

# A linguistic multi-criteria decision-aiding system to support university career services

J. Nguyen<sup>a</sup>, G. Sánchez-Hernández<sup>b</sup>, A. Armisen<sup>b</sup>, N. Agell<sup>b,\*</sup>, X. Rovira<sup>b</sup>, C. Angulo<sup>a</sup>

<sup>a</sup>*GREC Research Group, Universitat Politècnica de Catalunya (BarcelonaTech), Spain*

<sup>b</sup>*GREC Research Group, ESADE Business School, Ramon Llull University, Spain*

---

## Abstract

In this paper we introduce a linguistic multi-criteria decision-aiding model to support college students with the internship job market application. It considers a fuzzy ordered weighted averaging (FOWA) operator in the matching to capture the inherent uncertainty and vague nature of personnel selection processes. The decision model is integrated in a software tool able to capture data from university student resume and internship databases. The application assesses position characteristics implicitly by means of linguistic descriptions according to each student's preferences. The software tool is enabled with the ability to propose positions according to student preferences. The system selects a reduced list of alternatives from the set of job offers, helping students to decide on which positions to focus their applications.

*Keywords:* Decision support systems, Multi-criteria decision-aiding, Hesitant fuzzy linguistic terms, Fuzzy OWA operator, Personnel selection problem

---

## 1. Introduction

Organizations are challenged daily to make complex decisions. These decisions can be subjective, uncertain, and imprecise [1]. As data becomes continually available, these decisions become increasingly more complex, making

---

\*Corresponding author: Tel.: +34 932 806 162 (ext 2272)

Email address: [nuria.agell@esade.edu](mailto:nuria.agell@esade.edu) (N. Agell)

5 the role of decision support tools more important. Specifically, this notion can  
be observed within human resource personnel selection. In general, person-  
nel selection depends on a firm's specific targets, and the preferences of the  
hiring managers [2] and candidates. Therefore, Multi-Criteria Decision-Aiding  
(MCDA) methods and fuzzy logic are an appropriate solution to capture the  
10 multi-criteria and uncertainty aspects inherent in the process [2, 3].

For global organizations, human resources personnel selection can be chal-  
lenging as candidates are disperse and vary in level of knowledge of a topic.  
Their knowledge is difficult to qualify and changes frequently [4]. Personnel  
selection is subjective in nature with regards to assigning crisp values to the  
15 job requirements and evaluating candidate qualifications. Previous studies have  
extended MCDA methods to this problem to address its fuzziness [2, 3, 5].

Within universities, obtaining an internship is a specific personnel selection  
process. It may be the first time a student is applying for a position. Therefore,  
the terms used to describe the desired position may be unfamiliar making the  
20 job search process overwhelming. Students may not know which terms to use  
when searching for a specific position or for which position their skills are most  
relevant. Hence, the positions obtained in their search results may not be the  
best match for them. There are two different perspectives to personnel selection.  
The hiring company is looking for the best candidate to fill a position. On the  
25 other hand, the candidate is looking for a position which satisfies their interests.  
Knowing on which positions to focus their time is key to both the student and  
the hiring company.

The aim of this paper is to introduce a practical decision support system  
to assist students with identifying positions most related to their interests. A  
30 real case example is implemented with student and job information provided  
by a university's career services office. In terms of feature representation, the  
novelty of the application is two-fold. First, the requirements of a position are  
extracted in an implicit manner and represented via linguistic terms. Second,  
linguistic terms are also considered to represent students' interests. The model  
35 considered for linguistic descriptions is the hesitant fuzzy linguistic model. This

model was introduced by Rodriguez et al. in [6] and further developed in [7].

The rest of the paper is structured as follows. First, a review of current MCDA applications to the personnel selection problem is presented. Next, we discuss tools used in the design of a linguistic MCDA system which include linguistic descriptors, and fuzzy matching and aggregation. These tools are applied to a decision support system to help students in the selection of their internship, presented in Section 4. Following the explanation of the methodology, a real case is provided with the implementation of the proposed method. Lastly, conclusions are presented and future research directions are proposed in Section 6.

## 2. State-of-the-Art in Personnel Selection

The personnel selection problem has been studied quite extensively [8, 3, 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 has four components: a) environment, b) number of positions, c) number of candidates, and d) 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

65 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 1, most of the papers analyzed in the literature  
review implemented an illustrative case while only two papers had use cases. In  
the first group of papers, the authors selected positions, candidates, and features  
70 to suit their illustrative example. The features selected were estimated based  
upon their specific positions. Regarding the number of candidates considered  
in each paper, only two papers had 100 candidates while the others had six  
or fewer. The lower number of candidates may be to facilitate the illustrative  
example while the papers with 100 candidates had a full web implementation.

75 Our proposed method differs from existing methods for several reasons. As  
students, rather than positions, are the main focus of our method we propose to  
elicit the features from the students. We incorporate an existing automatic topic  
modeling technique to extract these features. Therefore, the number of features  
80 considered is determined through a process defined in [14] and is tailored to the  
students. Next, the required features are identified for each job description ap-  
plying a posterior distribution based on the previously defined features. Lastly,  
an automated matching process, based on an aggregation function defined by a  
FOWA operator, allows the simultaneous use of the relevant features without  
85 any filtering process. Specifically, each component of a student’s interests and  
position’s features are compared by a fuzzy matching operator and aggregated  
with an ordered weighted averaging operator (OWA), introduced by Yager and  
Kacprzyk [15], to obtain a fuzzy linguistic label.

