundancy. For these values, both for covering and for overlapping, we select an aggregation operator to obtain the degree of global covering of X and the degree of global redundancy on X. In particular, we select an averaging operator, the arithmetic mean. Therefore, $\varphi^T = \frac{1}{m}\sum_{i=1}^{m}\varphi_{im}$, $\phi^T = \frac{1}{m}\sum_{i=1}^{m}\phi_{im}$, where φ_{im} and ϕ_{im} are the degrees of covering and redundancy of the m-th item to the partition with n classes. m is the total number of items in the data set studied, in our case, $m = 54$. Table 2, summarizes the results;

As before, the partition with the highest degree of global covering and low degree of redundancy has been chosen without deteriorating interpretability. Therefore, we selected the partition with three clusters due to the following reasoning; (1) the greater interpretability of the partitions is the most relevant. If we increase the cluster numbers to four, the partition does not improve. A better partition is obtained with six clusters, but the changes are not strong, at least in the covering. (2) If there is a partition for which the index decreases, the immediately preceding partition is selected. The degree of redundancy does not change and the degree of coverage decreases. (3) The indexes together allowed establishing a selection criterion without omitting any of them.

4 Results

Figures 1 and 2 present the graph of the clusters for partitions with $\boldsymbol{c} = 3$ and $\boldsymbol{c} = 4$. The data sets are represented by points in the plot, using the first two main components.

From the procedure developed, it can be observed that (1) in the case of the selected economic sector, the best classification according to its solvency is given by three classes, which can be interpreted as solvent companies, insolvent companies,

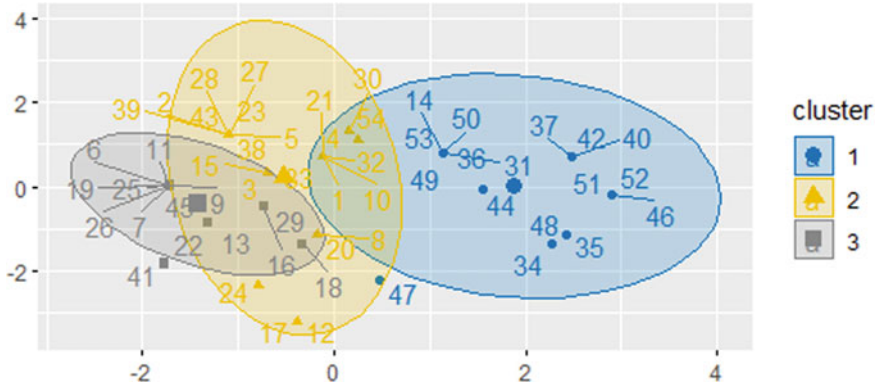


Fig. 1 Fuzzy grouping with three clusters. Cluster 1: insolvent companies. Cluster 2: companies in intermediate status. Cluster 3: solvent companies

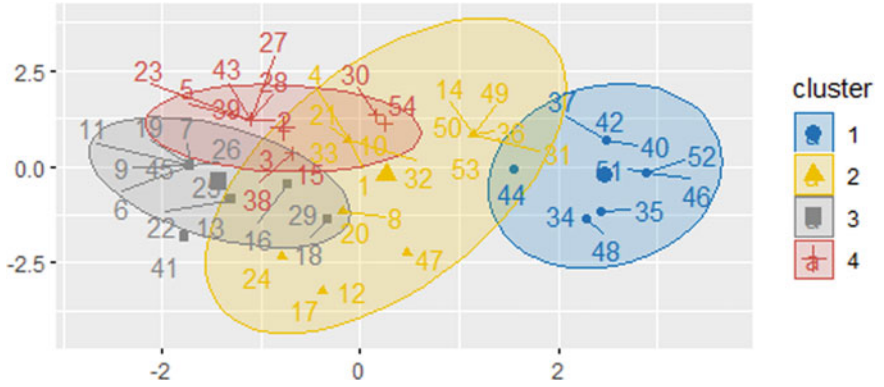


Fig. 2 Fuzzy grouping with four cluster. Cluster 1: highly insolvent. Cluster 2: insolvent. Cluster 3: solvent. Cluster 4: highly solvent

and companies in an intermediate state, (2) seventeen companies were identified as in financial insolvency status according to the selected financial ratios. (3) When the database of the Superintendent of Companies was consulted at the end of 2018, fifteen of these companies started a reorganization process under the Regime Judicial Insolvency of Colombia. At the beginning of the study, none of these companies were under such a process. (4) A classification with only two clusters considers as insolvent companies some that are closer to a solvency state.

5 Final Comments

In this paper, the notions of degree of global covering and degree of global redundancy were applied on a fuzzy classification system. This notion was generalized to any kind of property that is evaluated on a fuzzy partition. The use of fuzzy classification systems together with the global indices have great potential in the evaluation of the accuracy of a fuzzy partition because (1) it integrates two key properties to evaluate in a fuzzy partition. The properties are not observed separately and (2) allow both to evaluate the partition for each element of the set and for the total elements. A property is evaluated in each element and in the partition.

As future work, will be examined the relevance property based on the structure of fuzzy classification systems, understood as the necessity of including, or not excluding, a class or family of classes from the classification, following the ideas raised in (Castiblanco et al. 2018, 2019). Therefore, it is considered a property that allows the comparison process between classes. Likewise, the study of the property of Heterogeneity is proposed; classes can be clearly differentiated (patterns are different enough). Finally, it is proposed to validate the application of fuzzy classification systems in other clustering algorithms and determine some general criteria to select aggregation operators.

References

Alam, P., Booth, D., Lee, K., & Thordarson, T. (2000). The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: An experimental study. *Expert Systems with Applications, 18,* 185–199.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589–609.

Amo, A., Gómez, D., Montero, J., & Biging, G. (2001). Relevance and redundancy in fuzzy classification systems. *Mathware Soft Computing, 8,* 203–216.

Amo, A., Montero, J., Biging, G., & Cutello, V. (2004). Fuzzy classification systems. *European Journal of Operational Research, 156,* 495–507. https://doi.org/10.1016/S0377-2217(03)00002-X

Barboza, F., Kimura, H., & Altman, E. I. (2017). Machine learning models and bankruptcy prediction. *Expert Systems With Applications, 83,* 405--417

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4,* 71–111.

Bezdek, J. C. (1974). Cluster validity with fuzzy sets. *Journal of Cybernet, 3*(3), 58–73.

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms.* New York: Plenum Press.

Boyacioglu, A. M., Kara, Y., & Kaan, O. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications, 36*, 3355–3366

Campello, R.,& Hruschka, E. (2006). A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems, 157*, 2858–2875

Castiblanco, F., Gómez, D., Montero, J., & Rodríguez, J. T. (2017). Aggregation tools for the evaluation of classifications. In *Fuzzy system association of 9th international conference on software computer intelligent system (IFSA-SCIS)*, 2017 Jt. 17th World Congress International of IEEE, Otsu, Japan, 2017, pp. 1–5. https://doi.org/10.1109/IFSA-SCIS.2017.8023242.

Castiblanco, F., Montero, J., Rodríguez, J. T., & Gómez, D. (2017). Quality assessment of fuzzy classification: An application to solvency analysis. *FUZZY Economics Review, 22,* 19–31. https://doi.org/10.25102/fer.2017.01.02

Castiblanco, F., Franco, C., Montero, J., & Rodríguez, J. T. (2018). Relevance of classes in a fuzzy partition. A study from a group of aggregation operators. In: G. Barreto, R. Coelho (Eds.), *Fuzzy information processing.* NAFIPS 2018. Communications in computer and information science (vol. 831). Cham: Springer. https://doi.org/10.1007/978-3-319-95312-0_9

Castiblanco, F., Franco, C., Montero, J., & Rodríguez, J. T. (2019). Aggregation operators to evaluate the relevance of classes in a fuzzy partition. In: R. Kearfott, I. Batyrshin, M. Reformat, M. Ceberio, V. Kreinovich (Eds.), *Fuzzy techniques: Theory and applications.* IFSA/NAFIPS 2019 2019. Advances in intelligent systems and computing (vol. 1000). Cham: Springer. https://doi.org/10.1007/978-3-030-21920-8_2

Chen, M. (2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers and Mathematics with Applications, 62,* 4514–4524.

Dave, R. N. (1996). Validating fuzzy partitions obtained through c-shells clustering. *Pattern Recognition Letters, 17,* 613–623.

De Andrés, J., Lorca, L., de Coz Juez, & F. Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications, 38*, 1866--1875

Dombi, J. (1982a). Basic concepts for a theory of evaluation: The aggregative operator. *European Journal of Operational Research, 10,* 282–293. https://doi.org/10.1016/0377-2217(82)90227-2

Dombi, J. (1982b). A general class of fuzzy operators, the demorgan class of fuzzy operators and fuzziness measures induced by fuzzy operators. *Fuzzy Sets and Systems, 8,* 149–163. https://doi.org/10.1016/0165-0114(82)90005-7

Kaufman, L. Rousseeuw, P. J. (1990). Finding groups in data: an introduction to cluster analysis. New York: Wiley.

Krishnapuram, R., & Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems, 1,* 98–110. https://doi.org/10.1109/91.227387

Pessoa de Oliveira, A. (2016). Análisis de solvencia de las empresas: modelo estático vs modelo dinámico. https://zaguan.unizar.es; https://zaguan.unizar.es/record/48319/files/TUZ_0863_pessoa_analisis.pdf

Ravi, K. P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques—A review. *European Journal of Operational Research, 180*(1), 1–28.

Rojas, K., Gómez, D., Montero, J., & Tinguaro Rodríguez, J. (2013). Strictly stable families of aggregation operators. *Fuzzy Sets System, 228*, 44--63. https://doi.org/10.1016/j.fss.2012.12.010

Ruspini, E. H. (1969). A new approach to clustering. *Information and Control, 15,* 22–32. https://doi.org/10.1016/S0019-9958(69)90591-9

Sinkey, J. F. (1975). A multivariate statistical analysis of the characteristics of problem banks. *Journal of Finance, 30,* 21–36.

A Note on the Role of Government Incentives in Promoting Innovations

Javier Ignacio García-Fronti

Abstract This work aims to model the alternatives of government incentives for promoting innovation with social impact. Firstly, it analyses the R&D performed by a small technological firm and, secondly, the subsequent production to enter the market done by a big producer. This two companies must agree on how to share the output; it is a game between these two players, with a bigger negotiation power held by the big producer. If the big company want a big share, the government could incorporate incentives through push and pull subsidies to compensate the small one. Each one of these strategies has advantages and limitations in its application, and consequently, this paper proposes to use hybrid contracts, where the government can articulate the advantages of both subsidies.

Keywords R&D valuation · Join investment · Government incentives · Game theory

1 Introduction

To encourage the development of innovations that could solve social problems, the literature recognises two main categories of incentives to be used by policymakers: (i) *Push* or (ii) *Pull*. A government programme of push incentives is one that drives development by contributing with the cost of the development. This subsidy can be granted to people (scholarships or research awards) or to companies. An example is the cost-sharing subsidy plans, which encourage innovation in vaccine development by reducing the cost of research and expanding production. This scheme is effective in promoting R&D activities, but ineffective in guaranteeing the effort in the process.

On the other hand, pull programmes stimulates innovative research by promising to increase the future income generated by the developed product, which may consist

J. I. García-Fronti (✉)
Universidad de Buenos Aires. Facultad de Ciencias Económicas. IADCOM, Centro de Investigación en Métodos Cuantitativos Aplicados a la Economía y la Gestión, UBA, Av. Córdoba 2122, C1120AAQ Ciudad de Buenos Aires, Argentina
e-mail: javier.garciafronti@economicas.uba.ar

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_3

of an extension of the patent or a purchase commitment plan. The first one increases business incentive to carry out new research. Other governments prefer to opt for a price subsidy plan or directly a purchase commitment of a certain amount. The distinctive feature is that the government decides how much and how the discovery is distributed. The company waives its right to obtain a monopoly income, in exchange for a purchase commitment at an agreed price. For its part, Kremer (2002) concludes that pull subsidy programmes are more effective because they largely eliminate the agency problems between the developer and the developer.

One way to achieve the advantages of both incentive systems is to consider the possibility of *hybrid* plans, merging a subsidy of future income (*pull*) with a subsidy of expenses (*push*). For example, compensation contracts can be combined with a cost-sharing subsidy commitment to create hybrid contracts, which combine the positive attributes of both types of subsidies.

This work aims to model the alternatives of government incentives that promote the emergence of innovative products with social impact. To achieve this, it is structured in two sections. The first one analyses the stages of research and development necessary to achieve a patent and its subsequent production to enter the market. The stage of development is modelled by a small company with high professional skills that ends with the patent and the prototype of the product in question. This company must articulate with a large company with production and marketing capacity to reach the market. In this joint venture, clearly, the large company has greater bargaining power, which can strangle the income of the small one. It is modelled which are the restrictions that each company puts to participate in the project. In the second section, using the model proposed in the first, the role of government incentives is incorporated through push and pull subsidies. Each one has advantages and limitations in its application, so it is proposed to analyse the possibility of using hybrid contracts, where the government can articulate the advantages of both subsidies.

2 Joint Innovative Investments

Dixit and Pindyck (1994) argue that most investment decisions in innovative projects have three key features: (i) investments are mostly irreversible (fully or partially), (ii) uncertainty about the future return on the investment and (iii) investors have the option of waiting for better information about future prices (Pindyck 1991; Brennan and Schwartz 1985; Trigeorgis 1993; McDonald and Siegel 1986).

This section presents, in stylised form, a model for analysing investment decisions in innovative projects, particularly those that require collaboration between companies and, often, state subsidies. This section organised in two subsections. The first one analyses the R&D stages of an innovative project until the product is developed and discussed about the future market once the product is industrialised. Using backward induction, the second subsection proposes an algorithm that compute the net present value of a R&D investment.

2.1 R&D Stages and the Future Market

The stylised model proposed that this paper splits the investment project timeline in two main subprojects. One related with the *development* of the product, involving high skills professionals. In this subproject, the objective is to surpass different stages of the R&D process, to end up with a prototype and, possible, a patent of the product. The second one started with this mentioned knowledge and experience, aiming to *industrialise* the patent and *sell* the product efficiently to the market. Clearly, for this stage, usually some big producer is needed.

An R&D subproject is a process of sequential investments; in which it is possible to abandon it each time a new contribution is needed. In each stage, many tasks are carried out, and a learning process is taking place; the investor puts money to continue the development process. This is a process with high uncertainty that requires high professional skills. At the end of this stage, a prototype is built, and a patent form is submitted to the corresponding government office. Once the product has been developed and the factory installed for its manufacture, the new challenge is to conquer market.[1]

However, the decision to invest is shared. Both, the small company in charge of development analyses and the big company evaluates if they are willing to invest. The following subsection shows how the articulation between the two companies is modelled.

2.1.1 A Joint Investment Model

In every market, there are big players that have enough power and money to invest in new products. Some of they have the R&D built inside, others look for joint ventures with a small technological firm. This paper is interested in the last case.

Casparri and García Fronti (2010) presented a model for joint investment valuation. In their economy there are two companies. One is a small company that has specialisation in R&D (with no access to the consumer market). The second one is a national manufacturer with huge experience selling consumer products. They need each other to deliver a product with innovative technology to the market.

Before the project started, each agent knows how much money the project demands: K_N: capital contribution of the small firm and K_A: capital contribution from the big producer. Moreover, the expected income at the end of the project is Y, the probability of good outcome is p, the volatility is σ, and the discount factor is β.

The timeline is as follow (seeFig. 1.): At time "0", small firm (N) invests in R&D a quantity K_N. By time "1", the manufacturer invests K_A to start the production. At the end of the project, they receive, due to uncertainty,[2] a quantity above means (uY)

[1] Garcia Fronti (2017) develops a backward induction algorithm to calculate the present value of the project based on the works of Longstaff and Schwartz (2001) and Hsu and Schwartz (2008).

[2] Agents have zero risk aversion and the project has a failure probability of p, so u and d are calculated as follow:

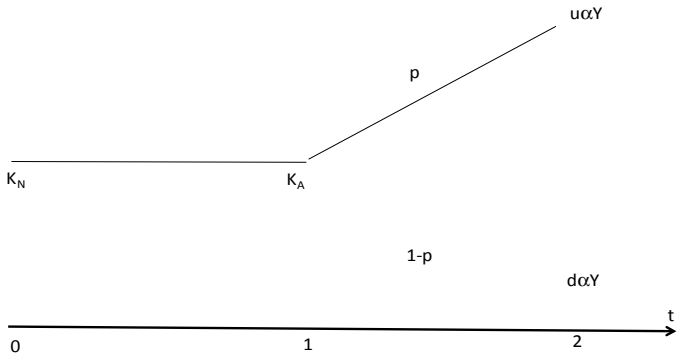


Fig. 1 Timeline and pay offs to small firm at time "2". *Source* García Fronti (2010)

or below (d*Y*). In any case, the final output is shared among companies, the small firm receives αY and the manufacturer $(1 - \alpha\ Y)$.

The basic game proposed to both players is as follow. At time zero, the small firm estimates the parameters of the project and its present value (with discount factor β). Using this information, he decides to invest in R&D an amount K_N at time "0" (in this case, the technology is ready for production at time "1"), or not. At time "1", the national producer values the project and decides if she wants to invest K_A (agreeing to pay Small firm a fraction α of the final income at time "2"). The project is subject to uncertainty and, at the end of the project, the outcome is realised, and each company receives the fraction previously agreed.

Each agent values the project from their point of view. The small firm will receive a fraction α of the outcome; so, he discounts it two periods to compare it with the initial investment in R&D. The national producer values the project knowing that she will receive the rest of the outcome and discounting it back one period. Therefore, the following participation constraints must be satisfied:

$$\text{Small firm } \beta^2[pu + (1-p)d]\alpha Y \geq K_N \tag{1}$$

$$\text{Producer } \beta[pu + (1-p)d](1-\alpha)Y \geq K_A \tag{2}$$

As far as the joint project is profitable, small firm will participate if receives an expected share (present value) bigger than the initial investment K_A, this imply asking for a minimum $\alpha \geq \alpha^d$. For its part, the producer will join if $\alpha \leq \alpha^u$. Therefore, the joint investment will take place if $\alpha \in \alpha^d, \alpha^u$. From (1) and (2), we obtain:

$$\alpha \geq \frac{K_N}{Y\beta^2[pu + (1-p)d]} = \alpha^d \quad \alpha \leq \frac{K_A}{Y\beta[pu + (1-p)d]} - 1 = \alpha^u \tag{3}$$

$$u = 1 + \sigma\sqrt{\frac{1-p}{p}}; d = 1 - \sigma\sqrt{\frac{p}{1-p}}$$

If the national producer has bigger bargain power than small firm, she could impose $\propto$, maximising her revenue into the game, forcing small firm out of the game and destroying the possibility of the joint project. Next section analyses how government intervention could help to keep both on board if the project has positive externalities for society.

If the present value of the project is low (or negative), none private investor would be interested. So, if the project has positive social impact, the government could intervene with subsidies to make the project more desirable. Next section analyses some forms of subsidies.

3 Government Intervention

If the manufacturer has enough bargain power to enforce $\alpha < \alpha^d$, the small firm opts out, due to a violation of his participation constrain. In this case where the project is not profitable for the small firm, the government could offer *push* or *pull* strategies to get it carried out. Formally, the government intervention extends the range of α values α^d, α^u for which both companies are willing to participate.

This section is organised in three subsections. The first one analyses how to formalise push incentives, the second one does the same with pull incentives, and the last one proposes an articulation of both strategies.

3.1 Push Incentive

In case $\alpha < \alpha^d$, government could propose push incentives. It consists of giving the small firm a proportion of his initial investment $(x_h K_N)$, to compensate producer negotiation power. Small firm participation constraint changes accordingly; while producer´s one remains unchanged:

$$\text{Small firm } \beta^2[pu + (1-p)d]\alpha Y \geq K_N - x_h K_N \tag{4}$$

$$\text{Producer } \beta[pu + (1-p)d](1-\alpha)Y \geq K_A \tag{5}$$

So,

$$\alpha \geq \frac{K_N(1-x_h)}{Y\beta^2[pu + (1-p)d]} = \alpha_h^d \quad \alpha \leq \frac{K_A}{Y\beta[pu + (1-p)d]} - 1 = \alpha^u \tag{6}$$

With the government subsidy, the small firm could accept $\alpha_h^d \leq \alpha^d$. Consequently, the new interval where both firms agree to produce α_h^d, α^u is extended, allowing that, although the producer presses for a small α, there is still the possibility of carrying

out the project. An important limitation of this type of policy is that the small business uses the money without making a real effort to develop the new product.

3.2 Pull Incentive

A different alternative of government intervention would be to promise a payment at the end of the project, in case it ends without the maximum expected profitability. To model this situation, let assume the government guarantee a minimum repayment of $(x_l + d)\alpha Y < uY$ at time "2" to the small firm, consequently his participation constraint changes accordingly (while the producer´s one remains unchanged) as follow:

$$\text{Small firm } \beta^2[pu + (1-p)(x_l + d)]\alpha Y \geq K_N \tag{7}$$

$$\text{Producer } \beta[pu + (1-p)d](1-\alpha)Y \geq K_A \tag{8}$$

So,

$$\alpha \geq \frac{K_N}{Y\beta^2(pu + (1-p)(x_l + d))} = \alpha_l^d \quad \alpha \leq \frac{K_A}{Y\beta[pu + (1-p)d]} - 1 = \alpha^u \tag{9}$$

Therefore, the joint investment will take place now if $\alpha \in \alpha_l^d, \alpha^u$, extending the interval where the game will take place. In the bad-case scenario, government paid at time "2" the amount of $x_l \alpha Y$, and in the good-case scenario, no payment is required.

3.3 Hybrid Incentives

As has been argued before, subsidies can be granted at the beginning of the project (*push*) or be a future promise once the project is completed (*pull*). On one hand, *push* incentives guarantee enough money upfront to start the development, on the other, *pull* grant are more effective to eliminate agency problems.

To have the positive attributes of both types of subsidies, governments could combine purchase contracts with cost-sharing subsidy commitment to create hybrid contracts. Formally, if the government proposes a hybrid contract, it should offer a payment at the beginning $(x_h K_N)$ and a future compensation $\left(\frac{x_l \alpha Y}{\beta^2}\right)$ in case the project ended with low profitability. For this new situation, participation constrains are as follow:

$$\text{Small firm } \beta^2[pu + (1-p)\max(d, x_l)]\alpha Y \geq K_N - x_h K_N \tag{10}$$

$$\text{Producer } \beta[pu + (1-p)d](1-\alpha)Y \geq K_A \tag{11}$$

So,

$$\alpha \geq \frac{K_N(1-x_h)}{Y\beta^2(pu+(1-p)\max(d,x_l))} = \alpha_h^d \quad \alpha \leq \frac{K_A}{Y\beta[pu+(1-p)d]} - 1 = \alpha^u \tag{12}$$

Government payment in case of a successful project is $P_u^G = x_h K_N$ and, in case of the bad scenario, $P_d^G = x_h K_N + \frac{x_l \alpha Y}{\beta^2}$. So, the expected present value is:

$$\text{Expected Goverment payment} = P_u^G + (1-p)P_d^G = x_h K_N + (1-p)\frac{x_l \alpha Y}{\beta^2} \tag{13}$$

4 Conclusion

The intervention of the state in the market has been analysed, through subsidies to companies. These subsidies can be granted at the beginning of the project (push) or be a promise if the project is successfully completed (pull). You can combine purchase contracts with cost-sharing subsidy commitment to create hybrid contracts, which combine the positive attributes of both types of subsidies. It is concluded that hybrid contracts are preferable to commitment purchase contracts that are in turn preferable to cost-sharing subsidy contracts. This framework could be used by policymakers to analyse incentives in the R&D market.

References

Brennan, M. J., & Schwartz, E. S. (1985). Evaluating natural resource investments. *The Journal of Business, 58*(2), 135–157.

