FoDRA - A New Content-Based Job Recommendation Algorithm for Job Seeking and Recruiting

Nikolaos D. Almalis
University of Piraeus
Dept. of Informatics
Piraeus, Greece
nikosalmalis@unipi.gr

Prof. George A. Tsihrintzis
University of Piraeus
Dept. of Informatics
Piraeus, Greece
geoatsi@unipi.gr

Nikolaos Karagiannis
MSc Software Engineer
Piraeus, Greece
nkaragiannis87@unipi.gr

Aggeliki D. Strati
University of Aegean
MSc Maths
Samos, Greece
stratiaggeliki@gmail.com

***Abstract*—In this paper, we propose a content-based recommendation Algorithm which extends and updates the Minkowski distance in order to address the challenge of matching people and jobs. The proposed algorithm FoDRA (Four Dimensions Recommendation Algorithm) quantifies the suitability of a job seeker for a job position in a more flexible way, using a structured form of the job and the candidate's profile, produced from a content analysis of the unstructured form of the job description and the candidate's CV. We conduct an experimental evaluation in order to check the quality and the effectiveness of FoDRA. Our primary study shows that FoDRA produces promising results and creates new prospects in the area of Job Recommender Systems (JRSs).**

Keywords—job recommendation systems; content-based filtering; job seeking and recruiting; matching people and jobs; similarities measures.

I. INTRODUCTION

Over the last few years, we are daily overwhelmed by vast amounts of information from various sources. This deluge of information makes the task of finding useful or suitable items/objects - such as newspapers, web sites, songs, movies, books or even jobs - a big challenge. For this reason, more and more applications have been broadly developed and new techniques have been emerged to support human decisions suggesting services, products and various types of information to customers [1] [2] [4]. One field of research in this direction is that of Recommender Systems (RSs) [5]. RSs are tools that use various techniques and algorithms to isolate irrelevant information from a huge amount of data and generate personalized suggestions of a small subset of them which a user can examine in a reasonable amount of time [6] [7].

Many web applications are based on RSs to suggest personalized items to users. These suggestions accommodate various decision-making processes, such as what items to buy, what news to read, what music to listen, what job to apply for, etc [3] [16]. Typical examples of services like these can be found on e-commerce service, such as *Amazon,* and on social networks, such as *Facebook* and *Linkedin*. Especially in the area of job seeking and recruiting, the study in [15] shows that 73% of personnel recruiters have made successful hires through social media and 59% of personnel recruiters think that candidates which were recommended through social networks are of the highest quality.

RSs have some specialties which differentiate them from other common scientific domains such as artificial intelligence, data mining, pattern recognition etc. From the above point of view Burke et al [8] consider two basic principles which distinguish RSs research:

> "A recommender system is personalized. The recommendations it produces are meant to optimize the experience of one user, not to represent group consensus for all.
>
> A recommender system is intended to help the user select among discrete options. Generally the items are already known in advance and not generated in a bespoke fashion"

Thus, RSs typically use one of the next four basic techniques, in order to achieve technical adjustment of the aforementioned two principles [9]:

Collaborative Filtering Recommenders (CFRs): Recommendations are generated by utilizing social knowledge (typical ratings of items by a community of users). More specifically, a CFR finds users with similar interests as the target user and suggests recommendations to him/her based on their liked items. The key function in CFRs is the computation of similarities among users [5] [9] [12]. The concept of CFRs is presented in Fig 1.

Content-Based Recommenders (CBRs): Items are recommended having similar content information to those a user has liked in the past. A CBR analyzes a set of characteristics of items which are rated by the target user and build the profile of the interests of this user based on the features of the items which are rated by him. The recommendation process matches up the user profile attributes against the set of properties of item content [1] [5] [9]. Fig. 2 illustrates the CBRs concept.

Knowledge-Based Recommenders (KBRs): Recommendations are generated by utilizing domain knowledge. These systems have the advantage of enhanced reliability since they usually contain less noise. However, a KBR requires considerable knowledge acquisition for setup

Authorized licensed use limited to: SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECH. Downloaded on September 14,2021 at 08:59:32 UTC from IEEE Xplore. Restrictions apply.

and maintenance during their lifetime [9] [13]. Fig 3 shows a KBR.

Hybrid Recommenders (HR): The approaches mentioned above have their limitations, such as the *cold start* and the *sparsity* problem. So, HRSs combine two or more techniques to overcome these limitations [6] [9]. The approach of HRs is presented in Fig 4.