We present a real case study with 275 students. These students were the  
90 actual internship candidates for a business school in 2016. Given that these  
students were from the same college with similar backgrounds it is expected  
that they would compete for the same positions. Therefore, this scenario is  
analogous to the personnel selection problem, which human resource managers  
face, with many candidates for a single position.

Table 1: 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 [9]	AHP	Predetermined by recruiter	Use case	3	100	4
Kabak et al., 2012 [8]	Fuzzy TOPSIS and fuzzy ELECTRE	Fuzzy ANP computations to determine weights	Illustrative example	1	6	10
Baležentis et al., 2012 [3],	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 hesitant fuzzy elements (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

### 95 3. Preliminaries

In this section, we briefly present the necessary tools to design a linguistic multi-criteria decision-aiding system, that is, the concept of fuzzy matching for linguistic descriptions and fuzzy aggregation operators for the selection of alternatives.

100 *3.1. Linguistic descriptions and Fuzzy matching*

To introduce a decision support system which proposes available positions to college students, there are some uncertainties that should be considered in evaluating the students' interests. The uncertainty is inherent in students' abilities to communicate their affinity for specific features of a position. Having had little  
105 experience with these features, it may be difficult to express their preferences as a single label. Given this uncertainty, as mentioned in the introduction, we propose the application of Hesitant Fuzzy Linguistic Term Set (HFLTS) [7] to manage the need for several labels to define preferences.

Other linguistic modeling techniques could have been considered such as  
110 multi-granular linguistic modeling [16], computing with words based on discrete fuzzy numbers [17], 2-tuple linguistic modeling [18], or linguistic modeling based on ordinal symbolic information [19]. In fact, our method could be considered a multi-granular linguistic model as it considers different levels of granularity in the linguistic assessments. However, in general, multi-granular linguistic  
115 modeling methods aggregate the opinions of experts across all of the alternatives prior to ranking them. In contrast, we propose to use a matching operator which enables matching student preferences to position features on an individual attribute and student level, and then computes an overall score. Secondly, with respect to computing with words based on discrete fuzzy numbers and 2-tuple  
120 linguistic modeling, experts would be required to provide additional information regarding the grade of the value contained in the semantic support as part of their qualitative or linguistic evaluation. In our method, we require less information from the participants because specifying a grade of a value would be difficult as students may not have this information. Finally, if we consider  
125 linguistic modeling based on ordinal symbolic information, as defined in [19], experts would be asked to pairwise compare features. In the context of this real case, students would not have the flexibility to express their preferences as "I don't know" which may be the case if they have had no experience with a feature.

130 The approach proposed in this paper relies on the use of linguistic terms

based on a qualitative absolute order-of-magnitude model [20, 21] that allows us to deal with the imprecision and hesitance involved in decision processes. We will express this model by means of HFLTS introduced by Rodriguez et al. [7].

Let  $\mathbb{S}_n$  be a finite set of totally ordered basic terms,  $\mathbb{S}_n = \{B_0, \dots, B_n\}$ ,  
 135 with  $B_0 < \dots < B_n$  and the hesitant fuzzy linguistic terms set,  $H_{\mathbb{S}_n}$ , be the set of all consecutive linguistic basic terms of  $\mathbb{S}_n$ , i.e.  $B_{ij} = \{x \in \mathbb{S}_n \mid B_i \leq x \leq B_j\} \forall i, j \in \{0, \dots, n\}$ , with  $i \leq j$ . In general, each term corresponds to a linguistic label, with  $B_0$  being the term “None”. For simplicity, we will denote the singleton  $B_{ii} = B_i$ . The total order in the set of basic terms,  $\mathbb{S}_n$ , allows us  
 140 to define a total order in  $H_{\mathbb{S}_n}$  based on the lexicographic order such that: given two linguistic terms,  $B_{ij}, B_{i'j'} \in H_{\mathbb{S}_n}$ ,  $B_{ij} \leq_L B_{i'j'}$ , iff  $i < i'$  or  $i = i'$  and  $j \leq j'$ .

For instance, let us consider  $n = 3$  and  $B_0 = \text{“None”} < B_1 = \text{“Low”} < B_2 = \text{“Medium”} < B_3 = \text{“High”}$ , then, terms  $B_{12}$  and  $B_{03}$  will represent  
 145 the linguistic labels “Low” or “Medium” and “Unknown” (“None”, “Low”, “Medium”, or “High”), respectively. From the lexicographic order, we get  $B_0 \leq_L B_{03} \leq_L B_1 \leq_L B_{12} \leq_L B_2$ .

From this point forward, we consider  $H_{\mathbb{S}_n^*}$ , a subset of  $H_{\mathbb{S}_n}$ , which corresponds to the HFLTS obtained when the set of basic elements is  $\mathbb{S}_n^* = \{B_1, \dots, B_n\}$ .  
 150 In addition, in  $H_{\mathbb{S}_n}$  we consider the subset inclusion to define the relation “to be more precise or equal to”. We say that  $B_{ij}$  is more precise or equal to  $B_{i'j'}$ ,  $B_{ij} \preceq B_{i'j'}$ , if and only if,  $B_{ij} \subseteq B_{i'j'}$ , i.e.  $i' \leq i$  and  $j \leq j'$ . For instance, in the previous example, we have  $B_1 \preceq B_{02}$  and  $B_{12} \preceq B_{13}$ . Finally, the connected union operator  $\sqcup$ , is considered in  $H_{\mathbb{S}_n}$  defined as  $B_{ij} \sqcup B_{i'j'} = B_{kl}$   
 155 where  $k = \min(i, i')$  and  $l = \max(j, j')$ . Following the previous example,  $H_{\mathbb{S}_3}$ ,  $B_{01} \sqcup B_3 = B_{03}$ .