Casparri, M. T., & Garcia-Fronti, J. (2010). A nanotechnology joint investment framework. *Computational Intelligence in Business and Economics*, 161–168

Dixit, A. K., & Pindyck, R. S. (1994). *Investment under uncertainty* (Vol. 15). NJ: Princeton University Press Princeton.

García Fronti, J. (2017). Stochastic valuation of innovative investments. In M. Peris Ortiz, D. Palacios Marqués, I. Gil Pechuán, S. Estelles Miguel, & C. Rueda Armengot (Eds.),*International conference on modelling and simulation* (pp. 183–190). AMSE, Editorial Universitat Politècnica de València

Hsu, J. C., & Schwartz, E. S. (2008). A model of R&D valuation and the design of research incentives. *Insurance: Mathematics and Economics, 43*(3), 350–367

Kremer, M. (2002). Pharmaceuticals and the developing world. *The Journal of Economic Perspectives, 16,* 67–90.
Longstaff, F., & Schwartz, E. (2001). Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies, 14*(1), 113.
McDonald, R., & Siegel, D. (1986). The value of waiting to invest. *The Quarterly Journal of Economics, 101*(4), 707.
Pindyck, R. S. (1991). Irreversibility, uncertainty, and investment. *Journal of Economic Literature, 29*(3), 1110–1148.
Trigeorgis, L. (1993). Real options and interactions with financial flexibility. *Financial Management, 22*(3), 202–224.

A Multi-agent MDSS for Supporting New Product Design Decisions

Juan Francisco Figueroa-Pérez, Juan Carlos Leyva-López, Édgar Omar Pérez-Contreras, and Pedro Sánchez-Sánchez

Abstract Market-driven product design decisions are receiving increasing attention in business practice, complemented by academic research on product design in a variety of disciplines. Mathematical and marketing research models are being integrated into decision-based design frameworks to represent customer behaviors and estimate the demand for design alternatives. The evolution of market structure eventually reshapes customer preference and competition, pushing the designers to rethink their design decision strategies. In this manner, it is necessary to develop decision support tools that would effectively provide support in new product design. In this paper, we argue that a combination of marketing decision support systems (MDSSs), multicriteria and multi-objective methodologies and agent technologies could be a very tool to support decision making in new product design applications. In this work, we propose a multi-agent architecture for a new product design MDSS and describe a system prototype for making new product design decisions.

Keywords Decision support systems and agent technology · New product design · Multi-objective evolutionary algorithms · Multicriteria decision analysis

J. F. Figueroa-Pérez (✉)
Universidad Autónoma de Sinaloa, Culiacán, México
e-mail: juanfco.figueroa@uas.edu.mx

J. C. Leyva-López · É. O. Pérez-Contreras
Universidad Autónoma de Occidente, Culiacán, México
e-mail: juan.leyva@udo.mx

É. O. Pérez-Contreras
e-mail: edgar.perez@uas.edu.mx

P. Sánchez-Sánchez
Universidad de Jaen, Jaen, España
e-mail: pedroj@ujaen.es

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_4

1 Introduction

With the globalization of the economy, most modern companies have found a new form of the global competition arena. Marketing managers are faced with a new changing and competitive market environment and are forced to make better decisions to become more competitive (Alexouda 2005). So, the way in which a company could adapt its marketing strategy will have a strong impact on the success or failure of the company's business (Ai et al. 2004).

The product design is one of the most important and expensive marketings' decisions. In the current competitive environment, the efficient development of new products is necessary for the survival of a company. The optimal product design problem is a vital element of the new product development problem and one of the most critical decisions for an enterprise (Alexouda 2005).

According to Matsatsinis et al. (2003), the design and development of new products is one of the most interesting and difficult decision-making cases in modern enterprises. Lei and Moon (2015) indicate that to maintain and improve the level of profitability in a competitive market an enterprise must redesign and reposition frequently its existing products or introduce new ones. At the same time, experts in this field have pointed out the importance of this for the viability of companies. Baril et al. (2012) said that the competitiveness of industries depends, among other things, on their ability to identify the customer's needs and to create products that meet these needs.

In recent years, new products developing is becoming in decentralized work. In this new environment, the task is distributed among different working groups, where the members are separated by discipline. With this, the industries are facing new challenges: (1) designing products in collaboration with distributed environment; (2) designing products that meet customer needs; and (3) reducing product development time and costs (Baril et al. 2012).

Ai et al. (2004) stated that building an MDSS is a valid way to increase the speed and effect of the decision, so as to improve the level of competitiveness of the company. An MDSS is defined as "a coordinated collection of data, models, analytic tools and computing power by which an organization gathers information from the environment and turns it into a basis for action" (Alexouda 2005).

This chapter proposes an MDSS to support new product design based on consumer preferences. The MDSS is developed using software agents, which are capable to respond to the requirements of the software related to the collection, choice, evaluation and utilization of information to support decision makers during the decision process (Matsatsinis and Siskos 2012). The models of MDSS are implemented through multi-objective evolutionary algorithms and are mainly based on the multicriteria ELECTRE III method.

It is believed that this combination will lead to the development of a tool that could effectively support this decision-making process according to the current nature of the new product design activities.

In the remainder of the chapter, it reviewed the existing related research work, the new product design methodology implemented by the software, the proposed multi-agent architecture and the developed prototype of the MDSS.

2 Product Design Decisions, MDSS and Software Agents

The product design process (PDP) in marketing environments "is a knowledge-intensive activity and unstructured decision-making process. In the whole process of product development, designers must make a decision according to a great amount of uncertain information" (Yu and Yan 2006). Modern information system technologies can support marketing managers to manage and take advantage of the large amounts of information currently available through different means.

According to Cassie (1997), MDSSs are "valuable tools to assist in making marketing decisions. MDSS can be used to support, rather than replace, decision makers in the complex, semi- or unstructured situations which are common in marketing problems. They incorporate the personal judgment and experience of the user to improve the effectiveness, rather than the efficiency of decision making."

As we can see in Ai et al. (2004, 1998); Matsatsinis et al. (2003); Banerjee et al. (2002); Liang and Huang (2002); Zha et al. (2003); Vahidov and Fazlollahi (2004); Yu et al. (2009); Morales and Ortega (2010); and Dostatni et al. (2015), many efforts have been made in recent years to develop MDSS to carry out PDP.

An analysis of marketing decision tasks and the characteristics of multi-agent system (MAS) carried out by Ai et al. (2004) showed that:

- Multi-agent technology "makes possible for the system to asynchronously increase or decrease the agent function without influencing other parts of the system" (2004), what is appropriate for an MDSS, which may require smartly increase or decrease according to the change of marketing circumstances.
- In a MAS, the "characteristics of distribution, mobility and self-adaptability help users to get the newest information relevant to the decision task from the network" and "each agent has its own knowledge base, making possible a distributed store of knowledge" (Ai et al. 2004). Both elements are appropriate for current marketing decision's requirements of information.

Based on the above, the nature of software agent technology makes them very suitable for marketing applications.

Some researchers are looking to automate product design-related activities using agent technologies (Ai et al. 2004, 1998; Matsatsinis et al. 2003; Banerjee et al. 2002; Liang and Huang 2002; Zha et al. 2003; Vahidov and Fazlollahi 2004; Yu et al. 2009; Morales and Ortega 2010; Dostatni et al. 2015). Matsatsinis et al. (2003, 1998) presented an MDSS that implements a new consumer-based methodology for product penetration strategy selection. Banerjee et al. (2002) proposed a method to address uncertain in the decisions involved in the new product and technology development and positioning with hybrid intelligent systems. It uses fuzzy inference systems

to model ambiguous conditions and ant colony optimization for searching optimal combined strategy to meet requirements. Liang and Huang (2002) described a collaborative multi-agent system and a procedure to develop modular products. Zha et al. (2003) described a hybrid decision model and a multiagent framework for collaborative decision support in the design process. Ai et al. (2004) designed a layered system architecture composed of decision customer layer, the decision core layer and decision resource layer to develop a distributed MDSS. Vahidov and Fazlollahi (2004) describe a multi-agent architecture for an e-commerce's MDSS a prototype system for making investment decisions. Yu et al. (2009) presented a multi-agent intelligent decision support system (IDSS) based on new product development. The system is composed of eight modules: data collection, data management, statistical analysis, breakeven analysis, risk prediction analysis, resource optimization scheduling, expert system and information exchange. Morales and Ortega (2010) proposed a distributed intelligent system to determine tendencies in customer's preferences from the application of surveys to clients. Dostatni et al. (2015) described the structure of a multi-agent decision support system (DSS) to support the decision-making process of the designers in eco-design.

While these developments in DSS to support new product design are very encouraging, they have the disadvantage that methods implemented in their models not taking into account the parameter of importance that is assigned by a customer to each criterion to evaluate a product, ignoring with it valuable information to model accurately consumer preferences.

3 Customer-Based Multicriteria Methodology for Product Design Decisions

A new product design multicriteria methodology (NPDMM) based on consumer preferences is implemented computationally by the MDSS proposed in this paper. Such methodology has as one of its main characteristics that it models consumer preferences using a more flexible approach than those used currently in the literature and takes into account the importance assigned by a customer to each criterion to evaluate a product. It also considers the preferences of the decision maker, since he is the one who decides what criteria can be approached in the design of the new product based on the actual enterprise condition, capacities and needs of the business. The methodology is based mainly on the multicriteria ELECTRE III method.

4 Relation Between Simon's Decision-Making Process and the NPDMM

The NPDMM is directly related to Simon's decision-making process (SDMP), which includes a choice between alternative plans of action. The choice includes facts and values. Every decision is composed of a logical combination of fact and value propositions. Facts mean existence or occurrence of something tangible and concrete which can be verified. Value is a matter of preference (Simon 1997).

Simon divides the process of decision making into three phases—intelligence: finding occasions calling for decision; design: identifying, developing and analyzing all possible alternative courses of action; and choice: selecting a particular course of action from available choices (Simon 1997).

The NPDMM is structured according to Simon's decision-making process and consists of several stages and tasks that are carried out to design a new product. The main stages and tasks of the methodology are described below:

1. Project Definition: Problem description, criteria, participants, alternatives and evaluation scales are defined.
2. Market Study: Once defined the problem, criteria, alternatives and scales, they will be used to design the survey which will be applied to the consumers to carry out the market study corresponding to the new product to be designed.
3. New Aggregation/Disaggregation Preference Method: It is an ELECTRE III-based method that represents the preferences of the set of consumers from the knowledge stored in the market study.
4. New Market Segmentation Method: It allows to segment the market from a multicriteria analysis point of view using an evolutionary algorithm. It will form classes of consumers similar between them.
5. New Brand Choice Model: It assigns a brand choice model that represents the consumer behavior per consumer preferences generated by the new aggregation/disaggregation preference method. The output of the model is the market share of the new product in a specific market segment.
6. What-If Analysis: It is an iterative process that will allow the decision maker to test different configurations of design of a new product in a specific market segment before being launched. It will give the ranking position of a new product under design with respect to its competitors in a particular market segment.

Table 1 summarizes the link between Simon's problem-solving model and the NPDMM described in this work.

The next section describes the proposed architecture for the MDSS that implements the NPDMM previously explained.

Table 1 Link between Simon's problem-solving model and the new product design multicriteria methodology

SDMP stages	NPDMM	Comments
Intelligence	Project definition	Finding occasions calling for a decision. Product idea/opportunity discovery. Problem description/design specifications: criteria, scales/identify participants
Design	Market studies, aggregation/disaggregation method, market segmentation method, brand choice model	Identifying, developing and analyzing all possible alternatives. Generate and evaluate rough design layouts/preliminary specifications of the layout. Get consumer preferences and market segments
Choice	Optimal product recommendation and what-if analysis	Selecting from the available choices. Get the specification of production from an optimal recommendation or what-if analysis

5 Architecture of Multi-agent MDSS

Vahidov and Fazlollahi (2004) stated that "agent technologies offer the highest-ever level of abstraction in computer science. This allows the designers to devise the overall architecture of a multi-agent system without specifying the implementation details" (Vahidov and Fazlollahi 2004).

The architecture design in Fig. 1 is based on a combination of the well-known proposal of agent types presented in Sycara et al. (1996) and the component integration based on Simon's decision-making process presented in Vahidov and Fazlollahi (2004).

First, agents are considered according to three different types (Sycara et al. 1996):

1. Information agents provide access to information sources. Their goal is to provide information from the system's general database or other agents. MDSS architecture has the following information agents: project, monitor and market study data manager.
2. Task agents support decision making by formulating problem-solving plans and interacting with other software agents to carry them out. The architecture shows the following task agents: initialization manager, aggregation/disaggregation preferences, market segmentation, automatic optimal recommendation and brand choice.
3. The interface agents interact with the user by receiving user input and showing results. Decision maker and facilitator are interface agents in the MDSS architecture.

As in Vahidov and Fazlollahi (2004), components of the architecture include intelligence team, design team and choice team. Its name corresponds to the phases

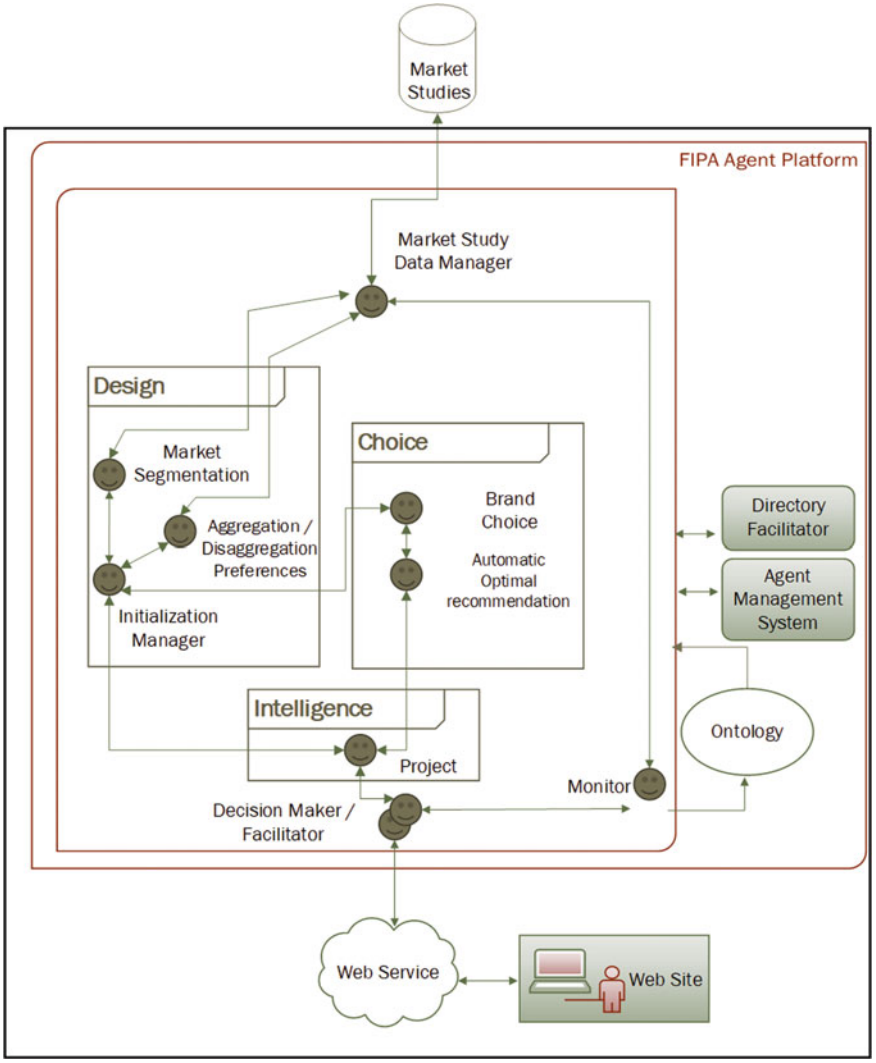


Fig. 1 MDSS multi-agent architecture

of Simon's decision-making stage that the corresponding agents aim to support (Vahidov and Fazlollahi 2004).

- The intelligence team has one information agent: the project agent. Its responsibilities are to define the decision problem and identify criteria to evaluate new product and alternatives of an existing product in the market.
- The design team incorporates information and task agents: initialization manager, aggregation/disaggregation preference agent and market segmentation agent. Its main responsibilities include the initialization of the environment for the new

product design. It obtains consumer behavior through multi-objective evolutionary algorithms based on ELECTRE III multicriteria method incorporated in the aggregation/disaggregation preference agent. Consumer characteristics are obtained by multi-objective optimization based on the non-dominated sorting genetic algorithm II (NSGA II). The output of both elements is used as input for the brand choice agent.

- The choice team also consists of information and task agents: automatic optimal recommendation agent and brand choice agent. Its main responsibility is to generate an optimal recommendation of the new product under study. The output of choice team agents will tell the position of a new product under design with respect to its competitors in a particular market segment. The recommendation is generated automatically from an evolutionary algorithm based on linear programming that determines the optimal configuration of a product with respect to the group of competitors in a specified market segment. Here can be carried out a what-if analysis. It is an iterative process that obtains the position of a new product under design with respect to its competitors in a market segment.

6 Multi-agent MDSS Prototype for Product Design Decisions

In order to show our proposal, we have developed an MDSS prototype to support new product design decisions. The interface is a PHP Web site that provides interaction between the user and MDSS multi-agent platform to show information, perform calculations and what-if analysis and carry out other related tasks. The models, data and user interface are traditional DSS components. Data are collected from applied surveys of market studies and are stored in a relational Microsoft SQL Server database. At this moment, the prototype incorporates the agents of intelligence, design and choice teams which was described above. The decision maker (DM) or facilitator (F) can accept it or go to evaluate different configurations until it finds a satisfactory one. Below are some sample screens of the prototype user interface.

Figure 2 shows a dashboard that is presented once the user logs in the system. Here, the list of projects in which the user participates is displayed. In this window, it is possible to create a new project by pressing the <Design New Product> button, or manage an existing one by clicking on the name of the desired project.

Figure 3 shows the automatic recommendation screen for the *Corn Oils* project, which is divided into three sections: *Setting Grid*—in this section are displayed the optimal values for each attribute of the new product in a grid format; *Market Share Grid*—in this section, the market share obtained for the new product is shown regarding the products evaluated by consumers in grid format; *Expected Market Share*—it shows the same information from the Market Share Grid section but in pie chart format in which the market share of the new product can be observed visually in relation to the products evaluated.

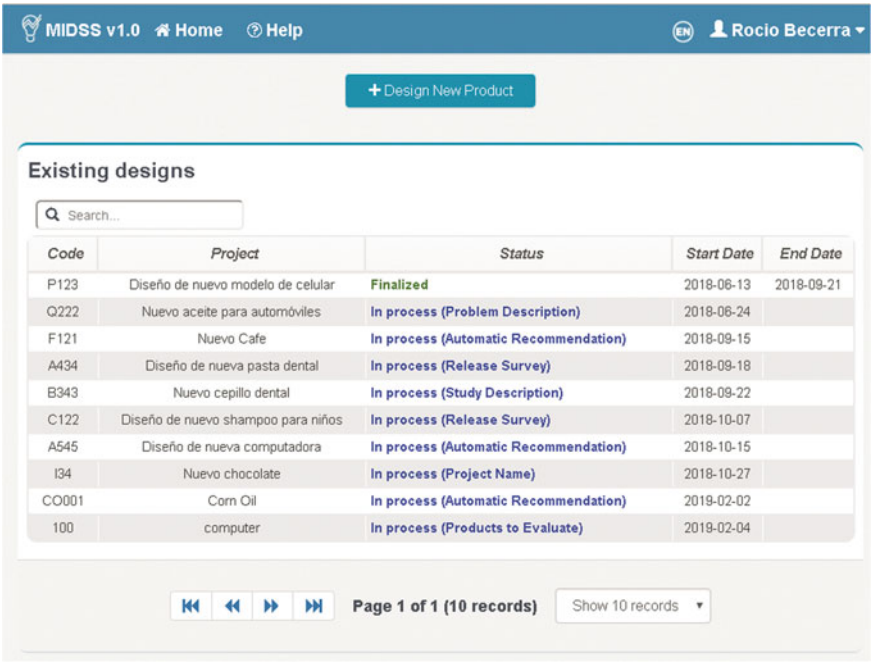


Fig. 2 MDSS new product design screenshot

7 Discussion and Conclusions

This paper presented a multi-agent architecture for a new decision support system to support new product design decision process in a real-world context. It is argued that a combination of DSS, agent technologies and multi-objective evolutionary algorithms based on multicriteria ELECTRE III methods to support decision making in marketing applications will prove to be a powerful tool. It described a multi-agent architecture for a novel MDSS which implements a new methodology for designing new products based on consumer preferences. Such methodology is substantiated on Simon's decision model. Their models are based on ELECTRE III multicriteria method and were built in the software through multi-objective evolutionary algorithms. Unlike other approaches in the literature, consumer preferences are modeled in the methodology using a more flexible approach which allows taking into account the importance assigned by a customer to each criterion to evaluate a product with what is intended to model more appropriately consumer preferences than in the current proposals.

The MDSS is built using a generic reusable multi-agent architecture which has three different agent types: interface agents, task agents and information agents. The major components of the architecture also reflect the phases of decision-making stage of Simon and include intelligence team, design team and choice team. Intelligent

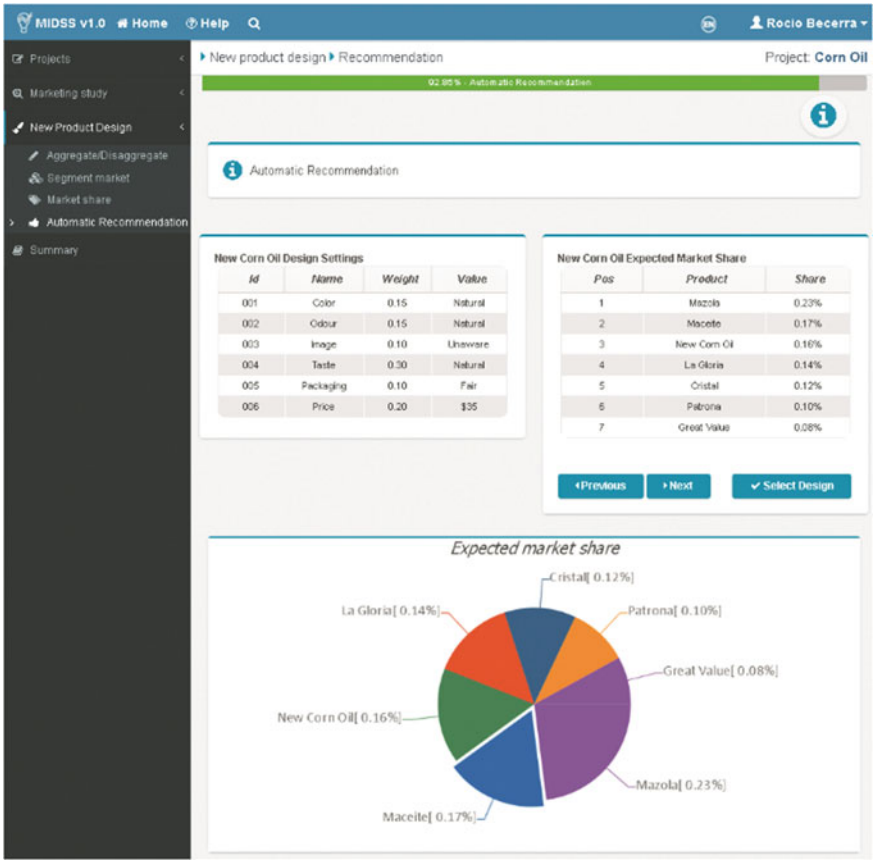


Fig. 3 MDSS automatic recommendation screenshot

software agents are capable to respond exactly to the requirements of collection, choice, evaluation and utilization of information to support decision makers during the product design process in distributed marketing environments. They allow to have a distributed store of knowledge, a flexible increase or decrease of the system without influencing other parts of software and a close cooperation between each intelligent agent for task's decomposing.