The techniques mentioned provide the general framework on RSs application development and research. In this paper, we present a new approach, based on the CBRs concept, which extends and updates the Minkowski distance [10], in order to address the challenge of matching people and jobs [14]. The proposed algorithm FoDRA (**Fo**ur **D**imensions **R**ecommendation **A**lgorithm) quantifies the suitability of a job seeker for a job position in a more flexible way, using a structured form of the job and the candidate's profile, produced from a content analysis of the unstructured form of the job description and the candidate's CV. More specifically, the problem is formulated as follows: a set of job positions are described by several required qualifications (e.g. education level, language skills, work experience etc.) with specific requirements. At the same time, a candidate for these jobs can be assigned a value for each one of the jobs' qualifications. The algorithm takes under consideration the four different categories of the measurable attributes/skills which depend on the values range (i.e. a specific value, a range with lower limit, a range with upper limit and a range with both lower and upper limit, providing a final score which shows the suitability of the job seeker for a specific job. We performed evaluations based on real world data, which show promising results.

The rest of the paper is organized as follows: Section 2 reports previous relevant work about Job Recommender Systems, while Section 3 analyzes our suggested algorithm and presents examples of its application. In Section 4, we conduct the experimental evaluation presenting the methodology and results. Finally, Section 5 provides a brief summary of the paper and discusses some of its possible future extensions.

II. RELATED WORK

A. *Job Recommender Systems (JRS)*

Recently, Job Recommender Systems (JRS) have started to play an important role on the recruiting process, especially on online recruiting websites. The difference between traditional RSs and JRSs is that the first ones recommend items to users while the second ones recommend one type of users (e.g. job applicants) to another type of users (e.g. personnel recruiters). More specifically, a JRS is designed to suggest either a list of job positions to a job applicant based on his/her preferences, or to retrieve a list of job candidates to an employer or recruiter based on the job requirements. It is worth noting that surveys have shown that the 93% of recruiters use or plan to use social recruiting tools such as *Linkedin* to support their recruiting efforts and only 18% of recruiters consider themselves to be experts at social recruiting [15]. Thus, online or offline JRSs seems to be an unquestionably strong necessity that leads the RSs researchers.

W. Hong et al. developed iHR an online JRS that classifies users into groups by utilizing both the historical behaviors of users and individual information and then uses the appropriate recommendation approach for each user group. This approach is suitable for the cases in which different users may have different attributes and a single recommendation approach may not be appropriate for all users [16].

Salton et al suggested a CB JRS that utilizes the Vector Space Model (VSM) [17]. In VSM, each job was represented by a vector in an n-dimensional space, where each dimension corresponded to a textual feature from the overall feature-set of the job set. Furthermore, each job was represented as a vector of feature weights, where each weight designated the degree of association between the feature and the job. Finally, a relevance score is computed for a candidate employee and each candidate job and then a list of top-k jobs with the highest score are suggested.

Based on the requirement that a good match between jobs and persons needs to take into account both the preferences of the candidate and the preferences of the recruiter, Malinowski et al. proposed in [14] an approach applying the following two distinct JRSs in order to improve the match between jobs and candidates. The first JRS, CV-recommender, recommends CVs which are similar to resumes previously selected by the same recruiter for a specific job description considered. The second Job-recommender considers the preferences of the candidate for a specific kind of job and recommends jobs to candidates based on their profiles.

An ontology-based hybrid approach has been presented by Fazel-Zarandi which effectively matches job seekers and job postings. The approach uses a deductive model to determine the type of match between a posting and a job seeker by applying a similarity-based approach to rank applicants [19].

The Austrian job board for graduates *Absolventen* [20] uses an RS to suggest appropriate jobs to applicants. This system considers the CV as the only input to create the user profile, which is compared with the available jobs. Moreover, the RS has been enhanced with implicit relevance feedback, allowing the system to figure out user preferences. Additionally, gradual forgetting factors are used for adapting the profile over time. All these data are included in the hybrid user profile, which is represented as a hyper dimensional vector and calculated by a linear combination of the preferred jobs and CVs.

Proactive is a web based JRS which helps job seekers to discover relevant opening jobs in multiple ways [21]. This system focuses on job related to information technology. At the beginning of each session, the system selects and displays jobs posted within the last 24 hours. When the user finds an interesting job, he/she assigns it as a favorite job. The recommendations are generated based on the properties of the user's favorite jobs. Whenever a user assigns a job as favorite, a new list of recommended jobs is created. Finally by analyzing the user's preferences, the RS suggests a set of jobs.

