

Applying Different Classification Techniques in Reciprocal Job Recommender System for Considering Job Candidate Preferences

Gözde Özcan
Istanbul Technical University
Department of Computer Engineering
Maslak, Istanbul, Turkey
gzdzcn@gmail.com

Şule Gündüz Ögüdücü
Istanbul Technical University
Department of Computer Engineering
Maslak, Istanbul, Turkey
sgunduz@itu.edu.tr

Abstract—In this paper, a reciprocal job recommendation system, CCRS (Classification - Candidate Reciprocal Recommendation), is proposed. With this proposed system, offering job advertisements in a sequence for candidates that they can get feedback reciprocally by using the user's profile, interaction and preference information is aimed all together. An approach has been used based on the preference information of the candidates to determine the jobs' order in the proposed list and the success of different classification methods has been compared to estimate the feedback rate of the advertisements for the target candidate. CCRS also addresses the cold start problem of new candidates joining the site by providing recommendations based on their profiles. The performance of the proposed method was evaluated by using various performance measurements on an actual data set received from an online recruiting website. Evaluation results show that the proposed method outperforms the compared methods for the top 10 ranked recommendations.

Keywords—Job Reciprocal Recommender; Classification Methods; Cold start

I. INTRODUCTION

Excessive information has been generated with the increase of the number of users on online social networks and it has become difficult to find and to decide the items users can choose among them.

Many websites have developed recommendation systems to help them find the items that could interest their users to overcome this problem[1]. The aim of this system is to recommend items that might interest users, such as: books, movies, pharmacy products and restaurants. There are different approaches used for these recommendations; such as content-based[3], collaborative filtering[4] and hybrid methods[5].

A special class of recommendation systems is also reciprocal recommendation systems. These systems are designed for practices focused on people-to-people recommendations and include recommendations that ensures satisfaction of both sides. Reciprocal recommendation systems are different from traditional user-item recommendation systems. While traditional recommenders only recommend products according to the preferences of the

person for whom the recommendations are generated, reciprocal recommendation systems recommend people-to-people according to the preferences of both sides [6]. Many systems include or use reciprocal recommendation systems. One of the most important examples is the job search website. These sites aim to help jobseekers find jobs that match their preferences and characteristics or recommend reasonable candidates that match their needs to recruiters [2].

The basic problem for reciprocal recommendation is to make a match that fulfils the mutual needs of both users. The relations between the users must be modeled considering the preferences of both sides. In this study, a reciprocal job recommendation method that handle the users' profiles, interaction features and preferences all together to make an recommendation for the candidate is presented. A summary of the contributions made by this model are below;

- A model using profile, preference and interaction information of both sides for reciprocal job recommendation is proposed. This system recommends a job list to the target candidate by using the applications made by the candidates that are similar to the target candidate and the preference information of these jobs.
- As a part of the proposed system, an approach based on the preference information of the candidates for the sequencing of the recommended job list is used.
- The proposed system has been evaluated by using a data set received from a job recruiting websites.

An important advantage of the proposed method is to avoid cold start problem for the candidates. With this method, there is no need for the candidate to make any entry other than profile information, to make a recommendation for the new candidate that signs up the system.

The rest of this study was organized as follows: Section 2 presents a literature review about reciprocal recommendation. The 3rd part includes the description of data set and the problem, the 4th part provides the structure of the proposed system and the description of the

methods used for the proposed systems, the 5th part presents the experimental results. Finally, section VI contains some conclusions and some idea about our future work.

I. RELATED WORKS

Reciprocal recommendation systems offer recommendation according to the users' mutual preferences. There are different application areas for the reciprocal recommendations, such as: online recruiting services, online dating services, online mentoring systems. Recommendation systems are basically divided into 4 categories[7]: collaborative filtering, content-based, information-based and hybrid approaches. A lot of research has been done for recommendation systems in academic and industrial fields using these approaches. The concept of reciprocal recommendation is first introduced in [2]. In this study, the characteristics of reciprocal recommendation systems have been analyzed in detail and compared with the traditional recommendation systems.

