Evaluating student-internship fit by using fuzzy linguistic terms and a fuzzy OWA operator

Jennifer Nguyen
GREC Research Group
Industrial Engineering School
Universitat Politècnica de Catalunya
Barcelona, Spain
Email: jennifer.nguyen@upc.edu

Germán Sánchez-Hernández, Núria Agell,
Albert Armisen, Xari Rovira
GREC Research Group
ESADE Business School
Ramon Llull University
Barcelona, Spain
Email: {german.sanchez, nuria.agell,
albert.armisen, xari.rovira}@esade.edu

Cecilio Angulo
GREC Research Group
Automatic Control Department
Universitat Politècnica de Catalunya
Barcelona, Spain
Email: cecilio.angulo@upc.edu

Abstract—**Personnel selection is a well-known problem that is made difficult by incomplete and imprecise information about candidate and position compatibility. This paper shows how positions, which satisfy candidate's interests, can be identified with fuzzy linguistic terms and a fuzzy OWA operator. A set of relevant positions aligned with a student's interests is selected using this approach. The implementation of the proposed method is illustrated using a numerical example in a business application.**

I. Introduction

Human resource management (HRM) is an important contributor to an organization's competitive strategy [1], [2]. With the growth of globalization, HRM has become more challenging, forcing many organizations to seek efficient tools to optimize their resources' performance and productivity [3]. One aspect of HRM where this is evident is the personnel selection process which aims to identify the best candidate for a position. Some papers indicate that job satisfaction is predictive of performance [3]. Job satisfaction has also been shown to be correlated with withdrawal behaviors such as turnover, absenteeism, and lateness. The area which best predicts job satisfactions is the nature of the work which refers to "intrinsic job characteristics", such as job challenge, autonomy, variety, and scope [3]. Therefore, our approach focuses on the candidate and matches positions to them rather than vice versa.

For global organizations, finding an expert can be challenging as experts are disperse and vary in level of knowledge of a topic. Their knowledge is difficult to qualify and changes frequently [4]. Expertise recommender systems (ERSs) help users to identify informed people. In this paper we explore the application of ERSs for personnel selection in an academic setting. First, we identify characteristics of each candidate and position. Then, we propose a fuzzy order weighted averaging (FOWA) methodology to group positions. Finally, we propose an intelligent system based on this approach.

Many organizations use e-recruitment systems to automate the process of attracting many candidates to each position [5]. Considering the time and resources required to manually evaluate each candidate, several systems have been developed to make the process more efficient. Multi-criteria decision-aiding (MCDA) methods have been applied to many studies in this area as they are well suited to address the complexity of the problem [1]. Other proposed methods include associating semantically equivalent concepts extracted from job descriptions and candidate resumes [6], and employing traditional information retrieval [7] and recommender systems [8].

This paper introduces a methodology which utilizes natural language processing to implicitly create position profiles and applies MCDA methods to propose positions according to a candidate's interests. Its aim is to facilitate the job search process by identifying positions which best match the interests of a candidate narrowing the focus of the candidate's search. Specifically, the system extracts semantic concepts from job descriptions and applies a fuzzy OWA operator to compute a candidate-job fit.

The rest of the paper is organized as follows. First, a review of related work is presented. Second, we propose an intelligent system which applies FOWA to develop our theory, and explain our methodology. An illustrative example of the proposed application is provided in a business application. Finally, conclusions and areas for future research are discussed.

II. Related Work

The personnel selection problem has been studied quite extensively [1], [5], [9]. In this section, we review and compare related research in personnel selection with specific attention to applications of MCDA to the problem. Nearly all of the papers reviewed assess candidates with respect to a position's requirement. As personnel selection is a two-sided problem, our study proposes to address the problem from the less studied point of view. Therefore, we define a support system for students to choose among a set of alternative internships. However, both sides of the problem share the main characteristics of defining applicant and job profiles, and an assignment process. We characterize the existing literature according to three dimensions that consider the *ranking method*, *feature weights*, and *case implementation*. The first dimension, ranking method, refers to the method by which the candidates for a position are ranked according to their qualifications. The second dimension, feature weights, considers how the importance of each feature for a position is assigned. The third dimension, case implementation, is based on four components: *environment*, *number*

of positions, *number of candidates*, and *number of features*. Environment refers to how the methodology was executed, number of positions refers to the number of jobs to which the case attempted to assign candidates, number of candidates refers to the number of candidates each case tried to assign to a position, and the number of features refers to the number of evaluation criteria assessed.

