CAREER AND SKILLS RECOMMENDATIONS USING DATA MINING TECHNIQUE: MATCHING RIGHT PEOPLE FOR RIGHT PROFESSION, IN PAKISTANI CONTEXT

MARIA KIRAN¹, HIRA ASIM¹, MALIK TAHIR HASSAN¹

¹Department of Software Engineering, University of Management and Technology, Punjab, Pakistan
 Email: tahir.hassan@umt.edu.pk

ABSTRACT. *There are a number of recommendation systems available on the internet for the help of jobseekers. These systems only generate job recommendations for people on the basis of input entered by user. The problem observed in Pakistani people is they are not clear in which field they should start or switch working. Before searching and applying for a job, one should be clear about his/her profession and important skills regarding selected profession. Based on above issues, there is a need to design such a system that can overcome the problem of profession selection and skills suggestions so that it can be easy for a jobseeker to apply for a specific job. In this research, the problem which is discussed above is resolved by proposing a model by using Association Rules Mining, a data mining technique. In this model, professions are recommended to job seekers by matching the profile of applicant or job seeker with those persons who have same profile like educational background, professional skills and the type of jobs which they are doing. The data collected for this research itself is a major contribution as we collected it from different sources. We will make this data publically available for others so that they can use for further research.*

Keywords: Association Rule Mining; Data Mining, Profession Recommendations; Skills Recommendations

1. Introduction. The information is growing day by day with the fast growth of internet. It is quite difficult to manage the information available online. For a jobseeker it is really tough and time taking process to search for a job online. What the jobseekers usually do, they search jobs on different job searching sites and apply for the suitable one. When they search for a specific job on a site, the system or site returns a list of multiple jobs against their search. Most of the times, jobseekers are not clear in which field they should start or switch working. They don't know which professions suit them best and which skills they should have or improve for a specific profession. Regarding this problem, we have proposed a model for making profession recommendations and skills recommendations for a specific profession. The proposed model focused on the educational background, working experience, skills and the professions of the people who are already doing jobs somewhere and by analyzing this training data, we have recommend professions to job seekers which will exactly match with their profile like their educational background and skills. The model is also able to recommend technical skills to those who have selected a profession and want to know about skills which will be helpful in context of that profession. The model is proposed by using some data mining techniques such as association rule mining.

I. Recommender Systems. Recommender systems are used where people have so many choices and they are confused to pick the best one. A number of recommender systems have been introduced in different fields like books, news, movies, novels, etc. Following are the major categories by using which recommender systems generate recommendations.

II. Content-Based Recommendations. Recommendations are generated on the basis of items that are similar to those items which user has liked in the past [12]. Description of the items as well as the user's interest matters a lot in content based recommendation [13].

- III. **Collaborative Recommendations.** Recommendations are made on the basis of users' preferences [14]. Collaborative filtering mainly consists on the user-based and item-based [15].
- IV. **Hybrid Recommendations.** This approach combines more than one recommendation technique for making recommendations [16].
- V. **Knowledge-Based Recommendations.** Knowledge based recommendations approach is applied when other approaches failed to apply. This approach based on the explicit knowledge regarding user preference, items, etc [17].

Figure-1: Major Categories in which recommender systems generate Recommendations

Data Mining Techniques for Recommender System. All the recommendation systems developed by using data mining techniques, make recommendations with the help of attributes and knowledge learned [17]. There are some data mining techniques that are mostly used in the development of recommender system. The techniques include:

- **Clustering.** This technique in recommender system is used to find the group of those people who have same preferences. Clustering is basically an unsupervised learning technique [11]. It can be used to improve the efficiency of recommender system.
- **Classification.** In this technique, a classifier or a model is built that is based on training data and then can be used for new items. It usually implemented with the help of machine learning approaches like neural networks and Bayesian network etc [18].
- **Association Rules.** The most useful data mining technique which is used in recommender system is association rule mining [18]. In recommender systems, it is also known as item-to-item correlation. Association rules are commonly used in recommender systems for the purpose of identifying the patterns among products and to make the recommendations for the users on the basis of items they have selected.

The aim behind the proposed model is to facilitate the job seeker who do not have much experience and are not clear about which profession is suitable for them, which skills are important to join a particular profession and for which job they can apply after selecting a profession.