[8] iHR+ method offers reciprocal recommendations for mobile applications. This method divides the profile information of job seekers and employers into 2 categories: self-description and preferences. It generate recommendation using cross-similarity among users' mutual characteristics. A content-based approach that uses the skills of the candidates and the requirements of employers to propose suitable candidates for a specific position is presented in [9]. But these two methods did not take the users' preference information and the interaction characteristics into account when generating recommendation. [1] proposes a new collaborative filtering method to improve user recommendations in reciprocal and bipartite social networks. This proposed method has been evaluated for online dating networks. However, this method did not use the content information to generate recommendation. [6] CCR method presents a new reciprocal content-collaborative based recommendation system for online dating systems. While the content-based part finds the similar users, the collaborative filtering based part uses the interaction of the similar users to produce reciprocal recommendations. This system also avoids the cold start problem.

Apart from the above mentioned methods, our recommended method proposes the content along with interaction and preference information of both sides. Unlike other methods, for the users' feedback information, it offers an approach that handles the preferences of the jobs with classification methods and estimates the feedback rate for the target candidate. Thus, the content and preference information for both users are used together to obtain the information of whether they would prefer each other. At the same time, the proposed method avoids the cold start problems of the candidates. Another difference from other methods is that it uses a different approach that uses the preferences of the candidates to sequence the recommendation list of jobs.

II. DEFINITIONS AND PROBLEMS

This section includes the contents of the data set used, and the definition of the problem for which solutions are proposed.

A. Dataset Introduction

The proposed method was generated on a data set received from Kariyer.Net[10], a job seeking website. Data received from this website consist of information not containing the identity and contact information of users. This data sets consists of 7455 applications in total, for 10 job postings by 5600 candidates between April 2015 and December 2015.

1) Candidate profile features:

Candidate profiles consist of 19 features. Those items are listed below:

- The Curriculum Vitae Status: Divided in 2 groups: True(Active) or False(Passive).
- Marital status: Candidate's marital status. Divided in 2 groups: Single or Married.
- Gender: Candidate's gender. Divided in 2 groups: Male or Female.
- Age: Candidate's age. Between 15 and 58.
- Military state: Candidate's military state. Divided in 5 groups: Continuing ,Privileged, Done, Probation, Not Done.
- Educational Status: Candidate's academic degree. Which consist of 6 groups: Doctorate, Postgraduate, Undergraduate, Technical School, High School, Primary School.
- Driving license: Candidate's driving license class info. Divide in 2 groups: True(has) or False(has not).
- Disability Status: Candidate's disability status. Divide in 2 groups: True(disabled) or False(not disabled).
- Country: The Country that candidate lives in.
- City: The City that candidate lives in.
- Work Status: Candidate is working or not. Divide in 2 groups: True or False.
- Employment Status: The information about the candidate's employment status in the course of application. Divide in 2 groups: True or False.
- Employment Preference: Candidate's employment preference. Divide in 5 groups: Full Time, Part Time, Volunteer, Intern and Project Based.
- Grades: Candidate's graduation degree.
- Foreign Languages: Candidate's knowledge of foreign language.

- Certificates: The information about the candidate's certificates.
- Experience: Total experience time of candidate.
- Business Area: The business fields of the candidate who preferred to work (IT, Reserach&Development, financ, etc.).
- Sector: The sectors of the candidate who preferred to work (Informatics, health, media,etc.)
- Skills: Extra knowledge/skills possessed by the candidate.

2) Candidates Interaction Features:

There are 9 more features in addition to those 19 features above in the data set used for the candidates. These features reflect the actions of the job announcing companies on the candidate's Curriculum Vitae (CV). These are as follows:

- CV Opened: CV is opened by the Employer or not.
- Drawer Thrown: CV is saved by the Employer to look again or not.
- Short Long List: CV url added to the list by the Employer or not.
- Printed CV: CV is printed by the Employer or not.
- CV Shared: CV is shared with others by the Employer or not.
- Job Offered: Candidate has been offered by employer or not.
- Interview: Candidate was invited to interview or not.
- CV Multiple Times Opened: CV is looked more than once or not.

3) Job profile features:

- Job Location: After getting the job, where the candidate will be working.
- Position Offered: Position offered for the job.
- Employer/Company: Company that is offering the job.
- Sector: major/discipline in which the job is offered.
- Experience: Required experience for the job.
- Education: Required education for the job.
- Military Situation: Required military situation for the job.
- Certificate: Required certificates for the job.

- Foreign Language: Required foreign language information for the job.
- Requirements: Required extra knowledge/skills for the job.