The above approaches suffer from a common drawback called *lack of intelligence*. This means, that some otherwise unexpected jobs, will not appear to a job seeker in the set of recommended jobs. The only jobs that will be suggested are the ones with the highest match against the seeker's profile.

Authorized licensed use limited to: SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECH. Downloaded on September 14,2021 at 08:59:32 UTC from IEEE Xplore. Restrictions apply.

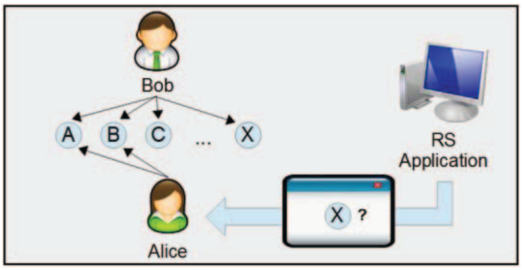

Fig. 1. Bob chose the objects A, B and X, Alice chose A and B, the CFR system will recommend to Alice object X.

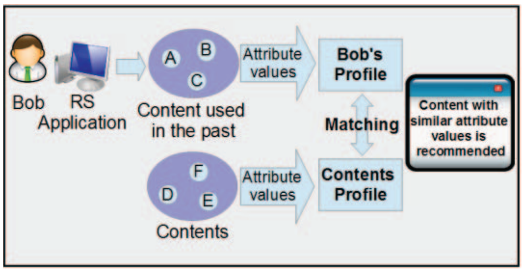

Fig. 2. The CBR uses object's characteristics to recommend objects, which have similar content information to those Bob has liked in the past.

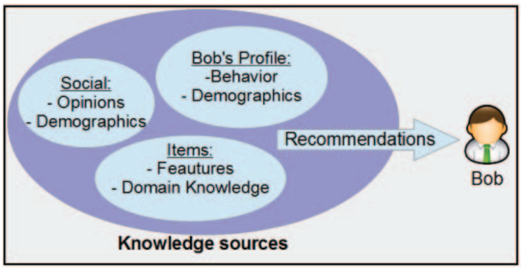

Fig. 3. KBR generates recommendations to Bob by utilizing domain knowledge from several sources.

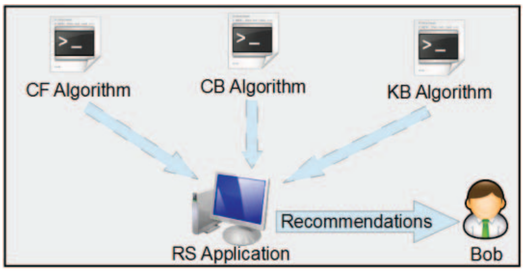

Fig. 4. HRS combines two or more techniques to provide better recommendations to Bob.

B. *Our Previous Content-Based Job Recommendation Approach*

In our previous work [11], we suggested a content-based job recommendation algorithm that deals with the recommendation of candidates for a specific job position. In proportion with the problem definition of our current approach, a specific job position was described by several required qualifications with specific values and a number of candidates had respective values for each of these qualifications. The job applicant with the *closer or even higher* of the respective values of the specific job requirements had higher probability to be recruited. Our algorithm utilized the Minkowski distance to perform a primary analysis in order to investigate how the job seeking and recruiting field could have benefited further. More specifically, our algorithm aimed at quantifying the suitability of a candidate employee for a specific job position.

The results of our approach were promising, yet the approach presented a significant limitation. The algorithm was taking under consideration for an attribute only values that were closer or higher to the required one, meaning that a candidate must be qualified at a desirable level or higher. However, as we will discuss in Section III, a qualification can be modeled in more ways, namely an attribute can be qualified if it is below a certain threshold (e.g. a job requires the candidate to be at most 40 years old), or between a predefined range (e.g. qualified candidates are between 18 and 40 years old). As we will further discuss in Section III, our analysis of the variation in qualifications results in four different dimensions and our initial CB approach did not capture all of them.

In this paper, we present the Four Dimensions Recommendation Algorithm (FoDRA) which extends our previous approach by covering the whole spectrum of attribute qualification types, i.e. (i) an attribute has an exact value, (ii) the attribute's value is below a certain threshold, (iii) the attribute's value is above a certain threshold and (iv) the attribute's value is between a range with specified both the upper and lower limits.

