

Fuzzy logic, Web scraping, Personnel Selection, AHP

Isaac FLORES-HERNÁNDEZ ^{[0000-0003-4620-9144]*},
Edmundo BONILLA-HUERTA ^{[0000-0003-2062-4219]*},
*Perfecto Malaquías Quintero-Flores**, *Oscar Atriano PONCE***
*José Crispín HERNÁNDEZ-HERNÁNDEZ**,

APPLYING INTELLIGENT TECHNIQUES FOR TALENT RECRUITMENT

Abstract

The objective of this research is to describe a system to aligned the hard and soft skills of the applicant to the current labor market. For this, a system was implemented which uses Web Scraping to get a general profile of an area, meanwhile for the evaluation of the applicant soft skills is used a Test Cleaver and for the hard skills fuzzy inference system is implemented. Therefore, the data is entered into an Analytic Hierarchy Process, with this, the applier is able to see which area is better to improve according to the hard and soft skills.

1. INTRODUCTION

The profile of graduates of the universities in the different sectors of industry and research worldwide are evolving and frequently changing. Therefore, the skills that are expected from them are more technical and specialize. This make cause that many of the vacant posts to not be fill in many different areas, which leads to the increase of the unemployment index in Mexico. From the employer point of view the selection process is very important but takes a lot of time and human resources for the evaluation filters and the hiring of the appropriate candidate for a specific position, this reveals a flaw in the communication between the universities and the industries on the development of the skills that the graduates require in the work sector.

* Instituto Tecnológico de Apizaco, Computer Systems Department, Fco I Madero, Barrio de San José, 90300 Apizaco, Tlaxcala, México, isaac_camus@hotmail.com, edbonn@walla.com, malakaz_81@hotmail.com, josechh@yahoo.com.

** Smartsoft America BA, Calle Adolfo López Mateos, Texcacoac, 90806 Chiautempan, México, oatriano@smartsoftamerica.com.mx

In order to solve this problem, several methodologies had been applied for instance: the decisional tree J48 and random forest (Algur, Bhat & Kulkarni, 2016), which are used in order to predict if a student will be admitted depending on factors such as: Average aggregate score of all the semesters in CGPA, Communication Skill, Placement Preparation Hours per Week, Breaks between 12th and Engineering in Years, Performance in Extra Curricular Activities, Performance in Cultural Activities, Number of Industrial Visits during the course.

In his paper (Koutra, Kardaras, Barbounaki & Stalidis, 2017) use the Analytic Hierarchy Process (AHP) method incorporation with the Correspondence Analysis (CA), by doing this forms qualitative evaluations scales based on the data without any previous supposition.

In the other hand, (Rianto, Budiyo, Setyohadi & Suyoto, 2017) designed an evaluation model using different criteria and sub-criteria, the weights of each one of this indicator was achieved through AHP, and then the range of alternatives of the student selection is calculated using TOPSIS. The considering criteria where: Academic ability, English Ability, Psycho test, Attitude, Soft Skill, Communication skills, Solve problems and think critically, Time management, Teamwork, Flexible & Adaptation.

In this paper a system based on Fuzzy Logic is proposed, Web Scraping and AHP (Analytic Hierarchy Process) to match the formation of the graduates with the demands of the industry.

2. METHODOLOGY

2.1. Test cleaver

The test cleaver is used by many Mexican companies for the recruiting and selection process of the personnel. It was designed in 1959 in Princeton New Jersey by J.P Cleaver & CO (Gil-Gaytán & Núñez-Partido, 2017). It is a specific technique to respond to the personnel selection needs, the main objective of this technique is to measure the work conduct to be able to place the appropriate person for the appropriate job. It is because this technique measures the behavior and skills of the evaluated candidate, this is because it provides a description of the natural and daily conduct of an individual at work, which allows to determine if the candidate possess the ideal work skills for a position or the ability to perform properly in a specific area.

2.2. Fuzzy Logic

Fuzzy logic was introduced in 1965 by Lofti A Zadeh in his paper “fuzzy sets” (Zadeh, 1965) which presented a way to process the information in a way that the data can have a partial membership degree associated to different sets, stating the fuzzy sets theory.