B. Problem Definition

The main problem in reciprocal job recommendation systems is to make a match that fulfils the mutual needs of both users (candidate and job). In these systems, the relations between the users must be modeled considering the preferences of both sides. In job recommendation systems, the job seekers look for positions that match their preferences and the employers search suitable candidates to fulfill the requirements of the job. A reciprocal job recommendation scenario consists of two separate set of users, C and J users to represent the set of candidates and jobs, depending on $c \in C$ and $j \in J$, the job list $R1(c)$ recommended for c candidate, the candidate list $R2(j)$ recommended for j job, to match the (c,j) couple reciprocally, c and j users need to be in each other's preference list [2].

$$RR : \{j: j \in R1(c) \cap c \in R2(j)\} \quad (1)$$

Where, RR: Reciprocal Recommendation.

In this study, a method is proposed to recommend jobs in an order, which candidate may prefer and which the candidate may be in the jobs preference list, to a new candidate signed up to the system in job recommendation systems. At the same time, this method avoids the cold start problem of the candidates.

III. FRAMEWORK OVERVIEW

Fig. 1 illustrates the CCRS general working structure of the reciprocal job proposal system. The recommended system consists of 4 components that are associated with each other.

- *Creation of the job list to be proposed to the candidate* : In order to make a recommendation to a candidate who is entering the system recently, the first 10% of the candidates similar to this candidate are to be found and a CandidateList is to be created. If the candidate is not a cold start, then this candidate himself/herself is added to the CandidateList as well. Afterwards, the previous jobs to which the candidates of the CandidateList applied are found and the JobList is created. Additionally, the first 1% of the jobs similar to the jobs in the JobList are added to the JobList. The list of the jobs which the target candidate may prefer to apply to is obtained in this way. In order to calculate the similarity between the users (candidate, job) with the users within their own set the cosine similarity formula is used. Cosine similarity between 2 vectors a & b is calculated by the following formula:

$$\text{Cosine similarity (a, b)} = (a \cdot b) / (|a| |b|) \quad (2)$$

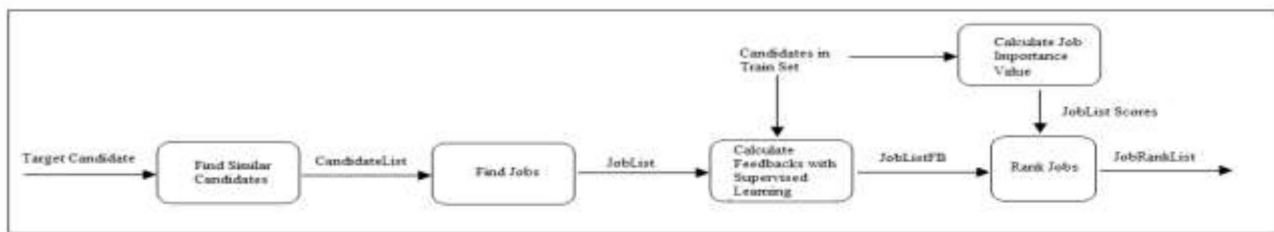


Figure 1. An overview of CCRS

The reason why 10% is accepted for similar candidates and 1% for job is the fact that these ratios provided the best results during validation tests.

- *Obtaining the return values of the jobs:* The label value indicating the return value of all jobs in the data set that replied to all candidates from whom they received an application is obtained, by using the interactions on the curriculum vitae of the candidates who previously applied to the jobs. These label values are trained for each job via methods of classification and the reply of the jobs in the JobList that may be sent to the target candidates is calculated in this way. The target candidate feedback value of the jobs has been designated as FB. In this way, the information whether they will be preferring the target candidate for each job in the JobList is obtained (JoblistFB).
- *Determining the Importance values of the jobs:* A confidence value is calculated for each candidate by using the applications in the system that were previously carried out by the candidates. The confidence value is the value obtained from the similarity of the jobs to which a candidate has applied to. If the similarity among the jobs the candidate applied to is very high, the confidence value of the candidate is high, if it is low then the confidence value of the candidate is low. Each candidate provides his/her confidence value to the job they apply to. The importance value of the jobs is determined through the sum of the candidates who are applying to it and these importance values are accepted as the score value of the jobs and names as SV. Thus, a score value is obtained for each job in the JobList (JobListScore).
- *Creation of the ranked job recommendation list:* The FB value and SV values of each job in the JobList are multiplied and the jobs are recommended to the target candidates in accordance with the obtained values, thus obtaining the JobRankList.