HFLTS can be used to compare individual’s preferences to object’s attributes to capture imprecision in decision processes. To this end, we will define an operator matching two basic terms and extend it to the entire set of HFLTS  
 160 catching all possible combinations of hesitancy in both descriptions.

**Definition 3.1.** *The fuzzy matching operator is the map*

$$* : H_{\mathbb{S}_n} \times H_{\mathbb{S}_n^*} \rightarrow H_{\mathbb{S}_n}$$

such that:

1.  $\forall B_i \in \mathbb{S}_n$  and  $\forall B_j \in \mathbb{S}_n^*$ ,  $B_i * B_j = B_{\min(n, n-(j-i))}$ ,
2.  $\forall B_{ij} \in H_{\mathbb{S}_n}$  and  $\forall B_{i'j'} \in H_{\mathbb{S}_n^*}$ ,  
 $B_{ij} * B_{i'j'} = \bigsqcup \{B_k * B_l, i \leq k \leq j \text{ and } i' \leq l \leq j'\}$ .

165 Note that, 2. coincides with 1.  $\forall B_i \in \mathbb{S}_n$  and  $\forall B_j \in \mathbb{S}_n^*$ .

**Example 3.1.** *Let us consider that a candidate's preferences are represented by  $H_{\mathbb{S}_n^*}$  and the features of each position are represented by  $H_{\mathbb{S}_n}$ , then given the previously considered HFLTS,  $H_{\mathbb{S}_n^*}$ , with  $n = 3$ , the results of the fuzzy matching*

170 *operator for the basic terms are shown in Table 2.*

Table 2: 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 ( $B_3$ )

Interpreting the table, it can be seen that when the candidate has a "Low" preference for a feature, a position with the same value or higher for the feature is a "High" match. It is considered that the candidate's preference has been met or exceeded. A position having a value one step lower than the candidate's preference is considered a "Medium" match as the feature partially meets the candidate's preference. A value two steps lower is a "Low" match because the preference of the candidate is barely met. Looking at the far right side of the table, when the candidate's preference is "High" but the position value is "None", the difference is three steps lower and the position does not contain this feature



180 resulting in a "None" match leaving the preference unmet for this feature. Continuing with the candidate's preference of "High", a position with a "Medium" value partially meets and a "Low" value barely meets the candidate's preference. Therefore, the match qualities are "Medium" and "Low", respectively.

From Example 3.1, it can be seen that the fuzzy matching operator deliberately returns the value "High" in half of the situations in order to capture the positions with features which meet or exceed student preferences.

**Example 3.2.** To demonstrate how the  $*$  operator works with non-basic labels let us consider,  $B_{02}$  and  $B_{12}$  along with Table 2.  $B_{02} * B_{12} = \sqcup\{B_0 * B_1, B_0 * B_2, B_1 * B_1, B_1 * B_2, B_2 * B_1, B_2 * B_2\} = \sqcup\{B_2, B_1, B_3, B_2, B_3, B_3\} = B_{13}$ .

190 **Proposition 3.1.** The fuzzy matching operator  $*$  fulfills the following properties:

1.  $\forall B_{ij}, B_{i'j'} \in H_{\mathbb{S}_n^*}$ , then  $B_{ij} * B_{i'j'} \neq B_{i'j'} * B_{ij}$ .
2.  $\forall B_{ij} \in H_{\mathbb{S}_n}$  and  $\forall B_{i'j'} \in H_{\mathbb{S}_n^*}$ , with  $B_{i'j'} \leq_L B_{ij}$ , then,  $B_n \preceq B_{ij} * B_{i'j'}$ .
3.  $\forall B_{ij} \in H_{\mathbb{S}_n}$ ,  $B_{ij} * B_n = B_{ij}$ .

195 From Property 1 we can infer that the order always matters when matching two different terms in  $H_{\mathbb{S}_n^*}$ . If the first one is greater than or equal to the second one, the result is less precise than  $B_n$ . In addition, whenever the first label,  $B_{ij}$  is matched with a second label of  $B_n$ , the result is always  $B_{ij}$ . It follows that the element  $B_n$  is neutral with respect to  $B_{ij}$ .

### 200 3.2. Fuzzy aggregation and alternatives selection

Given two  $k$ -dimensional different vectors,  $X = (X_1, \dots, X_k) \in (H_{\mathbb{S}_n})^k$  and  $Y = (Y_1, \dots, Y_k) \in (H_{\mathbb{S}_n^*})^k$ , 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 3.2.** Given  $X \in (H_{\mathbb{S}_n})^k$  and  $Y \in (H_{\mathbb{S}_n^*})^k$ , the fuzzy matching between  $X$  and  $Y$  is defined as:

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

205 **Example 3.3.** Continuing with Example 3.1, given the vectors  $X \in (H_{\mathbb{S}_3})^5$  and  $Y \in (H_{\mathbb{S}_3^*})^5$ ,  $X = (B_2, B_1, B_3, B_0, B_2)$ , and  $Y = (B_2, B_2, B_1, B_3, B_1)$ , the match is  $X * Y = (B_3, B_2, B_3, B_0, B_3)$ . In the same way, if  $X = (B_{02}, B_{12}, B_1, B_{13}, B_0)$ , and  $Y = (B_2, B_1, B_{13}, B_{12}, B_2)$  the match is  $X * Y = (B_{13}, B_3, B_{13}, B_{23}, B_1)$ .