The fuzzy sets allow to formalize linguistic excretions that have some ambiguity degree, in other words, provides a method to mathematically express concepts like: tall, cold, fast. Etcetera; that in everyday life are used consistently, but are not precise in themselves, this is why the membership degree of an element to a set is determined by a membership function that may contain all the real values between the interval [0,1]. The formal definition of a fuzzy set is:

A fuzzy set A in X is expressed as a set of ordered pairs:

$$A: \{x, \mu_A(x) \mid x \in X\}, \quad (1)$$

where: A – Represent the Fuzzy Set,

$\mu_A(x)$ – Represent the Membership Function,

$x \in X$ – Universe of universe of discourse.

Thus a fuzzy set is totally characterized by the membership function (MF) that indicates the degree in which every element of the universe belong to a given universe. There are many membership functions but the commonly used are Triangular MF, Trapezoidal MF and the Gaussian MF.

The Fuzzy rules are conditional sentences shape as IF-THEN, where de fuzzy propositions of the premises are related with the consequent by the implication, for example:

$$IF \ x \text{ is } A \ THEN \ y \text{ is } B$$

Where A and B are linguistic values defined by fuzzy sets in the universe X and Y. The (IF x is A) part is called premise or antecedent, meanwhile the (THEN y is B) part is called consequent or conclusion. For example

$$IF \ road \ is \ wet \ THEN \ driving \ is \ dangerous$$

In the classic rules systems, if the premise is true then the consequent is true. In the fuzzy systems this is different, because the premise is fuzzy variable, this why the rules are partially executed, therefore the consequent is true in a certain degree.

A fuzzy logic system used the inference like a calculus mechanism for a system where the input and output are numbers. A basic inference system is composed by 5 elements:

- Set of rules IF-THEN,
- A data base,
- Decision-making unit,
- Fuzzification unit,
- Defuzzification unit.

One of the most used inference systems methods based in linguistic rules was proposed in 1975 (Mamdani & Assilan, 1975), in an attempt to try to control a combination of the steam motor and boiler by a set of rules. This process has 4 steps:

- Fuzzification of input variables: Consist on taking the crisp values of the input and determine the membership degree to the fuzzy sets associated.
- Evaluation of the rules: Takes as the input the fuzzified values and it applied to the premises in the fuzzy rules. If a rule has multiple premises the operator And/Or are used to get only one number that represent the evaluation and its applied to the conclusion.
- Aggregation of the output rules: Is a unification process of the output of all the rules, which means the membership function of all the consequents are combined (By union) to get and only fuzzy set for each output variable.
- Defuzzification: Is the final result usually expressed by a crisp value.

2.3. Web Scraping

Data mining is currently a powerfull tool, because is one way to get information form a set of data (Broucke & Baesens, 2018). A part of the data mining is web content mining which has 4 extraction ways:

- Not structured Data mining,
- Structured Data mining,
- Semi-structured Data mining,
- Multimedia data extraction.

Web Scraping is a technique which is part of not structured Data mining, which consist in the extraction of one or many web pages from a same site, to manipulate, process and analyze part of its content.

2.4. Analytic Hierarchy Process

The Analytic Hierarchy Process was developed in 1980 by Thomas L. Saaty (Saaty, 1980). Is a multicriterial method that provides an evaluation for the alternatives to problems that contain multiple criteria, depending on the relative importance of each one of them, and then specifies the preference according to each one of the decision alternatives for each criteria. The result is the establishment of a hierarchy according to the priorities that shows the global preference for each one of the decision options. Is a worldwide use process in a wide variety if situation on different fields such as, health, business, government, education, etc.

3. IMPLEMENTATION

The proposed model is composed by 3 parts that are shown (figure 1).