III. The Problem Of Matching People And Jobs

The problem of *matching people and jobs* has concerned several researchers the last decades [14] [16] [19]. The papers presented in Section 2 show exactly this trend. Meanwhile, several approaches and techniques have been proposed in order to confront this recommendation challenge having pros and cons each one. Specially, the CBR systems generally adopt two common approaches in order to define the interest for an item, the heuristics-based and the model based approach. Both try to find *similarities* among discrete options [23] [24].

The above approaches may be efficient for some kind of items such as books, magazines, songs, movies etc. but for the domain of matching people and jobs, they are not successfully responding. The reason is that the domain of people and jobs has some specificity about the desirable values of the job's features. More specifically, there are attributes depending on job description and their values present the above four different forms: (i) may have an exact value, (ii) a range with a lower limit, (iii) a range with an upper limit or (iv) a range with both lower and upper limit. Consequently, our approach overcomes the limits occurred in Boolean and smallest distance calculation of classic CBR approaches [22]. The above approaches manage to take under consideration only one of the four cases, that of the exact value. This limitation was the motive to our research and the proposed algorithm FoDRA tries to extend and update the recommendation's efficiency on the domain of job seeking and recruiting.

A. *Definition*

We first present a general formalized definition of matching people and jobs. This definition will allow us to investigate and

Authorized licensed use limited to: SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECH. Downloaded on September 14,2021 at 08:59:32 UTC from IEEE Xplore. Restrictions apply.

explore the domain in a controllable procedure and a within bounds defined way. More formally, the recommendation problem of matching people and jobs can be formulated as follows:

Let J be the set of all jobs and let P be the set of all possible candidate employees (people) that can be recommended. Let f be a utility function which measures the suitability of candidate employee *p* with a job J that is f: J x P → R, where R is a totally ordered set (for example, real numbers within a certain range). For each job $j \in J$ we want to choose such a candidate employee $p' \in P$ that maximizes the job's suitability.

From the above definition, three basic fields stand out that distinguish the CBR systems of matching people and jobs research.

Types of relationships: We distinguish three types of relationships based on what is recommended to whom. Thus, we have: (i) a job and many candidates to be recommended (ii) a candidate and many jobs to be recommended and (iii) both many jobs and many candidates to be matched (Fig 5).

Domain: One of the key issues in content-based job recommendation is the feature quality. Every job and candidate should be described at the same level of detail and the feature set would contain descriptors that correlate with the discriminations made by job seekers or recruiters.

Suitability function: There are a variety of algorithms which calculate the suitability of a candidate employee for a job. It is mentioned that we confront the problem *matching people and jobs* using our proposed algorithm, as suitability function.

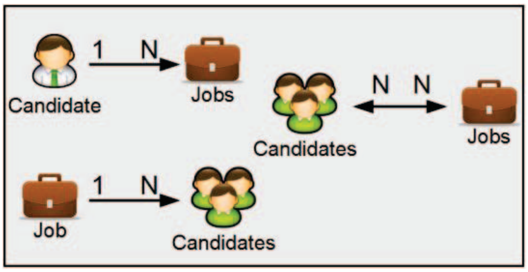

Fig. 5. According to the definition of the problem of *matching people and jobs* we distinguish three types of relationships based on what is recommended to whom: i) a job and many candidates to be recommended ii) a candidate and many jobs to be recommended, iii) both many jobs and many candidates to be matched after the recommendation process.

IV. The Four Dimensions Recommendation Algorithm

A. Formulation

Consider a vector J to be a specific job of the form J=$(j_1, j_2, \dots, j_n)$ with attributes $j_1, j_2, \dots, j_n, n \in N^+$ and each attribute belongs to one of the four referred types: an exact value (E), a range with lower limit (L), a range with upper limit (U) and a range with both lower and upper limit (LU). Furthermore we define a vector W=$(w_1, w_2, \dots, w_n)$, w ∈ [0, 1] where each element w_i is the weight of the attribute j_i of vector A, $i \in N^+$. Also, let vector C=$(c_1, c_2, \dots, c_n)$ with the attributes $c_1, c_2, \dots, c_n$, $n \in N^+$ be a candidate employee that will be compared with the job J (vector J and E have the same structured form), in order to find the suitability of candidate employee C for the job J. We follow the next three steps:

Step 1

We construct four vectors J_x, $x = 1,2,3,4$, where their attributes belong to the initial set J of attributes $j_1, j_2, \dots, j_n$ of job J and their classification is done based on the type of their attributes. If E is the set of the attributes with exact value, L the set with lower limit, U the set with upper limit and LU the set with both lower and upper limit, it should be that:

- $E \cup L \cup U \cup LU = J$
- $E \subseteq J, L \subseteq J, U \subseteq J, LU \subseteq J$
- $E \cap L = \emptyset, E \cap U = \emptyset, E \cap LU = \emptyset, L \cap U = \emptyset,$ $L \cap U = \emptyset,\ U \cap LU = \emptyset$.