As previously mentioned, we apply an OWA introduced by Yager and Kacprzyk  
 210 [15] to our specific context, to obtain a fuzzy linguistic label from a vector of  $(H_{\mathbb{S}_n})^k$ .

**Definition 3.3.** Given  $Z = (Z_1, \dots, Z_k) \in (H_{\mathbb{S}_n})^k$  we define its weighted average index as:

$$\mu^Z = \sum_{i=1}^k w_i \cdot \varphi(Z_{(i)})$$

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  $\leq_L$ ,  $\varphi : H_{\mathbb{S}_n} \rightarrow \mathbb{R}$ ,  
 215 such that  $\varphi(B_s) = s, \forall s \in \{0, \dots, n\}$ .

For our purpose, we consider the regular increasing monotone (RIM) function, introduced by Yager [22], guided by the linguistic quantifier ‘most of’, expressed as:

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

Note that a RIM function must be used to obtain positive weights  $w_i$ , and  $Q(x) = x^\alpha$  should be defined with  $\alpha \in [0, 1]$  to obtain a concave operator able to model those aggregations with importance associated with them.

**Definition 3.4.** Given  $Z = (Z_1, \dots, Z_k) \in (H_{\mathbb{S}_n})^k$  we define the fuzzy ordered weighted average operator  $\Phi : (H_{\mathbb{S}_n})^k \rightarrow H_{\mathbb{S}_n}$  is defined as follows:

$$\Phi(Z_1, \dots, Z_k) = B_{\mu_1^Z \mu_2^Z}$$

where  $\mu_1^Z$  and  $\mu_2^Z$  are the rounded and ceiling values, respectively.

220 **Definition 3.5.** Given  $X \in (H_{\mathbb{S}_n})^k$ ,  $Y \in (H_{\mathbb{S}_n^*})^k$ , we define the degree of fitness of  $X$  to  $Y$  by means of the composition between the operator  $*$  and the function  $\Phi$  defined previously, i.e.:  $\phi_Y(X) = \Phi(X_1 * Y_1, \dots, X_k * Y_k)$ .

**Example 3.4.** Continuing with Example 3.3, we can consider the increasing function:  $\varphi(B_{sl}) = s + \frac{l-s}{3+1-s}$ ,  $\forall s, l \in \{0, 1, 2, 3\}$  for our example. The function  
 225 chosen for  $\varphi(B_{sl})$  could be defined differently in other contexts. In addition, to define the set of weights,  $w_i$ , we consider the RIM function, guided by the linguistic quantifier ‘most of’, expressed as:

$$w_i = Q\left(\frac{i}{5}\right) - Q\left(\frac{i-1}{5}\right), \quad i = \{1, \dots, 5\}, \quad (2)$$

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

Then, given the matching vector,  $X * Y = (B_{13}, B_3, B_{13}, B_{23}, B_1)$ , between  
 230 vectors  $X$  and  $Y$ , and applying Definitions 3.3, 3.4, and 3.5, the degree of fitness of  $X$  to  $Y$  is  $\phi_Y(X) = \Phi(X_1 * Y_1, \dots, X_5 * Y_5) = \Phi(B_{13}, B_3, B_{13}, B_{23}, B_1) = B_{23}$ . This result comes from the fact that:  $B_3 \geq_L B_{23} \geq_L B_{13} \geq_L B_1$ ,  $\varphi(B_3) = 3$ ,  $\varphi(B_{23}) = \frac{5}{2}$ ,  $\varphi(B_{13}) = \frac{5}{3}$ ,  $\varphi(B_1) = 1$ , and  $w_1 = \sqrt{\frac{1}{5}}$ ,  $w_2 = \sqrt{\frac{2}{5}} - \sqrt{\frac{1}{5}}$ ,  $w_3 = \sqrt{\frac{3}{5}} - \sqrt{\frac{2}{5}}$ ,  $w_4 = \sqrt{\frac{4}{5}} - \sqrt{\frac{3}{5}}$ ,  $w_5 = 1 - \sqrt{\frac{4}{5}}$ .

#### 235 4. Proposed Multi-Criteria decision-aiding system to support university career services

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 [23, 24]. When these factors are related to user  
 240 preferences but not easily measurable, the introduction of fuzzy and linguistic descriptions brings an appropriate framework [25, 26]. 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  
 245

preferences. In this section we introduce a MCDA system to support college students with the internship job market application process.

#### *4.1. System Description*

Much like online job boards, university career services have 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. Therefore, it is difficult for a student to search for any position by keyword alone.

The proposed system caters to the interests of students rather than the requirements of a position. Specifically, 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 preferences and features of each, respectively. Preferences are student interests elicited from each student and features are requirements determined from each position. Student's preferences are compared with each position's features. The outputs of the decision-making model are internship positions sorted in a manner which represents students preferences. A diagram of the process follows in Figure 1 and detailed descriptions of the individual steps are given below.

#### *4.2. Determine Features from All CVs*

Before the process begins all of the curriculum vitae (CV) 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. Although there may be small changes in the features selected from year to year, extracting features specific to the student participants enables the system to better discern between positions. This is particularly important if in a given year all the positions are closely related, e.g. being in a single type of position.