References

Ai, W.-G., Sun, J. S. J., & Li, H. L. H. (2004). A distributed marketing decision support system based on multi-intelligent-agent. In *Proceedings of 2004 international conference on machine learning and cybernetics (IEEE Cat. No.04EX826)* (vol. 1, pp. 26–29)

Alexouda, G. (2005). A user-friendly marketing decision support system for the product line design using evolutionary algorithms. *Decision Support Systems, 38,* 495–509.

Banerjee, S., Abraham, A., & Grosan, C. (2002). Intelligent decision making for new product development and market positioning using soft computing. *International Journal of System Management*, 1–25

Baril, C., Yacout, S., & Clément, B. (2012). An interactive multi-objective algorithm for decentralized decision making in product design. *Optimization and Engineering, 13,* 121–150.

Cassie, C. (1997). Marketing decision support systems. *Industrial Management & Data Systems, 97,* 293–296.

Dostatni, E., Diakun, J., Grajewski, D., Wichniarek, R., & Karwasz, A. (2015). Multi-agent system to support decision-making process in Ecodesign. In *10th international conference on soft computing models in industrial and environmental applications* (pp. 463–474)

Lei, N., & Moon, S. K. (2015). A decision support system for market-driven product positioning and design. *Decision Support Systems*

Liang, W. Y., & Huang, C. C. (2002). The agent-based collaboration information system of product development. *International Journal of Information Management, 22,* 211–224.

Matsatsinis, N., Siskos, Y. (2012). *Intelligent support systems for marketing decisions* (vol. 54). Springer Science & Business Media

Matsatsinis, N., Mora, P., Psomatakis, V., & Spanoudakis, N. (1998). Intelligent software agents for products penetration strategy selection. In *MAAMAW*

Matsatsinis, N., Moraïtis, P., Psomatakis, V., & Spanoudakis, N. (2003). an Agent-based system for products penetration strategy selection. *Applied Artificial Intelligence, 17,* 901–925.

Morales V. L., & Ortega, O. L. (2010). Direct marketing based on a distributed intelligent system. In *Marketing intelligent systems using soft computing* (pp. 255–271). Springer

Simon, H. A. (1997). *Administrative behavior* (4th ed.). Free Press

Sycara, K., Pannu, A., Willamson, M., Zeng, D., & Decker, K. (1996). Distributed intelligent agents. *IEEE Expert, 11,* 36–46.

Vahidov, R., & Fazlollahi, R. (2004). A multi-agent DSS for supporting e-commerce decisions. *Journal of Computer Information Systems, 44,* 87–94.

Yu, G. Y. G., & Yan, H. Y. H. (2006). A New decision making method for product development based on multiple neural network. In *2006 6th world congress on intelligent control and automation* (vol. 2, pp. 6792–6795)

Yu, T. Y. T., Zhou, J. Z. J., Xu, F. X. F., Gong, Y. G. Y., & Wang, W. W. W. (2009). Decision support system of product development based on multi-agent. In*International conference on information technology and computer science* (vol. 2, pp. 0–3)

Zha, X. F., Sriram, R. D., & Lu, W. F. (2003). Knowledge intensive collaborative decision support for design process. In *ASME 2003 international design engineering technical conferences and computers and information in engineering conference* (pp. 425–438)

Profile Information Analysis of Twitter Social Network

Angélica Urrutia and Carolina Nicolas

Abstract The vertiginous development of technology and knowledge globalization has generated a high interest on social networks within organizations, where its presence has multiplied exponentially in recent years. That is why, it is proposed in this paper to analyze data extracted from social networks, specifically from Twitter, aiming to obtain different data elements that allow management and analysis of the opinions provided by users on different. This information is very useful for client management and acknowledging preferences of brands and organizations. In particular, this work responds to the following research questions: (1) How does the Twitter user behave in different brands? and (2) How do opinions on the network affect the company's Twitter profile? In this manuscript, we present the tweet user profile for information analysis via a practical software architecture proposal, which is composed by four layers (extraction of the data source, ETL processes (extraction, transformation, and loading), selection of database, and visualization of the results). The implementation and dashboards of this architecture come from the study case of different types of organizations: banking, telephony, shopping, and supermarkets. The processing of the data corresponds to the extraction of tweets generated by the Twitter users of the organizations. Then, the ETL process is obtained via the useful Spoon from Pentaho Data Integration. The processed data is employed to build the final database, and finally, the generated information is visualized by utilizing dashboards from Qlik Sense Desktop. The results of this study evidence that it is possible to implement a practical architecture to analyze the model information of the Twitter user profile through dashboard; consequently, the organizations can opportunely realize better decisions.

A. Urrutia
Facultad de Ciencias de la Ingeniería, Universidad Católica del Maule, Casilla 617, 3480112 Talca, Chile
e-mail: aurrutia@ucm.cl

C. Nicolas (✉)
Escuela de Ingeniería Comercial, Facultad de Economía y Negocios, Universidad Santo Tomás, Av. Ejercito 146, 8370003 Santiago, Chile
e-mail: cnicolas@santotomas.cl

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_5

Keywords Data management · Tweet analysis · Opinion in social networks

1 Introduction

Keeping record of the opinions and comments that are being shared on social networks, which are part of the Web pages of public and private organizations, especially in companies of shopping, allows to manage this information and thus give direct answers to questions or criticisms from the clients, or analyze the data, or promote products, or carry out market studies, among other options, this being fundamental for the management of information. This topic has been treated by different authors (Ghiassi Skinner and Zimbra 2013; Jansen and Chowdury, 2009; Lin and Ryaboy 2013; Mostafa 2013; StrateBi 2013) by applying different techniques of information analysis and opinion assessment of tweet users.

Nowadays, organizations urgently demand useful, fast, and reliable information in order to get their goals and hence desired success. The traditional information systems generally show a static structure, for instance, great rigidity in extracting data, long response periods, and lack of integration. As a consequence, there is the need to develop models that support the organization sustainably and continuously to improve its competitiveness, by providing the necessary information for decision making (Cano 2007). In this context, we use the tweet information of some Chilean organizations: banks, shopping, supermarkets, and telephony, to create personalized and interactive views (dashboards). This information responds to the following research questions: (1) How does the Twitter user behave in different brands? and (2) How do opinions on the network affect the company's Twitter profile?

The main computational tools with free license to business intelligence model are: Python (Anaconda Data Science Platforms), RStudio, RapidMiner, and Qlik Sense. On the other hand, by employing private licenses, the following tools are highlighted: Tableau, QlikView, and Power BI (Visvizi et al. 2019). Via these tools may analyze social networks, such as Twitter and Facebook, and with the assistance of data mining (Chae 2015), the results can be depicted to intelligence dashboard (Kurnia and Suharjito 2018).

In this work, we use a tweet user profile for information analysis based on the need to analyze the value of the experience of Twitter clients with respect to different organizations. A free architecture software for its implementation is considered. And to obtain dashboard, which supports the decision-making process, the following sections were developed: user profile proposal, software architecture, analysis of profile data tweets, and results of the analysis of the proposed profile.

2 User Profile Proposal

The analysis of the data extracted from Twitter for this research focuses on organization profiles: banking, telephony, shopping, and supermarkets, selecting the current accounts. The interest of the information extracted from each page focuses on the display of the profile of the type of Twitter, considering the following characteristics:

- **Leadership**, meaning that we want to visualize which tweet is the one that is being talked about more recurrently and to whom it belongs. For this, it is necessary to have the messages issued and with information such as the amount of "retweet" and "I like" that the tweets that refer to the different profiles have.
- **Experience**: This approach seeks to inform how participative or expert is the user who has made the tweet, either because it has a large number of messages posted, a large number of people who are interested in their profile, the number of profiles of interest or the amount of "I like" that performs to different tweets. For this, the information that is required is related to the user's profile, where the relevant data for this analysis corresponds to the date of creation of the said account, the number of tweets, number of followers, amount of followers, and quantity of likes owned by the user.
- **Origin**, which uses the information related to the user's location, which is entered by himself when creating his account. This restricts to obtaining an effective analysis, since the location is presented in different formats or even in an incorrect way, as well as because it is not mandatory information, we find users without defining their location.

As a summary, Fig. 1 shows a diagram with the mentioned requirements and the characteristics required in the analysis of the user profile of a social network, such as Twitter.

For analyzing the data, the collection of different tweets from the Web pages of the selected companies is taken into account (Martínez-Cámara et al. 2011; Sebastiani and Esuli 2006). Data associated with the characteristics is extracted for each profile. Subsequent implementation is done using the practical tool, Qlik Sense Desktop [that uses the associative data indexing (QIX) engine (Badr et al. 2018)] with the purpose of generating personalized and interactive visualizations, dashboard, of data, which in turn seeks to answer the following research questions:

- How does the Twitter user behave in different brands?
- How do opinions on the network affect the company's Twitter profile?

3 Software Architecture

For implementing and obtaining analysis results of user's profile, it is of crucial importance of the existence of a properly defined software architecture, which allows to extract and analyze data from Twitter (Ikeda et al. 2013; Jansen et al. 2009; Nigam

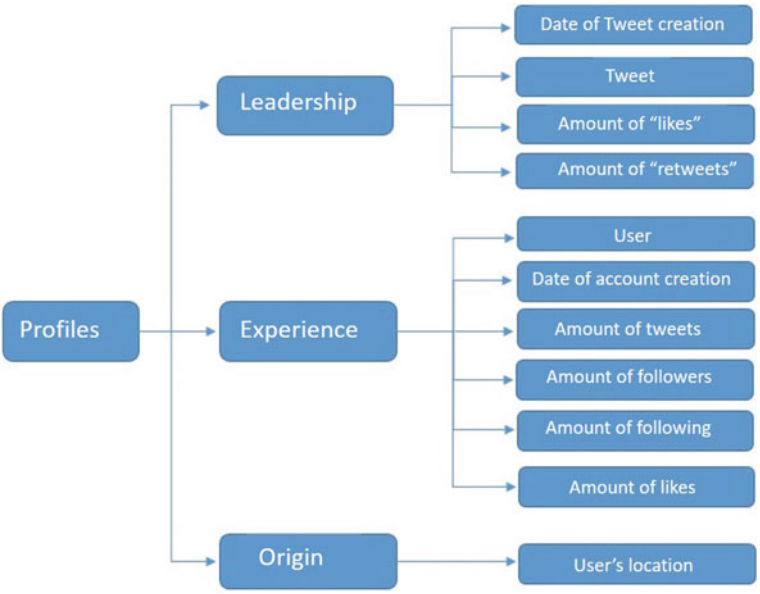


Fig. 1 Proposed features for user profile Twitter

Lafferty and McCallum 1999; Sebastiani 2002). Figure 2 shows the proposal arranged in four layers based on data extracted by Twitter to analyze user profile.

Figure 2 shows a proposal of a software architecture and tools composed of four layers of processes described below, and each of them requires specific tools

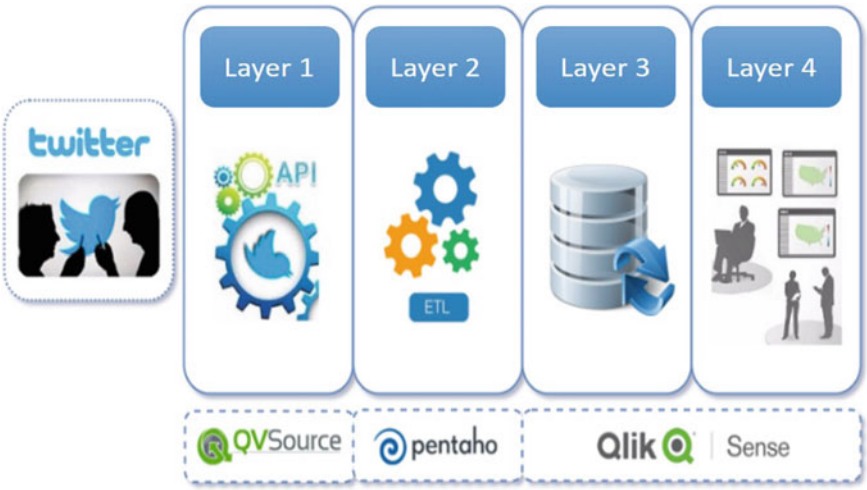


Fig. 2 Software architecture proposal to analyze profile Twitter

associated with different tasks and processes, such as: QVSource, Pentaho, and Qlik Sense Desktop.

Layer 1 Extraction of the data source. It is the first layer that performs the extraction of data from Twitter, making reference to the three proposed profiles: leadership, experience, and origin. For the extraction process, we use the QVSource tool, which allows to connect Qlik Sense Desktop to the large number of social and business APIs available on the Web. In this case, QVSource is used as a connector with the Twitter Search API in an interactive way, thus allowing to obtain the data of all the tweets that refer to each user and specific information of the user who had made those tweets.

Layer 2 ETL processes. In this second layer, the Spoon tool of Pentaho Data Integration is used to carry out the transformation and data loading. The data obtained is converted with the purpose to make it as clear and clean as possible, and then generate a single file with the data of the twelve Chilean profiles selected for further analysis: Banco de Chile, Banco Estado, Santander, Falabella, Ripley, Paris, Jumbo, Lider, Unimarc, Claro, Entel, and Movistar. The data was first extracted and then transformed (by including users' data and by adjusting the creation date of the tweet, the creation date of the user, and location). Finally, we created a single file in CSV format that possesses all information of the 12 Twitter profiles of the organization users considered in the work.

Layer 3 Selection of database. This layer acquires importance, since it is where the Qlik Sense Desktop interactive software starts to be used for the analysis of the Twitter user profile. As a first task, we loaded the file generated in Layer 2 is loaded, and then select the attributes or data of interest, to carry out the information analysis model, already proposed in Fig. 1. In this way, the data base is obtained with the elements of analysis required, where 14 characteristics (creation, information, amount of both retweets and likes, user name, data, year, user location, realized tweets, number of followers and following, profile) were selected.

Layer 4 Visualization of the results. In this last layer is where the results obtained from the previous layers of the proposed architecture are visualized. From the analysis of the generated database and the different visualization options provided by Qlik Sense Desktop, dynamic and easily comprehensible dashboards are done, in order to provide answers to the proposed management indicators in a way that will be useful for making decisions regarding what Twitters think (Lewis 1998; Mostafa 2013; Saxena and Gadhiya 2014).

4 Data Analysis of Tweet Profile

There architecture obtained (Fig. 2) allows implementation for data analysis extracted from social tools (Twitter), related to the value of the customer experience, through

the "user profile" model (Fig. 1), for which the Qlik Sense Desktop program is used as a data visualization tool. This type of solution has a great advantage that is it is low cost and, in addition, it has the particularity that can be used, by any person that has the need to solve this model for the development and optimization of their organization, without the need to count necessarily with experts in the area of business intelligence.

The proposed solution complies with the fundamental stages of a business intelligence model, since it provides a model of information analysis and different visualizations that allow to deliver answers to the concerns raised by companies to make the best decisions based on relevant information.

Obtaining information provided by users through social networks is of high value and utmost importance for organizations, since it is through them that people expose claims, opinions, experiences, or needs that are relevant to them and that constitute a valuable information for all types of entities that wish to remain connected with the market.

To perform this information analysis, the social network taken as an example is Twitter, since it is one of the most used networks, with a large number of active users. For this reason, the choice of this network to develop the user profile model was considered as the most appropriate and ideal, mainly due to its particular characteristic of having text messages limited to 280 characters (Priya et al. 2019), and thanks to the fact that the information provided by Twitter is of public access, as are also the applications that allow you to connect to the API of this social network to obtain the desired information.

In this way, the possibility of accessing data is generated by users of social networks associated with organizational Webs, thus providing a great and valuable opportunity for them to approach adequately and with relevant information to their clients, since this possibility has allowed us to know more in-depth customer opinions.

This alternative is nowadays crucial in the globalized world with increasingly demanding and informed organizations and clients. For this reason, organizations are constantly confronted with challenges, in which making the right decisions can make the difference between achieving success or simply failure. In this sense, access to useful, relevant, and reliable information allows the company to make strategic decisions and guide its business operations focusing on the commercial management of its clients that will positively affect the corporate image of the company and thereby be competitive. As well as consolidate its position in the market.

5 Results of Twitter Profile Analysis

The analysis of the data and its graphical representation is relevant to this research, as shown in the information displayed in Figs. 3 and 4 generated. These results were observed from 42,341 tweets extracted since different sources (banks, telephony, supermarkets, selected for this case) during a compilation period of 35 days and

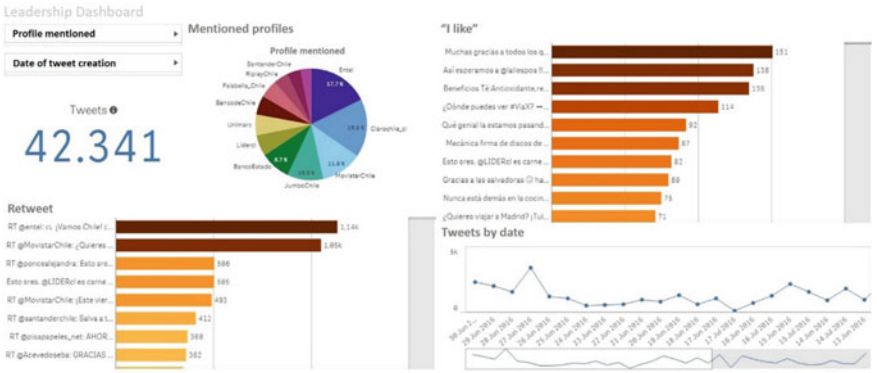


Fig. 3 Dashboard in relation to the leadership approach

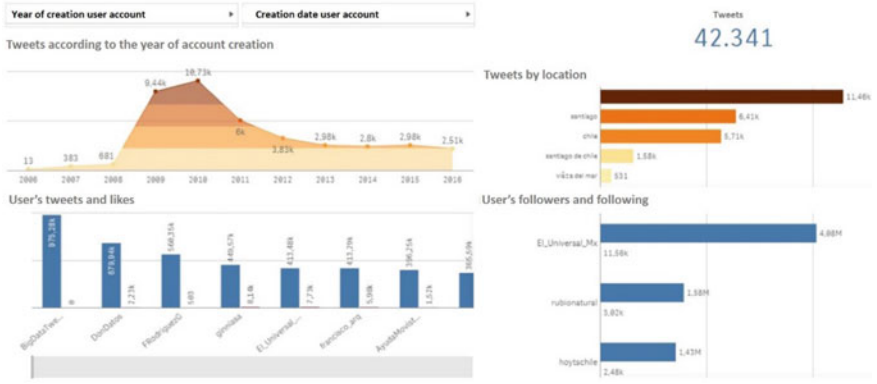


Fig. 4 Dashboard in relation to the experience and location approach

considering the items of the Twitter user profile proposal (Fig. 1) and implementing each of the layers of the architecture proposal (Fig. 2).

Figure 3 shows a dashboard in relation to the leadership approach of the proposed user profile. The pie chart gives a visualization corresponding to the percentage of tweets that refer to each profile studied by each of the companies; in the upper bar graph, they visualize the different tweets together with the amount of "Like" they have, excluding those that do not have, with the aim of obtaining the tweet that is most liked by users for analysis according to their content. In the lower bar graph, the different tweets are displayed together with the number of times they have been retweeted, excluding those that were not, with the aim of obtaining the tweet that has been most widely spread by the users for its later analysis according to its content, in addition, with the purpose of knowing the day in which the greater amount of tweets is made, and a line graph is created that allows to observe how it behaves over time and the total amount of tweets generated.

Figure 4 shows a dashboard in relation to the experience and location approach of the proposed user profile. In the case of the number of tweets made according to the year of creation of the user's account, a line graph is displayed at the top showing the number of tweets made by the users depending on the year in which they had created their Twitter account. The graphs of the lower part show the results of the amount of tweets and likes or "likes" made by the user, information that is produced by the user itself, that is, the amount of tweet and likes of the account, plus, the number of followers and followings that the user has. To do this, a bar graph is generated that shows the number of followers that the user has, in order to visualize the user with more followers, who follows the largest number of users or compare between them.

In the upper right part of Fig. 4, the number of messages per location is represented, for which a bar graph is drawn indicating the number of tweets generated according to the user's location, in case it is registered.

The dashboards of Figs. 3 and 4 allow information to be managed among the visualizations, being a relevant topic the date of creation of the user's account, in this way the number of tweets, likes, followers, and following that the user had during the time he has been interacting on Twitter. For example, a user with thousands of followers who created his account 10 years ago is not the same as a user with the same number of followers incorporated into the social network just a couple of years ago, since the latter has greater participation and popularity that has a large number of users who are interested in your profile in a shorter period.

For a more thorough analysis, the information that indicates the number of tweets that the user has made is used, visualizing in this manner the users that have greater participation in the social network for the short term. From the management of the information generated by the proposed user profile and architecture, the silver questions can be answered at the beginning. How does a Twitter user behave in the different brands? And how do opinions on the network affect the company's Twitter profile?

Although most of the users do not enter their location or do but incorrectly, it is decided in the same way, to carry out an analysis where the number of tweets made by the user's location is visualized, to have a look at this information and generate the possibility of analyzing if it will be useful according to the needs. For these dashboards, a filtering panel has been used, which allows to visualize the information, according to the year in which the user created his Twitter account, and more specifically, by the day and month of the creation of said account. Finally, the information and data analysis have a number of applications and ways of visualizing that users will support the management of the companies based on the user profile proposed in Fig. 1.

6 Conclusions

In this paper, an architecture of the value of the tweet customer experience was implemented. This was done by using the model of the user profile via the data visualization

tool known as Qlik Sense Desktop. This proposal of business intelligence model and information analysis is characterized by the data extraction since Twitter. After the extraction, transformation, and loading (ETL) steps, the interest data is stored and then loaded in the Qlik Sense Desktop software. Finally, different visualizations are carried out to obtain clear answers to the organization questions. Therefore, organizations can realize best decisions based on their relevant information.

In the manuscript, the results demonstrate that the information given by Twitter users represents a high value item and, consequently, a fundamental point to the business intelligence in organizations. This occurs owing to this social network, people reveal their inquiries, opinions, experiences, or requirements, given topic information to any organization, who wish to remain connected to the current markets.

Over the course of the research carried out for this paper, some interesting topics that may arise from its review have been identified. Some future works that can be done are proposed as follows.

Algorithm to Perform Sentiment Analysis of Tweets: This research provides the basis for obtaining different analyses of the value of the experience of Twitter clients, which can be complemented by other factors associated with the messages to obtain more specific studies.

Information Management: With the data stored by each tweet, you can focus on more specific research questions, associated with the management of brand satisfaction of each client who accesses the Web page of the organizations and gives their opinion on it. A pending task is to expand social tools such as Facebook among others.

Acknowledgements This paper is part of the project supported by "Red Iberoamericana para la Competitividad, Innovación y Desarrollo" (REDCID) project number 616RT0515 in "Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo" (CYTED), and the Catholic University of Maule with the research group in databases TRICAHUE.

References

Badr, Y., Mohamed, N., & Mohamed, A. (2018). Recent trends in big data analytics towards more enhanced insurance business models. *International Journal of Computer Science and Information Security, 15*(12), 39–45.

Cano, J. L. (2007). *Business intelligence: competir con información.* Madrid: Banesto, Fundación Cultur [i.e. Cultural].