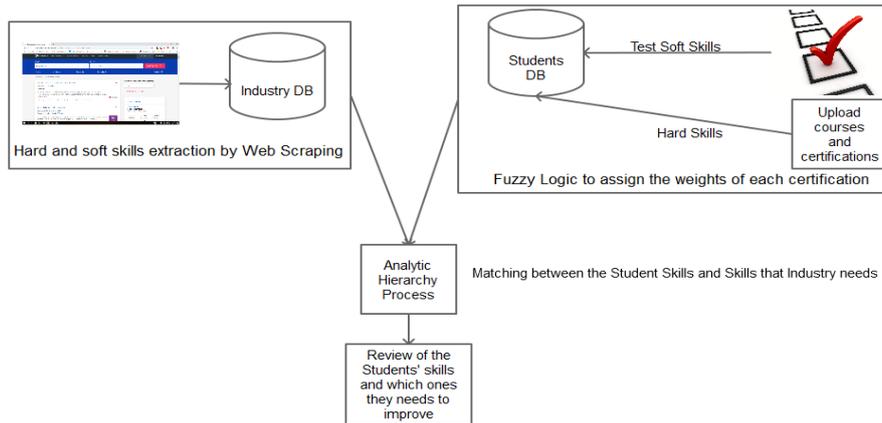


Fig. 1. Proposed System

The first part implemented was the web scraping in order to extract the hard and soft skills from job offers. To be able to do this the web site for the data mining was selected: OCCMundial (www.occ.com.mx) thus the web page structure had to be thoroughly analyzed for the purpose of obtaining the data to be extracted.

Once the design analysis was done it was determined to extract all the text referring to the job offer and the search the hard and soft skills required in the job market was done using previously elaborated dictionaries. The design of the dictionary for the soft skills was done taking into account the different writing styles for these offers, this is because the verbs can be conjugated in different indicative tenses in Spanish, which makes the search more difficult. To solve this issue the use of regular expressions was required in a way that the base form of the verb was looked for without caring the indicative tense used. 29 soft skills were considered which can be seen in table 1.

Tab. 1. Soft Skills

responsib[\w]	stress work under pressure	positi[\w]
organiz[\w]	commit[\w]	initiative
tolera[\w]	interest*[\w]	flexib[\w]
analy[\w]	sincer[\w]	honest*[\w]
teamwork	adaptab[\w]	punctual*[\w]
collaborat[\w]	enthusias[\w]	critical
[\w]*active proactivity	reliable confidence	social[\w]
availab[\w]	integrity honest person	dedicat[\w]
attitude	patien[\w]	assertive*[\w]
communication easy word	decisive	

For the hard skills this process was easier, because of the lack of conjugation of the word use, the 20 hard skills can be seen in table 2.

Tab. 2. Hard Skills

java	html 5
.net	css
visual studio	jquery
asp	bootstrap
php	sql
javascript	mvc
sql server	rest
postgresql	ajax
json	ios
c sharp c#	android

Once the results had been obtained are analyze, and as a result we get a general profile for certain specific job, for instance what are the skills that a programmer needs according to the requirements in the working market.

For examples, the soft skills more sought for a programmer are:

- Teamwork,
- Adaptable,
- Responsible,
- Integrity,
- Analytical.

The hard skills more require for a programmer are:

- iOS,
- SQL,
- Java,
- REST.
- .Net.

The second part is address to the evaluation of soft and hard skills of the applier. To evaluate soft skills a psychological test designed under the Cleaver Model was used, providing a description of the applier emphasizing the skills. The skills to be evaluated in the test match the ones in the dictionary of the first stage in the system see (Fig. 1), and the interface can be seen in figure 2.

Choose from each group of characteristics, the one that best describes you (+) and the one that least describes you (-):