As can be seen in Table I, most of the papers implement an illustrative case. In these instances, the position features are created for illustration purposes. With regards to the feature weights, most of the papers rely on the recruiter or decision maker to provide these values. The range of approaches considered for the matching method is very wide, varying from crisp to fuzzy methods.

Our proposed method differs from existing methods for two reasons. First, it extracts position features from job descriptions without requiring the need for manual entry. Second, an automated matching process, based on an aggregation function defined by a fuzzy OWA operator, allows the simultaneous use of the relevant features without any filtering process.

III. The Proposed Intelligent System

Multi-criteria decision-aiding systems are designed to help users in situations where there are several decision factors that may cause controversy or complexity in decision processes [14], [15]. When these factors are related to user preferences but not easily measurable, the introduction of fuzzy and linguistic descriptions brings an appropriate framework [16], [17]. Multi-criteria decision support systems are comprised of several steps. First, the set of alternatives to be considered are introduced into the system. Second, the user or decision maker (DM) introduces his/her preferences with regards to different criteria. Finally, the system ranks or selects the alternatives that are closest to the user preferences. In this section we introduce a MCDA system to support college students with the internship job market application process.

A working version of the system is currently being tested as well as its interface, designed. It has been developed in R. The main contribution of our work is the design and development of a decision support system based on fuzzy OWA to aggregate uncertain information.

A. System Description

Much like online job boards, university career services own a database of available positions. Companies post internship offerings for the upcoming year that can be reviewed by students online. Each internship has a record with information about the position such as its title, organization, and requirements, all of which are qualitative values. Each piece of information is provided in a free text field making the information unstructured and difficult for a student to search for any position by keyword alone. Therefore, the system is intended to help students identify internship offerings which best match their individual interests. To accomplish this task profiles are created for each student and position to represent interests and features of each, respectively. Interests are student preferences elicited from each student and features are requirements determined from each position. Student's interests are compared with each position's features. The outputs of the

decision-making model are internship positions sorted in a manner which represents student's preferences.

B. Linguistic Definition of Preferences, Requirements, and Profiles

To create the student and position profiles, we consider linguistic terms based on a qualitative absolute order-of-magnitude model [18], [19] which allows us to deal with the imprecision and hesitance involved in decision processes.

Let $\mathbb{S}_n$ be a finite set of totally ordered basic terms, $\mathbb{S}_n = \{B_0, \ldots, B_n\}$, with $B_0 < \ldots < B_n$. In general, each term corresponds to a linguistic label, such as "None" < "Low" < "Medium" < "High" ($n = 3$). The total order in the set of basic terms, $\mathbb{S}_n$, is based on the lexicographic order such that: given two linguistic terms, $B_i, B_j \in \mathbb{S}_n$, $B_i \leq_L B_j$, iff $i < j$ or $i = j$.

From this point forward, we consider $\mathbb{S}_n^*$, a subset of $\mathbb{S}_n$, when the set of basic elements is $\mathbb{S}_n^* = \{B_1, \ldots, B_n\}$. In addition, in $\mathbb{S}_n$ we consider the subset inclusion to define the relation *"to be more precise or equal to"*. We can say that Bi is *more precise than or equal to Bj*, $B_i \preceq B_j$, if and only if, $Bi \subset B_j$.

C. Creating Profiles

Before the process begins all of the curricula vitae (CVs) of the participating students for the internship cycle are collected. From these CVs a set of features are determined to represent the main interests of the student body and define features for positions. To obtain these features, Latent Dirichlet allocation (LDA) is applied to the entire set of CVs. Originally developed by Blei et al. [20], LDA is an unsupervised topic modeling method. It is a generative probabilistic model of a collection of documents. Each document is represented as a mixture of latent features based on keywords. Once the features have been obtained the system user interface is updated to reflect them.

This interface is made available to students to perform a search to positions matching their interests. When students enter the system they will see a set features available to them. Each student selects the features which reflect his/her interests. Then, for each feature, the student also indicates his/her level of interest. Therefore, for each student, Y_j, the vector $Y_j = (Y_{j1}, ..., Y_{jk}) \in (\mathbb{S}_n^*)^k$, with $k \leq K$, is setted corresponding to his/her selected interests expressed in fuzzy linguistic terms.

Initially the entire collection of internship postings are possible alternatives for every student. In order to be able to match these positions with the interests of each student, the features of each position needs to be determined. One output of the LDA performed is a set of keywords related to each feature. Each position is then valued according to this set of keywords and a fuzzy linguistic term for each feature is computed.