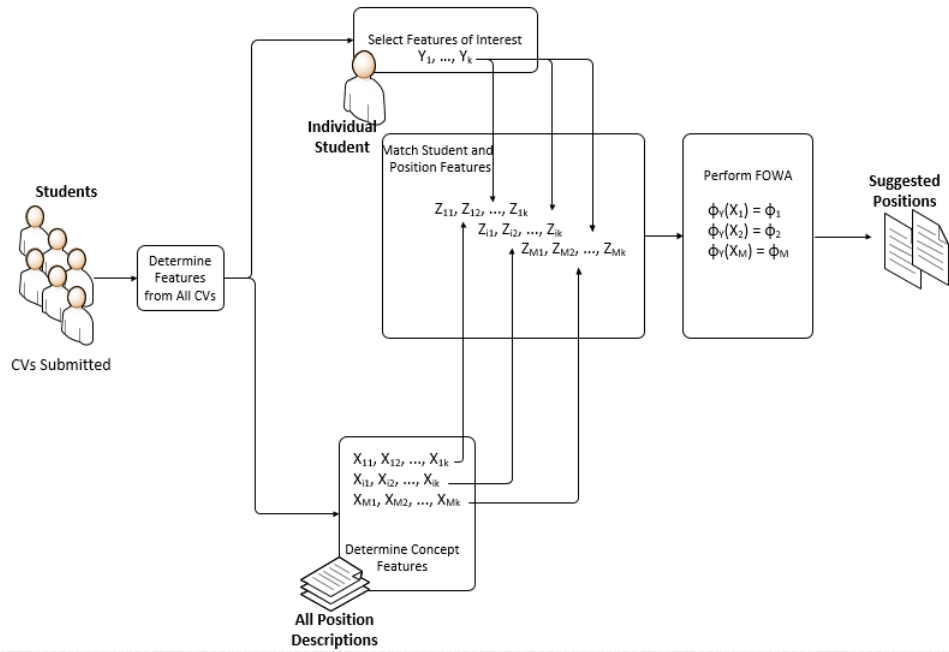


Figure 1: System process flow

275 To obtain these features, Latent Dirichlet allocation (LDA) is applied to  
the entire set of CVs. Originally developed by Blei et al. [27], 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. The number of features,  $K$ , is determined  
280 using a qualitative approach, following [14]. This method consists of varying  
the number of features until an expert can recognize each feature from the  
keywords representing it (e.g. the keywords finance, economical and model are  
associated with finance). When two experts concur on the recognizable features  
the number of features is determined. Once the features have been determined,  
285 the student user interface is updated to reflect the considered features and the  
decision process begins.

### 4.3. Determine Features for Positions and Student

Initially the entire collection of internship postings are possible alternatives for every student. In order to be able to match these positions with the preferences of each student, the features of each position needs to be determined. One output of the LDA performed in Section 4.2 is a set of keywords related to each feature as shown in Table 3. For each student CV, there is a probability distribution of all possible features determined. Because this method seeks to provide equal value to all features, it normalizes each feature according to its distribution across students and jobs, respectively, by applying cutoff values. These cutoff values are then translated into a linguistic term (ie. “None” when  $Feature_i < 10\%$ , “Low” when  $10\% \leq Feature_i < 50\%$ , “Medium” when  $50\% \leq Feature_i < 90\%$ , “High” when  $Feature_i \geq 90\%$ ).

When students enter the system, they will see the features determined in Section 4.2 available to them. The user interface is personalized to reflect each student’s preference and level of preference based on the results of LDA. As LDA was applied to the entire collection of CVs to obtain the underlying features overall, it also computes the probability distribution of these features for each document. Each student may then adjust the feature preferences and levels presented to them, as necessary (e.g. change a feature preference from “Low” to “Medium”-“High”). Therefore, for each student,  $Y_j$ , the vector  $Y_j = (Y_{j1}, \dots, Y_{jk}) \in (H_{S_n^*})^k$ , with  $k \leq K$ , is setted corresponding to his/her selected preferences expressed in hesitant fuzzy linguistic terms, as introduced in Section 3. The following figures detail the system’s user interface. Specifically, from Figure 2 the student selects his/her interests and corresponding level. Note that a student may select a level that corresponds to two label categories (e.g. Jenn’s preference for Sales and Marketing is between “Low” and “Medium”), or at least or at most some level of interest (e.g. Jenn has “at most” a medium preference for Strategy).

Welcome, Jenn.

Begin the search for your future job by entering your interests here and I'll take care of the rest.

inspiring futures **E**

SELECT HOW TO DEFINE YOUR INTERESTS:

Read them from my CV

Let me choose specific interests

SELECT AT LEAST ONE INTEREST AND ITS LEVEL:

Sales and Marketing	Between	Low	Medium
Client and Team Oriented	Single Term	Low	
Technical Skills	Between	Medium	High
Strategy	At most	Medium	
Finance	Between	Low	High

Figure 2: Interest selection user interface

315 4.4. Match Student Interests and Position Features to Propose Positions

Each internship opportunity is an alternative for a specific student,  $Y_j$ . Therefore, to perform a match, we need to create the position profile, expressing the relevance of each feature determined from the collection of CVs. Once the student and position profiles have been created, a matching is performed

320 between the preferences of the student and the features of each position. The system performs the matching process of Section 3.2. The process concludes with a proposed list of positions which best match the interests of the student as shown in Figure 3.

The match is performed between the preferences of the student and the

325 features of the position, where only the features of each position representing the preferences of the student are retained (i.e. the features that the individual identified as "None" are removed). Therefore, the position's vector is redefined as  $X_i = (X_{i1}, \dots, X_{ik})$  and is compared to the student's preferences,  $Y_j = (Y_{j1}, \dots, Y_{jk})$ . The outcomes of the matching are linguistic labels,  $H_{S_n}$ , that

330 are assigned to a matching vector,  $Z = (Z_1, \dots, Z_k)$ , based on the position's  
ability to satisfy the interests of the student.