Methods used in the proposed model;

A. *Determination of the return values of the jobs related to the applications:*

The data set received from Kariyer.Net [10] contains information about which jobs candidates applied and the actions the companies which announced the job were taken on the curriculum vita(CV) of the candidates who applied for that

job. However, data set shows that 10 candidates in total are hired and the amount of hiring for each job is not uniformly distributed.

In order to evaluate the proposed method, the candidates hired by the job announcing companies must be known. Yet, the fact that the companies mostly do not specify the candidates they hired to the system, there occurs sparse data in the system. As a solution for this problem the activity information of the companies are used. If the job announcing company takes action for at least 4 on the CV of the candidate applying for the job, these candidates are assumed to be hired for the job and these applications are labeled with the "1" class whereas the opposite cases are labeled with "0". As a result of these operations label 1 is used for 748 of applications in total. These label values are assumed to be the real return values of the jobs and used in the estimation of accuracy of this model proposed as the real label values. In this work, the focus is not on which assumption is truer, but on the elimination of the sparse data of the return values in order for the methods to be evaluated, resulting in the usage of interactions of the candidates.

B. *Obtainment of the Confidence Values of the Candidates and the Importance Values of the Jobs:*

The confidence values of the candidates and the importance values of the jobs are being used for the determination of the place of the recommended jobs that have been recommended to the candidates. The confidence value of the candidates determines the similarity among the applied jobs. If the candidates applies to dissimilar jobs then the confidence value is high. The confidence value of the candidates is determined by the sum of the similarity values among the jobs that have been applied to by the candidates.

$$AGD(C) = \frac{1}{m.(m-1)} \sum_{i=1}^m \sum_{k=i+1}^m (sim(J_k, J_i)) \quad (3)$$

Where, AGD(C): Confidence value of candidate C,
Sim(J_k,J_i): Similarity value between the i and j jobs
m: Number of jobs candidate C applied to,

The importance value of the jobs is obtained via the sum of the confidence values of the candidates that applied for the job.

$$IOD(J) = \sum_{c=1}^n AGD(C) \quad (4)$$

Where, IOD(J): Importance value of the J job,
AGD: Confidence value of candidate C,
n : Number of candidates applying for the J job.

V. EXPERIMENTAL EVALUATION

A. Data Acquisition

In order to measure the performance of the proposed method, data set obtained from Kariyer.Net[10] website has been used. In this data set, candidates and job postings consist of qualities specified in Section 3. Data set consists of 7455 applications in total to 10 job postings from 5600 candidates, and includes only job postings issued for software engineering positions. Label information obtained in Section 4 have been used to measure the success of methods. 1000 applications were used to assign the parameters of classification methods that were used to predict return values of job postings. Accuracy of the method was measured for both cold start and non cold start candidates by creating two evaluation sets. Excluding the 1000 applications used in determining parameters of the classification methods, both evaluation sets were divided into two sets as train and test. Accordingly, train set for cold start candidates consisted of 3216 applications from 1336 candidates who made at least 2 applications, while test set consisted of applications made by 3239 candidates with just 1 application. Test set was divided into three equal pieces, and the average of success values in these pieces was calculated. For non cold start candidates, train and test sets were created by dividing 3216 applications, made by candidates with at least 2 applications, into two. Accordingly, education set consists of 2145 applications, and test set consists of 1071 applications.

B. Evaluation Metrics

Proposed method was programmed with java, using weka library[11] for classification methods. Two measurements are used for evaluating the success of the proposed method.

1) *Ranking Evaluation*: Mean averaged precision(map) measurement is used to measure the accuracy of recommendation ranks of applications made by the candidates, and the candidates were accepted to. Methods used for this are precision@k and recall@k. These methods generate a success value according to the rank of actually recommended applications on the top-k recommendation list [12].

$$ap@q = \sum_{i=1}^Q P(k)/\min(r, q) \quad (5)$$

Where, P(k) : Precision value for first k number of applications proposed,

r: Number of applications that should be actually proposed,

q: Number of actually proposed applications on the recommendation list.