D. Identify Positions of Interest

Once the student and position profiles have been created, a matching is performed between the interests of the student and the features of each position where only the features of each position representing the interests of the student are retained. Therefore, the position's vector is redefined as

TABLE I: Applications of MCDA to personnel selection

Paper	Ranking Method	Weights	Case Implementation			
			Environment	# Po	# Ca	# Fe
Canós and Liern, 2008 [10]	OWA and parametric aggregation	Learned weights and FWA	Illustrative example	1	5	6
Güngör et al., 2009 [11]	Comparison of fuzzy AHP and Yager's weighted method	Predetermined by recruiter	Illustrative example	1	6	17
Faliagka et al., 2012 [5]	AHP pairwise comparison	Predetermined by recruiter	Use case	3	100	4
Kabak et al., 2012 [9]	Fuzzy TOPSIS and fuzzy ELECTRE	Fuzzy ANP computations to determine weights	Illustrative example	1	6	10
Baležentis et al., 2012 [1],	MULTIMOORA for group decision making using FWA operator	Predetermined by recruiter	Illustrative example	1	4	8
Yu et al., 2013 [12]	GHFPWA and GHFPWG operator used to aggregate HFE	Prioritized average (PA) operator	Illustrative example	5	5	4
Faliagka et al., 2014 [13]	Learning to rank	N/A	Use Case	3	100	4

$X_i = (X_{i1}, \ldots, X_{ik}) \in (\mathbb{S}_n)^k$ and is compared to the student's interests, $Y_j = (Y_{j1}, \ldots, Y_{jk}) \in (\mathbb{S}_n^*)^k$. We analyze the existing matching between these vectors, comparing each component, by means of the fuzzy matching operator $*$, and a FOWA (fuzzy ordered weighted average).

Definition 1: The fuzzy matching operator is the map

$$* : \mathbb{S}_n \times \mathbb{S}_n^* \to \mathbb{S}_n$$

such that: $\forall B_i \in \mathbb{S}_n$ and $\forall B_j \in \mathbb{S}_n^*$, $B_i * B_j = B_{\min(n, n-(j-i))}$. The system is currently set with $n = 3$. The results of the fuzzy matching operator are shown in Table II

Therefore, the fuzzy matching between X and Y is defined as:

$$X * Y = (X_1 * Y_1, \ldots, X_k * Y_k) \in (\mathbb{S}_n)^k$$

TABLE II: Fuzzy matching operator *

*	Low (B_1)	Medium (B_2)	High (B_3)
None (B_0)	Medium (B_2)	Low (B_1)	None (B_0)
Low (B_1)	High (B_3)	Medium (B_2)	Low (B_1)
Medium (B_2)	High (B_3)	High (B_3)	Medium (B_2)
High (B_3)	High (B_3)	High (B_3)	High (B_3)

The outcomes of the matching are linguistic labels, $\mathbb{S}_n$, that are assigned to a matching vector, $Z = (Z_1, \ldots, Z_k)$, based on the position's ability to satisfy the interests of the student.

E. Fuzzy OWA operator

Definition 2: Given $Z = (Z_1, \ldots, Z_k) \in (\mathbb{S}_n)^k$ we define the fuzzy ordered weighted average operator $\Phi : (\mathbb{S}_n)^k \to \mathbb{S}_n$ as follows:

$$\Phi(Z_1, \ldots, Z_k) = B_r$$

where r is the rounded value of $\sum_{i=1}^{k} w_i \cdot \varphi(Z_{(i)})$, with: $Z_{(i)}$ having the same terms as Z_i ordered from the largest to the smallest by means of the total order $\leq_L$, a set of decreasing weights, w_i, such that $w_i \in [0, 1]$ and $\sum_{i=1}^{k} w_i = 1$, and an increasing function with respect to $\leq_L$, $\varphi : \mathbb{S}_n \to \mathbb{R}$, such that $\varphi : (B_s) = s, \forall s \in \{0, \ldots, n\}$.

Definition 3: Given $X \in (\mathbb{S}_n)^k$, $Y \in (\mathbb{S}_n^*)^k$, we define the degree of fitness of X to Y by means of the composition

between the operator $*$ and the function Φ defined previously, i.e.: $\phi_Y(X) = \Phi(X_1 * Y_1, \ldots, X_k * Y_k)$.

To define the set of weights, w_i, we consider the Regular Increasing Monotone (RIM) function guided by the fuzzy linguistic quantifier 'most of', expressed as:

$$w_i = Q\left(\frac{i}{k}\right) - Q\left(\frac{i-1}{k}\right), i = \{1, \ldots, k\}$$

where $Q(x) = x^{\frac{1}{2}}$.