Once the matching vector is obtained, a fuzzy order weighted average is  
computed. The FOWA, introduced in Section 3.2, is applied to aggregate the  
linguistic terms from the matching step in order to emphasize the features with  
335 the greatest match between students and positions. The resulting level of sat-  
isfaction is a fuzzy linguistic term set  $\phi_Y(X) = \Phi(X_1 * Y_1, \dots, X_k * Y_k)$  obtained  
via the weighted average. Positions falling within the highest level of satis-  
faction are proposed. Note that the number of positions proposed can vary  
between students depending on the student preferences and their match with  
340 each position's features.

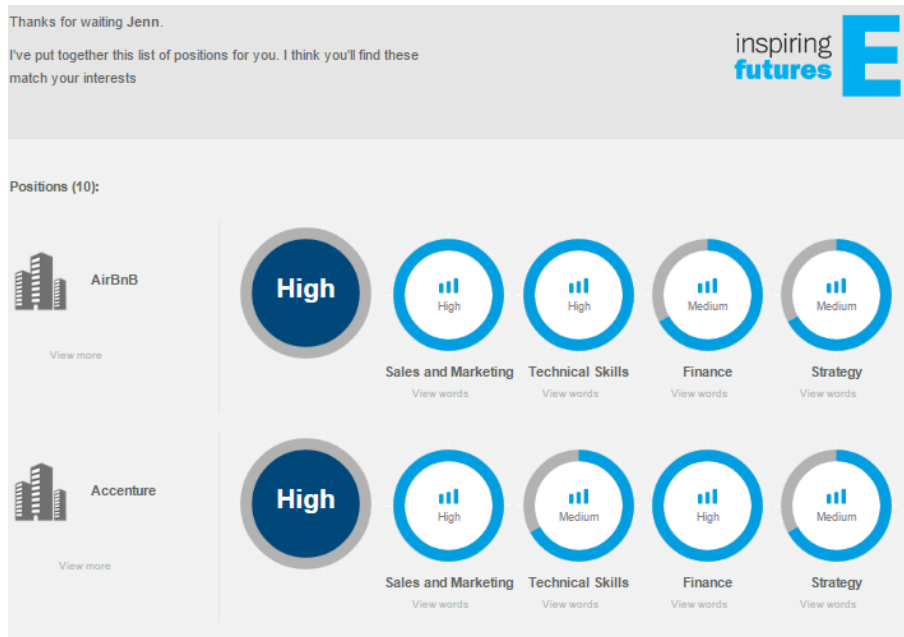


Figure 3: Positions with highest level



## 5. A Real Case Example

In this real case example, the 2016 internship program for the Bachelor of Business Administration at ESADE Business School in Barcelona, Spain, was used to apply the proposed method. This program provides students with the opportunity to gain professional experience at an organisation. For some students, this may be their first-time working in their future profession.

### 5.1. Data Sets

The data set was composed of 275 student resumes and 1063 available internships. All resumes and internship descriptions in English were considered. The final data set consisted of 275 students and 549 internships. Student information was limited to the resumes provided for the purposes of the 2016 internship cycle. Internship positions included national and international postings.

### 5.2. Implementation and Results

We applied Latent Dirichlet allocation (LDA) to extract features from the set of 275 CVs following the steps in Section 4.2. Five features were defined, as shown in Table 3, according to two experts as described in the method in [14].

Table 3: Features defined from collection of 275 CVs

Feature	Top 10 Keywords	Distribution
Client and Team Oriented	user, international, team, social, sales, sports, students, program, service, association	20%
Strategy	intern, project, consulting, strategy, competition, innovation, development, services, projects, case	20%
Sales and Marketing	marketing, sales, market, assistant, brand, managed, social, events, manager, collaborated	21%
Technical Skills	excel, word, office, access, powerpoint, marketing, spss, point, united, power	18%
Finance	financial, analysis, participated, team, companies, research, finance, investment, students, analyst	21%

With these features the system created the student profiles. The distribution of each feature was considered across all student resumes. Given this distribution, the percentiles 10th, 50th, and 90th were determined. For any student and feature, a value below the 10th percentile was discarded as it is assumed that the student would not have selected this feature. The linguistic terms “*Low*”, “*Medium*”, and “*High*” were assigned to the remaining features for each student. Therefore, the linguistic term set for this case includes the basic labels (“*Low*”, “*Medium*”, and “*High*”) and its associated non-basic labels. Students are able to adjust the initial basic labels according to their preferences and apply basic or non-basic labels for each feature.

The rest of the case implementation follows the system description in Section 4.1. Finally, for each student, linguistic values are assigned expressing the fitness between the student and the position. The set of positions with a degree of satisfaction equal to “*High*”, according to the operator defined in Section 3.2, is shown to the student. Of the 549 positions, an average of 22 positions with a median of 13 were proposed to each student. The distribution of the variable, “number of positions selected for each student”, is represented in Figure 4.

### 5.3. Evaluation of the method

As can be seen from Figure 4, using our proposed method, the number of positions for the student to review has been significantly reduced. By narrowing the focus for the student’s internship search, he/she saves considerable time and can work more effectively with only the positions which match his/her interests. Overall, this efficiency is passed directly to the career services office. In a real life scenario, students would be able to refine their search by modifying their preference parameters, thus reducing the number of returned position results.