Chae, B. K. (2015). Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics, 165,* 247–259. https://doi.org/10.1016/J.IJPE.2014.12.037

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications, 40*(16), 6266–6282. https://doi.org/10.1016/J.ESWA.2013.05.057

Ikeda, K., Hattori, G., Ono, C., Asoh, H., & Higashino, T. (2013). Twitter user profiling based on text and community mining for market analysis. *Knowledge-Based Systems, 51*(1), 35–47. https://doi.org/10.1016/j.knosys.2013.06.020

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology, 60*(11), 2169–2188. https://doi.org/10.1002/asi.21149

Kurnia, P. F. (2018). Business intelligence model to analyze social media information. *Procedia Computer Science, 135,* 5–14. https://doi.org/10.1016/J.PROCS.2018.08.144

Lewis, D. D. (1998). *Naive (Bayes) at forty: The independence assumption in information retrieval*. https://doi.org/10.1007/BFb0026666

Lin, J., & Ryaboy, D. (2013). Scaling big data mining infrastructure. *ACM SIGKDD Explorations Newsletter, 14*(2), 6. https://doi.org/10.1145/2481244.2481247

Martínez-Cámara, E., Martín-Valdivia, M., Perea-Ortega, J., & López, L. (2011). Técnicas de clasificación de opiniones aplicadas a un corpus en español. In *Procesamiento de Lenguaje Natural* (Vol. 47).

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications, 40*(10), 4241–4251. https://doi.org/10.1016/J.ESWA.2013.01.019

Nigam, K., Lafferty, J., & McCallum, A. (1999). Using maximum entropy for text classification. *IJCAI-99 Workshop on Machine Learning for Information Filtering.*

Priya, S., Sequeira, R., Chandra, J., & Dandapat, S. K. (2019). Where should one get news updates: Twitter or Reddit. *Online Social Networks and Media, 9,* 17–29. https://doi.org/10.1016/J.OSNEM.2018.11.001

Saxena, A., Gadhiya, S. (2014). A Survey on frequent pattern mining methods-Apriori, Eclat, FP growth. *International Journal of Engineering Development and Research (IJEDR)*, *2*(1). Retrieved from https://www.academia.edu/6435809/A_Survey_on_frequent_pattern_mining_methods-Apriori_Eclat_FP_growth.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*(1), 1–47. https://doi.org/10.1145/505282.505283

Sebastiani, F., & Esuli, A. (2006). Determining term subjectivity and term orientation for opinion mining. In *Procedings Proceedings of EACL-06, 11th Conference of the Europen Chapter of the Association for Computational Linguistics*, 193–200. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.60.8645.

StrateBi. (2013). *Nuevas Tendencias en Business Intelligence Del Big Data al Social Intelligence.* Retrieved from https://www.emotools.com/contents/articulos-y-blogs/stratebinuevas-tendencias-en-business-intelligence/.

Visvizi, A., Jussila, J., Lytras, M. D., Ijäs, M. (2019). Tweeting and mining OECD-related microcontent in the post-truth era: A cloud-based app. *Computers in Human Behavior*.https://doi.org/10.1016/J.CHB.2019.03.022

Fuzzy Control of Morelia's Manufacturing Companies' Innovation Capabilities

Gerardo G. Alfaro-Calderón, Artemisa Zaragoza-Ibarra, and Víctor G. Alfaro-García

Abstract Innovation has considered organizations' competitive advantage. The purpose of this study is to visualize the effects of the behavior of innovation level at innovation capability change. This work presents a fuzzy controller design using logic tables and a generalized ordered weighted averaging (GOWA) index to model the internal innovation phenomenon of manufacturing enterprises in Morelia, Michoacán, Mexico. The linguistic rules were programmed using a fuzzy design module in MATLAB software, and the controller was simulated using the Simulink tool. The results show that the values of at least two inputs have to change in order for the innovation value to change, and two static input values are enough to restrict the minimum innovation value. This paper presents an original methodology for visualizing the behavior of the internal innovation of companies based on control and fuzzy set theory that allows us to capture the dynamics of the phenomenon.

Keywords Fuzzy control system · IGOWA operator · Fuzzy analysis · Innovation capabilities

1 Introduction

Throughout history, innovation has been defined in different ways. Schumpeter (1934) conceives innovation as a dynamic concept of economic development, while the Oslo manual (OECD 2005) defines innovation as the implementation of a new or significant product, good or service.

G. G. Alfaro-Calderón · A. Zaragoza-Ibarra · V. G. Alfaro-García (✉)
Facultad de Contaduría y Ciencias Administrativas, Universidad Michoacana de San Nicolás de Hidalgo, 58030, Morelia, Mexico
e-mail: victor.alfaro@umich.mx

G. G. Alfaro-Calderón
e-mail: ggalfaroc@gmail.com

A. Zaragoza-Ibarra
e-mail: artemisazaragoza@hotmail.com

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_6

Currently, evidence suggests that there is a dependency between enterprises' competitive success and innovation management processes. However, how to quantify and evaluate innovation and its impacts is too complex for many companies at present (Adams et al. 2006).

Worldwide, Mexico is ranked 56th out of 124 countries in the Global Innovation Index. With 72% efficiency, Mexico rose two places from its previous rating and has *"human capital and innovation"* as a strength and *"business sophistication"* as a weakness (Cornell University, INSEAD, and WIPO 2018). Michoacán, on the other hand, ranks 28th out of 32 states in the National Innovation Index with an efficiency of 0.90%. Morelia has an efficiency of 0.58% and is ranked 17th in the Medium City Innovation Index (Venture Institute, CONACYT, and Secretaría de Economía 2013).

This study presents a fuzzy control model that allows one to observe the innovation level behavior of companies in Morelia when there is a change in their innovation capabilities. The preliminaries section describes the roles of enterprises' innovation capabilities in generating internal innovation. In the methodology section, OWA's results were used to build a logic table, in order to acknowledge the change on innovation capabilities. The input and output ranges were defined, and the membership functions were constructed using the fuzzy module designer on MATLAB. Using the logic table as a baseline, a set of IF–THEN-type rules were established and programmed on the same module, which conclude the fuzzy control system design. To work with the controller designed, the inputs were settled at the medium possible value, to begin the analysis. The input values were changed synchrony incrementally, and the behavior of the output was observed. The results section of this paper presents the analysis of the innovation behavior when companies' innovation capabilities change. Finally, the concluding section summarizes the findings of the research and propose future research lines.

2 Preliminary

2.1 Innovation Capabilities

Innovation capabilities have taken an important place among enterprises' strategies. Knight and Cavusgil (2004) are positively linked to their development (Calantone 2002). Previous work shows that innovation allows enterprises to survive in a volatile environment (Calantone 2002).

An interesting approach to assess innovation management measurement is presented in Adams et al. (2006) here, the authors compare other authors' work on the subject, conciliated their opinions and design an enterprise innovation capability measurement model. Seven variables are included in this model: *inputs,* internal drivers which are the entry systems and tools for innovation process which provide a competitive advantage; *knowledge management which* includes the management of explicit and implicit knowledge within organizations and the process of collection

and use of such information; *innovation strategy, which* helps to reduce processes inefficiencies and is linked to key business objectives, leadership and pro-activity in the commitment to innovation; *organization and culture,* which are used to know the intensity with which companies maintain their organizational structure aligned with their project management processes; *portfolio management* considered a determinant key of competitive advantage; *project management* is commonly measured in terms of cost, duration and return on investment; *and commercialization,* external drivers measure the intensity which company launches its products to the market (Alfaro-García et al. 2017).

Based on that new method, Alfaro-García et al. (2017) performed an assessment of the innovation capabilities of the enterprises in Morelia, Michoacan, Mexico. This study focused on the manufacturing enterprises of the region and was executed with fuzzy methods and an IGOWA operator. Alfaro's work presented a methodological structure for an innovation management measure using a fuzzy approach. The principal advantage of this operator is that the information aggregation considers the decision makers' highly complex attitudes, which is truly useful for innovation management and has been used as a baseline for the present work.

2.2 *Fuzzy Control Systems*

A system is *"a complex collection of highly related elements that accomplish a specific objective"* (Valdivia-Miranda 2012) with inputs that can be manipulated and outputs that can be observed. Those systems are generally affected by external and internal factors that cannot be manipulated, which are called *disturbances*. A control system manipulates the outputs using one rule or a set of rules, even the occurrence of disturbances (Valdivia-Miranda 2012).

Over time, fuzzy control has become quite an active area in the research and application of fuzzy set theory (Chuen Chien 1990). Mamdani, who was motivated by Zadeh's articles, is a pioneer of the discipline. Some applications of fuzzy set theory are qualitative phenomena modeling, pattern recognition, information processing, decision making, management, finance, among others. Particularly, fuzzy control has become one of the most successful applications due to its effectiveness in nonlinear complex system control (Feng 2006).

A characteristic of fuzzy control is that it incorporates a knowledge-based system (KBS), which allows for increased reliability, robustness and performance by incorporating knowledge of phenomena that cannot be included in an analytic model or an algorithm (Driankov et al. 1996).

Unlike a classic controller, a fuzzy control is built using a set of symbolic IF–THEN-type rules that are compiled using basic numeric objects and algorithms as true tables, interpolations, comparators and others, which are then used as the baselines to the controller's speed. Fuzzy control systems are thus heuristic forms that define nonlinear systems and are based on tables (Driankov et al. 1996).

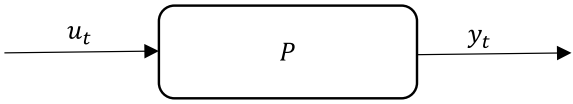


Fig. 1 Basic fuzzy control system diagram. Self-made figure

Identical to a conventional system, a fuzzy system is formed by a plant module (P) that contains the process or phenomenon that is to be represented. The inputs u_t are the independent variables that are required for the plant to operate, and the outputs y_t are the dependent variables that will be obtained once the inputs are processed by the plant (Hooda and Raich 2017) (Fig. 1).

In this controller, the algorithm is the linguistic rule set interpreter, and the decision will be based on fuzzy set theory (Mamdani et al. 1974).

3 Methodology

We base our methodology on Alfaro-García et al.'s results from their article *"On Ordered Weighted Logarithmic Averaging Operators and Distance Measures"* (Alfaro-Garcia et al. 2016), in which a function that generates and establishes the weight of each innovation capability is obtained. According to it, a fuzzy control model that allows the observation of the behavior of innovation based on the changes of enterprises' innovative capabilities is built.

3.1 Inputs and Outputs

These controller inputs are defined using Adam's innovation capability model as follows (Adams et al. 2006):

1. Innovation strategy,
2. Knowledge management,
3. Project management,
4. Portfolio management,
5. Inside factors,
6. Organization and structure, and
7. Outside factors.

Innovation is the only output.

Alfaro-García et al. at their article *"A Fuzzy Methodology for Innovation Management Measurement"* (Alfaro-García et al. 2017) defined the OWA index of innovation capabilities for Morelia's manufacturing companies as is shown in Fig. 2.

This index shows the weight of each capability to generate innovation in the company. With this index, a logic table was built. Each capability was placed in columns. Then, all combinations of the presence or absence of capabilities was set.

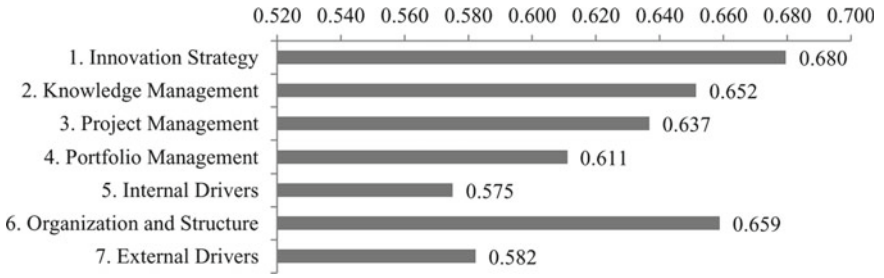


Fig. 2 IGOWA index for Morelia manufacturing companies (Alfaro-García et al. 2017)

A one was used to determine the presence of a capability and a zero to show the absence of it. To calculate the innovation value for each combination, the one or zero was multiplied by the weight of the correspondent capability and then added together, based on the way that innovation indexes are commonly calculated (Venture Institute, CONACYT, and Secretaría de Economía 2013). At the end, 128 combinations were set (Figs. 3 and 4).

After innovation values have been calculated, the rules get translated into linguistic labels. One becomes a "Hi", zero becomes a "Low", and the range of innovation gets divided into four intervals ("Null", "Low", "Medium" and "Hi"), between zero and 4.396, the minimum and maximum value obtained when all the innovation capabilities are absent or present, respectively (Fig. 5).

Once the linguistic rules have been settled, the plant of the fuzzy control is designed, using the fuzzy control module from MATLAB.

Following Mamdani's fuzzy controller design method (Mamdani et al. 1974), we characterize all variables as necessary because of fuzzy set theory. Hence, each one of the variables is going to be divided based on a linguistic form of the labeled levels. The number of levels will define the number of rules for the controller. Larger numbers of

P1	P2	P3	P4	P5	P6	P7
1	1	1	1	1	1	1
1	1	1	1	1	1	0
1	1	1	1	1	0	1
1	1	1	1	1	0	0

Fig. 3 Example of the logic table. Self-made table

P1	P2	P3	P4	P5	P6	P7	INNOVATION
0.68	0.652	0.637	0.611	0.575	0.659	0.582	4.396
0.68	0.652	0.637	0.611	0.575	0.659	0	3.814
0.68	0.652	0.637	0.611	0.575	0	0.582	3.737
0.68	0.652	0.637	0.611	0.575	0	0	3.155

Fig. 4 Example of the table set to calculate the innovation final values. Self-made figure

P2	P3	P4	P5	P6	P7	Innovation
Hi	Hi	Hi	Hi	Hi	Hi	Hi
Hi	Hi	Hi	Hi	Hi	Low	Hi
Hi	Hi	Hi	Hi	Low	Hi	Hi
Hi	Hi	Hi	Hi	Low	Low	Hi
Hi	Hi	Hi	Low	Hi	Hi	Hi

Fig. 5 Example of the logic table with linguistic labels. Self-made figure

rules imply a higher resolution, but the controller's robustness also increases, which will result in a slower controller that is useless for dynamic phenomena. Under these restrictions, each one of these variables was set to the levels of "Hi" and "Low". Each level was defined by a membership function with a 0–1 range (Fig. 6).

The output (innovation) is divided into four levels:

1. Null,
2. Low,
3. Medium, and
4. Hi.

The different levels are shown in Fig. 7.

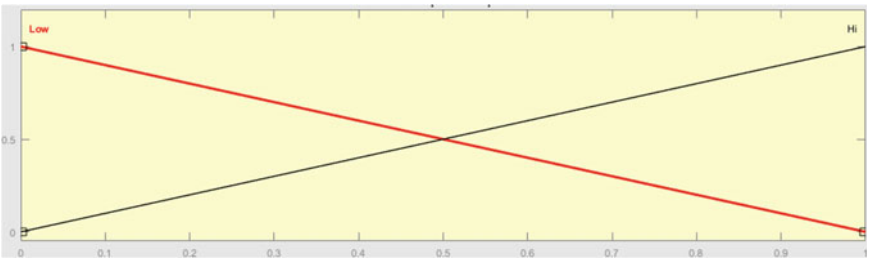


Fig. 6 Input structure. Self-made figure

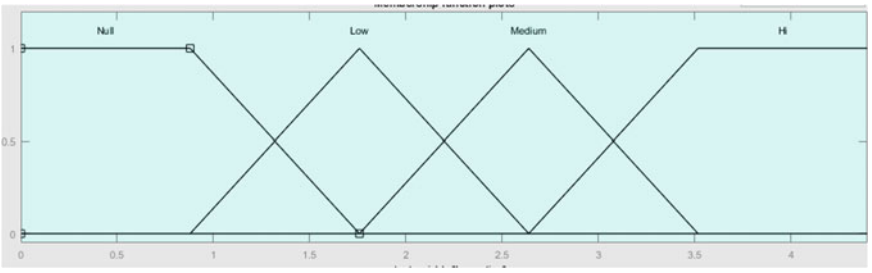


Fig. 7 Output structure. Self-made figure

Fig. 8 Example of the input area calculation. Figure from the MATLAB fuzzy control module

Fig. 9 Example of the output area calculation. Figure from MATLAB fuzzy control module

3.2 Range Definition

Once the linguistic labels have been settled, the plant of the fuzzy control is designed, using the fuzzy control module from MATLAB to set the rules of operation. These are IF–THEN-type rules that indicate to the controller which has to be the output, if one of the known conditions occurs. For example:

If (InnovationStrategy is Hi) and (KnowledgeManagement is Hi) and (Project-Management is Hi) and (PortafolioManagement is Hi) and (InternalDrivers is Hi) and (OrganizationAndStructure is Hi) and (ExternalDrivers is Hi), then (Innovation is Hi).

We use 128 total rules because of the number of combinations that define the innovation level. Then, for the unknown conditions according to Mamdani's method, the input value on each capability gives the area under the curve of the membership function, and as all the rules used were built with an AND operator, the resulting area for each rule is the intersection of the seven capabilities' areas. For the final result, OR operator has been applied to blend all the resulting areas and the center of area of the final figure has been calculated (Mamdani et al. 1974) (Fig. 8).

At the end, the control plant has been added at the simulation software "Simulink". The inputs have been set on 0.5 as a starting point, to give the output a chance to move above and under the set value (Fig. 9). The simulation allows us to observe the innovation-level changes that are caused by the observed capability changes over continuous time (Fig. 10).

4 Results

For this work, a large amount of testing was conducted that allowed us to observe the behavior of the innovation level due to the changes in the capabilities. In the first test, medium values were assigned to six variables, and an oscillatory value was assigned to one of the variables. Regardless of the variable which changes, the innovation level was a constant output value (Fig. 11).

In the second test, two inputs are altered, and five inputs maintain medium values. A fluctuation of the innovation value over a constant value is observed, and it is directly proportional to the change in the inputs. The minimum innovation value is

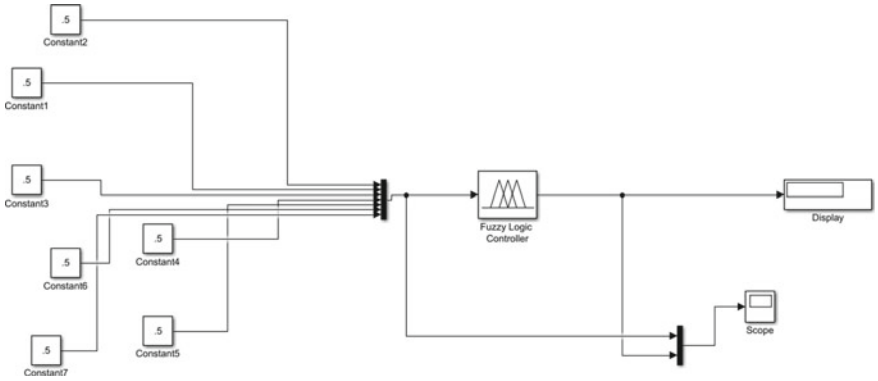


Fig. 10 Innovation capability fuzzy controller. Self-made figure

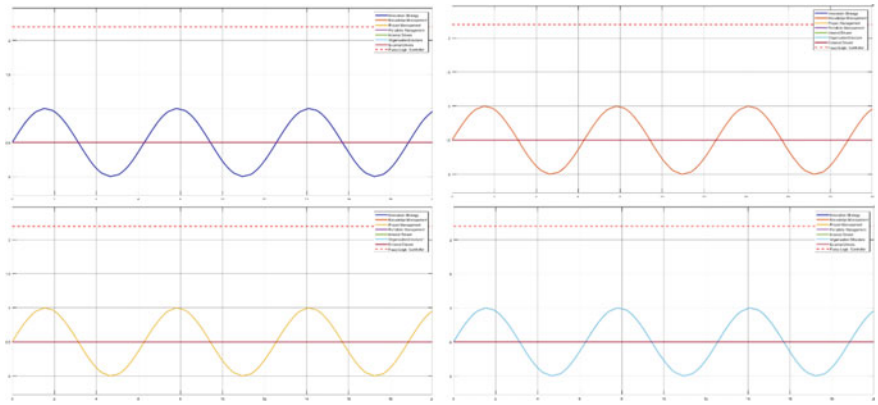


Fig. 11 Only one oscillating input. The innovation value is represented as a dotted line. Self-made figure

set by the constant input value, and when the input changes are lower, the innovation value is not affected. A change is only observed when the input values are higher than the static inputs (Fig. 12).

The next test was conducted using three oscillating inputs. In this case, innovation follows inputs and the lowest value is not set by the static variables. The results with four oscillating variables are similar (Fig. 13).

When five inputs are oscillating, the innovation value still behaves the same as the input values. However, a deformation in the output wave that is observed implies that the static input variables are influencing. The maximum innovation values that are reached are higher than those of the previous tests. When the six input variables are oscillating, we observe how the lowest value of innovation can be achieved. The restrictive static variable effect remains and increases as the number of static inputs increases (Fig. 14).

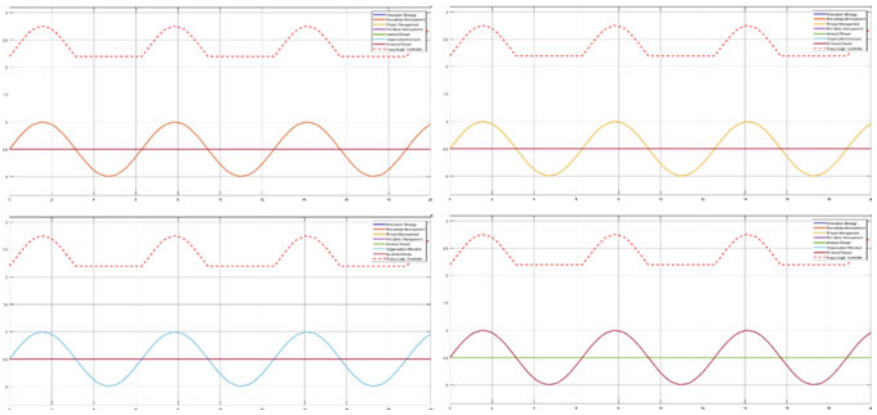


Fig. 12 Two oscillating inputs. The innovation value is shown as a dotted line. Self-made figure

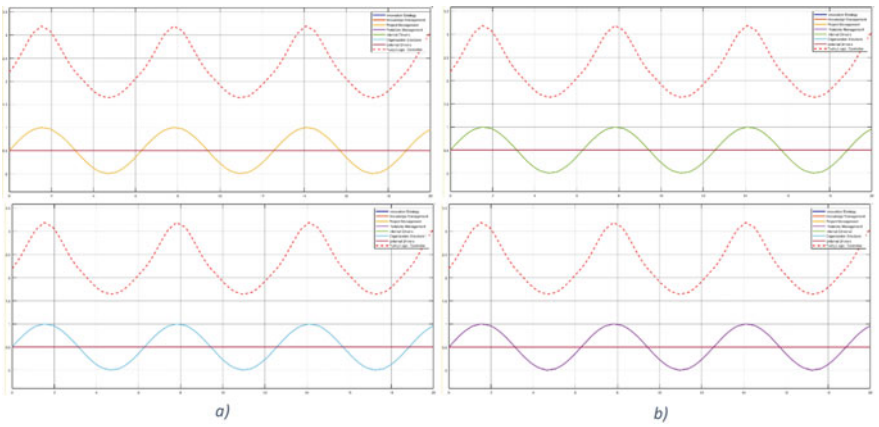


Fig. 13 **a** Three oscillating inputs. **b** Four oscillating inputs. The innovation value is shown as a dotted line. Self-made figure

Finally, the test is conducted with all input values changing. The innovation output value now completely follows the input values, and its fluctuation achieves the highest and lowest values that are originally established by the output design (Fig. 15).