The degree of satisfaction is obtained from the fitness function in Definition 3. Positions falling within the highest level of satisfaction are proposed. Note that the number of positions proposed can vary between students depending on the student interests and their match with each position's features. Figure 1 displays the output of the system.

IV. NUMERICAL EXAMPLE

Suppose that the student internship hiring season has begun. In order to identify positions which match a student's interest, he/she enters his/her interests into the system. For each interest, the student may select a level of interest as $B_0 =$ "None", $B_1 =$ "Low" $< B_2 =$ "Medium" $< B_3 =$ "High". Consider a student has selected the following interests and level of interests:

1) Marketing (High)
2) Located in Spain (Low)
3) Strategy (High)
4) Business conducted in English (Medium)

The student profile is represented in vector Y_j as in Table III. Students are given six features from which they may select their level of interests. Here, the student is interested in just four of the six features.

TABLE III: Student interest in all features

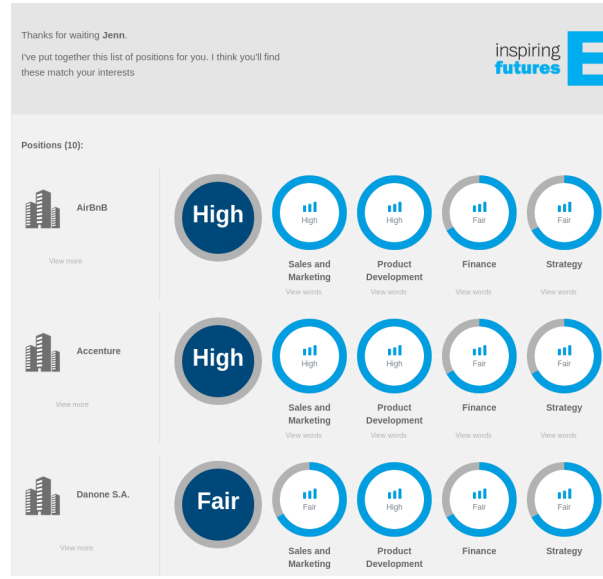| Marketing	Located in Spain	Strategy	Bus. Cond. English	Finance	Consulting
High	Low	High	Med	None	None

Fig. 1: Positions sorted by the level of satisfaction

The method described in Section III is applied to solve the problem and is computed as follows:

Step 1: Consider the fuzzy linguistic term computed for each position X_i and feature k as in Table IV.

TABLE IV: Position features

	Market	Loc. in Spain	Strategy	Bus. Cond. English	Finance	Consult
pos 1	Low	High	Med	None	None	None
pos 2	High	High	Med	Low	None	High
pos 3	Med	Low	Med	High	None	High
pos 3	Med	Low	Med	High	None	High
pos 4	None	High	Low	None	Med	High
pos 5	Low	None	Low	High	None	High

Step 2: Perform a matching between the interests of the student and the features of each position retaining only the features of each position related to the interests of student as in Table V.

TABLE V: Fuzzy Matching

	Market	Loc. in Spain	Strategy	Bus. Cond. English
pos 1	1	3	2	1
pos 2	3	3	2	2
pos 3	2	3	2	3
pos 4	0	3	1	1
pos 5	1	2	1	3

Step 3: Aggregate the fuzzy matching matrix using FOWA over the features with the weighting vector as in Table VI. Compute the weighting vector from Definition 3 as $w = \{0.500, 0.207, 0.159, 0.134\}$.

From Table VI, the positions available to this student are separated into two levels: "Medium" and "High". The system displays the highest level. In this case positions 2 and 3 are shown to the student.

TABLE VI: Fuzzy OWA

					Average	Rounded
pos 1	1.5	0.414	0.159	0.134	2.21	2
pos 2	1.5	0.621	0.318	0.268	2.71	3
pos 3	1.5	0.621	0.318	0.268	2.71	3
pos 4	1.5	0.207	0.159	0	1.87	2
pos 5	1.5	0.414	0.159	0.134	2.21	2

V. CONCLUSION AND FUTURE RESEARCH

In this paper, a new method for sorting internship postings according to student interests has been introduced. This methodology improves existing methods in several ways. First, it proposes to perform a matching between students and internships from the perspective of the job candidate rather than the position. This is the reverse of the more popular matching to find the best candidate for a position. More specifically, the system is directed at students or new graduates with very little experience. Their interests may be a better representation of themselves since they have less relevant experience than seasoned professionals. In addition, as students may have had limited exposure to their fields of interest, they may not be aware of which keywords to use or they may not be aware of what types of available positions match their interests. A system such as this can facilitate the search process by narrowing the list of positions to the ones that best satisfy student interests. Second, the method considers a FOWA operator in the matching to capture the inherent uncertainty of personnel selection. Futhermore, the FOWA operator avoids filtering but simultaneously considers several relevant variables for the aggregation process.

