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

J. Nguyen\textsuperscript{a}, G. Sánchez-Hernández\textsuperscript{b}, A. Armisen\textsuperscript{b}, N. Agell\textsuperscript{b,∗}, X. Rovira\textsuperscript{b}, C. Angulo\textsuperscript{a}

\textsuperscript{a}GREC Research Group, Universitat Politècnica de Catalunya (BarcelonaTech), Spain
\textsuperscript{b}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...
the role of decision support tools more important. Specifically, this notion can be observed within human resource personnel selection. In general, personnel selection depends on a firm’s specific targets, and the preferences of the hiring managers and candidates. Therefore, Multi-Criteria Decision-Aiding (MCDA) methods and fuzzy logic are an appropriate solution to capture the multi-criteria and uncertainty aspects inherent in the process.

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

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 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 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 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 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
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 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.

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 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 applying 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 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 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

<table>
<thead>
<tr>
<th>Paper</th>
<th>Ranking Method</th>
<th>Weights</th>
<th>Case Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canós and Liern, 2008</td>
<td>OWA and parametric aggregation</td>
<td>Learned weights and FWA</td>
<td>Illustrative example 1 5 6</td>
</tr>
<tr>
<td>Güngör et al., 2009</td>
<td>Comparison of fuzzy AHP and Yager’s weighted method</td>
<td>Predetermined by recruiter</td>
<td>Illustrative example 1 6 17</td>
</tr>
<tr>
<td>Faliagka et al., 2012</td>
<td>AHP</td>
<td>Predetermined by recruiter</td>
<td>Use case 3 100 4</td>
</tr>
<tr>
<td>Kabak et al., 2012</td>
<td>Fuzzy TOPSIS and fuzzy ELECTRE</td>
<td>Fuzzy ANP computations to determine weights</td>
<td>Illustrative example 1 6 10</td>
</tr>
<tr>
<td>Baležentis et al., 2012</td>
<td>MULTIMOORA for group decision making using FWA operator</td>
<td>Predetermined by recruiter</td>
<td>Illustrative example 1 4 8</td>
</tr>
<tr>
<td>Yu et al., 2013</td>
<td>GHFPWA and GHFPWG operator used to aggregate hesitant fuzzy elements (HFE)</td>
<td>Prioritized average (PA) operator</td>
<td>Illustrative example 5 5 4</td>
</tr>
<tr>
<td>Faliagka et al., 2014</td>
<td>Learning to rank</td>
<td>N/A</td>
<td>Use Case 3 100 4</td>
</tr>
</tbody>
</table>

#### 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.
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 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 multi-granular linguistic modeling [10], 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 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 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 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.

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 $S_n$ be a finite set of totally ordered basic terms, $S_n = \{B_0, \ldots, B_n\}$, with $B_0 < \ldots < B_n$ and the hesitant fuzzy linguistic terms set, $H_{S_n}$, be the set of all consecutive linguistic basic terms of $S_n$, i.e. $B_{ij} = \{x \in S_n \mid B_i \leq x \leq B_j\} \forall i, j \in \{0, \ldots, n\}$, 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, $S_n$, allows us to define a total order in $H_{S_n}$ based on the lexicographic order such that: given two linguistic terms, $B_{ij}, B_{i'j'} \in H_{S_n}$, $B_{ij} \leq_L B_{i'j'}$, iff $i < i'$ or $i = i'$ and $j \leq j'$.

For instance, let us consider $n = 3$ and $B_0 = “None” < B_1 = “Low” < B_2 = “Medium” < B_3 = “High”, then, terms $B_{12}$ and $B_{03}$ will represent 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 $H_{S_n}$, a subset of $H_{S_n}$, which corresponds to the HFLTS obtained when the set of basic elements is $S_n = \{B_1, \ldots, B_n\}$.

In addition, in $H_{S_n}$ 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 $B_{i'j'}$, $B_{ij} \preceq_L B_{i'j'}$, 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 $H_{S_n}$ defined as $B_{ij} \sqcup B_{i'j'} = B_{kl}$ where $k = \min(i, i')$ and $l = \max(j, j')$. Following the previous example, $H_{S_3}$, $B_{01} \sqcup B_3 = B_{03}$.