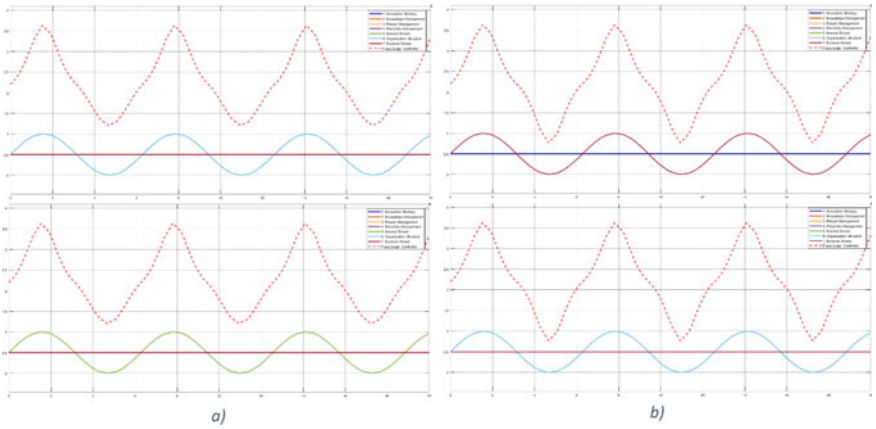


Fig. 14 **a** Five oscillating inputs. **b** Six oscillating inputs. The innovation value is shown as a dotted line. Self-made figure

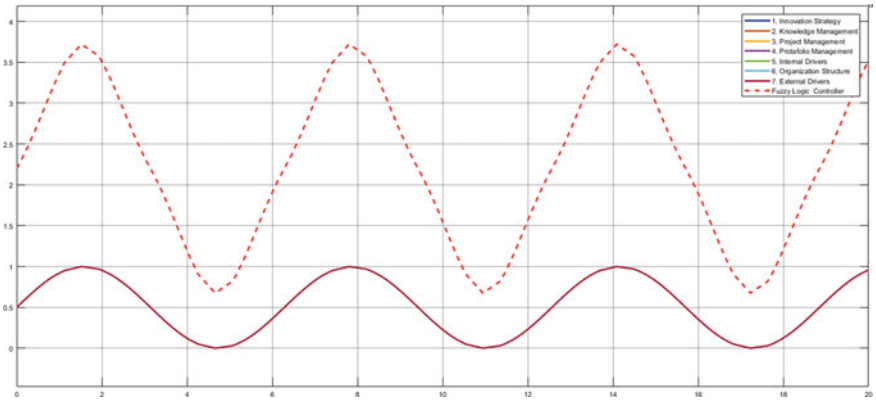


Fig. 15 Seven oscillating inputs. The innovation value is shown as a dotted line. Self-made figure

5 Conclusions

This investigation presents a fuzzy controller system for analyzing the innovation level inside enterprises according to the changes in the innovation capabilities. Using Alfaro's OWA index as a baseline, a logic table was constructed that defines the model's behavior.

An open-loop controller system was built that allowed the visualization of the changes in the internal innovation value of the enterprise according to the changes in their innovation capabilities. The innovation capabilities are synchronously and incrementally placed in order to observe the effect of enterprise innovation.

The results support the necessity of at least two capability changes in order to change the enterprises' innovation level. It is also shown that a minimum innovation level is defined by the values of two static variables inside the controller. Finally, it is observed that the maximum and minimum innovation values are only possible to achieve when all capabilities reach their maximum and minimum values, respectively.

The limitation of using a fuzzy control to study this phenomenon is that when the number of linguistic labels increases, also the number of rules, then it complicates the programming and slows the response of the controller, which makes it less useful for dynamic phenomena.

The union of the three themes, OWA's, control and innovation capabilities, has allowed to model innovation inside the companies with techniques that includes the opinion of the decision makers and visualized the possible behavior true time when a change in innovation has been made, different from the actual works where the model portrays only a moment in time.

Future investigation requires an analysis of the changes in the innovation level as innovation capabilities nonsynchronically change and even follow simultaneous inverse changes.

The fuzzy controller design was based on fuzzy logic because this tool allows one to model and describe phenomena with insufficient information to use a classic logic equation.

References

Adams, R., Bessant, J., & Phelps, R. (2006). Innovation management measurement: a review. *International Journal of Management Reviews, 8*(1), 21–47.https://doi.org/10.1111/j.1468-2370.2006.00119.x

Alfaro-Garcia, V. G., Gil-lafuente, A. M., & Alfaro-Calderón, G. G. (2016). Quantification of the Incidence in Innovation Capacities by the Promotion of Specialized Economic Sectors. In ASME (Ed.), *Lectures on Modelling and Simulation* (pp. 1–10). Santiago, Chile: ASME

Alfaro-García, V. G., Gil-Lafuente, A. M., & Alfaro Calderón, G. G. (2017). A fuzzy methodology for innovation management measurement. *Kybernetes*, *46*(1), 50–66.https://doi.org/10.1108/K-06-2016-0153

Calantone, R. J., Cavusgil, S. T., & Zhao, Y. (2002). Learning orientation, firm innovation capability, and firm performance. *Industrial Marketing Management, 31*(6), 515–524.

Chuen Chien, L. (1990). Fuzzy logic in control systems: Fuzzy logic controller. I. *IEEE Transactions on Systems, Man, and Cybernetics*, *20*(2), 404–418. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=52551&isnumber=1891%0A

Cornell University, INSEAD, & WIPO. (2018). *The global innovation index 2018: Energizing the world with innovation. Global innovation index 2018: Energizing the world with innovation*

Driankov, D., Hellendoorn, H., & Reinfrank, M. (1996). *An introduction to fuzzy control*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-03284-8

Feng, G. (2006). A survey on analysis and design of model-based fuzzy control systems. *IEEE Transactions on Fuzzy Systems, 14*(5), 676–697. https://doi.org/10.1109/TFUZZ.2006.883415

Hooda, D. S., & Raich, V. (2017). *Fuzzy logic models and fuzzy control: an introduction.* Oxford, United Kingdom: Alpha Science International Ltd.

Knight, G. A., & Cavusgil, S. T. (2004). Innovation, organizational capabilities, and the born-global firm. *Journal of International Business Studies, 35*(2), 124–141. https://doi.org/10.1057/palgrave.jibs.8400071

Mamdani, E. H., Eng, M. S., Ph, D., & Mem, I. E. E. E. (1974). Control and science application of fuzzy algorithms for control of simple dynamic plant, *121*(12), 1585–1588

OECD. (2005). Oslo manual: guidelines for collecting and interpreting innovation data, 3rd ed. *The Measurement of Scientific and Technological Activities*. Paris: OECD Publishing.

Schumpeter, J. A. (1934). *The theory of economic development*. Cambridge, MA: Harvard University Press

Valdivia-Miranda, C. (2012). *Sistemas de control continuos y discretos.* (M. J. López Raso, Ed.) (1a ed.). Madrid, España: Paraninfo

Venture Institute, CONACYT, & Secretaría de Economía. (2013). *Índice Nacional De Innovación*

Evaluation Scale of the Development and Quality Dimension in Software Development with an Exploratory Factorial Analysis

Roberto Carlos Valdés-Hernández, José Luis Arcos-Vega, and Ernesto León-Castro

Abstract Nowadays, the software environment has reached an unthinkable development in the world; companies in this area with the objective of competing aim to sustain software development and quality, influencing these two extents in the profitability of small and medium companies. Small- and medium-sized enterprises (SMEs) dedicated to software development are present in the software industry in Mexico and other countries, which is why they seek to improve their software development processes in order to offer quality services and products. This study has as main goal to evaluate the factorial structure and reliability of two extents: software quality and development, starting from measure and the factorial analysis of SMEs concept in the software area.

Keywords Factorial analysis · Small- and medium-sized enterprises (SMEs) · Competitiveness · Development · Software quality

1 Introduction

Nowadays, companies seek to maintain themselves in an increasingly competitive world, and in an economic environment known for its globalization and internationalization of economies, in a dynamic environment produced by continuous changes

R. C. Valdés-Hernández (✉)
Universidad Autónoma de Baja California, Mexicali, Mexico
e-mail: carlos.valdes@uabc.edu.mx

J. L. Arcos-Vega
Universidad Politécnica de Baja California, Mexicali, Mexico
e-mail: jlarcosv@upbc.edu.mx

E. León-Castro
Universidad Católica de la Santísima Concepción, Av. Alonso de Ribera 2850, #37150 Concepción, Chile
e-mail: eleon@delasalle.edu.mx

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_7

that impact in companies' management, having the need to establish comprehensive management diagrams in order to improve processes, considering the needs of employees who develop software and classifying according to the hierarchy model (López et al. 2012; Taormina and Gao 2013). In Mexico, SME's represent 95% of the total companies (Ortiz and Arredondo 2018), by having in mind companies of small magnitude; it is frequently considered that this is because of their incapacity to develop themselves and puts them in a disadvantage against large companies. During the first year of companies, approximately 50% of them perish, 25% during the second year, and 15% before reaching the fifth year, with only 10% surviving and growing as small companies (INEGI 2014). It is important to reflect on the fact that, in the software development process, the use of techniques and models from SMEs is a main core, since these are aimed to assure software quality (Strub 2015). SMEs have created their own work line (Sánchez et al. 2017) regarding quality management, and in other words, they work over the software production process in order to provide quality in their products and, thus, satisfying their clients' needs, which is why using certifications that allow to prove the processes efficiency, seeking to improve the ways to achieve objectives (Esterkin and Pons 2017; Montoya 2016). In this sense, software quality can be measured in terms of imperfection density (Fenton and Bieman 2014), which is why companies dedicated to software development (Puello et al. 2016) must apply control methods that allow the developed product to comply with the regulations and quality standards, which must satisfy the clients' expectations. Considering the aforementioned, it is important to assure the software product quality emphasizing human talent (Chavarría et al. 2016), considering that people must commit to the quality of the planned software, and being this one of the components that plays a main role nowadays in the industry. It also generates an added value, which must meet the highest quality standards (Acuña et al. 2015; Rojas et al. 2015).

Therefore, the information technology and software industries in Mexico (Ortiz and Arredondo 2018) are constituted by 2134 companies, of which 91% are SMEs and the remaining 9% are large companies, which is why the software industry in Mexico is defined by the entrance of companies with a high acquisitive power, with the possibility of innovating their development processes in order to generate utilities, considering competitive environment, innovation capacity, clients' needs, among others (Rocha and Freixo 2015). Considering the aforementioned, in Mexico, the companies dedicated to developing software are few, being SMEs in a large number (Guadarrama and Casalet 2012). A big step to success for a small company is to know and understand the motives that originate failure, establishing strategies to support, minimize, eliminate, or avoid them, achieving growth, development, and maturing (Nuño 2012). In this sense, the company's owner has the responsibility of creating the company and the development it can have, starting from its own capacities for an adequate management, taking the company to be competitive or be beaten by other companies, or seeking the company's productiveness in order to have a sustainable development, allowing it to obtaining a higher development compared to others (Acuña et al. 2015; Mendoza 2008). In addition, SMEs destine their utilities to covering basic needs for sustainability (López et al. 2012), which

are in time insufficient to invest in innovation or new working methods in order to make development processes more efficient (Mendoza 2008; Pressman 2010), since, in Mexico, software industry is in position to compete worldwide, having geographic location and technology development as advantages (Bañales and Adam 2007).

2 Method Description

The present research shows a factorial analysis, which was selected in order to look into an organizational phenomenon starting from the statistical analysis via variable correlation. The information technologies industry was selected, the number of companies was obtained through the information of the Economic Census created by the National Institute of Statistics and Geography (INEGI 2014). Population is formed of SMEs established in the state of Baja California, specifically in the cities of Mexicali, Ensenada, and Tijuana. The sample size was determined considering an error margin lower than 0.03 points with a 95% trust level.

3 Instruments

The instrument applied consists of a 15 questions survey, divided in three sections: the first consisted of seven questions regarding software development process carried out by the company, the second consisted of five questions about software quality, and the third one consisted of three questions about companies' competitiveness. The Likert classification scale was used, with the purpose of characterizing development, quality, and competitiveness of the SME dedicated to offer software products and services in the cities of Ensenada, Mexicali, and Tijuana in the state of Baja California. Likewise, the results obtained correspond to 52 SMEs dedicated to software development, which allowed the identification of the relation between development, quality, and competitiveness, seeking to contribute to the continuous improvement of the software development process.

4 Procedure

Data obtaining. The data collection technique was carried out through an instrument (survey) sent via e-mail, and in person in some cases, to the SMEs manager during June to September of 2016. Finally, a sample of 52 companies dedicated to software development was obtained. This sector was selected for its high growth potential as a development sector, for the state of Baja California.

Data analysis process. In order to evaluate the scale factorial structure, a factorial exploratory analysis (AFE) was used as extraction method for main axis with varimax

rotation. Afterwards, the internal consistency of the identified factors was evaluated, using the Cronbach alpha reliability coefficient. The data analysis was made with the statistical package SPSS version 21.

5 Results

The TI sector was selected, specifically small and medium companies dedicated to developing software. The company number was obtained from the Economic Census information elaborated by the National Institute of Statistics and Geography (INEGI 2014). The information presented in Table 1 indicates the classification according to the number of employees the company has, which are mainly micro type (0–10) with 59.6%, followed by 34.6% that said they were a small company (11–50), and with lower percentage levels of 5.7% for the ones that informed they are a medium company.

In order to determine the pertinence of the factorial exploratory analysis to estimate the instrument's validity, the statistical sample adaptation of Kayser-Meyer-Olkin (KMO) was calculated, which resulted in a value of 0,879, and the Barlett sphericity test that resulted statistically significant, 375.664; $p < 0.001$, showing both the pertinence of this analysis. In this sense, where the Cronbach alpha average obtained was 0.854, the result of the Cronbach reliability analysis through the SPSS software version 21 can be observed in Table 2. Based on the results, it is concluded that the instrument is reliable. Regarding fall contrast criteria or scree plot, the sedimentation graphic indicated the presence of three factors, Fig. 1. Since it presents significant factorial charges in all items, a significant charge is understood as the one that presents a value superior to 0.30, which is the minimum suggested value as threshold for accepting a factorial charge as suitable (Hair 2005).

With the objective of analyzing the instrument's reliability, a bivariate Pearson correlation analysis was carried out, between development items (1, 2, 4, 5, 10, 13,

Table 1 Number of the employees of the company

Classification *n* percentage		
Micro (0–10)	31	59.6
Small (11–50)	18	34.6
Medium (51–250)	3	5.7
Total	52	100

Table 2 Abstract of the reliability analysis (Cronbach alpha) for each aspect

Aspect	Cronbach alpha
Software development	0.884
Software quality	0.849
Competitiveness	0.831

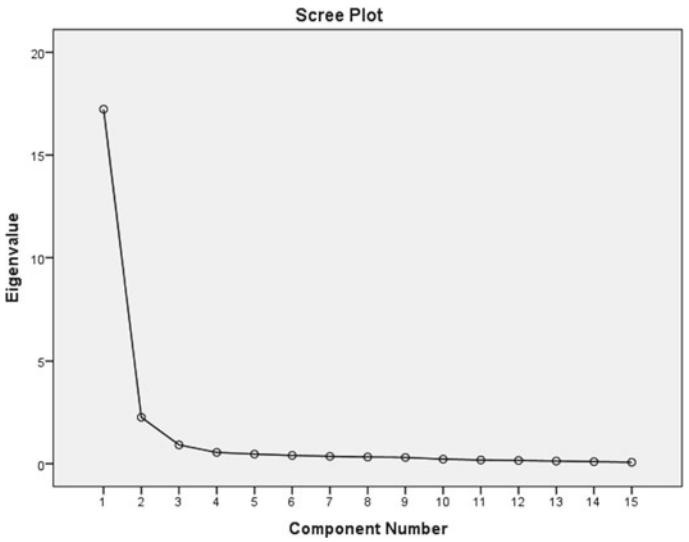


Fig. 1 Sedimentation graphic for the development, software quality, and competitiveness. Considering 15 items

and 15), quality (3, 6, 7, 11, 12), and competitiveness (8, 9, and 14). In this sense, the calculation was made for each analyzed factor. The result can be observed in Table 2, where the Cronbach alpha average was 0.854. The result of the Cronbach reliability analysis can be observed through SPSS software version 21. Based on these results, it is concluded that the instrument is reliable. Information in Table 3 shows the communalities after and before the extraction; the components analysis works in the initial case in which all variations are common; therefore, before extraction, the communalities are all 1. The extraction reflects the common variability in data structures. Hence, for instance, it can be said that 85.7% of variability associated to question 8, being the highest percentage, is a common or shared variance. The variance quantity in each variable that can be explained by retained factors is represented in communalities after extraction.

In this sense, the data on Table 4 show the values associated with each lineal component (factor) before extraction, after the extraction, and after rotation. Before extraction, 15 components were identified inside the data combination. The values associated to each factor represent the variance explained by this particular lineal component, where the value is also shown in terms of the variance percentage explained, and thus, factor 1 explains 17.231% of the total. The values of factors after rotation are shown, where rotation has the effect of optimizing the factorial structure, and a consequence for this data is that relative importance of the four factors is even. Before the rotation, factor 1 represented a variance considerably higher than the three remaining ones (72.771% compared to 9.543, 3.871, 2.315, and 1.976%), nonetheless, after the extraction only it represents 13.117% of variance, compared to 3.587, 2.772, 2.147, and 1.578%, respectively. In regards to these results, the composition of

Table 3 Communalities

		Initial	Extraction
1	The analysis and specification of requirements allow knowing the elements needed to define the software project	1.000	0.723
2	The company has a requirement analysis suitable for the software development process	1.000	0.687
3	The company uses a quality model for software requirements	1.000	0.708
4	The company has measures of the software development process	1.000	0.832
5	Building a quality and efficient system, with minimum error, is the objective for software development	1.000	0.702
6	Thinks it is necessary to be certified in some regulation regarding software quality	1.000	0.653
7	Thinks that a software quality certification is useful and beneficial	1.000	0.595
8	Due to the high competitiveness of the sector, it is necessary to offer added value services to clients, which can differentiate us from the competition	1.000	0.857
9	The company considers suppliers and actual competition in the sector in the analysis of the competitive environments to clients	1.000	0.765
10	The company evaluates the performance of the developed software program, detecting defects, and proceeding to elimination before delivering	1.000	0.692
11	The quality of development processes is considered as a success factor of their company	1.000	0.759
12	The products and services' continuous improvement responds in great extent to the needs and expectations detected in clients	1.000	0.839
13	The company has registries generated by processes or procedures	1.000	0.780
14	In their company, flexibility for developing new products or services (innovation capacities) is considered a competitive advantage	1.000	0.690
15	The company formally documents their software development processes	1.000	0.765

Extraction Method Main component analysis

five factors solution was analyzed for these 15 items, showing that this time all items presented factorial charges on the threshold. An evaluation of the composition of six factors solution was made using the extraction method of main axis and varimax crosswise rotation. In this solution, all items showed charges over 0.30. In Table 5, the factor to which every item was assigned is indicated in red.

The five factors identified were configured in the following way:

Factor I: Formed by items 1, 2, 3, 10, 11, 13, and 15 (ordered from high to low charge), making reference to the relation of quality models used by the company in software development processes. After analyzing the internal consistency of the instrument, a Cronbach alpha coefficient of $\alpha = 0.89$ was obtained.

Factor II: It contains items 6, 7, and 14, which refer to the perception had by SMEs managers regarding the importance of being certified in a quality regulation in order

Table 4 Total variance explained

Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	17.231	72.771	72.771	17.231	72.771	72.771	3.106	13.117	13.117
2	2.260	9.543	82.314	2.260	9.543	82.314	0.849	3.587	16.704
3	0.917	3.871	86.185	0.917	3.871	86.185	0.656	2.772	19.476
4	0.548	2.315	88.501	0.548	2.315	88.501	0.436	2.147	21.623
5	0.468	1.976	90.477	0.468	1.976	90.477	0.676	1.578	23.201
6	0.404	1.705	92.182						
7	0.357	1.508	93.689						
8	0.329	1.390	95.079						
9	0.305	1.288	96.367						
10	0.221	0.931	97.299						
11	0.183	0.772	98.071						
12	0.156	0.660	98.731						
13	0.130	0.551	99.281						
14	0.101	0.425	99.706						
15	0.070	0.294	100.000						

Extraction Method Principal component analysis

Table 5 Matrix of development, quality, and competitiveness configuration (15 items) obtained through the main axis analysis with varimax rotation

#Item	Statement	I	II	III	IV	V
2	The company has a requirement analysis suitable for the software development process	*0.845*	0.168	0.273	−0.172	−0.230
3	The company uses a quality model for software requirements	*0.778*	0.237	−0.164	0.078	−0.176
15	The company formally documents their software development processes	*0.756*	−0.296	0.036	−0.391	0.147
1	The analysis and specification of requirements allow knowing the elements needed to define the software project	*0.735*	−0.173	0.051	0.370	0.033
13	The company has registries generated by processes or procedures	*0.706*	0.290	0.206	0.154	0.069
11	The quality of development processes is considered as a success factor of their company	*0.642*	−0.232	0.163	0.043	−0.073
6	Thinks that it is necessary to be certified in some regulation regarding software quality	0.175	*0.792*	0.285	−0.163	−0.181
7	Thinks that a software quality certification is useful and beneficial	0.142	*0.743*	−0.255	0.192	0.154
14	In their company, flexibility for developing new products or services (innovation capacities) is considered a competitive advantage	0.387	*0.652*	0.176	0.222	0.071
9	The company considers suppliers and actual competition in the sector in the analysis of the competitive environments to clients	0.186	0.159	*0.865*	−0.030	0.194
8	Due to the high competitiveness of the sector, it is necessary to offer added value services to clients, which can differentiate us from the competition	0.261	−0.184	*0.775*	0.219	−0.045
5	Building a quality and efficient system, with minimum error, is the objective for software development	0.152	−0.149	0.103	*0.891*	0.097
12	The products and services' continuous improvement responds in great extent to the needs and expectations detected in clients	0.072	−0.131	0.239	−0.078	*0.877*
4	The company has measures of the software development process	0.160	−0.380	0.474	0.092	*0.680*
10	The company evaluates the performance of the developed software program, detecting defects, and proceeding to elimination before delivering	0.166	0.196	−0.152	0.195	*0.879*

Rotation Method Varimax with Kaiser normalization

to keep innovating to maintain themselves in a competitive market. After analyzing its reliability, a Cronbach alpha coefficient of $\alpha = 0.87$ was obtained.

Factor III: Includes items 8 and 9, alluding to the perception had by companies' managers about the need for the company to implement technology in their processes to keep offering added value products and services to clients. Their reliability was $\alpha = 0.84$.

Factor IV: Constituted by item 5, that makes reference to the importance of developers' skills to communicate with the users in the software development process, denominating the factor in this way, and finding a Cronbach alpha coefficient of $\alpha = 0.83$.

Factor V: Formed by items 4 and 12, related to the needs and expectations of clients and making a perfect software product, where reliability for this factor was $\alpha = 0.86$.

6 Discussion

After analyzing the exploratory factorial of the evaluation scale of software development and quality and its incidence in SME competitiveness, five factors were identified, in which factor I refers to quality models in the software development process, factor II refers to the certification in some quality regulation and its impact in the company's innovation, factor III makes reference to the implementation of technology in software development processes and, thus, to offer added value products and services, factor IV refers to the importance of developers' skills in the software development process, and factor V refers to the competitive environment of the company. In this sense, these factors explain why all companies dedicated to developing software show that the scale used in this research is a reliable instrument capable of obtaining information about the perception had by companies' owners or managers. It is concluded that this topic should be researched further, since software development in this industry impacts directly in the Mexico's northeastern economic region.

7 Conclusions

The results have confirmed that in SME dedicated to software development exists a good joint between quality and software development. There has been a small amount of innovation in processes, which implies there is low investment in technology regarding software development, being one of the main factors that impact on SME organizational development and competitiveness. Nonetheless, it does not affect the company's profitability significantly. In addition, evidence was obtained regarding the influence of quality in the software development process and how it impacts in quality assurance of such process, especially from a client's satisfaction perspective.

For future research, a different approach based on software process improvement want to be done, where selected organizations shall have the characteristics considered in the literature of this research, taking into account innovation as a fundamental aspect for software development. The result would increase the relation of high direction strategies involved in software development process based on innovation. Also, defining new aspects is to analyze parting from factorial analysis in order to measure the impact of competitiveness, as well as the company's profitability parting from sales yield, in addition to the return of resources in short term, being these important to measure the company's competitive success.