2.Literature survey. The term recommender system was recognized first time by Resnick and Varian in 1977. Many researchers are working to improve the quality and scalability of recommendations by several means. In the context of person-job fit, the techniques of simple keyword based search and filter are not satisfactory because selection decisions usually depend on fundamental attributes like personal characteristics and skills [1].

Malinowski proposed a hybrid system and the thought depended on the best match between the applicants or jobseekers and the jobs that needs to consider both the inclinations of the recruiters and the applicants [2]. In the result of proposed system, the prediction of the matches between jobseekers and jobs were done according to the applicant's CV, employer descriptions and requirements as well as previous rating information. Another recommendation technique that is Apriori algorithm was used for the purpose of web recommendations [3]. There was a limitation of this proposed system. It is very difficult for the people to rate a job without working on that. An idea of joining the conceptual and practice information was exploited in hybrid web recommender system in order to improve a reinforcement learning agenda, primarily planned for web recommendations based on web usage data.

Fuzzy association rule mining is another method that is proposed for web recommendation system [3]. This method describes some standards algorithms that are helpful in the development of web recommender systems. The method incorporates the practice of fuzzy association rule mining algorithm for association rule mining. It helps the people to find those web sites that are relevant according to their taste.

Recommender system methods are divided into four major classes. They are content based filtering, collaborative filtering, knowledge based and hybrid [4]. In content based filtering, we use keywords for describing the items. We also built a user profile which shows the type of item that user likes [5]. According to previous work [6], a job recommender system comprises of a jobseeker or candidate subsystem that is intended for job applicants and an e-recruiting subsystem. Apart from the old fashioned or traditional recommender systems, which only focus on the one-sided preference like the preference of a user on the items, this system is developed on user clustering which generates recommendations separately for each group.

Another job recommender system was developed using profile matching with the help of web crawling for TPO (the person who manually match profiles with jobs) [7]. In this recommender system, better result or recommendation is produced by providing two types of matching, semantic matching and tree based knowledge matching. Key word based searching is done by using web crawling. Semantic matching is implement on the attributes like technical skills, projects and tree based matching is performed on the attributes like qualification etc. As matching is complete, a preference list is generated as result of matching. In [8] paper, they have proposed another method that is reciprocal job recommender system (a type of recommender systems which has not gained much attention) in mobile environment for online recruiting. The work is done by extracting the elementary features of users, and then calculated the similarities among users after preprocessing is done. For categorization of features into self-description and preference, some strategies of feature preprocessing are supposed to apply. A Chinese word segmentation LuceneI KAnalyzer was adopted for the purpose of feature extraction and transformation of the features into term vectors was done with the help of TF-IDF.

A baseline term base recommender system was proposed in which recommendations are generated by the concept of overlapping between all the terms in applicants' position history and advertisements of jobs [9]. Cosine similarity is used for similarity measures and score range is from 0 to 1. As the value get high, means large number of common terms are present between the applicants; position history. Recommendations will only be generated if common terms exit in history. This recommender system includes some independent modules like Title Transition module which works on the evaluation of the similarity of advertised job title and applicants' job title. Other modules are Title Description Matching Module and Description-Description Matching module.

A unique method was proposed based on finding similarity between job offers and user profiles in social network environment [10]. They considered three categories of documents which contain numerous fields with information in textual form. Three categories are job offers, job categories and user profiles on social network. They applied field to field matching approach means the match between the user profile and offered job. Moreover they have also applied similarity functions for the purpose of cross matching.

A forecasting model in context of career recommendations is proposed in [20]. This paper described the problem of receiving too much or few number of job applications for a specific job post. The proposed model solved this problem by providing predictions of quantity of applications till the date of job adds expiration. They have also presented the design architecture for prediction of job applications received on LinkedIn. The system is then evaluated by some experiments and results show the significance of the proposed model.

A comprehensive student centric career recommendation system is proposed in [21] which is purely aimed to help out students to choose right career. The proposed career recommendation system employs several dimensions from three-dimensional model which is. The three-dimensional model is fuzzy logic, influence and preference. These measurements are then integrated with some weights which are obtained by a decision system named as Analysis Hierarchical Process. The system is then evaluated by computing desired score of student regarding engineering stream and the results show that the propose system suggests more suitable career path as compared to other approaches.