After average precision values for each candidate on the recommendation list are obtained, mean average precision is used to calculate precision value of the recommendation queues for each candidate. This method takes the average of average precision values obtained for each candidate.

$$MAP@q = \sum_{i=1}^Q ap@qi/Q \quad (6)$$

Where, Q: Candidate number

i: Candidate's precision value

2) *Evaluation of Accuracies*: True negative rate (TNR) measurement is used to calculate the accuracy of applications that are actually made by the candidate and the candidate was rejected are not on the recommendation list[12]. For candidate C, actual rejected applications M, and applications not recommended to the candidate N, TNR is calculated as following;

$$TNR = \frac{MnN}{M} \quad (7)$$

C. Experimental Results

In this section, the effects of the classification methods on the accuracy of the proposed method has been evaluated and the proposed method has been compared to previously proposed methods. Classification methods used for the proposed method are Naive Bayes (NB), Support Vector Machine (SVM) with rbf kernel classifier, Multilayer Perceptron (MLP) and C4.5. Detailed descriptions of these classification methods can be found in [13]. By applying the 10-fold cross-validation, the best parameter values found from a validation set of 1000 applications are: for SVM complexity parameter(c) = 10, kernel parameter(γ) = 0.07; for MLP seed(s)= 3, hiddenLayers(h) = 10; for C4.5confidenceFactor= 0.01. And k is set as 10.

The methods used to compare the proposed method are content-collaborative reciprocal recommender (CCR) [6] and a mobile reciprocal job recommender system (iHR+) [8]. CCR is a reciprocal content-collaborative recommendation system proposed for online dating systems. The Content-based part allows similar users to be found by user profiles, and the collaborative filtering part produces reciprocal recommendation using interactions of similar users. Using similar users to provide recommendations, and producing recommendations among cold start users, CCR shows similarities with the proposed CCRS system. Therefore, CCR can be defined as a special condition of the CCRS system. iHR+ is a mobile application system that provides reciprocal job recommendations. In this method, attributes of job seekers and recruiters are separated into two categories: self-description and preferences. The similarity between these attributes is calculated by cross-similarity and reciprocal recommendations are generated. Since the iHR+ method uses profile information of users to make recommendations and also considers data sparsity, it is a method that can be applied to the CCRS system. Therefore it can be compared to the proposed CCRS method.

In Table I, different classification methods for the proposed method on the evaluation set generated for cold start candidates, and MAP and TNR results obtained for the compared methods are shown. According to this, the best TNR value for the cold start data set is obtained by MLP, while the best MAP value is obtained by C4.5. The proposed method has achieved higher success for MAP and TNR from CCR and iHR+ methods, for all classification methods tried.

In Table II, the results obtained from non cold start candidates are shown. According to this, the best TNR value is obtained by MLP, while the best MAP value is obtained by SVM. For both data sets, the TNR value for the iHR+ method is zero. This is because all applications are proposed to the candidates since the similarity between candidates and 10 applications are all higher than zero.

TABLE I. MAP AND TNR VALUES OF COMPARED METHODS PERFORMANS FOR COLD START DATA SET

	NB	C4.5	MLP	SVM	CCR	iHR +
MAP	0.42	0.49	0.43	0.48	0.43	0.37
TNR	0.56	0.55	0.57	0.56	0.48	0.04

TABLE II. MAP AND TNR VALUES OF COMPARED METHODS PERFORMANS FOR NON COLDSTART DATA SET

	NB	C4.5	MLP	SVM	CCR	iHR+
MAP	0.52	0.51	0.48	0.56	0.48	0.47
TNR	0.58	0.57	0.61	0.53	0.45	0

In Fig. 2 and Fig. 3, precision values of the methods for cold start and non cold start data sets are shown. For every k value in the cold start data set, similar precision values were obtained for both C4.5 and SVM classification methods. For every k value in the non cold start data set, the best result was obtained for the SVM method.

According to these results, SVM method can be used in this data set for MAP evaluation, since C4.5 and SVM values are very similar for the MAP values obtained from the cold start data set, and SVM method is significantly better than the remaining methods in the non cold start data set. For TNR, however, MLP obtained the highest values for both data sets.