Step 2

Case 1 -*Exact value*- The suitability S_E will be given by the metric:

$$S_E = -\left(\sum_{i=1}^{n} w_i |j_i - c_i|^p\right)^{\frac{1}{p}}, \qquad p \in R^* \tag{1}$$

Case 2 -*Lower limit*- The suitability S_L will be given by the sum of the following two metrics S_{L+} and S_{L-}, where:

$$S_{L+} = +\left(\sum_{k=1}^{n} w_k |j_k - c_k|^p\right)^{1/p}, p \in R^* \tag{2}$$

$$S_{L-} = -\left(\sum_{l=1}^{n} w_l |j_l - c_l|^p\right)^{1/p}, p \in R^* \tag{3}$$

$$S_L = S_{L+} + S_{L-} \tag{4}$$

with $c_k \geq j_k$ for (2), $c_l < j_l$ for (3) and $(k \neq l)$.

Case 3 -*Upper limit*- The suitability S_U will be given by the sum of the following two metrics S_{U-} and S_{U+}, where:

$$S_{U-} = -\left(\sum_{u=1}^{n} w_u |j_u - c_u|^p\right)^{1/p}, p \in R^* \tag{5}$$

$$S_{U+} = +\left(\sum_{v=1}^{n} w_v |j_v - c_v|^p\right)^{1/p}, p \in R^* \tag{6}$$

$$S_U = S_{U+} + S_{U-} \tag{7}$$

with for $c_u \geq j_u$ for (5), $c_v < j_v$ for (6) and $(u \neq v)$.

Case 4 -*Lower and upper limit*- The suitability S_{LU} will be given by the metric:

Authorized licensed use limited to: SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECH. Downloaded on September 14,2021 at 08:59:32 UTC from IEEE Xplore. Restrictions apply.

$$S_{LU} = -\left(\sum_{m=1}^{n} w_m |j_m - c_m|^p\right)^{\frac{1}{p}}, \qquad p \in R^* \tag{8}$$

where j_m is the lower limit if $c_m \leq j_m$ and j_m is the upper limit if $c_m \geq j_m$ (for $j_{mMax} \leq c_m \leq j_{mMin}$ the $S_{LU} = 0$, as the value is within the desirable range)

Step 3

Overall, the suitability S_{CJ} of the candidate employee C for the job J is defined as follows:

$$S_{CJ} = S_E + S_L + S_U + S_{LU} \tag{9}$$

B. Example

In order to give a comprehensive case study for our contend-based job recommendation approach, we present a simplified form of the problem of *matching people and jobs*. Let $S(a, b)$, be a random job seeker with attributes A and B and ratings 2 and 3 respectively. Also let $J_1\,\{4, 5\}$, $J_2\,\{1, 4\}$, $J_3\,\{5, 1\}$, $J_4\,\{[2-4], [4-6]\}$ be four jobs with the same attributes A and B, the job seeker and the candidate jobs have the same structured form, and rating as shown above. Then we consider the weight of each attribute is $W_A = 1$, $W_B = 1$ and the variable p = 2 (Euclidean Distance). Also in job J_1 both attributes have exact values, in job J_2 both attributes have lower bound, in job J_3 the attribute A has an upper limit and the attribute B has exact value. Last in job J_4 both attributes are in a range with lower and upper limit.

Then we calculate the distances between the point S and the four points J_1, J_2, J_3, J_4 based on FoDRA which is translated to our problem as *the suitability of the person S for the job* J_x, *x=1, 2, 3, 4*.