1.-	(+)	(-)	2.-	(+)	(-)	3.-	(+)	(-)
Persuasive	<input checked="" type="radio"/>	<input type="radio"/>	Agressive	<input type="radio"/>	<input type="radio"/>	Pleasant	<input type="radio"/>	<input type="radio"/>
Gentile	<input type="radio"/>	<input type="radio"/>	Partying	<input type="radio"/>	<input type="radio"/>	Religious	<input type="radio"/>	<input type="radio"/>
Humble	<input type="radio"/>	<input type="radio"/>	Apathetic	<input type="radio"/>	<input type="radio"/>	Tenacious	<input type="radio"/>	<input type="radio"/>
Original	<input type="radio"/>	<input type="radio"/>	Fearul	<input type="radio"/>	<input type="radio"/>	Atractive	<input type="radio"/>	<input type="radio"/>
4.-	(+)	(-)	5.-	(+)	(-)	6.-	(+)	(-)
Cautios	<input type="radio"/>	<input type="radio"/>	Docile	<input type="radio"/>	<input type="radio"/>	Prepared	<input type="radio"/>	<input type="radio"/>
Determinated	<input type="radio"/>	<input type="radio"/>	Dared	<input type="radio"/>	<input type="radio"/>	Desirous	<input type="radio"/>	<input type="radio"/>
Convinving	<input type="radio"/>	<input type="radio"/>	Loyal	<input type="radio"/>	<input type="radio"/>	Agreement	<input type="radio"/>	<input type="radio"/>
Good Natured	<input type="radio"/>	<input type="radio"/>	Charming	<input type="radio"/>	<input type="radio"/>	Enthusiastic	<input type="radio"/>	<input type="radio"/>
7.-	(+)	(-)	8.-	(+)	(-)	9.-	(+)	(-)
Willpower	<input type="radio"/>	<input type="radio"/>	Trusted	<input type="radio"/>	<input type="radio"/>	Impartial	<input type="radio"/>	<input type="radio"/>
Open mind	<input type="radio"/>	<input type="radio"/>	Agreeable	<input type="radio"/>	<input type="radio"/>	Precise	<input type="radio"/>	<input type="radio"/>
Helpul	<input type="radio"/>	<input type="radio"/>	Tolerant	<input type="radio"/>	<input type="radio"/>	Nervous	<input type="radio"/>	<input type="radio"/>
Spirited	<input type="radio"/>	<input type="radio"/>	Assertive	<input type="radio"/>	<input type="radio"/>	Cheerful	<input type="radio"/>	<input type="radio"/>

Fig. 2. Interface of Test Cleaver

The evaluation of the hard skills is done when the applier tape in the data about certifications and current courses. This data is entered into a fuzzy inference system, which stabilshes if the hard skill is good enough. An example of the interface can be seen in figure 3:

Hard Skills

Certification/ courses area:

Expedition date: Expiration date:

Company or Institution that gave the certification/ courses:

Fig. 3. Interface of System Fuzzy

Where the area of certification is selected by the user selects the certification area (Programming, Network and DataBases), the beginning and the end of the certification course; and finally the company or institution that gave it.

Once the data had been entered this are arranged in the 2 fuzzy sets proposed, the first one regarding the certification time, for this purpose 3 linguistic variables are considered, which can be seen in figure 4.



Fig. 4. Validation Fuzzy Set

The second fuzzy set is a singleton according to the certifying company, which have a value of 1 (Google, Amazon, Microsoft, IBM, Cisco, ITIL, ISACA), 0.75 (CompTIA, PSP) y 0.25(Other companies). That can be seen in the figure 5.

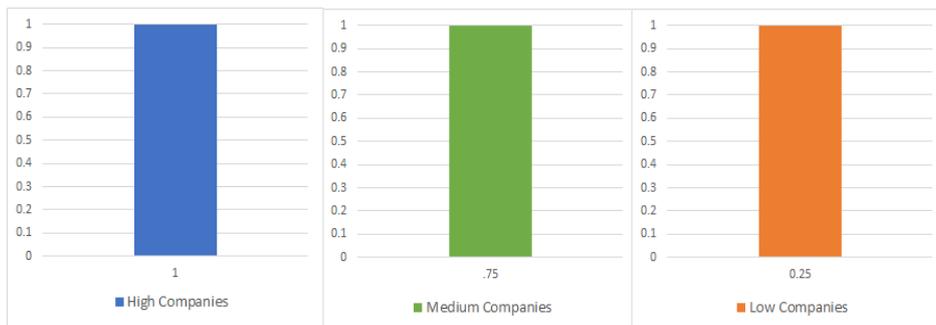


Fig. 5. Singleton Fuzzy Set

In this way for the 2 inputs (Months, Company) membership degree to the fuzzy sets is determined and associated. Immediately after this, the fuzzified data is enter in the 9 previously designed fuzzy rules, which are:

- IF validation = Excellent and company = High THEN Hard_skill = High,
- IF validation = Excellent and company = Medium THEN Hard_skill = Medium,
- IF validation = Excellent and company =Low THEN Hard_skill = Medium,
- IF validation = Good and company = High THEN Hard_skill =High,
- IF validation = Good and company = Medium THEN Hard_skill = Medium,
- IF validation = Good and company = Low THEN Hard_skill = Low,
- IF validation = Average and company = High THEN Hard_skill = Medium,
- IF validation = Average and company = Medium THEN Hard_skill = Low,
- IF validation = Average and company = Low THEN Hard_skill = Low.