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 catching all possible combinations of hesitancy in both descriptions.
Definition 3.1. The fuzzy matching operator is the map

\[ * : H_{S_n} \times H_{S_n^*} \rightarrow H_{S_n} \]

such that:

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

Note that, 2. coincides with 1. \( \forall B_i \in S_n \) and \( \forall B_j \in S_{n^*}^n \).

Example 3.1. Let us consider that a candidate’s preferences are represented by \( H_{S_n^*} \) and the features of each position are represented by \( H_{S_n} \), then given the previously considered HFLTS, \( H_{S_n^*}^n \), with \( n = 3 \), the results of the fuzzy matching operator for the basic terms are shown in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Low (( B_1 ))</th>
<th>Medium (( B_2 ))</th>
<th>High (( B_3 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (( B_0 ))</td>
<td>Medium (( B_2 ))</td>
<td>Low (( B_1 ))</td>
<td>None (( B_0 ))</td>
</tr>
<tr>
<td>Low (( B_1 ))</td>
<td>High (( B_3 ))</td>
<td>Medium (( B_2 ))</td>
<td>Low (( B_1 ))</td>
</tr>
<tr>
<td>Medium (( B_2 ))</td>
<td>High (( B_3 ))</td>
<td>High (( B_3 ))</td>
<td>Medium (( B_2 ))</td>
</tr>
<tr>
<td>High (( B_3 ))</td>
<td>High (( B_3 ))</td>
<td>High (( B_3 ))</td>
<td>High (( B_3 ))</td>
</tr>
</tbody>
</table>

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.
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. $B_{02} * B_{12} = \bigcup\{B_0 * B_1, B_0 * B_2, B_1 * B_1, B_1 * B_2, B_2 * B_1, B_2 * B_2\} = \bigcup\{B_2, B_3, B_2, B_3, B_3\} = B_{13}$.

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

1. $\forall B_{ij}, B_{i'j'} \in H_{S_n^*}$, then $B_{ij} * B_{i'j'} \neq B_{i'j'} * B_{ij}$.
2. $\forall B_{ij} \in H_{S_n}$ and $\forall B_{i'j'} \in H_{S_n^*}$, with $B_{i'j'} \leq L B_{ij}$, then, $B_n \preceq B_{ij} * B_{i'j'}$.
3. $\forall B_{ij} \in H_{S_n}$, $B_{ij} * B_n = B_{ij}$.

From Property 1 we can infer that the order always matters when matching two different terms in $H_{S_n^*}$. 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 $B_{ij}$.

### 3.2. Fuzzy aggregation and alternatives selection

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

$$X * Y = (X_1 * Y_1, \ldots, X_k * Y_k) \in (H_{S_n})^k$$
Example 3.3. Continuing with Example 3.1, given the vectors $X \in (H_S^3)^5$ and $Y \in (H_S^3)^5$, $X = (B_2, B_1, B_3, B_0, B_2)$, and $Y = (B_2, B_2, B_1, B_3, B_1)$, the match is $X \ast Y = (B_3, B_2, B_3, B_0, B_3)$. 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 $X \ast Y = (B_{13}, B_3, B_{13}, B_2, B_1)$.

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

Definition 3.3. Given $Z = (Z_1, \ldots, Z_k) \in (H_{S_n})^k$ 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 $\leq_L$, $\varphi : H_{S_n} \rightarrow \mathbb{R}$, such that $\varphi(B_s) = s, \forall s \in \{0, \ldots, n\}$.

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

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

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 $Z = (Z_1, \ldots, Z_k) \in (H_{S_n})^k$ we define the fuzzy ordered weighted average operator $\Phi : (H_{S_n})^k \rightarrow H_{S_n}$ is defined as follows:

$$\Phi(Z_1, \ldots, Z_k) = B_{\mu_{1}^Z \mu_{2}^Z}$$

where $\mu_{1}^Z$ and $\mu_{2}^Z$ are the rounded and ceiling values, respectively.
Definition 3.5. Given \( X \in (H_{S_n})^k \), \( Y \in (H_{S^*})^k \), 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.: \( \phi_Y(X) = \Phi(X_1 * Y_1, ..., X_k * Y_k) \).