Finally we distinguish four separate cases. In each case below we present the job and the candidate (the same in all cases) and the calculations for the steps of the algorithm. At the end the results show the suitability of the selected candidate employee for the specific job. Let's examine the cases:

- Job: $J_1\,(4, 5)$ - Candidate: $S\,(2, 3)$

  Attribute A: exact value, Attribute B: exact value

  Step 1→ we compute the metric $S_{J_1S} = -2{,}828$

- Job: $J_2\,(1, 4)$ - Candidate: $S\,(2, 3)$

  Attribute A: lower limit, Attribute B: lower limit

  Step1→ we construct the points $S_1\,(2, 4)$ and $S_2\,(1, 3)$

  Step2→ we compute the metrics

  $S_{J_2S_1} = 1$ and $S_{J_2S_2} = -1$

  Step3→ the final distance is $S_{J_2S} = 0$

- Job: $J_3\,(5, 1)$ - Candidate: $S\,(2, 3)$

  Attribute A: upper limit, Attribute B: exact value

  Step1: we construct the points $S_1\,(2, 1)$ and $S_2\,(5, 3)$

  Step2: we compute the metrics

  $S_{J_3C_1} = 3$ and $S_{J_3C_2} = -2$

  Step3: the final distance is $S_{J_3C} = 1$

- Job: $J_4\,\{(2-4), (4-6)\}$ - Candidate: $S\,(2, 3)$

  Step1: we construct the point $J_{4a}\,(2, 4)$

  Step2: we compute the metric $S_{J_{4a}S} = -1$

Analyzing the results of the final step in each case we conclude that we have a primary approach of the suitability in a quantitative way. So we can assume that the candidate is below the requirements of the job J_1 and J_4, above the requirements for the job J_3 and exactly equal to the requirements for the job J_2. More detailed analysis is presented in Section V.

V. Experimental Evaluation

In this section we provide the experimental results of our evaluation using FoDRA. Firstly, we present the dataset and the content analysis methodology. Then we execute the proposed recommendation algorithm for the given dataset and finally we evaluate the results.

A. Dataset And Content Analysis Methodology

In order to evaluate the quality and the effectiveness of the CBR strategy we conducted our experimental procedure using real world data; available at the online website *www.kaggle.com/c/job-recommendation/data*. The job information consists of: job title, job description, requirements, skills etc. and the job applicant profile consists of: education level, working experience, job history etc.

At this point we focus on measurable skills and following the content-analysis methodology which is based on the extraction of a structured model from the initial unstructured form of the job description. This model represents the jobs in a manageable and flexible way which is proper for our CBR algorithm to manipulate it.

In detail the structured form of the job is described as follows: (i) the job's attributes represent the required skills the job applicant should have, (ii) attributes' values reflect the level of each skill (the value may have one of the four below forms: exact value, range with lower limit, range with upper limit and range with both lower and upper limit) and (iii) attributes' weights mirror the importance of each skill for the specific job.

Similarly the job seeker's profile must have the same structure with the job in order the algorithm be able to conduct the comparison and suitability to be found. Basically, the job seeker's profile is produced according to the attributes which are extracted from the content of the jobs' models. The above concept is presented in Fig. 6.

B. Execution and results

We chose 100 different jobs from the IT category (analyst, health IT, project management, network administration, system designer, programming, software engineering, technical support etc) and we identified 8 skills: education level (S_1), past experience (S_2), language skill (S_3), programming

Authorized licensed use limited to: SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECH. Downloaded on September 14,2021 at 08:59:32 UTC from IEEE Xplore. Restrictions apply.

knowledge (S_4), network knowledge (S_5), technical support knowledge (S_6), development experience (S_7) and project management experience (S_8). Also we discriminate the skills into the four cases: {Exact value (E), Lower limit (L), Upper limit (U), both Lower and Upper limit (LU)}, which the FoDRA needs to calculate the suitability. Finally, the output of the content analysis is mapped into a manageable shape in order for the algorithm to be able to manipulate it. This means that an 8-dimensional vector is used to represent the jobs (J). Each element of the vector is mapped to an attribute and each attribute has a value (j_i) and a weight (w_i), where i=1, 2,…,8 and $e_i = j_i \times w_i$ are the elements of $E = (e_1, e_2, e_3, \ldots, e_8)$. Note that the aforementioned 8 skills S_1 to S_8, were rated based on a scoring table defined by experts on job seeking and recruiting domain. Specially, the value weight variable defined to be 1 for each skill.

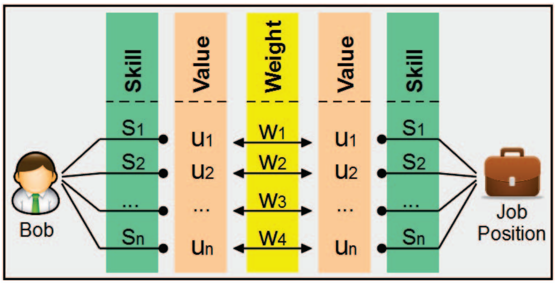

Fig. 6. The job seeker's model has the same structure as the job does. The form is structured on: i) the job's attributes, ii) the attributes values and iii) the attributes weights.