Our methodology can be extended to both sides of the general personnel assignment problem making the process more efficient. A position which is closely aligned with the interests of a job candidate may lead to better job loyalty. Therefore, as future research, we propose to adapt our methodology to

other personnel selection environments like headhunting firms, online job boards, and industry human resources to uncover the interests of a job candidate prior to the interview process.

REFERENCES

[1] A. Baležentis, T. Baležentis, and W. K. Brauers, "Personnel selection based on computing with words and fuzzy multimoora," *Expert Systems with applications*, vol. 39, no. 9, pp. 7961–7967, 2012.

[2] S. Bouajaja and N. Dridi, "A survey on human resource allocation problem and its applications," *Operational Research*, pp. 1–31, 2016.

[3] L. M. Saari and T. A. Judge, "Employee attitudes and job satisfaction," *Human resource management*, vol. 43, no. 4, pp. 395–407, 2004.

[4] M. T. Maybury, "Discovering distributed expertise," *Regarding the âĂIJIntelligenceâĂİ in Distributed Intelligent SystemsMITRE*, 2007.

[5] E. Faliagka, A. Tsakalidis, and G. Tzimas, "An integrated e-recruitment system for automated personality mining and applicant ranking," *Internet research*, vol. 22, no. 5, pp. 551–568, 2012.

[6] M. Mochol, H. Wache, and L. Nixon, "Improving the accuracy of job search with semantic techniques," in *International Conference on Business Information Systems*. Springer, 2007, pp. 301–313.

[7] A. Singh, C. Rose, K. Visweswariah, V. Chenthamarakshan, and N. Kambhatla, "Prospect: a system for screening candidates for recruitment," in *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010, pp. 659–668.

[8] J. Malinowski, T. Keim, O. Wendt, and T. Weitzel, "Matching people and jobs: A bilateral recommendation approach," in *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, vol. 6. IEEE, 2006, pp. 137c–137c.

[9] M. Kabak, S. Burmaoğlu, and Y. Kazançoğlu, "A fuzzy hybrid mcdm approach for professional selection," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3516–3525, 2012.

[10] L. Canós and V. Liern, "Soft computing-based aggregation methods for human resource management," *European Journal of Operational Research*, vol. 189, no. 3, pp. 669–681, 2008.

[11] Z. Güngör, G. Serhadlıoğlu, and S. E. Kesen, "A fuzzy ahp approach to personnel selection problem," *Applied Soft Computing*, vol. 9, no. 2, pp. 641–646, 2009.

[12] D. Yu, W. Zhang, and Y. Xu, "Group decision making under hesitant fuzzy environment with application to personnel evaluation," *Knowledge-Based Systems*, vol. 52, pp. 1–10, 2013.

[13] E. Faliagka, L. Iliadis, I. Karydis, M. Rigou, S. Sioutas, A. Tsakalidis, and G. Tzimas, "On-line consistent ranking on e-recruitment: seeking the truth behind a well-formed cv," *Artificial Intelligence Review*, vol. 42, no. 3, pp. 515–528, 2014.

[14] S. Greco, J. Figueira, and M. Ehrgott, "Multiple criteria decision analysis," *Springer's International series*, 2005.

[15] R. Ureña, F. Chiclana, J. A. Morente-Molinera, and E. Herrera-Viedma, "Managing incomplete preference relations in decision making: a review and future trends," *Information Sciences*, vol. 302, pp. 14–32, 2015.

[16] C. Carlsson and R. Fullér, "Fuzzy multiple criteria decision making: Recent developments," *Fuzzy sets and systems*, vol. 78, no. 2, pp. 139–153, 1996.

[17] F. J. Cabrerizo, I. J. Pérez, F. Chiclana, and E. Herrera-Viedma, "Group decision making: Consensus approaches based on soft consensus measures," in *Fuzzy Sets, Rough Sets, Multisets and Clustering*. Springer, 2017, pp. 307–321.

[18] L. Travé-Massuyès, F. Prats, M. Sánchez, and N. Agell, "Relative and absolute order-of-magnitude models unified," *Annals of Mathematics and Artificial Intelligence*, vol. 45, no. 3, pp. 323–341, 2005.

[19] N. Agell, M. SáNchez, F. Prats, and L. Roselló, "Ranking multi-attribute alternatives on the basis of linguistic labels in group decisions," *Information Sciences*, vol. 209, pp. 49–60, 2012.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.