References

Acuña, et al. (2015). La vinculación Universidad-Industria como instrumento de mejora para la competitividad de las empresas de desarrollo de Software del NEA. JUI 2015, 9º Jornadas de Vinculación Universidad-Industria (pp. 31–36)

Bañales, D., & Adam, M. (2007). Factores críticos de éxito de la industria del software y su relación con la orientación estratégica de negocio: Un estudio empírico-exploratorio. *JISTEM-Journal of Information Systems and Technology Management, 4*(1), 47–70.

Chavarría, A., Oré, S., & Pastor, C. (2016). Aseguramiento de la Calidad en el Proceso de Desarrollo de Software utilizando CMMI, TSP y PSP. *RISTI-Revista Ibérica De Sistemas E Tecnologias De Informação, 20,* 62–77.

Esterkin, V., & Pons, C. (2017). Evaluación de calidad en el desarrollo de software dirigido por modelos. *Ingeniare. Revista Chilena De Ingeniería, 25*(3), 449–463.

Fenton, N., & Bieman, J. (2014). *Software metrics: A rigorous and practical approach.* CRC Press

Guadarrama, V. y Casalet, M. (2012). La industria electrónica y la evolución de la industria de software en dos regiones de México. En Comisión Económica para América Latina y el Caribe (CEPAL), Deslocalización de servicios y cadenas globales de valor: ¿Nuevos factores de cambios estructurales en América Latina y el Caribe? Santiago, Chile, 18–19 (2012). Disponible en. https://www.eclac.cl/comercio/tpl/contenidos/Industria_electronica_y_software_Guadarrama_Casaletl.pdf. Consulta 12 junio 2019

Hair, et al. (2005). *Multivariate data analysis* (6th ed.). New Jersey: Prentice Hall

INEGI. (2014). Instituto Nacional de Estadística, Geografía E informática. Censo Económico del Directorio Estadístico Nacional de Unidades Económicas (DENUE)

López, A., Contreras, R., & Espinosa, R. (2012). The impact of financial decisions and strategy on small business competitiveness. *Global Journal of Business Research, 6*(2), 93–103.

Mendoza, S. (2008). Supervivencia de la micro, pequeña y mediana empresa (MIPYME). *Revista Del Centro De Estudios Superiores Del Noreste, 1*(1), 36–48.

Montoya, J. (2016). Comparación de metodologías ágiles y procesos de desarrollo de software mediante un instrumento basado en CMMI. *Scientia et technica, 21*(2), 150–155.

Nuño, P. (2012). *Administración de pequeñas empresas.* México: Red Tercer Milenio.

Ortiz, C., & Arredondo, E. (2018). Competitividad de las empresas desarrolladoras de software en Morelia Michoacán. *Red Internacional de Investigadores en Competitividad, 8*(1)

Pressman. Ingeniería del Software (2010). Un enfoque práctico", McGrow-Hill 7ta edición, México

Puello, J., San Feliu, T., & Calvo-Manzano, J. (2016). Una aproximación basada en metamodelado del área de proceso de Validación del CMMI: Un caso de estudio. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, 17, 26–40

Rocha, Á, & Freixo, J. (2015). Information architecture for quality management support in hospitals. *Journal of Medical Systems, 39*(10), 1–11.

Rojas, M., Pino, F., y Martínez, J. (2015). Proceso de pruebas para pequeñas organizaciones desarrolladoras de software. *Facultad de Ingeniería*, *24*(39), 55–70

Sánchez, M., Amezcua, A., O'Connor, R., & Larrueca, X. (2017). A standard-based framework to integrate software work in small settings. *Computer Standards & Interfaces, 54*(Part 3), 162–175

Strub, A. (2015). Métricas estadísticas asociadas al proceso de desarrollo y mejora continua de software. In *Congreso Virtual sobre Tecnología, Educación y Sociedad*, Centro de Estudios e Investigaciones para el Desarrollo Docente. CENID A.C., vol. 1, No. 5

Taormina, R., & Gao, J. (2013). Maslow and the motivation hierarchy: Measuring satisfaction of the needs. *The American Journal of Psychology, 126*(2), 155–177.

Analysis of Business Growth in Mexico Using Weight of Evidence. Period: 2008–2017

Deyanira Bernal-Domínguez and Inés Fernando Vega-López

Abstract The growth of Mexican companies is measured by increasing sales, and some of their determinants are, according to the literature review, accounting and financial data such as total assets, financing alternatives, and profitability. We seek to know the weight of the evidence of each of the mentioned variables in order to identify if they are good predictors of growth. The applied methodology uses the weight of evidence and information value to rank variable importance. The most important variables are used to generate three different predictive models. Our results show that the main variables affecting growth are EBIT, consolidated net worth, and total assets. The best predictive model is artificial neural networks.

Keywords Profitability · Investments · Neural networks

1 Introduction

Business growth is explained by factors such as leadership, company culture, and business model strategy such as income, costs, investments, and financing, argue Amat and Lloret (2014). Juarez (2018) coincides and presents preliminary results of his research on the importance of total assets as determinants of business growth, but classifying them by economic relevance in each industry and according to their size. The cubic function is the one that best fits the industry-company-size combination.

Likewise, the growth in companies has been studied from different methodological approaches, both qualitative and quantitative. Just as there are different findings on definitions, explanatory variables, and business growth measures, other research is also oriented to the methodologies used. The different types of research strategies and empirical evidence applied in the business growth studies were presented by Muhos

D. Bernal-Domínguez (✉) · I. F. Vega-López
Universidad Autónoma de Sinaloa, Culiacán, Mexico
e-mail: berde@uas.edu.mx

I. F. Vega-López
e-mail: ifvega@uas.edu.mx

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_8

(2015). This author performs a meta-analysis in which he states that the most common type of research in the sample between the first studies and the central ones was the case study with a descriptive, longitudinal-retrospective approach, increasing the individual case with an explanatory and transversal exploration approach. In recent studies, the survey method has increased. For example, Benáček and Michalíková (2016) carried out a study with panel data analysis applying the method of estimation of square fixed effects, with econometric tests to know the growth factors of family businesses.

Machek and Machek (2014) propose and test a model of multiplicative form with a logarithmic decomposition of four indicators: productivity, labor intensity, capital intensity, and frequency of visits, to examine the individual contribution of each one in the total sales growth. The indicators used for the application of the model are calculated based on the quantitative variables that determine and measure the growth of companies, in other words, sales, employees, total assets, and the target market.

The indicators that measure business growth, according to Blázquez et al. (2006), state that those that stand out the most based on the revision of documents published from 1967 to 2004 are total assets, the number of employees, the volume of sales, the weight of sales, added market value, added economic value or added value, profitability, own funds, and multicriteria value. They affirm that the growth of the company must be related to the demand of the market, and for this, an analysis of the profitability must be made and that this justifies the investments made, that is to say, the profitability indexes must justify the increase of the productive capacity.

Among the indicators used the most in growth measurement, according to García-Manjón and Romero-Merino (2010), highlight the market value, the number of employees, sales, the value of production or added value; argue that there is a direct relationship between growth and investment because implementing new technologies, in research, in updating plants in manufacturing companies, significantly benefits production and increases in sales, resulting in profitable growth for the company. Without a doubt, the success of any company depends on its investment to expand its marketing territory. When investing in a project, the company expects to make a profit in the future and is not limited to that, rather it will ensure that your investment is profitable. Therefore, decisions to choose a suitable project are crucial for the achievement of financial objectives (Vecino et al. 2014).

Some coincidences can be observed with Weinzimmer et al. (1998), cited by Gielnik et al. (2017), since they assert that the growth measures of business are sales, employees, and assets; Profitability, return on assets and the value of the company are alternative indicators of growth, according to Achtenhagen et al. (2010), Delmar et al. (2003), Dobbs and Hamilton (2007), cited by Gielnik et al. (2017), who conclude that the age of small business managers negatively affects business growth through the opportunities approach. Likewise, business development understood as the search for growth opportunities is presented by Voeth et al. (2018), who propose four key dimensions: product, processes, markets, and business model. They claim that the main challenges are the shortage of human and financial resources.

In base to this, it is observed that the studies on business growth show that the factors that determine it are multivariable, highlighting access to private financing,

demand, technological improvements, government support, job offers, the macroeconomic environment, fixed investments, the age and size of the company, the gender issue, profitability, among others. Therefore, the hypothesis that guides this research is that the growth of sales depends on the increase in investments, financing, and profitability.

It should be noted that most of the studies are aimed at smaller companies, and in the case, this study is based on the information collected from 83 large companies listed on the Mexican Stock Exchange. The data of balance sheet accounts and results for the period 2008–2017 were downloaded from the economics software in November 2018.

In summary, the general research question is: Is the weight of the evidence of investment, financing, and profitability accounts good predictors of sales growth?

Specific questions:

Is the sales growth related to the increase in total assets, current assets, inventory, property, branch and equipment, total liabilities, current liabilities, consolidated equity, gross profit, earnings before interest and taxes (EBIT), profit net, consolidated profit, capital expenditures (Capex), invested capital, earnings before interest, taxes, depreciation and amortization (EBITDA)?

Is the increase in the debt and profitability ratios related to the sales growth?

What is the ideal predictive model for sales growth based on the results of the analysis of the weight of the evidence?

2 Methodology

Historically, techniques have been proposed to determine the importance of decision variables when estimating the value of an objective variable by means of regression techniques. In particular, the weight of evidence and the importance of information have been studied at least since the 1960s for that purpose (Good 1985). In essence, the weight of the evidence of a variable is the amount of information provided by knowing its value to infer the value of a dependent variable. Usually, the weight of the evidence is applied to discrete variables and is defined as the logarithm of the odds ratio that given a particular value of the independent variable, a specific event occurs in the target variable.

Therefore, the method for the development of this research is the weight model of the evidence. The financial information used to estimate the target variable for business growth was the accounting data collected from the annual financial reports of the issuers listed on the Mexican stock exchange in the period 2008–2017. We built our dataset with data downloaded using Economatica in November of 2018. Economatica is a, licensed, Web-based application storing financial information from companies in Brasil, Chile, Perú, Colombia, México, Spain, The United States, and Venezuela.

The research variables are the following:

Dependent variable is business growth which is measured by the increase in the accounting account called sales revenue.

Independent variables:

(1) Investments are equal to the accounts of total assets, current assets, inventory, property, branch, and equipment.
(2) Financing is equal to the accounts of total liabilities, current liabilities, consolidated net worth. Two indicators or ratios were also calculated to measure the indebtedness, with the formulas: Total liabilities/Total assets and Consolidated equity/Total liabilities.
(3) Profitability is equal to the net profit, EBIT, and EBITDA accounts. Likewise, the applied profitability ratios were: Net income/consolidated equity or stockholders' equity, EBIT/Total assets, EBIT/Consolidated equity, EBIT/Capex, EBIT/Capital employed, EBITDA/Consolidated equity, EBITDA/Total assets.

The growth rates were calculated as follows: $g = ((\text{data } 2009 - \text{data } 2008)/\text{data } 2008) * 100$ and so on until 2017.

The calculation of the Information Value (IV) was done with the statistical tool R (2018) and the woeBinning package (Eichenberg 2018). With the selected variables, predictive models were generated with a dependent variable called growth. This variable indicates whether a company grew or not with respect to the previous year. Growth is defined as true if the sales index is bigger than zero; otherwise, it is said that there is no growth.

The predictive models were constructed using the statistical tool R (2018). For the logistic regression and for the support vector machine models, the e1071 package was used (Meyer et al. 2018), whereas for the neural network model, the nnet package was also required (Venables and Ripley 2002). Finally, the ROCR package (Sing et al. 2005) was used to select the ideal predictive model.

ROC curves have been used for more than 40 years to evaluate the decision-making process (Rakotomamonjy 2004). These are curves that illustrate the behavior of a binary classification model and indicate the probability of correctly classifying the positive examples (class A) against the rate of incorrectly classifying the negative examples (class B), in other words, the performance of the model by comparing the false positive rate against the true positive rate. The false positive rate is calculated as the probability that the model indicates that a company will grow when it does not. On the other hand, the rate of true positives indicates the probability that the prediction of the model, on the growth of a company, is correct. Ideally, the receiver-operating characteristics (ROC) curve would be two straight lines that go from point (0,0)–(0,1) and from there to point (1,1). This would indicate that the model can make predictions with true positive rates equal to 100%, without incorporating false positives. That is to say, a model capable of not making mistakes in its predictions.

3 Analysis and Presentation of Results

From the dataset described above, we proceeded to calculate the Information Value of each of the index variables. Since the dataset has records marked NA (not available) that prevent the correct calculation of the IV, a first step was the elimination of all records with NAs. So before calculating the IV of each variable, all the NAs were eliminated. We proceeded this way for each variable independently. Variables showing a high correlation index were also eliminated; although some variables such as EBITDA have a high information value (0.237), they were not selected because they are correlated with others (in this case EBIT).

From this initial processing, it was determined that the variables with the highest information value were EBIT, consolidated net assets, and total assets. With values of 0.318, 0.194, and 0.180, respectively, it is indicated that the interval from 0.3 to 0.5 means strong and from 0.1 to 0.3 is medium. Once these three variables were selected, we proceeded to the detection and elimination of data with atypical values (outliers). Once the atypical data were eliminated, the three predictive models were constructed: logistic regression, neural networks, and support vector machines. They were constructed indicating that growth should be predicted (objective variable) using EBIT, consolidated net assets, and total assets as predictors (independent variables). Figure 1 shows the ROC curve, and it is observed that in red, the behavior of a classifier based on an artificial neural network is shown, in green, the logistic regression model, and in blue color, the model that uses support vector machines.

It is analyzed that the models based on logistic regression and support vector machines do not have a significant difference in their behavior. On the other hand, it is observed that the neural network model is clearly superior. It is necessary to

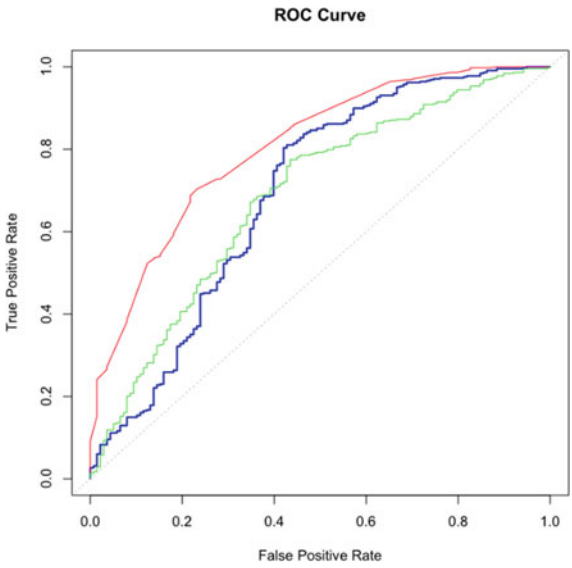


Fig. 1 ROC curve of the predictive models evaluated. *Source* Own elaboration

mention that the artificial neural network model has a hidden layer architecture with ten neurons. Naturally, the input layer of the network has three neurons, and the output layer has only one neuron (because it is a binary classification model).

The logistic regression model is, perhaps, the most widely used predictive model. This is because the simple linear combination of variables makes this model easy to understand. In our case, the linear model that we inferred from our dataset has the following expression.

$$y = 0.0034715 * \text{EBIT} + 0.0593914 * \text{Activo\_Total} + 0.0004158 * \text{PNC} + 0.1398631$$

Artificial neural networks are derived from the field of artificial intelligence. They are a network of interconnected nodes, where each node performs a simple computation. Each of these nodes is known as perceptron. Consequently, this kind of network is also known as multilayer perceptron. Each perceptron is a linear classifier. It computes, from its inputs, a value known as activation that is passed down to the nodes on the next layer of the network. When creating a predictive model based on a neural network, an architecture must be defined. This architecture dictates the number of nodes in the networks. As the name implies, perceptron nodes are grouped in layers. Nodes between consecutive layers are fully connected, but nodes in the same layer are not. If two nodes are connected, the activation value (output) of one of these nodes becomes the input for the following node. Each of the input values to a perceptron is associated to a weight that indicates the strength of the connection. The weights associated with each connection are not know beforehand; an algorithm known as backpropagation is used to infer the sets of weights from a (training) dataset. The goal of the algorithm is to adjust the weights in order to minimize the classification error with respect the training dataset. The following figure illustrates the architecture of the artificial neural network used in this work. While we experimented with fewer perceptrons in the hidden layer, using ten nodes gave us the best accuracy. We omit the connection weights on this figure for the sake of clarity. See Fig. 2.

Finally, it should be mentioned that the accuracy observed in these models, measured as the percentage of successes of each model when predicting growth, was the following: 76.62% for neural networks, 70.48% for logistic regression, and 72.01% for support vector machines. Perhaps, a better understanding of the performance of these models can be obtained from the confusion matrix. We present the confusion matrix for the three predictive models that we built in Table 1.

4 Conclusions and Discussion

The Mexican companies in this study report that the behavior of their growth is related to a greater information value in the profitability accounts through profit

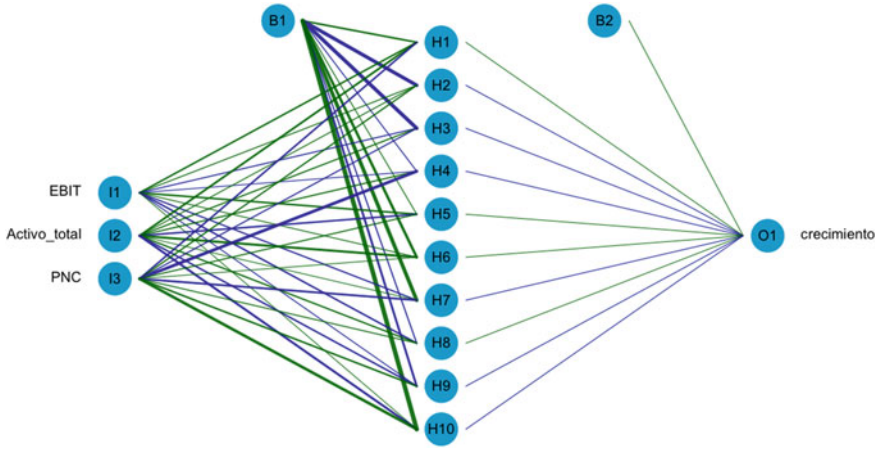


Fig. 2 Neural network model. *Source* Own elaboration

Table 1 Confusion matrix

		Predicted condition					
		Logistic regression		ANN		SVM	
		G	NG	G	NG	G	NG
True condition	Growth	351	29	340	40	363	17
	No growth	144	62	97	109	147	59

Source Own elaboration

before interest and taxes (EBIT) and consolidated net worth; and with the investments represented with the total assets. The information value or weight of the evidence of each one is 0.318, 0.194, and 0.180, respectively; these results coincide in part with the statements by the authors Blázquez et al. (2006), Manjón and Romero-Merino (2010), Gielnik et al. (2017), and Juárez (2018). Investment and profitability are decisive in business growth. However, for the ratios or index of indebtedness and profitability, an information value of less than 0.1 was obtained, which is why they are considered to be very weak predictors, and their use in predictive models is not recommended.

It is concluded that in the first place, the profitability, in second place, the consolidated net worth or stockholders' equity consisting of social capital and retained earnings, and in third place, the investment measured by the total assets are determining factors in the growth of the Mexican companies and that the iIdeal predictive model is neural networks.

References

Achtenhagen, L., Naldi, L., & Melin, L. (2010). Business growth - do practitioners and scholars really talk about the same thing? *Entrepreneurship Theory and Practice, 34*(2), 289–316.

Amat Salas, O., & Lloret Millan, P. (2014). *Avanzando: Claves para sobrevivir y crecer.* España. Profit Editorial, segunda edición

Benáček, V., & Michalíková, E. (2016). The factors of growth of small family businesses—A robust estimation of the behavioural consistency in panel data models. *Prague Economic Papers, 25*(1), 85–98. https://doi.org/10.18267/j.pep.538

Blázquez, F., Dorta, J. A., & Verona, M. C. (2006). Concepto, perspectivas y medida del crecimiento empresarial. *Cuadernos de administración*, *19*(31), 165–195

Delmar, F., Davidsson, P., & Gartner, W. B. (2003). Arriving at the high-growth firm. *Journal of Business Venturing, 18*(2), 189–216.

Dobbs, M., & Hamilton, R. T. (2007). Small business growth: recent evidence and new directions. *International Journal of Entrepreneurial Behaviour & Research,13*(5), 296–322.

Eichenberg, T. (2018). WoeBinning: Supervised weight of evidence binning of numeric variables and factors. R package version 0.1.6. https://CRAN.R-project.org/package=woeBinning

García-Manjón, J. V., & Romero-Merino, M. E. (2010). Efectos de la inversión en I+D sobre el crecimiento empresarial. *GCG GEORGETOWN UNIVERSITY—UNIVERSIA, 4*(2), 16–27. https://doi.org/10.3232/GCG.2010.V4.N2.01

Gielnik, M. M., Zacher, H., & Schmitt, A. (2017). How small business managers' age and focus on opportunities affect business growth: A Mediated moderation growth model. *Journal of Small Business Management, 55*(3), 460–483. https://doi.org/10.1111/jsbm.12253

Good, I. J. (1985). Weight of evidence: A brief survey. *Bayesian Statistics, 2,* 249–270.

Juárez, F. (2018). The growth of companies as a function of total assets. *WSEAS Transactions on Business and Economics, 15*(29), 301–310. Recuperado 01 de Julio de 2018 en. https://www.wseas.org/multimedia/journals/economics/2018/a585907-011.php

Machek, O., & Machek, M. (2014). Factors of business growth: A decomposition of sales growth into multiple factors. *WSEAS Transactions on Business and Economics, 11*(1), 380–385. https://www.scopus.com/inward/record.url?eid=2-s2.0-84903835753&partnerID=40&md5=6bec31f1b987a59b49c6174d636e4d0a

Meyer, D., Evgenia, D., Kurt, H., Andreas, W., & Friedrich, L. (2018). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-0. https://CRAN.R-project.org/package=e1071

Muhos, M. (2015). Review of business growth models: Methodology and the assumption of determinism. *International Journal of Management and Enterprise Development, 14*(4), 288–306. https://doi.org/10.1504/ijmed.2015.073810

R Core Team (2018). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria. https://www.R-project.org/

Rakotomamonjy, A. (2004). Optimizing area under ROC curve with SVMs. In: Proceedings of the 1st workshop on ROC analysis and artificial intelligence, pp. 71–80, Valencia, Spain, August 22, 2004.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: Visualizing classifier performance in R. *Bioinformatics, 21*(20), 7881. https://rocr.bioinf.mpi-sb.mpg.de

Vecino, C. E., Rojas, S. C., & Munoz, Y. (2014). Prácticas de evaluación financiera de inversiones en Colombia. Elsevier España, S.L.U., 31, 41–49

Venables, W. N., & Ripley, B. D. (2002) *Modern applied statistics with S* (4th ed.). New York: Springer. ISBN: 0-387-95457-0

Voeth, M., Poelzl, J., & Eidhoff, A. T. (2018). It's all about growth—An empirical status report of business development. *International Journal of Entrepreneurship and Innovation Management, 22*(1–2), 4–32. https://doi.org/10.1504/ijeim.2018.089713

Weinzimmer, L. G., Nystrom, P. C., & Freeman, S. J. (1998). Measuring organizational growth: Issues, consequences and guidelines. *Journal of Management, 24*(2), 235–262.