A recommender system is proposed named as occupation recommendation (OCCREC) in [22] which aimed to determine student's career as early as possible so that they have some guidance to improve their skills regarding selected occupation. The proposed recommendation system is actually a hybrid system which integrated the approaches of content based filtering and collaborative filtering. The information which included in this system is student's profile, student's interest and behavior of student. And the data of student's profile is obtained from Facebook. The recommendations are generated on the basis of five similarity measures which include Cosine, Pearson, Euclidean, Jaccard and Intersection. The system is then evaluated by some experimentation and show significant results.

3. Data collection and visualization. Data plays a vital role in the field of research as we need it for the analysis and results. Data mining purely consists on data, manipulation of data and finding trends, hidden patterns and knowledge from data by applying different data mining techniques. In this research, we have applied some data mining techniques on collected data and tried to find some patterns which can help us for Profession Recommendations and Skills Recommendations.

A. Type of Data Required. This research is based on profession and skills recommendations. For the analysis purpose, it required educational background, working experience, skills and professions of different people for the purpose to apply some data mining techniques and find patterns to make recommendations for jobseekers. After spending much time in brainstorming and discussions about, we came to a point that instead of asking one by one to everyone to tell your educational background, working experience, skills and profession, it would be better to circulate a survey questionnaire to get the data.

B. Data Collection Approach. Google form is an easiest and simple way of online data collection from multiple resources with in minimum time frame. A survey form was built which contained some questions/attributes with their possible answers.

Table 1. Attribute and their Possible Values

Attribute	No. of Possible Values
Gender	2
Under-Graduate Stream	12
Under-Graduate Institute	26
Master's Stream	12
Master's Institute	26
Graduate Stream	12
Graduate Institute	26
Doctorate Stream	12
Doctorate Institute	26
Working Experience	6
Profession	30
Skills	45

Table 1 is showing all the attributes which were asked by the people and number of possible values. All the attributes in our dataset are of nominal/categorical type attributes.

C. Data Collection Sources. All the sources from where we collected the data for this research are shown in figure 2. We have collected the data from our friends, colleagues, and social media and from alumni UMT by sharing the link of Google survey form.

Figure-2. Data Collection Sources

D. Data Pre-Processing. The data which we collected for the research was about 300 records of working people. In order to remove inappropriate records, we have done data cleansing by removing the records which are not correct or irrelevant. The responses which we got from different sources were preprocessed by replacing all the options of “Other” selected by people with “NA” as there is no information gained by the option “Other”. We removed all responses in which educational background, profession and skills were not in accordance with each other. After the task of data pre-processing, we were left with 250 numbers of records.

4. **Proposed methodology.** This research is concerned in the field of *Data Mining*, so a well-known data mining technique named *Association Rule Mining* is used for making recommendations.

A. Association Rule Mining (ARM). In Association Rule Mining, we find the interesting patterns in our data which are most frequently occurred. An Association Rule basically consists of two parts which are called *Antecedent* and *Consequent* [19]. The items which exist in our database or data collection are called antecedent. And the items which are found in the combinations of antecedent are called consequent. For example, {Tea} => {Biscuits} is an AR, tea is antecedent in this rule and biscuits are consequent. A support threshold is set before finding association rules. All the item sets which have support equivalent to the threshold or greater than threshold are called frequent item sets.

In order to measure the quality, interestingness and accurateness of an association rules, some metrics like support, confidence and lift are used in data mining. For example if is an association rule: Support is actually the probability of how many times an item is occurred in total number of transactions. Confidence is the conditional probability between association rules’ items on both sides. Lift is the proportion of observed collective support against individual support of items which infer the independent nature of items.

B. Recommendation Process Step by Step. The most common tools for ARM we found were *WEKA* and *R Studio*. These tools are mostly used for performing tasks related to Data Mining. From these two tools, the one which we selected was *R Studio*.

The following seven steps which are shown in the figure 3 were followed for generation of Association Rules in R language.

- In very **1st step**, the data file was converted into .csv format so that it can be loaded into R Studio for further process.
- In **2nd step**, the data file which was converted into .csv format, loaded into R Studio.
- In **3rd step**, association rules mining package named *arules* was downloaded and installed in R Studio. Without this package, association rules could not be generated.
- In **4th step**, we applied *Apriori* algorithm (which is used in data mining for generating association rules) on the dataset which was already loaded.
- In **5th step**, we got the output of implementation of Apriori algorithm which was in the form of association rules.

- In 6th step, we sorted the rules on the basis of lift and filtered them.
- In last step when all the rules were sorted and filtered, we saved them in different text files according to the length of rule.