To evaluate the advantages and drawbacks of our proposed method, we will compare it to: 1) TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) and 2) a ranking method based on Hellinger distance. The first method is a ranking method based on a multi-granular fuzzy linguistic modeling that ranks alternatives based on comparing distances to a optimal alternative as

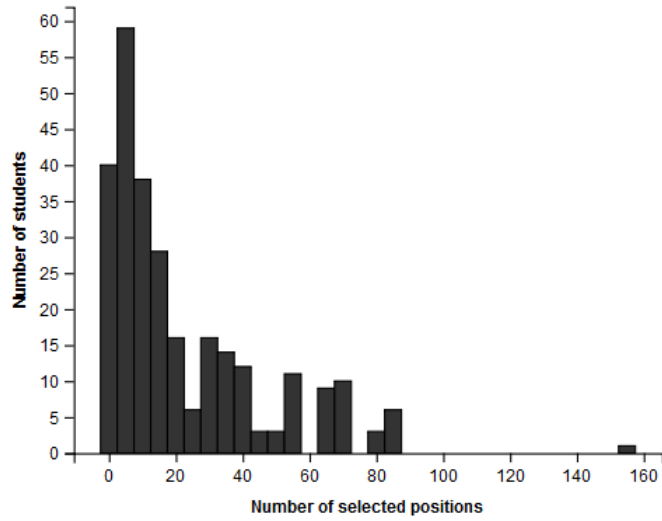


Figure 4: Distribution of number of positions selected per student using proposed method

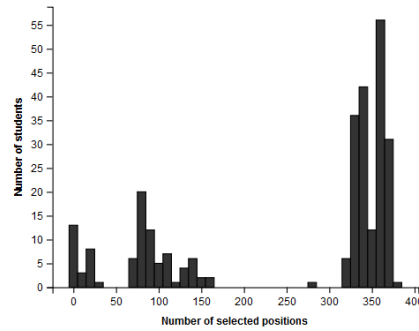
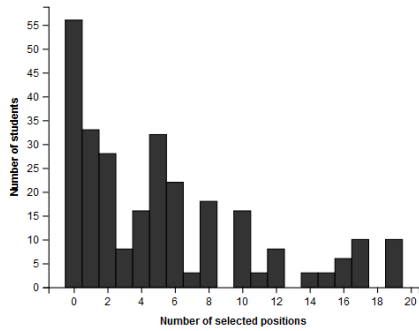


Figure 5: Distribution of number of positions selected per student using TOPSIS method      Figure 6: Distribution of number of positions selected per student using Hellinger distance

defined in [28]. The second method is based on the classic Hellinger distance [29, 30] that does not convert attributes into linguistic terms but uses the frequency distribution of variables.

390

In order to compare these methods to ours, we used the same cut-off value for a recommendation to the user. In this case a "High" linguistic term (i.e.

90%). The results of the first comparison are depicted in Figure 5. This method recommended fewer jobs to individuals than our proposed method. In fact, our method recommended zero positions to at most 40 users while the TOPSIS method recommended zero positions to at most 55 users, demonstrating that more students received recommendations with our method. The results obtained with the second comparison method is based on the Hellinger distance used to determine the distance from a student’s preferences to a position’s features. As can be seen in Figure 6, this method recommended 65 or more positions to the majority of the students, while our method provided more reasonable (i.e. 1–40) recommendations to most students. Our method has a main advantage of an asymmetric matching of student preferences and position requirements that captures position features which meet or exceed student preferences. The main drawback is the loss of information due to the fact that the sorting method proposed is not symbolic and requires translation to numerical values to be computed and as the computation is with numerical values, the results need to be translated to linguistic terms.

## 6. Conclusions

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. Furthermore, the FOWA operator avoids filtering but simultaneously considers several relevant variables for the aggregation process. Lastly, the interests and features of the students and positions are represented as HFLTS, reflecting human tendency to opine with imprecision and hesitance in making decisions.

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.

Regarding enhancements to the methodology, we plan to evaluate our method with a symmetric Sugeno Integral which is based only on min/max operations. The Sugeno Integral is a generalization of OWA and useful to model situations where dependence of criteria are not certain [31]. In our specific problem context, features from which students select their interests are determined implicitly from their CVs. Therefore, the relationships between the features cannot be determined beforehand, making Sugeno Integral an interesting alternative. We would like to note that although the method is employing numeric operators with numeric weights, it does not match each of the labels to a numeric value, rather, it considers different levels of precision labels to be mapped to binary numerical values. The mapping of each of the labels is to a pair of numeric values in order to consider different levels of precision. The result obtained by applying a FOWA operator considers a lexicographic order among all labels. In this context, we expect label translations to be acceptable as the method is seeking to sort and group positions according to preferences rather than to identify the position having the best match. To that end, an extension of the method could include techniques which do not require label translation in order

to better preserve human communicated preferences.

### Acknowledgements

455 This research has been partially supported by the INVITE Research Project (TIN2016-80049-C2-1-R and TIN2016-80049-C2-2-R (AEI/FEDER, UE)), funded by the Spanish Ministry of Science and Information Technology. The results of the research performed was partially funded by l’Obra Social la Caixa.