Electronic Leadership a Multifunctional Perspective: A Proposal Based on the Theory of the Structure of Initiation and Consideration of the Leadership and Adaptive Structures

Andres Ivan Canal-Carrillo, luiz Vicente Ovalles-Toledo, Eleazar González-Álvarez, and Zahira Moreno-Freites

Abstract The objective of this article is to describe and elaborate a contextualization of how nowadays electronic leadership must become a strategic tool for organizations and since the beginning of the twenty-first century has evolved in a process of social influence mediated by Information and Communication Technologies, of such that it has come to produce changes in attitudes, feelings behaviors even in the performance of individuals within organizations. Modern leadership has had to adapt and learn to communicate with your work teams no longer traditionally face to face but through a communication mediated by different electronic elements from laptops, cell phones, tablets, and other items from anywhere in the world. The documentary methodology allowed us through the consultation of multiple documents and a broad contrast of the information to make a proposal that structures the leadership electronically and fundamentally the actions of leaders in virtualized environments such as those of today .

A. I. Canal-Carrillo (✉)
Calle Selene 2042 APTO 9 Colonia Obrero Campesina, CP, 80013 Culiacan, Sinaloa, Mexico
e-mail: andrescanal@gmail.com

V. Ovalles-Toledo (✉)
Privada los Almendros, 2237, Fraccionamiento Los Almendros, CP, 80026 Culiacán, Sinaloa, México
e-mail: luiz.ovalles@uas.edu.mx

E. González-Álvarez (✉)
Calle Fray Pedro de Gante 4208 fraccionamiento Carlos V, CP, 80058 Culiacan, Sinaloa, Mexico
e-mail: eleazargonzalez.fca@uas.edu.mx

Z. Moreno-Freites (✉)
Universidad del Norte, Escuela de Negocios, Km.5 Vía Puerto Colombia, Barranquilla, Atlántico, Colombia
e-mail: debenvenutoz@uninorte.edu.co

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_9

Resumen El objetivo del presente artículo es describir y elaborar una contextualización de como hoy en día el liderazgo electrónico se debe convertir en una herramienta estratégica para las organizaciones y como desde inicios del siglo XXI ha evolucionado en un proceso de influencia social mediado por las Tecnologías de la Información y la Comunicación, de tal manera que ha llegado a producir cambios en las actitudes, sentimientos comportamientos incluso en el desempeño de los individuos al interior de las organizaciones. El liderazgo moderno ha tenido que adaptarse e ir aprendiendo a comunicarse con sus equipos de trabajo ya no de forma tradicional cara a cara sino a través de una comunicación mediada por diferentes elementos electrónicos, desde laptops, celulares, tabletas, y otros elementos desde cualquier lugar del mundo. La metodología documental nos permitió a través de la consulta de múltiples documentos y un amplio contraste de la información hacer una propuesta que estructure del liderazgo electrónico y fundamente el accionar de los líderes en ambientes tan virtualizados como los de hoy en día.

Keywords Electronic leadership · Tics · Virtual teams

Palabras clave Liderazgo electrónico · Tics · Equipos Virtuales

1 Introduction

The administration has now moved in environments marked by classical theories. This can be evidenced in the use of transactional leadership models. Nowadays in a globalized world and in constant connection, changes are being generated in the organizational structures, allowing new ways of working where its members interact through technological resources, but in different parts of the world, giving way to virtualization of the work teams and with it a substantial change in leadership, where new responsibilities must be assumed and new processes structured in virtual environments.

For Gonzalez and Sotelo (2016), new paradigms have motivated a review of the role of the traditional leader, the world economy is experiencing an important transition that is based on how organizations build new markets through the tics, allowing new ways of working and creating value both physically and in the virtual world. Avolio et al. (2001) have been stating that these new transformations have generated a new style of electronic leadership, which must be trained to meet the challenges of modern globalization.

Today's organizations depend heavily on multiple forms of digital communication, many of the processes that were previously done in person have been developed digitally, with the advantages offered by new technologies in order to achieve strategic objectives. If modern organizations do not enter into this dynamic of virtualization of many of their processes and the correct use of these technologies, they could be left behind losing great business opportunities.

Recent developments in the field of leadership have stimulated the need for them to reach these new business opportunities or strengthen a position within some area. It is very important to have a leadership capable of managing these advantages that modern technologies produce and that this in turn is ideal in guiding the teams in the scope of the strategic objectives of the organization in the middle of the virtualization of the processes. Starting from this base, today a new concept of e-leader or e-leadership is created with the aim of developing leadership proposals that are increasingly relevant to the modern environments of organizations.

For the development of this analysis, we will initially present a contextualization of electronic leadership and the main theories that we have used for the analysis.

1. Contextualization of leadership and Information and Communication Technologies

Until the end of the twentieth century, we continued talking about leadership in its traditional field, but they were Avolio et al. (2001) and a team of the first that began using the term electronic leadership at the end of the last century, beginning of the twenty-first century. They argued that technology has an effect on organizations according to their context and how users of these technologies adapt, resist, or reject technology based on structures that organizations create. The theory of adaptive structuring as indicated by Orlikowski (1992) highlights that human action is guided by structures, which define rules and resources that serve as the basis for planning and fulfilling tasks and structures that also arise when a group of work acts on these and produces new information that serves for interaction.

Avolio et al. (2000) define electronic leadership as "the process of social influence mediated by advanced information technologies to produce changes in attitudes, feelings, thoughts, behaviors, and performance among individuals, groups, and organizations". Lee (2009) defines it as an exercise outside the traditional environment highly mediated by technology, and it is the leadership exercised in the virtual context of business environments.

Modern leadership must coexist with current information technologies as these are fundamental parts of change and organizational transformation processes. Esguerra and Contreras (2016) point out that within these processes of organizational transformation, we must take into account that these they become very influential elements in the organization and that affect all levels of it, both individual, group, and organizational, modifying the internal dynamics of the processes and in equal measure the exercise of leadership that results from the interactions with these technologies.

Some of the first to define the structure of information and communication technologies were DeSanctis and Poole (1994), which differentiated two important characteristics; structural characteristics, which refer to the real characteristics of how the information is collected and managed by the users and spirit characteristics that are the structural characteristics but in reference to the underlying intention or purpose with which these technological structures are used. Regarding these spiritual characteristics, Avolio et al. (2001) point out that the repeated appropriation of technologies generates or transforms social structures, and over time, they are institutionalized.

In this sense, electronic leadership is based on the extent to which social structures accept, adapt, or reject these technologies.

The structural and spiritual characteristics have changed over time, but the objectives of leadership remain the same, creating vision, giving direction, motivating, inspiring and building trust, however, electronic leadership needs to implement these technology-mediated objectives (Avolio et al. 2001; Gonzalez and Sotelo 2016). However, these rapid changes affect leadership and individuals, for the amount of effort in order to adopt these new technologies. For Van Wart et al. (2017), most individuals tend to be reluctant to these changes requiring greater leadership effort. Given this, it is very important that the leadership is willing to understand that, although most of the effort is short-term in the implementation, the benefits of these new adaptations are long-term in reducing time, costs and an increase in the productivity of human capital.

Modern leadership must adapt to being mediated by technology, and therefore, it must learn to develop increasingly effective communication strategies and better structure its influence processes, whether by procedures, setting standards, or policies, generating a positive perception, where the organization is going, without losing motivation and increasingly adapting to the use of tics. In the same way, it is essential not to forget that leadership is influence and that influence must be materialized in the achievement of the objectives. It is of no use to have the best technology if the strategic objectives of the organization are not achieved; this process is analogous to as if it is done virtually or physically in the organization, the electronic leadership that can occur at all hierarchical levels of the organization, you must be able to motivate in such a way that in all areas the greatest number of objectives is achieved.

2 Virtual Team Communication

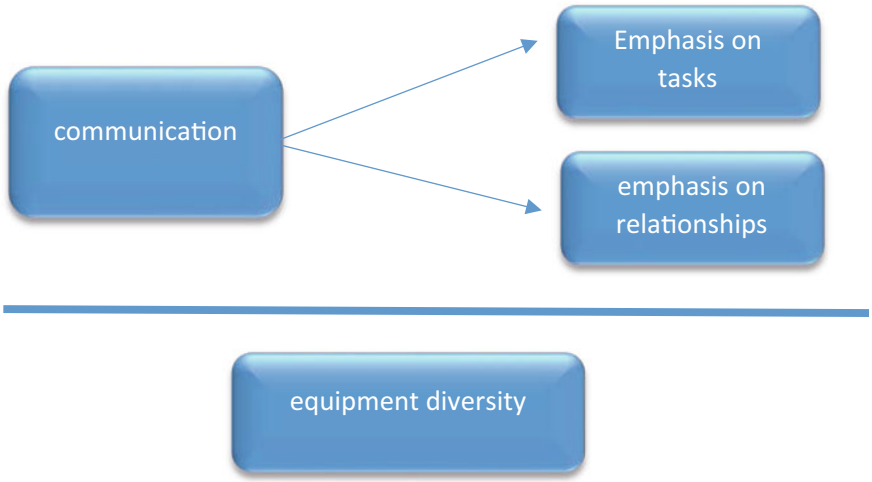
One of the characteristics of virtual teams is given in their communication, which occurs mainly through virtual tools affecting the behavior of the team. The investigations of Marlow et al. (2017) in their findings suggest that virtual teams take longer to complete tasks due to a variety of problems such as the lack of adaptation to the tics and the little synchrony between communication and virtual tools; similarly Hiltz et al. (1986) in the early beginnings of studies in highly virtualized teams found that they had a more focused orientation to tasks and communication than traditional teams.

It is clear that communication plays a fundamental role in electronic leadership. Marlow et al. (2017) argues that when communication is carried only through electronic means, it is more frequent and is greater in relation to face-to-face equipment, but there is a phenomenon of a decrease in its efficiency; this is because virtual teams must spend more time trying to filter out that communication that is irrelevant. For Chandler and Sweller (1991), the overabundance of unnecessary shared information can result in cognitive overload, which decreases performance, that is, once

an individual's cognitive load reaches a certain threshold, learning and information processing seem affected.

The communication has two emphasis: the first in relation to its content, a communication with high emphasis on the delegation of tasks, that is, one that focuses on how to assign and carry out the tasks, and another with relational emphasis where it can generate cohesion, trust in teams and that they are able to share relational information through virtual tools (Çekmecelioğlu and Özbağ 2016). Of the same, we cannot leave behind the diversity of the teams; diversity can promote high levels of performance if high quality communication occurs. Integration and understanding of the unique perspectives of team members can provide valuable information to achieve higher levels of performance. However, if the quality of communication is low, people may have difficulty coordinating their work with team members, due to a lack of understanding, especially related to different aspects of the task (Horwitz and Horwitz 2007; Marlow et al. 2017).

The communication in the virtual teams must contain these two emphasis mentioned above, but these must be supported on the basis of the diversity of the teams in all aspects (cultural, nationality, religious, among others). These aspects feed the communication and provide multiple points of view that allows to create alternative solutions to the problems; in the same way, it should also be taken into account that the misunderstandings in virtual teams are much more difficult to handle than in face-to-face teams. In this way, the leader must be clear about all these aspects so that his communication is more effective and timely.



Source self-made

Based on Horwitz and Horwitz (2007), Marlow et al. (2017)

On the other hand, leaders can influence the appropriation of technology, using legitimate models such as the formalization of processes, the establishment of norms and non-legitimate policies or procedures such as the manipulation of organizational

structures (Esguerra and Contreras 2016) of this. For a better adoption of technological changes, leadership must develop analytical skills, flexibility, create a need for achievement, motivate continuous learning, analyze, and support technical skills (Van Wart et al. 2017).

Avolio et al. (2001) argue that the nature of traditional leadership (face-to-face) and that of virtual leadership is essentially the same and that the greatest challenge that this leadership entails in itself is learning to manage the temporal, physical, and cultural dispersions that they exist today in modern work equipment. One of the challenges that this type of leadership must assume according to Avolio et al. (2014) is to learn to communicate with their work teams no longer in a traditional face-to-face way but through communication mediated by some type of computer system, which includes digital media, video conferencing environment, email, chat software, among others. The effectiveness of these media will depend on the way in which the leader manages them and makes the most of virtual environments. This point becomes a challenge, leading in virtual environments which is not easy, since you have to be updated, but about a high capacity to adapt to environments that change so rapidly today.

A survey conducted by the Society for Human Resource Management (SHRM) indicated that 46% of multinational organizations use virtual teams, (Minton-Eversole 2012). Globalization has greatly influenced the increase of these percentages and the development of new work teams. Traditional teams have the advantage of real-time personal communication and share physical space in many cases, which becomes a more effective process, while in virtual environments their collaborators must communicate and work asynchronously through technology; these mean that leadership must achieve synchrony in its work activities (Avolio et al. 2014; Kahai et al. 2003).

3 Characteristics of Electronic Leadership

The study of virtual leadership has focused on two areas: leadership behavior and leadership traits (Gilson et al. 2015; Liao 2017), returning to a behavioral approach (Blake and Mouton 1964; Bowers and Seashore 1966; Judge and Robbins 2013; Katz et al. 1950; Likert 1961) one could study virtual leadership by focusing the behaviors of leaders who are task oriented, that is, they create an initiation structure or those oriented to relationships, which contain a high degree of consideration. For electronic leadership, it is very important that leaders possess these two characteristics are task oriented, structuring and specifying roles and responsibilities; but without leaving aside the leadership also oriented toward relationships, emphasizing the well-being of the group, these two approaches were analyzed in behavioral theories and could characterize the way in which electronic leadership should behave.

3.1 Task-Oriented Leadership

The use of virtual teams has grown as organizations widely adopt team structures to carry out their work which are constantly integrated by different backgrounds and cultures, leading to understand that leadership and information technologies are evolving naturally, which will have cultural and even anthropological implications (Avolio et al. 2014).

The Ohio state studies conducted in 1950 within the framework of behavioral theories of leadership could contribute a little in how electronic leadership should behave, and these studies focused on trying to identify the different behaviors developed by leaders, from the observation of real situations. In the first of the two studies that were conducted at this University, under the direction of Hemphill (1956) and Fleishman (1951), it was tried to identify the different behaviors carried out by the leaders. For this, questionnaires were developed. These different behaviors were grouped into two dimensions or two types of behaviors, which were named as initiation of structure and consideration (López Martínez 2013; Shartle 1979; Vázquez 2010).

The Ohio state agency identified two types of behaviors that described the leader, the behavior of initiation of the structure and the behavior of consideration that describe the behavior of leadership within these theories. An important point also to take into account in virtual teams is the condition of having a vertical leader due to multiple geographical locations, this has allowed the emergence of natural leaders, that is, members willing to assume leadership roles, informal leaders within the team which facilitate the operation of the equipment allowing greater cohesion and stability.

For Robbins (2004) and Shartle (1979), taking up the studies of the Ohio state company, he considers that the initiation of the structure refers to the degree to which the leader defines and structures his role and those of his subordinates in order to achieve the goals. A leader who has a high qualification in the category of initiation of the structure is the one who assigns specific tasks to the members of his team and expects the members of his team to have a defined performance and insists that the deadlines be met.

In virtual teams, the establishment of objectives and a structure is probably one of the most important points of electronic leadership, clear rules, and practices that must be established since due to the little face-to-face interaction, or the multiple cultural differences, as schedules, among others, can hinder the scope of the objectives.

3.2 Relationship Oriented Leadership

Another characteristic that electronic leadership must possess is an orientation to relationships without leaving as explained above, the orientation to the structure. This orientation covers part of the personality of the leader; Hoch and Dulebohn

(2017) identified five important features that virtual leadership must have, the five central dimensions are extroversion, kindness, awareness, openness to experiences, and emotional stability. In addition to these five fundamental characteristics, agility and trust have also been shown to have a highly positive impact.

Participatory and consultative electronic leadership has demonstrated efficiency, rather than a managerial or authoritative style, has demonstrated high levels of efficiency when clear objectives are established, and there is transparency in decision making; when the leader provides virtual feedback generating trust, it promotes self-management of team. In terms of continuing to contribute to the construction of a more accurate definition of electronic leadership, the models of transformational leadership applied to virtual teams by their characteristics can favor the construction of trust and the development of individual creativity that contributes to the cohesion of groups (Esguerra and Contreras 2016).

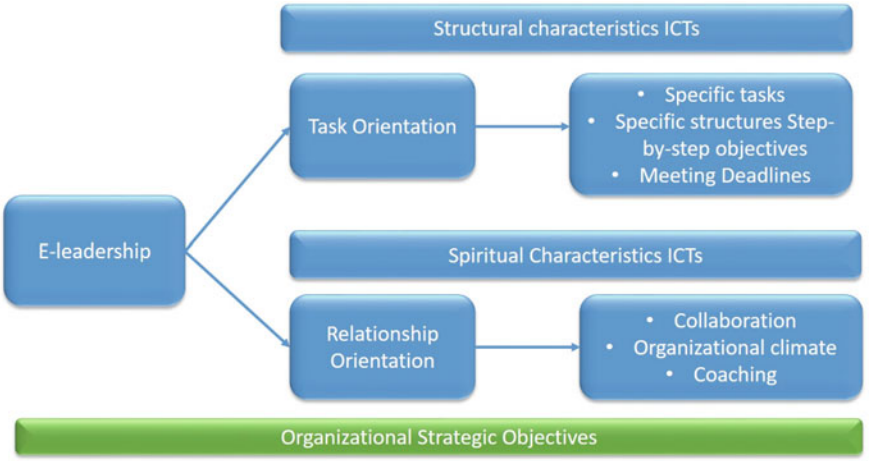
Returning to the studies of the state-owned Ohio, Robbins (2004) describes that the second style of leadership "consideration" which is the second orientation in behavioral theories is described as the degree to which the leader is likely to have working relationships characterized by mutual trust, respect for the ideas of subordinates, and their feelings. A very considerate leader is the one who helps his work team even with his personal problems, is friendly, and accessible; leaders with a high qualification in structure and consideration initiation tend to get their employees to achieve greater performance and satisfaction.

Now, similar to leadership in face-to-face contexts, electronic leadership can transmit through traits (who is one), behaviors (what one does), cognitive (what one thinks), and affective (what one feels), these mechanisms provide unique forms of leadership influence and are well received by collaborators (Hernandez et al. 2011).

Behavioral theories that initially tried to guide the way in which transactional leadership was exercised can be retaken and applied today to electronic leadership. A more promising approach would be to create an amalgam between behavioral models and modern approaches such as transformational leadership that perfect and give a more modern vision of what electronic leadership could be. More specifically take elements of behavioral models and transformational elements that give presence to the role of the leader, even if physically not found and everything is mediated by virtual communication.

On the basis of all of the above, Hambley et al. (2007) confirm what has been set out, where the establishment of transactional goals bony a high orientation to the tasks, and also the use of the facets of transformational leadership positively encourage group creativity and their efficiency.

In the following graph, we propose how the structure of electronic leadership could be, you can see the orientations that electronic leadership should have, and how these are complemented by the structure of the tics and their spirit approach according to the theory of structures placing as a basis the organizational objectives.



Based on Avolio et al. (2001), DeSanctis and Poole (1994), Gonzalez and Sotelo (2016), Robbins (2004)

4 Methodology

This study aimed to determine that when developing a theoretical, critical, and documentary review in which it has been possible to develop an analysis in view of the proposed theories of electronic leadership and its impact on work teams that are largely mediated by the technologies of the information and the communication.

5 Conclusions and Discussions

Leadership has evolved, but it has not lost its essence, in this research a proposal was drawn up of what could be the approaches and characteristics that electronic leadership should focus on and that can be adapted to the reality of organizations capacity. To disseminate and share information through technology, organizations have begun to redefine by allowing teams to separate in time and space, and elements such as culture, schedules, and other aspects affect to a certain extent the functioning of the teams and the leadership (Ziek and Smulowitz 2014).

Kahai et al. (2003) examine transformational versus transactional leadership in virtual team environments, finding that groups that work with a transactional leader had higher levels of group effectiveness and task satisfaction (task orientation), but also the effects of transformational leadership (consideration) were considered necessary; for Kirkman et al. (2002), the transformational/transactional leadership styles predict team processes and results, but none of these styles outperforms the other. This

strengthens the argument that we have been working on that the best way to strengthen electronic leadership is by giving behavioral tools, but also transformational ones.

The time and space variables are increasingly shorter, there are new relationships in the work teams, the same relationships of organizations with others make them increasingly complex, and in this way the traditional leadership only focuses on face-to-face teams. It is insufficient for the modern demands of organizations. Although a wide variety of definitions for electronic leadership have been suggested, we consider that the most important is to structure the way in which leaders must react to the changing ways of working, and the best way is an amalgam between transactional leadership styles and transformational, based on the theory of adaptive structuring. Traditional paradigms and models alone do not meet the operational needs of organizations, but a good interrelationship of these paradigms coupled with new models of exercising leadership can contribute to organizations reaching a greater degree of efficiency.

References

Avolio, B., Kahai, S., & Dodge, G. E. (2000). E-leadership. *The Leadership Quarterly, 11*(4), 615–668. https://doi.org/10.1016/S1048-9843(00)00062-X

Avolio, B. J., Kahai, S., & Dodge, G. E. (2001). E-leadership: Implications for theory, research, and practice. *The Leadership Quarterly, 11*(4), 615–668.

Avolio, B. J., Sosik, J. J., Kahai, S. S., & Baker, B. (2014). E-leadership: Re-examining transformations in leadership source and transmission. *The Leadership Quarterly, 25*(1), 105–131.

Blake, R. R., & Mouton, J. S. (1964). *The managerial grid*. Houston. TX: Gulf

Bowers, D. G., & Seashore, S. E. (1966). Predicting organizational effectiveness with a four-factor theory of leadership. *Administrative Science Quarterly*, 238–263

Çekmecelioğlu, H. G., & Özbağ, G. K. (2016). Leadership and creativity: The impact of transformational leadership on individual creativity. *Procedia Social and Behavioral Sciences, 235,* 243–249. https://doi.org/10.1016/j.sbspro.2016.11.020

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science, 5*(2), 121–147.

Esguerra, G. A., & Contreras, F. (2016). Liderazgo electrónico, un reto ineludible para las organizaciones de hoy. *Estudios Gerenciales, 32*(140), 262–268. https://doi.org/10.1016/j.estger.2016.08.003

Fleishman, E. A. (1951). *"Leadership Climate" and supervisory behavior: A study of the leadership role of the foreman in an industrial situation.* Ohio State University Personnel Research Board.

Gilson, L. L., Maynard, M. T., Jones Young, N. C., Vartiainen, M., & Hakonen, M. (2015). Virtual teams research: 10 years, 10 themes, and 10 opportunities. *Journal of Management, 41*(5), 1313–1337.

Gonzalez, E., & Sotelo, S. (2016). Liderazgo Electronico ,retos y habilidades una revision de la literatura *Revisiones Teoricas sobre ciencias administrativas*(1 edicion), 355

Hambley, L. A., O'Neill, T. A., & Kline, T. J. B. (2007). Virtual team leadership: The effects of leadership style and communication medium on team interaction styles and outcomes. *Organizational Behavior and Human Decision Processes, 103*(1), 1–20. https://doi.org/10.1016/j.obhdp.2006.09.004

Hemphill, J. K. (1956). *Group dimensions: A manual for their measurement*. Ohio State University. Bureau of Business Research. Research Monograph.

Hernandez, M., Eberly, M. B., Avolio, B. J., & Johnson, M. D. (2011). The loci and mechanisms of leadership: Exploring a more comprehensive view of leadership theory. *the Leadership Quarterly, 22*(6), 1165–1185.

Hiltz, S. R., Johnson, K., & Turoff, M. (1986). Experiments in group decision making: Communication process and outcome in face-to-face versus computerized conferences. *Human Communication Research, 13*(2), 225–252.

Hoch, J. E., & Dulebohn, J. H. (2017). Team personality composition, emergent leadership and shared leadership in virtual teams: A theoretical framework. *Human Resource Management Review*. https://doi.org/10.1016/j.hrmr.2016.12.012

Horwitz, S. K., & Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management, 33*(6), 987–1015.

Judge, T., & Robbins, S. (2013). *Comportamiento Organizacional 15 edicion.* Mexico: Pearson.