On the other hand, the profile of the job's seeker is represented with an 8-dimensional vector (S), in which each element is mapped into the same attribute as the jobs require. The values of the attributes u_i, i=1, 2, 3 …, 8 use the same scoring table. Finally, the corresponding vector is defined as $S = (s_A, s_B, \ldots, s_H)$. Table I shows the job seeker's representation while Table II presents the skills' values of the top 5 jobs with their corresponding category (i.e. E, L, U, LU).

TABLE I. CONTENT ANALYSIS RESULTS FOR THE JOB SEEKER

Attribute names (required skill)	Attribute values (required skill level)
S_1	25
S_2	20
S_3	11
S_4	20
S_5	29
S_6	19
S_7	21
S_8	21

TABLE II. THE SKILLS VALUES OF THE TOP 5 JOBS

Skills	Job_5	Job_{11}	Job_{14}	Job_{22}	Job_{35}
$S_1(E)^a$	26	25	27	23	25
$S_2(E)^a$	20	22	22	20	21
$S_3(E)^a$	10	12	11	9	9
$S_4(L)^b$	25	21	22	23	20
$S_5(L)^b$	33	35	30	31	35
$S_6(U)^c$	19	18	19	18	19
$S_7(U)^c$	17	19	15	20	16
$S_8(LU)^d$	10-20	10-20	15-30	20-30	15-20

[a] Exact value (E)

[b] Lower Limit (D)

[c] Upper Limit (U)

[d] Lower and Upper Limit (LU)

After the execution a sorted list is produced where the top 5 jobs are retrieved for further processing (see Table III). The algorithm is executed for the same dataset but for different values of the *p* variable namely 1 and 2 (p=1 Manhattan distance and p=2 Euclidean distance). Thus, we have two different sorted lists, for each *p* value. According to an expert on job seeking and recruiting domain, the choice of *p* value does not play such an important role, both choices produce accepted results based on his/her experience. Of course, this is a primary research and must be evaluated by other field experts.

TABLE III. THE TOP 5 JOBS AFTER THE CONTENT-BASED RECOMMENDATION ALGORITHM FODRA EXECUTION FOR DIFFERENT VALUES OF *p* VARIABLE

Sorted list of jobs for different values of p			
$p = 1^e$		$p = 2^f$	
Job	**Suitability Value**	**Job**	**Suitability Value**
Job_5	10	Job_5	10,585786
Job_{35}	7	Job_{35}	7,763932
Job_{11}	6	Job_{11}	6,763932
Job_{14}	5	Job_{14}	6,171573
Job_{22}	3	Job_{22}	4,171573

[e] p=1 is Manhattan distance

[f] p=2 is the Euclidian distance

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose a content-based recommendation approach which extends and updates Minkowski distance in order to address the challenge of matching people and jobs. The results confirm that the JRS play an important role in the field of job seeking and recruiting. The experimental steps navigate the companies through a coherent procedure to use effectively our suggested algorithm as a part of a JRS and to take the advantage of its benefits: fewer job positions for in depth evaluation, improvement of candidate description, incensement of companies' internal operation level, personnel evaluation, etc.

The methodology we introduce presents promising performance and encourages us to improve and investigate further its capabilities. So, further research will be conducted in order to improve the performance, in terms of time response

Authorized licensed use limited to: SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECH. Downloaded on September 14,2021 at 08:59:32 UTC from IEEE Xplore. Restrictions apply.

and reliability. The algorithm must be tested on other datasets coming from various fields. Furthermore, different values of the *p* variable will be examined. Directions of future research also consist of FoDRA's testaments using other similarity measures (i.e. cosine similarity, Pearson correlation, etc). These ideas and other related research is currently conducted and a new study will be published in the near future.

REFERENCES

[1] R. J. Mooney and L. Roy, "Content-Based Book Recommending Using Learning for Text Categorization," in In Proceedings of DL '00: Proceedings of the Fifth ACM Conference on Digital Libraries, New York, NY, 2000.

[2] G. Shani and A. Gunawardana, "Evaluating Recommendation Systems," Springer, pp. 257–298, 2011.

[3] G. Linden, B. Smith and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," Published by the IEEE Computer Society, IEEE Internet Computing, vol. [7] , no. 1, pp. 76-80, 2003.

[4] A. Das, M. Datar, A. Garg and S. Rajaram, "Google News Personalization Scalable Online Collaborative Filtering," in In Proceedings of the 16th International Conference on World Wide Web, WWW '07, Banff, Alberta, CANADA, 2007.