Example 3.4. Continuing with Example 3.3, we can consider the increasing function: \( \phi(B_{s,l}) = s + \frac{l-s}{3+1-s}, \forall s,l \in \{0,1,2,3\} \) for our example. The function chosen for \( \phi(B_{s,l}) \) 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:

\[
 w_i = Q \left( \frac{i}{5} \right) - Q \left( \frac{i-1}{5} \right), \; i = \{1, \ldots, 5\}, \quad (2)
\]

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

Then, given the matching vector, \( X \ast Y = (B_{13}, B_3, B_{13}, B_{23}, B_1) \), between vectors \( X \) and \( Y \), and applying Definitions 3.3, 3.4, and 3.5, the degree of fitness of \( X \) to \( Y \) is \( \phi_Y(X) = \Phi(X_1 \ast Y_1, ..., X_5 \ast Y_5) = \Phi(B_{13}, B_3, B_{13}, B_{23}, B_1) = B_{23} \).

This result comes from the fact that: \( B_3 \geq_L B_{23} \geq_L B_{13} \geq_L B_{13} \geq_L B_1 \), \( \varphi(B_3) = 3, \varphi(B_{23}) = \frac{5}{7}, \varphi(B_{13}) = \frac{5}{7}, \varphi(B_1) = 1 \), and \( w_1 = \sqrt{\frac{4}{5}}, \; w_2 = \sqrt{\frac{2}{5}} - \sqrt{\frac{1}{5}}, \; w_3 = \sqrt{\frac{3}{5}} - \sqrt{\frac{2}{5}}, \; w_4 = \sqrt{\frac{4}{5}} - \sqrt{\frac{3}{5}}, \; w_5 = 1 - \sqrt{\frac{4}{5}} \).

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 \cite{23,24}. When these factors are related to user preferences but not easily measurable, the introduction of fuzzy and linguistic descriptions brings an appropriate framework \cite{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.
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 vitaes (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.
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 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, 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 (i.e. “None” when Feature$_i$ < 10%, “Low” when 10% ≤ Feature$_i$ < 50%, “Medium” when 50% ≤ Feature$_i$ < 90%, “High” when Feature$_i$ ≥ 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 $Y^j = (Y^j_1, ..., Y^j_k) \in (H_S)^k$, 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).
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 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 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 $X_i = (X_{i1}, \ldots, X_{ik})$ and is compared to the student’s preferences, $Y_j = (Y_{j1}, \ldots, Y_{jk})$. The outcomes of the matching are linguistic labels, $H_{Si}$, 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.

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 the greatest match between students and positions. The resulting level of satisfaction is a fuzzy linguistic term set \( \phi_Y(X) = \Phi(X_1 \ast Y_1, \ldots, X_k \ast Y_k) \) obtained via the weighted average. 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 preferences and their match with 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>
<thead>
<tr>
<th>Feature</th>
<th>Top 10 Keywords</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client and Team Oriented</td>
<td>user, international, team, social, sales, sports, students, program, service, association</td>
<td>20%</td>
</tr>
<tr>
<td>Strategy</td>
<td>intern, project, consulting, strategy, competition, innovation, development, services, projects, case</td>
<td>20%</td>
</tr>
<tr>
<td>Sales and Marketing</td>
<td>marketing, sales, market, assistant, brand, managed, social, events, manager, collaborated</td>
<td>21%</td>
</tr>
<tr>
<td>Technical Skills</td>
<td>excel, word, office, access, powerpoint, marketing, spss, point, united, power</td>
<td>18%</td>
</tr>
<tr>
<td>Finance</td>
<td>financial, analysis, participated, team, companies, research, finance, investment, students, analyst</td>
<td>21%</td>
</tr>
</tbody>
</table>
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
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

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


[4] M. T. Maybury, Discovering distributed expertise, Regarding the Intelligence in Distributed Intelligent SystemsMITRE.