### References

- 460 [1] R. Montes, A. M. Sánchez, P. Villar, F. Herrera, A web tool to support decision making in the housing market using hesitant fuzzy linguistic term sets, *Applied Soft Computing* 35 (2015) 949–957.
- [2] A. Kelemenis, D. Askounis, A new topsis-based multi-criteria approach to personnel selection, *Expert systems with applications* 37 (7) (2010) 4999–  
465 5008.
- [3] A. Baležentis, T. Baležentis, W. K. Brauers, Personnel selection based on computing with words and fuzzy multimooora, *Expert Systems with applications* 39 (9) (2012) 7961–7967.
- [4] M. T. Maybury, Discovering distributed expertise, Regarding the Intelli-  
470 gence in Distributed Intelligent SystemsMITRE.
- [5] L.-S. Chen, C.-H. Cheng, Selecting is personnel use fuzzy gdss based on metric distance method, *European Journal of Operational Research* 160 (3) (2005) 803–820.
- [6] R. M. Rodríguez, L. Martínez, F. Herrera, Hesitant fuzzy linguistic term  
475 sets, in: *Foundations of Intelligent Systems*, Springer, 2011, pp. 287–295.
- [7] R. M. Rodriguez, L. Martinez, F. Herrera, Hesitant fuzzy linguistic term sets for decision making, *IEEE Transactions on Fuzzy Systems* 20 (1) (2012) 109–119.

- 480 [8] M. Kabak, S. Burmaoğlu, Y. Kazançoğlu, A fuzzy hybrid mcdm approach for professional selection, *Expert Systems with Applications* 39 (3) (2012) 3516–3525.
- [9] E. Faliagka, A. Tsakalidis, G. Tzimas, An integrated e-recruitment system for automated personality mining and applicant ranking, *Internet research* 22 (5) (2012) 551–568.
- 485 [10] L. Canós, V. Liern, Soft computing-based aggregation methods for human resource management, *European Journal of Operational Research* 189 (3) (2008) 669–681.
- [11] Z. Güngör, G. Serhadlıoğlu, S. E. Kesen, A fuzzy ahp approach to personnel selection problem, *Applied Soft Computing* 9 (2) (2009) 641–646.
- 490 [12] D. Yu, W. Zhang, Y. Xu, Group decision making under hesitant fuzzy environment with application to personnel evaluation, *Knowledge-Based Systems* 52 (2013) 1–10.
- [13] E. Faliagka, L. Iliadis, I. Karydis, M. Rigou, S. Sioutas, A. Tsakalidis, G. Tzimas, On-line consistent ranking on e-recruitment: seeking the truth behind a well-formed cv, *Artificial Intelligence Review* 42 (3) (2014) 515–528.
- 495 [14] M. S. Evans, A computational approach to qualitative analysis in large textual datasets, *PloS one* 9 (2) (2014) e87908.
- [15] R. R. Yager, J. Kacprzyk, *The ordered weighted averaging operators: theory and applications*, Springer Science & Business Media, 2012.
- 500 [16] J. A. Morente-Molinera, I. J. Pérez, M. R. Ureña, E. Herrera-Viedma, On multi-granular fuzzy linguistic modeling in group decision making problems: a systematic review and future trends, *Knowledge-Based Systems* 74 (2015) 49–60.

- 505 [17] S. Massanet, J. V. Riera, J. Torrens, E. Herrera-Viedma, A new linguistic computational model based on discrete fuzzy numbers for computing with words, *Information Sciences* 258 (2014) 277–290.
- [18] F. Herrera, L. Martínez, A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision-making, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 31 (2) (2001) 227–234.
- 510 [19] J. L. García-Lapresta, A general class of simple majority decision rules based on linguistic opinions, *Information sciences* 176 (4) (2006) 352–365.
- [20] L. Travé-Massuyès, F. Prats, M. Sánchez, N. Agell, Relative and absolute order-of-magnitude models unified, *Annals of Mathematics and Artificial Intelligence* 45 (3) (2005) 323–341.
- 515 [21] N. Agell, M. SáNchez, F. Prats, L. Roselló, Ranking multi-attribute alternatives on the basis of linguistic labels in group decisions, *Information Sciences* 209 (2012) 49–60.
- 520 [22] R. R. Yager, Quantifier guided aggregation using owa operators, *International Journal of Intelligent Systems* 11 (1) (1996) 49–73.
- [23] J. Figueira, S. Greco, M. Ehrgott, *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer Verlag, Boston, Dordrecht, London, 2005.
- 525 [24] R. Ureña, F. Chiclana, J. A. Morente-Molinera, E. Herrera-Viedma, Managing incomplete preference relations in decision making: a review and future trends, *Information Sciences* 302 (2015) 14–32.
- [25] C. Carlsson, R. Fullér, Fuzzy multiple criteria decision making: Recent developments, *Fuzzy sets and systems* 78 (2) (1996) 139–153.
- 530 [26] F. J. Cabrerizo, I. J. Pérez, F. Chiclana, 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.



- [27] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, *Journal of machine Learning research* 3 (Jan) (2003) 993–1022.
- 535 [28] I. Beg, T. Rashid, Topsis for hesitant fuzzy linguistic term sets, *International Journal of Intelligent Systems* 28 (12) (2013) 1162–1171.
- [29] D. Blei, J. Lafferty, *Text mining: Classification, clustering, and applications*, chapter Topic Models, Chapman & Hall/CRC.
- [30] V. Rus, N. Niraula, R. Banjade, Similarity measures based on latent dirich-  
540 let allocation, in: *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, 2013, pp. 459–470.
- [31] V. Torra, Y. Narukawa, The interpretation of fuzzy integrals and their application to fuzzy systems, *International journal of approximate reasoning* 41 (1) (2006) 43–58.