[5] M. Balabanovic and Y. Shoham, "Fab: Content-based, Collaborative Recommendation. Communications of the ACM," vol. 40, no. 3, pp. 66-72, 1997.

[6] R. Burke, "Hybrid Recommender Systems:Survey and Experiments," User Model. User-Adapt. Interact, vol. [12], no. 4, pp. 331–370, 2002.

[7] R. Burke, "Hybrid Web Recommender Systems," The Adaptive Web: Methods and Strategies of Web Personalization, vol. [4321], pp. 377-407, 2007.

[8] Burke, R., A. Felfernig, and M. H. Goker. "Recommender systems: an overview". AI Magazine, 32(3), 13–18, 2011.

[9] Ramezani, M., Bergman, L., Thompson, R., Burke, R. and Mobasher, B., "Selecting and applying recommendation technology". In Proc. of International Workshop on Recommendation and Collaboration, in Conjunction with 2008 International ACM Conference on Intelligent User Interfaces. Canaria, Canary Islands, Spain, 2008.

[10] Minkowski, H. Allgemeine Lehrsatze uber die konvexe Polyeder, Nachr. Ges. Wiss., Gottingen, 1897, 198–219 (=Ges. Abh., vol 2, pp. 103-121, Leipzig-Berlin, 1911).

[11] Almalis, N.D., Tsihrintzis, G.A. and Karagiannis, N., "A content based approach for recommending personnel for job positions", in Proceedings of the 5th International Conference on Information, Intelligence, Systems and Applications, IISA 2014, pp. 45-49, Chania, Greece, 2014.

[12] BadulSarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," Proceedings of the 10th International Conference of World Wide Web, pp. 285-295, 2001.

[13] Burke, R. "Knowledge-based recommender systems". In A. Kent (ed.), Encyclopedia of Library and Information Systems. Volume 69, Supplement 32. Marcel Dekker, 2000.

[14] J. Malinowski, T. Keim, O. Wendt, T. Weitzel, "Matching People and Jobs: A Bilateral Recommendation Approach," in Proceedings of the 39th Annual Hawaii International Conference on System Sciences, Hawaii, USA, 2006.

[15] Jobvite 2014. Social Recruitment Survey. August 2014. http://recruiting.jobvite.com

[16] W. Hong, S. Zheng, H. Wang, J. Shi, "A Job Recommender System Based on User Clustering", Journal of Computers, vol. 8, no. 8,pp. 1960-1967, 1, Aug. 51 2013.

[17] Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Communications of ACM, Vol. 18, No. 11, pp. 613 – 620, 1975.

[18] Guo, Xingsheng, Jerbi, Houssem, O' Mahony, Michael P. : An Analysis Framework for Content-based Job Recommendation. 22nd International Conference on Case-Based Reasoning (ICCBR), Cork, Ireland, 29 September - 01 October 2014, 2014.

[19] Fazel-Zarandi, M. and Fox, M. S. Semantic Matchmaking for Job Recruitment: An Ontology-Based Hybrid Approach. In Proceedings of the 8th International Semantic Web Conference, 2009.

[20] M. Hutterer, "Enhancing a job recommender with implicit user feedback," In Fakultät für Informatik, Technischen Universität Wien, 2011.

[21] D. H. Lee and P. Brusilovsky, "Fighting information overflow with personalized comprehensive information access: A proactive job recommender," in ICAS '07: Proceedings of the Third International Conference on Autonomic and Autonomous Systems, (Washington, DC, USA), p. 21, IEEE Computer Society, 2007.

[22] S. T. Al-Otaibi and M. Ykhlef, "A survey of job recommender systems," International Journal of the Physical Sciences, vol. 7(29), pp. 5127-5142, July, 2012.

[23] Lathia, N., Hailes, S., and Capra, L., The effect of correlation coefficients on communities of recommenders. In *SAC '08: Proceedings of the 2008 ACM symposium on Applied computing*, pages 2000–2005, New York, NY, USA, 2008. ACM.

[24] Spertus, E., Sahami, M., and Buyukkokten, O., Evaluating similarity measures: A large-scale study in the orkut social network. In *Proceedings of the 2005 International Conference on Knowledge Discovery and Data Mining (KDD-05)*, 2005.

Authorized licensed use limited to: SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECH. Downloaded on September 14,2021 at 08:59:32 UTC from IEEE Xplore. Restrictions apply.