B. Profession Recommendations. In order to make profession recommendations, first of all we set the values of support and confidence. As it is described earlier that the value of confidence is fixed that is 0.7. As we have twelve attributes in our dataset so we made the rules of all lengths till twelve attributes occur. The different values of support threshold and number of rules generated are given in below table.

Figure-3. A step by step Association Rules generation process

Table 2. No.of Rules Generated on Support Threshold

Length of Rule	Support Threshold	No. of Rules Generated
2	0.004	167
3	0.008	182
4	0.008	278
5	0.008	177
6	0.008	64
7	0.008	16
8	0.008	2
9	0.004	768
10	0.004	232
11	0.004	44
12	0.004	4

Here are some sample recommendations which are generated in the result of profession recommendations. For example length 4 rule is recommending the profession of quality assurance to someone who is a female having working experience of 3 to 4 years and expert in the skills of document test cases, understanding of QA tools and techniques and technical expertise.

Table 3. Sample Profession Recommendations

Rule's length: 2 Rule: Skills=Managerial skills; Case Solving; Communication Skills; Presentation skills; Conflict Resolution => Profession= <i>Legal Advisor</i> Support: 0.004, Confidence: 1, Lift: 244
Rule's length: 4 Rule: Gender=Female, Working Experience=3 to 4,Skills=Document Test Cases; Understanding of QA Tools and Techniques; Technical Experties => Profession= <i>Quality Assurance</i> Support: 0.008, Confidence: 1, Lift: 48.8
Rule's length: 6 Rule: Gender=Male, UnderGraduateStream=Law, UnderGraduateInstitute=University of Punjab, GraduateStream=Law, GraduateInstitute=University of Lahore => { Profession= <i>Lawyer</i> Support: 0.008, Confidence: 1, Lift: 30.5
Rule's length: 8 Rule: Gender=Male,UnderGraduateStream=Applied Sciences, MastersStream=Applied Sciences,GraduateStream=Applied Sciences, Doctorate=Applied Sciences and Humanities,WorkingExperience => 5,Skills=Managerial skills;Presentation Skills;Communication Skills} => Profession= <i>Teaching</i> Support: 0.008, Confidence: 1, Lift: 6.256

C. Skills Recommendations. For the purpose of making skills recommendations, same technique and procedure is followed as for profession recommendations. Apriori algorithm is used for making association rules and recommendations. The algorithm is applied on 250 numbers of records collected from multiple sources. The support threshold value for skills recommendations is taken 0.004 and the confidence value is taken 0.4. Numbers of association rules or recommendations which are generated on given threshold are 62 Rules.

Here are some sample skills recommendations which are taken from the result file. Recommended skills for the profession of teaching are communication skills and presentation skills. Similarly for the profession of quality assurance, recommended skills are technical expertise and understanding of QA tools and techniques.

Table 4. Sample Skills Recommendations

Rule: Profession=Teaching => Skills= <i>Presentation skills</i> Support: 0.012, Confidence: 0.206, Lift: 5.299
Rule: Profession=Teaching => Skills= <i>Communication Skills</i> Support: 0.012, Confidence: 0.206, Lift: 2.902
Rule: Profession=Doctor => Skills= <i>Patients Counseling</i> Support: 0.02, Confidence: 0.273, Lift: 13.455
Rule: Profession=Finance Officer => Skills= <i>SPSS Tool</i> Support: 0.005, Confidence: 0.2, Lift: 4.736
Rule: Profession=Database => Skills= <i>Familiarity with database technologies</i> Support: 0.005, Confidence: 0.214, Lift: 31.714

These are some sample recommendations shown in above table along with support, confidence and lift metrics. The question which comes in mind after reviewing these recommendations is why the values of these threshold metrics are very low. The reason behind this is small amount of data which we have used for providing recommendations. If we had a huge amount of data, the threshold values will automatically be increased.

5.Limitations. The outcomes provided by the model which we have proposed for professions and skills recommendations have some limitations. As the recommendations are made on the basis of the data which we have collected from different sources, so the results which we got may be biased against some attribute. As we do not have huge amount of data of all population in this research so it may violate accuracy of rules but the working of proposed model is absolutely correct. The biasness factor will be reduced when we get more data.