After the activation of the corresponding rules, the aggregation of the selected output is done, this means to make a combination of the membership function of all the consequents of the rules to get a unique fuzzy set for each output variable. The final result is shown by a crisp value which represent the hard skill of the applier.

Having obtained this data is enter to AHP to obtain a comparison between the applier skills and the skills that the market requires. An example of this is shown in figure 6.

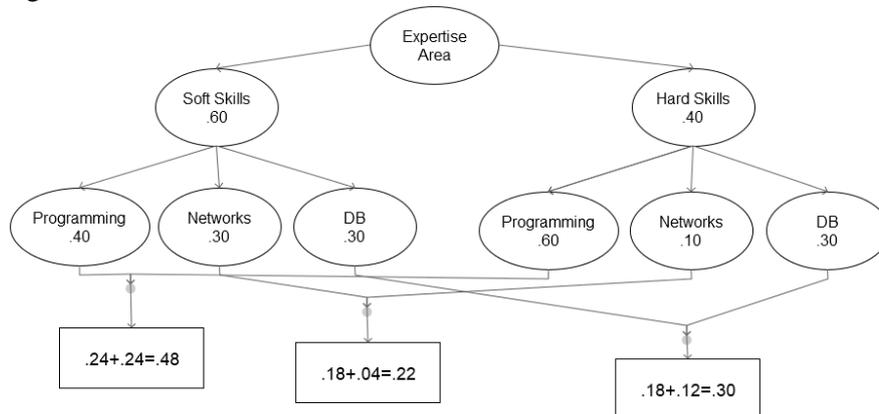


Fig. 6. Data obtained entered into the AHP

Through this process the applier is able to see which area is more appropriated according to the hard and soft skills displayed, most importantly it will show which skills the applier needs to improve according to the current work market.

4. CONCLUSIONS

According to this, the set of methodologies implemented in the proposed system is considered a good alternative, because it's possible to guide an applicant to the requirements of the labor market considering their soft and hard skills, this helps to the applicant to know their weaknesses and in which skills are better. Besides the implemented methodologies complement each other in order to do this type of evaluation; The web Scraping help to create the general profile, while the Test Cleaver and Inference Fuzzy System help to know the applicant skills. Thus, with the data, the applicant is evaluated by AHP according to the current work market.

Therefore, the system fulfills its purpose, which is given to the applicant a profile considering their soft and hard skills according to the current labor market.

REFERENCES

- Algur, S., Bhat, P., & Kulkarni, N. (2016). Educational Data Mining: Classification Techniques for Recruitment Analysis. *International Journal of Modern Education and Computer Science (IJMECS)*, 8(2), 59–65. doi:10.5815/ijmeecs.2016.02.08
- Broucke, S., & Baesens, B. (2018). *Practical Web Scraping for Data Science: Best practices and Examples with Python*. New York: Apress. doi: 10.1007/978-1-4842-3582-9
- Gil-Gaytán, O. L., & Núñez-Partido, A. (2017). Rasgos de personalidad de exportadores mexicanos con éxito. *Revista Academia & Negocios*, 3(1), 23–34.
- Koutra, G., Barbounaki, S., Kardaras, D., & Stalidis, G. (2017). A multicriteria model for Personnel selection in Maritime Industry in Greece. In *2017 IEEE 19th Conference on Business Informatics (CBI)* (pp. 287–294). Thessaloniki. doi:10.1109/CBI.2017.52
- Mamdani, E. H., & Assilan, S. (1975). An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. *International Journal of Man-Machine Studies*, 7(1), 1–13. doi:10.1016/S0020-7373(75)80002-2
- Rianto, Budiyanto, D., Setyohadi, & Suyoto. (2017). AHP-TOPSIS on Selection of New University Students and the Prediction of Future Employment. In *2017 1st International Conference on Informatics and Computational Sciences (ICICoS)* (pp. 125–130). Semarang. doi: 10.1109/ICICoS.2017.8276349
- Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353. doi:10.1016/S0019-9958(65)90241-X