Kahai, S. S., Sosik, J. J., & Avolio, B. J. (2003). Effects of leadership style, anonymity, and rewards on creativity-relevant processes and outcomes in an electronic meeting system context. *The Leadership Quarterly, 14*(4–5), 499–524.

Katz, D., Maccoby, N., & Morse, N. C. (1950). *Productivity, supervision, and morale in an office situation.* Part I

Kirkman, B. L., Rosen, B., Gibson, C. B., Tesluk, P. E., & McPherson, S. O. (2002). Five challenges to virtual team success: Lessons from Sabre, Inc. *The Academy of Management Executive, 16*(3), 67–79

Lee, M. R. (2009). Effective virtual project management using multiple e-leadership styles. *Encyclopedia of E-Business Development and Management in the Global Economy*, 445–454

Liao, C. (2017). Leadership in virtual teams: A multilevel perspective. *Human Resource Management Review, 27*(4), 648–659. https://doi.org/10.1016/j.hrmr.2016.12.010

Likert, R. (1961). *New patterns of management*. MacGraw Hill: New York

López Martínez, E. (2013). La importancia del liderazgo en las organizaciones

Marlow, S. L., Lacerenza, C. N., & Salas, E. (2017). Communication in virtual teams: A conceptual framework and research agenda. *Human Resource Management Review, 27*(4), 575–589. https://doi.org/10.1016/j.hrmr.2016.12.005

Minton-Eversole, T. (2012). Virtual teams used most by global organizations, survey says. Retrieved from Society for human Resource management. https://www.shrm.org, https://www.shrm.org/ResourcesAndTools/hr-topics/organizational-and-employee-development/Pages/VirtualTeamsUsedMostbyGlobalOrganizations,SurveySays.aspx

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science, 3*(3), 398–427. https://doi.org/10.1287/orsc.3.3.398

Robbins, S. (2004). *Comportamiento Organizacional*. Mexico: Pearson Prentice Hall.

Shartle, C. L. (1979). Early years of the Ohio State University leadership studies. *Journal of Management, 5*(2), 127–134.

Van Wart, M., Roman, A., Wang, X., & Liu, C. (2017). Integrating ICT adoption issues into (e-)leadership theory. *Telematics and Informatics, 34*(5), 527–537. https://doi.org/10.1016/j.tele.2016.11.003

Vázquez, J. F. S. (2010). *Liderazgo: teorías y aplicaciones*: Publicaciones Universidad Pontificia.

Ziek, P., & Smulowitz, S. (2014). The impact of emergent virtual leadership competencies on team effectiveness. *Leadership & Organization Development Journal, 35*(2), 106–120.

The Use of Big Data in the Modern Biology: The Case of Agriculture

Juan Campos-Guillén, Vicente Moreno-Andrade, Miguel A. Rico-Rodriguez, Bernardo Bañuelos-Hernández, Armando Ceja-Bravo, Josué Bermeo-Escalona, and Andrés Cruz-Hernández

Abstract There is a data revolution in the agriculture that involves the analysis of this data, the development of new models and the application of these models in a practical and simple way. All new technologies allow to process big volumes of data with new mathematical algorithms, checking the information in real time, creating new value standards and changing the conception of agriculture. The aim of this chapter is to show some alternatives for the improvement of the crops using the big data, produced with the new technologies for massive sequencing in two specific examples: microbial massive analysis and molecular breeding.

Keywords Agrogenomics · Genomes · Metagenomics · Molecular breeding · Sequencing

1 Introduction

To now there is a big challenge for agriculture, the production of food for 9000 people in a minor cultivable area, with less personal and facing problems with the water availability. Thirty percentage of food waste is produced around the world by several reasons, including problems in production, transport and storage. The use of technology at the field must be a key in order to establish an efficient productive process. In 2045, the land for food production must be limited and agriculture will

J. Campos-Guillén · V. Moreno-Andrade · M. A. Rico-Rodriguez
Facultad de Química, Universidad Autónoma de Querétaro. Cerro de las Campanas s/n, C.P., 76010 Santiago de Querétaro, QRO, México

B. Bañuelos-Hernández · A. Ceja-Bravo · J. Bermeo-Escalona · A. Cruz-Hernández (✉)
Universidad de la Salle Bajío, campus Campestre. Av Universidad 602, Col Lomas del Campestre. C.P., 37150 León, GTO, México
e-mail: acruz@delasalle.edu.mx; andrex1998@hotmail.com

A. Cruz-Hernández
Escuela de Agronomía, Universidad de la Salle Bajío, campus Campestre. Av Universidad 602, Col Lomas del Campestre. C.P., 37150 León, GTO, México

E. León-Castro et al. (eds.), *Intelligent and Complex Systems in Economics and Business*, Advances in Intelligent Systems and Computing 1249,
https://doi.org/10.1007/978-3-030-59191-5_10

have the necessity to develop and apply new technology in every part of the food chain.

Agriculture is used to work with open systems, with several variables, and a high competition at global level, with diverse agents participating at the food production and distribution chain. Also, it needs to improve their efficiency in a continuous way, with a chance to automate many of the operations and decisions. The first step is to register and digitize all the activities for production, transformation, storage, marketing and distribution of the products. Also, it is important to use all the available data and public information in combination with the internal data from every participant in order to take the right decisions.

Starting from there, the big data and analytical data will help to improve the agriculture in several ways: planning for the production, improving the products, planning the transformation process, marketing and distributing the products, preventing diseases or plagues, reducing environmental impacts and in the development of new cultivars, using the new genomic available data.

All the new strategies are related to the capture of the data in the different processes, integrate the external data and help in the decisions of the different steps in the agricultural value chain. There is a data revolution in the agriculture that involves the analysis of this data, the development of new models and the application of these models in a practical and simple way. All new technologies allow to process big volumes of data with new mathematical algorithms, checking the information in real time, creating new value standards and changing the conception of agriculture. The aim of this chapter is to show some alternatives for the improvement of the crops using the big data, produced with the new technologies for massive sequencing in two specific examples: microbial massive analysis and molecular breeding.

2 Metagenomics

2.1 Bacterial Identification

Analysis and identification of microorganisms are important in food industry, pharmacology, agriculture and environmental sciences. Bacteria is used for development of vaccines, plant breeding and synthesis of new proteins in food industry (Glick 2014; Schüürmann et al. 2014). Their identification in the food is a preventive condition in order to avoid the transmission of potential pathogens for humans (Acharya et al. 2016).

In the last 15 years, bacterial identification has used molecular techniques, such as polymerase chain reaction (PCR) that is based on the primer design which amplifies nucleotide sequences associated with a specific organism, allowing its identification by phylogenetic analysis using bioinformatics resources.

There is another type of bacterial identification using their proteins and lipids, using mass spectroscopy (MALDI-TOF) (Schäfer et al. 2014). Those techniques

based on sequencing require culture of isolated bacteria that sometimes are unable to grow because of its specific conditions of temperature, pH, CO_2 or a particular culture media.

2.2 Metagenomic Analysis

Metagenomics is the study of microorganisms from a environmental sample through their genomes with modern genomic tools, even if the microorganisms are culturable or not (Chen and Pachter 2005; Thomas et al. 2012). From a environmental sample, just 1% of the present organisms are able to be cultured by traditional methods. The metagenomic analysis involves two techniques based on the presence and function of microorganisms (Morgan and Huttenhower 2012). Both techniques need the extraction of DNA, a cloning vector and the transformation and detection of clones. The difference between the techniques is the amplification region and the design of specific primers for the genes of interest (Tamaki et al. 2011; Bashir et al. 2014).

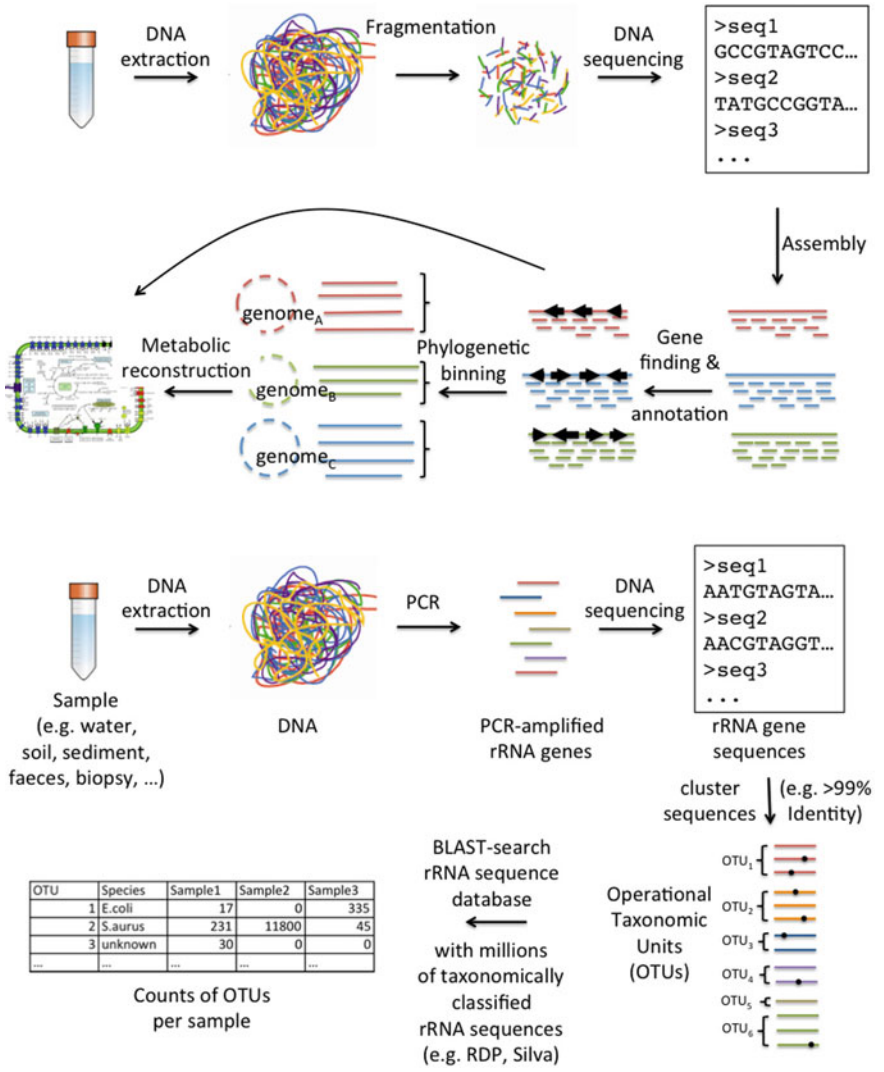
The metagenomic analysis starts with the DNA extraction from the ecological sample of interest (Fig. 1) that needs a specific method that avoids the DNA contamination or degradation, and a library of genomic sequences is constructed by DNA random fragmentation (shotgun) or PCR amplification (amplicon), followed by the sequencing of the fragments and finally the processing and analysis of the crude data with bioinformatic specific programs (Morgan and Huttenhower 2012).

2.3 Sequencing

The metagenomic analysis has been developed with the new techniques for massive genomic sequencing (MGS), and several platforms were developed in order to sequence millions of DNA molecules at the same time (Whiteley et al. 2012), allowing to analyze the genetic information from the bacterial communities in the environmental sample (Escobar-Zepeda et al. 2015). One of the most used platforms in the metagenomic analysis is Illumina, because of its sequencing yield, low mistake rate and economy. The platform is based on the reversible synthesis with fluorescent label nucleotides (Escobar-Zepeda et al. 2015).

2.4 Data Analysis

The taxonomical classification of the sequences in the metagenomic analysis will depend of the bioinformatics analysis. Some bioinformatics applications, i.e. Kaiju and MG-Rast, are easy to access. Kaiju was developed for the classification of pathogenic species, based on the protein data (Menzel et al. 2016; Wilke et al.



OTU	Species	Sample1	Sample2	Sample3
1	E.coli	17	0	335
2	S.aurus	231	11800	45
3	unknown	30	0	0
...	...	...	...	...

Fig. 1 Metagenomic analysis schematic representation. From the sample DNA, a library of genomic sequences is constructed by DNA random fragmentation (shotgun) (upper figure) or PCR amplification (amplicon) (down figure), followed by the sequencing of the fragments and finally the processing and analysis of the crude data with bioinformatic specific programs (https://envgen.github.io/metagenomics.html)

2015). There are other software programs for the microbial sequence analysis, such as QIIME and Geneouis and MEGAN (Caporaso et al. 2010; Huson et al. 2011; Escobar-Zepeda et al. 2015); QIIME is the most used programs for the metagenomic analysis (Schmieder et al. 2011). Those programs use different methods for sequence alignment (i.e., Uclust and Blast) and classification, such as the Bayesian classification RDP, from Qiime and Geneious (Velsko et al. 2018). The final taxonomical classification depends of the databases available, and some of these databases are Greengenes and Silva and are possible to establish particular databases from the National Center of Biotechnology (NCBI) (Huson et al. 2011).

2.5 Metagenomic Analysis Examples

Metagenomic analysis has been used in different biological areas such as medicine, for the analysis of microbial genes associated with the human intestine, bacterial disease diagnostic and the discovery of new antibiotics and enzymes (Nakamura et al. 2008; Parker and Chen 2017). In environmental sciences it has been used in biorremediation, through the identification of microorganisms from soil, water and air, even with the identification of microorganisms from the environment (George et al. 2010; Delmont et al. 2011; Kwong et al. 2014).

The identification of microbial diversity in a free media sample is now possible with the new genomic sequencing techniques. This technique was possible to identify molecules with biological activity, such as celluloses, lipases, amylases, proteases and DNAses with an important industrial, pharmacological and food use (Bashir et al. 2014).

3 Agrogenomics (Molecular Breeding)

Plant breeding is so antique such as agriculture, and it had been practiced since thousands years ago in plants with economical importance to the men. For some plants, several collections of genes had been identified and associated with interesting agronomic traits, but the number of these traits is limited and just for some genes the structure, sequence and mechanism expression are well known.

The development of molecular biology and the new technologies for massive genome sequencing (MGS) open new perspectives on the basic studies on plants and on the applications, including plant breeding.

All the traits are included in the genome, and all the genetic information is included at the chromosomes and is organized in functional units called genes. The interaction of the genome and the environment defines the biology of an individual plant. The DNA content in plants include 100 millions to 100 billions of symbols, organized in 20,000–50,000 genes (Ekblom and Galindo 2011).

Genomics could be defined as the study of the genomes, and its objectives are to catalogue all the genes in an organism, to study the organization and structure of each one of them and to discover its function and the mechanisms related in the regulation, and the interaction between different genes. To now, the costs of genome sequencing are in the order of thousand dollars, new bioinformatics resources for sequence analysis and the first plant genomes have been published (Table 1).

Plant genome sequencing will define their genetic structure. That will be fundamental for the design of new generation of crop plants. The ultimate sequencing generation technologies allow to sequence million of bases in a time. This new genomic knowledge will help to develop new plant varieties with an open number of uses.

Plant genomics could associate the phenotypic characteristics with agronomical interest with gene analysis, through the study of plants with specific genetic characteristics. Genomic research allows to determine with precision the allelic combinations, each form of a gene, and isolate the specific gene.

The genome for a plant was sequenced by first time 15 years ago, and it was a model plant. To now, at least 50 plants have been sequenced. They include food crops, i.e., banana, potato and tomato (Potato Consortium 2011; Tomato Consortium 2012). Also, the sequence of grape, papaya and rice had been sequenced (Table 1). Several international efforts have included the sequences of corn, wheat, sunflower, carrot, sugarcane or citrus (Iorizzo et al. 2016; Garsmeur et al. 2018).

Recently, it developed the first project with the application of tomato sequencing in order to identify the genetic characteristics associated with the like and the flavor of in the fruit (Tieman et al. 2017). The chemical compounds were quantified in 398 tomato accessions from wild, creoles and modern plant materials, including their

Table 1 Commercial crops sequenced

Common name	Scientific name	Botanical family	Chromosome number	Genomic content (Mbp)
Arabidopsis	*Arabidopsis thaliana*	Brassicaceae	5	125
Rice	*Oryza sativa*	Poaceae	12	389
Grape	*Vitis vinifera*	Viteaeae	19	487
Papaya	Carica papaya	Caricaceae	9	372
Corn	Zea mayz	Poaceae	10	2300
Cacao	*Theobroma cacao*	Sterculiaceae	10	430
Potato	*Solanum tuberosum*	Solanaceae	12	844
Tomato	*Solanum lycopersicum*	Solanaceae	12	760
Carrot	Daucus carota	Apiaceae	18	473
Sugarcane	*Saccharum spp.*	Poaceae	10	382

genomic sequencing. The authors found that modern varieties are lacking of some key compounds that the more antique accessions contain. The genome sequence allows the identification of some loci related to the licking and flavor of tomato, including sugar, fatty acids and volatile synthesis. With plant transformation, they were able to recover the liking and flavor through molecular breeding (Tieman et al. 2017).

4 Earth BioGenome Project

An international scientific consortium is trying to sequence, catalogue and analyze the genomes of all the **eucariots** species identified in the planet, animals and plants (Lewin et al. 2018). The Earth BioGenome Project has the objective of conservation of the planet biodiversity. To now in some other efforts, scientists have been sequenced less than 15,000 species, mainly microbia. The project pretends to sequence and analyze all the eucariots species genomes. In the next 10 years with a budget of 4700 millions of american dollars (3800 millions of euros), with a capacity of digital storage of 200 petabytes.

There is also a similar project called Earth Microbiome that includes more than 500 scientists in order to sequence the genomes of bacteria and archaebacteria around the world. Genomics have been helping to develop new medicines and new removable energy sources, to feed the people, to protect the environment and to support the **survivor** and the **human good will**.

5 Conclusions

A new revolution in agriculture involves the production and analysis of big data, the development of new models and its application of these models in a practical and simple way. All new technologies allow to process big volumes of data with new mathematical algorithms, checking the information in real time, creating new value standard and changing the conception of agriculture. One of these applications include the decoded genome in plant species and the use of this information for the improvement of agronomical characteristics, the detection of beneficial soil microorganisms and the detection of new emerging pathogens. It will give to the agronomical studies a very important perspective.

Acknowledgements Thanks to Universidad De la Salle Bajío Campus Campestre and Universidad Autónoma de Querétaro for financial support.

References

Acharya, A., Poudel, A., Sah, A. K., Maharjan, D., Tibrewal, S., Mandal, P. K., & Maharjan, D. (2016). Isolation and identification of bacteria from meat processing units of Kathmandu Valley. *International Journal of Microbiology and Allied Sciences, 2*(3), 33–37.

Bashir, Y., Pradeep-Singh, S., & Kumar-Konwar, B. (2014). Metagenomics: An application based perspective. *Chinese Journal of Biology, 2014,* 1–7.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 7*(5), 335.

Chen, K., & Pachter, L. (2005). Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Computational Biology, 1*(2), e24.

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., et al. (2011). Accessing the soil metagenome for studies of microbial diversity. *Applied and Environmental Microbiology, 77*(4), 1314–1324.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity, 107,* 1–15.

Escobar-Zepeda, A., de León, A. V. P., & Sanchez-Flores, A. (2015). The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics, 6,* 1–15.

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., Jenkins, J., Martin, G., Charron, C., Hervouet, C., Costet, L., Yahiaoui, N., Healey, A., Sims, D., Cherukuri, Y., Sreedasyam, A., Kilian, A., Chan, A., Van Sluys, M.-A., Swaminathan, K., Town, C., Bergès, H., Simmons, B., Glaszmann, J. C., van der Vossen, E., Henry, R., Schmutz, J., & D'Hont, A. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications, 9*, 2638. https://doi.org/10.1038/s41467-018-05051-5

George, I. Stenuit, B., & Agathos, S. (2010). *Application of metagenomics to bioremediation. Metagenomics: Theory, methods and applications* (1st ed., pp. 119–140). Norfolk: Caister Academic Press

Glick, B. R. (2014). Bacteria with ACC deaminase can promote plant growth and help to feed the world. *Microbiological Research, 169*(1), 30–39.

Huson, D. H., Mitra, S., Weber, N., Ruscheweyh, H. J., & Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN 4. *Genome Research, 21*(9), 1552–1560.

Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., et al. (2016). A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nature Genetics., 48,* 657–666. https://doi.org/10.1038/ng.3565

Kwong, W. K., Engel, P., Koch, H., & Moran, N. A. (2014). Genomics and host specilization of honey bee and bumble bee gut symbionts. *Proceedings of the National Academy of Sciences, 111*(31), 11509–11514.

Lewin, H. A., Robinson, G. E., Kress, W. J, Baker, W. J., Coddington, J., Crandall, K. A, Durbin, R, Edwards, S. V, Forest, F., Gilbert, M. T. P., Goldstein, M. M, Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., van Sluys, M. A., Soltis, P. S., Xu, X., Yang, H., Zhang, G. (2018). Earth biogenome project. Sequencing life for the future of life. In *Procedings of the National Academy of Sciences USA, 115*(17), 4325–4333

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications, 7,* 11257.

Morgan, X. C., & Huttenhower, C. (2012). Human microbiome analysis. *PLoS Computational Biology, 8*(12), e1002808.

Nakamura, S., Maeda, N., Miron, I. M., Yoh, M., Izutsu, K., Kataoka, C., et al. (2008). Metagenomic diagnosis of bacterial infections. *Emerging Infectious Diseases, 14*(11), 1784–1786.

Parker, J., & Chen, J. (2017). Application of next generation sequencing for the detection of human viral pathogens in clinical specimens. *Journal of Clinical Virology, 86,* 20–26.

Potato Genome Consortium. (2011). Genome sequence and analysis of the tuber crop potato. *Nature, 475,* 189–197. https://doi.org/10.1038/nature10158

Schäfer, M. O., Genersch, E., Fünfhaus, A., Poppinga, L., Formella, N., Bettin, B., & Karger, A. (2014). Rapid identification of differentially virulent genotypes of *Paenibacillus larvae*, the causative organism of American foulbrood of honey bees, by whole cell MALDI-TOF mass spectrometry. *Veterinary Microbiology, 170*(3), 291–297.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics, 27*(6), 863–864.

Schüürmann, J., Quehl, P., Festel, G., & Jose, J. (2014). Bacterial whole-cell biocatalysts by surface display of enzymes: Toward industrial application. *Applied Microbiology and Biotechnology, 98*(19), 8031–8046.

Tamaki, H., Wright, C. L., Li, X., Lin, Q., Hwang, C., Wang, S., & Liu, W. T. (2011). Analysis of 16S rRNA amplicon sequencing options on the Roche/454 next-generation titanium sequencing platform. *PLoS ONE, 6*(9), 1–6.

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics-a guide from sampling to data analysis. *Microbial Informatics and Experimentation, 2*(1), 3.

Tieman, D., Guangtao, Z., Resende Jr. Marcio F. R., Tao, L., Cuong, N., Dawn, B., Luis, R. J., Stephanie, O. B. K., Taylor, M., Bo, Z., Hiroki, I., Zhongyuan, L., Josef, F., Itay, Z., Antonio, M., Dani, Z., Antonio, G., Matias, K., Sanwen, H., & Klee, H. (2017). A chemical genetic roadmap to improved tomato flavor. *Science, 355*(6323), 391–394. https://doi.org/10.1126/science.aal1556

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature, 485,* 635–641. https://doi.org/10.1038/nature11119

Velsko, I. M., Frantz, L. A., Herbig, A., Larson, G., & Warinner, C. G. (2018). Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research. *Msystems, 3,* 1–28.

Whiteley, A. S., Jenkins, S., Waite, I., Kresoje, N., Payne, H., Mullan, B., et al. (2012). Microbial 16S rRNA Ion Tag and community metagenome sequencing using the ion torrent (PGM) platform. *Journal of Microbiological Methods, 91*(1), 80–88.

Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K. P., & Chaterji, S. (2015). The MG-RAST metagenomics database and portal in 2015. *Nucleic Acids Research, 44*(1), 590–594.