Conclusion. In this research, problem of Pakistani people regarding selection of professions and skills improvement is resolved by proposing a model so that it can be easy for a jobseeker to apply for specific jobs. The model basically generates recommendations and suggestions to jobseekers on the basis of educational backgrounds, working experience, profession and skills of those people who have joined some professions. A well-known technique of Data Mining that is Association Rules Mining with Apriori algorithm is used to build the model.

Future work. In future, we have decided to extend this research by providing Job Recommendations. Job openings along with number of vacancies, job description and job requirements will be collected from different job portals, company sites and job ads. The person who is in search of job will provide the details about educational background, experience, profession in which he/she is searching job and technical skills, the system will match the input data with the collection of job openings. Closely matched results will be extracted and the jobs will be recommended to the jobseeker.

REFERENCES

- [1] Malinowski, J., Keim, T., Wendt, O., & Weitzel, T. (2006, January). Matching people and jobs: A bilateral recommendation approach. In *null* (p. 137c). IEEE.
- [2] Lu, Y., El Helou, S., & Gillet, D. (2013, May). A recommender system for job seeking and recruiting website. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 963-966). ACM.
- [3] Kumar, A., & Thambidurai, P. (2010). Collaborative web recommendation systems based on an effective fuzzy association rule mining algorithm (FARM). *Indian Journal of Computer Science and Engineering, Engg Journals Publications*, 1(3), 184-191.
- [4] Wei, K., Huang, J., & Fu, S. (2007, June). A survey of e-commerce recommender systems. In *Service systems and service management, 2007 international conference on* (pp. 1-5). IEEE.
- [5] Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009.
- [6] Hong, W., Zheng, S., Wang, H., & Shi, J. (2013). A Job Recommender System Based on User Clustering. *JCP*, 8(8), 1960-1967.
- [7] Musale, D. V., Nagpure, M. K., Patil, K. S., & Sayyed, R. F. (2016). Job Recommendation System Using Profile Matching And Web-Crawling. *INTERNATIONAL JOURNAL*, 1(2).
- [8] Wenxing, H., Yiwei, C., Jianwei, Q., & Yin, H. (2015, July). iHR+: A mobile reciprocal job recommender system. In *Computer Science & Education (ICCSE), 2015 10th International Conference on* (pp. 492-495). IEEE.
- [9] Heap, B., Krzywicki, A., Wobcke, W., Bain, M., & Compton, P. (2014, December). Combining career progression and profile matching in a job recommender system. In *Pacific Rim International Conference on Artificial Intelligence* (pp. 396-408). Springer, Cham.
- [10] Malherbe, E., Diaby, M., Cataldi, M., Viennet, E., & Aufaure, M. A. (2014, August). Field selection for job categorization and recommendation to social network users. In *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 588-595). IEEE Press.
- [11] Amatriain, X., Jaimes, A., Oliver, N., & Pujol, J. M. (2011). Data mining methods for recommender systems. In *Recommender systems handbook* (pp. 39-71). Springer, Boston, MA.
- [12] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering*, (6), 734-749.
- [13] Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325-341). Springer, Berlin, Heidelberg.
- [14] Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325-341). Springer, Berlin, Heidelberg.
- [15] Cai, Y., Leung, H. F., Li, Q., Min, H., Tang, J., & Li, J. (2014). Typicality-based collaborative filtering recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 26(3), 766-779.
- [16] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4), 331-370.

- [17] Trewin, S. (2000). Knowledge-based recommender systems. *Encyclopedia of library and information science*, 69(Supplement 32), 180.
- [18] Schafer, J. B. (2005). Application of Data Mining to Recommender Systems. In *Encyclopedia of data warehousing and mining* (pp. 44-48). IGI Global.
- [19] Kaur, J., & Madan, N. (2015). Association rule mining: A survey. *International Journal of Hybrid Information Technology*, 8(7), 239-242.
- [20] Borisyuk, F., Zhang, L., & Kenthapadi, K. (2017, August). LiJAR: A system for job application redistribution towards efficient career marketplace. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1397-1406). ACM.
- [21] Verma, P., Sood, S. K., & Kalra, S. (2017). Student career path recommendation in engineering stream based on three dimensional model. *Computer Applications in Engineering Education*, 25(4), 578-593.
- [22] Ochirbat, A., Shih, T. K., Chootong, C., Sommoool, W., Gunarathne, W. K. T. M., Wang, H. H., & Ma, Z. H. (2018). Hybrid occupation recommendation for adolescents on interest, profile, and behavior. *Telematics and Informatics*, 35(3), 534-550.