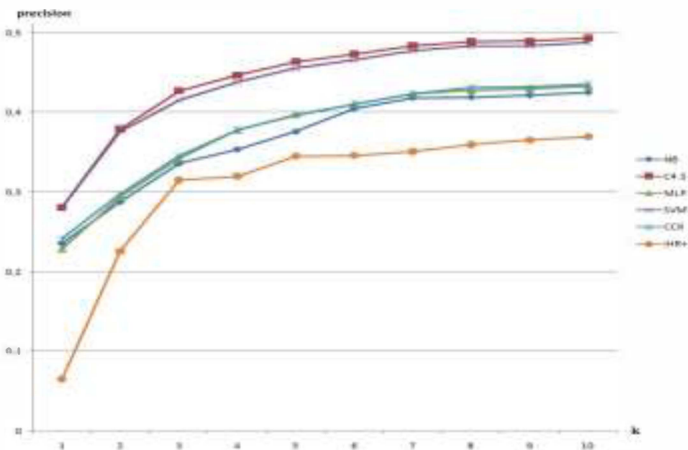


Figure 2. Precision@k for cold start data set

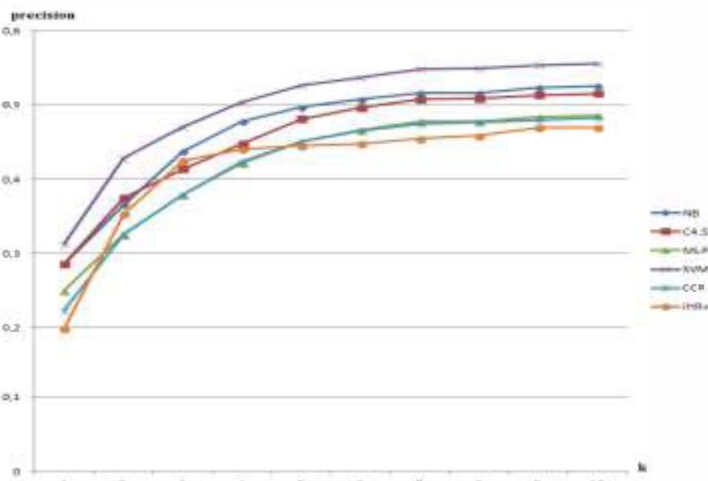


Figure 3. Precision@k for non cold start data set

VI. CONCLUSION

In this study, a reciprocal job recommendation system (CCRS) has been proposed. In order to recommend jobs to the candidates, the proposed system takes advantage of applications of similar candidates, and preference information on jobs. Created system has been tested using different classification methods. As a result, it was accepted that the proposed system has the best performance for MAP measurement with SVM classification methods. Experimental results show that the performance of the proposed system is better than other compared systems.

The reason for classification methods to give values close to each other might be insufficient data. As part of our future studies, plans are being made on assessment of the proposed method on bigger data sets.

REFERENCES

- [1] K. Zhao; X. Wang; M. Yu; B. Gao: User Recommendations in Reciprocal and Bipartite Social Networks--An Online Dating Case Study, IEEE Intelligent Systems, vol.29, Pp. 27 - 35, March/April 2014
- [2] Pizzato, L., et al. RECON: a reciprocal recommender for online dating. in Proceedings of the fourth ACM conference on Recommender systems. 2010. ACM.
- [3] M.J. Pazzani and D. Billsus. Content-based recommendation systems. The Adaptive Web, 4321:325, 2007.
- [4] J.A. Konstan, B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, and J. Riedl. Grouplens: applying collaborative filtering to usenet news. Commun. ACM, 40(3):7787, 1997.
- [5] R. Burke. Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4):331370, 2002.
- [6] J. Akehurst, I. Koprinska, K. Yacef, L. Pizzato, J. Kay, and T. Rej. Ccr-a content-collaborative reciprocal recommender for online dating. In Proc. of IJCAI, pages 2199–2204, 2011.
- [7] Adomavicius, G., A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 2005. 17(6): p. 734-749. [
- [8] H. Wenxing, C. Yiwei, Q. Jianwei and H. Yin iHR+: A Mobile Reciprocal Job Recommender System. In Computer Science & Education (ICCSE), 2015 10th International Conference on. 2015. IEEE
- [9] N. D. Almalis; G. A. Tsihrantzis; N. Karagiannis: A content based approach for recommending personnel for job positions. In Information, Intelligence, Systems and Applications (IISA), 2014 5th International Conference on. 2014. IEEE
- [10] Kariyer.net Available: www.kariyer.net
- [11] (2008). Weka Machine Learning Project, University of Waikato, <http://www.cs.waikato.ac.nz/ml/weka/>.
- [12] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.
- [13] Ethem Alpaydin, Introduction to Machine Learning, The MIT Press, 2010.