

Feature Selection for Job Matching Application using Profile Matching Model

Leah G. Rodriguez
Graduate Programs
Technological Institute of the Philippines
Manila, Philippines
e-mail: mam_leah@yahoo.com.ph

Enrico P. Chavez
Graduate Programs
Technological Institute of the Philippines
Manila, Philippines
e-mail: vez3@yahoo.com

Abstract—The advancement of technology has created big changes to speed up the job hiring process nowadays. With more tools developed, companies have embraced tools that help them recruit talents. This is a research-in-progress of developing a job matching application system for Job Recruitment Agency specifically in the province of Pangasinan, Philippines. This paper proposes an approach for the agency to extract the relevant information from resumes and analyze it based on the different attributes. With the identification of the attributes, the proposed system is directed to adopt a clustering algorithm to match the profile of the job seekers against the requirements of the job posted by the prospect employers. This helps the agency to find suitable candidates for a particular job in different companies and make more informed decisions.

Keywords—data mining; job matching application; profile matching

I. INTRODUCTION

Every communication over internet happens through the exchange of some data. Due to the increased use of the internet, data nowadays is growing at a huge velocity which also makes data handling a concern. One of the main difficulties in handling such a large amount of data is the increase in volume in comparison with the advancement in computing resources.

People in many countries consider the internet an effective communication medium and rate the internet as an important recruitment channel. Examples of internet services provided are vacancy announcements, job postings, interviewing, filling online application forms and communicating with recruitment agencies [1].

Job Matching is an approach of controlling exactly how a job register and a job candidate are correctly paired altogether [2]. The job matching procedure, in general, examines the data presented in the resume and match against the data registered in the list of open vacancies. One of the most challenging tasks of this type of job matching is that there is typically a bulk of information to coordinate against. Furthermore, these data usually submitted in a free form, as each individual has their own preference to prepare the data. However, a match among job seekers and employment needs to be combined as it needs taking into consideration the preferences of the company and the preferences of the candidate. Moreover, great strength is contributed in evaluating employment requirement and source, and procedures are taken at many levels to match one with the

other. However, to be able to identify the attributes (features) that will be extracted from the job seekers resume and job requirements, a feature selection process must first be made. This means that a collection of features are tested for significance during model training. Thus, this paper aims to conduct a feature selection to select the significant variable needed for job matching of Job Recruitment Agency through actual data. This leads to developing the match involving persons and jobs using the job seekers profile and company profile parameters.

II. RELATED WORKS

According to Hiers (2014) [3], job matching is a specific method of defining how a job register and job seekers are suitably paired altogether. In job matching, matches are defined based on real qualifications and resumes, which yield much more accurate matches than keyword searches.

Greenberg (2010) [4], defines job matching being the method of matching the exact person to the exact job based upon the individual's essential motivational strengths. It involves a complete understanding of the job and the person under consideration.

A. Understanding Feature Selection

Feature selection is a process that selects a subset of original features. Feature selection involves three stages. Firstly, screening. This step will remove insignificant inputs and records, or cases such as input fields with too many missing values or with too much or too little variation to be useful. The second step is ranking. This will sort remaining inputs and allocate ranks established importance. And the third step is selecting. This step identifies the subgroup of features to employing subsequent models—for instance, by keeping merely the best significant inputs and filtering or excluding all others [5].

Feature selection is one of the important and frequently used techniques in data preprocessing for data mining. It reduces the number of features, removes irrelevant, redundant, or noisy data, and brings the immediate effects for applications: speeding up a data mining algorithm, improving mining performance such as predictive accuracy and result comprehensibility [6]. Moreover, feature selection can help focus on relevant parts of data and improve our ability to process data. As feature selection is one step in data preprocessing, changes need to be made for classic

algorithms that require multiple database scans and/or random access to data [6].

B. Cosine Similarity

The similarity measure is the measure of how much alike two data objects are. Similarity measure in a data mining context is a distance with dimensions representing features of the objects. If this distance is small, it will be the high degree of similarity where large distance will be the low degree of similarity. The similarity is subjective and is highly dependent on the domain and application [7].

Computing similarity scores between two profiles. For the similarity score, the values of common attributes in both profiles are extracted and their similarity scores are computed [8]. The new similarity score is computed as follows:

$$\text{sim}^{\wedge}(J1.a_i, J2.a_i) = \frac{2 \times \text{sim}(J1.a_i, J2.a_i) \times w(a_i)}{1 + (\text{sim}(J1.a_i, J2.a_i) \times w(a_i))} \in [0, 1]$$

where:

- a_i an attribute used to describe a profile,
- $J1.a_i$ and $J2.a_i$ are two values of an attribute a_i in Profile P1 and Profile P2,
- $w(a_i)$ the computed/assigned weight of an attribute $\in [0, 1]$,
- $\text{sim}(J1.a_i, J2.a_i)$ the similarity score computed between the values of an attribute in $J1$ and $J2 \in [0, 1]$,
- $\text{sim}^{\wedge}(J1.a_i, J2.a_i)$ the new similarity score computed between the values of an attribute in $J1$ and $J2 \in [0, 1]$

C. Profile Matching

Profile matching alludes to choice dependent on candidate likeness to a pre-indicated example of remaining over a few commonly considered identity measurements. In Profile Matching process, it can be a procedure of looking at the person's capability in occupation competency. In this way, it tends to be the distinction capability (it is also called as gap), the smaller the gap generated the weight of large value, it implies that it have more noteworthy open doors for worker to take that position. In other words, profile matching is a system to settle on a choice that accept a dimension of variable forecast which ought to be completed by employees [9].

III. METHODOLOGY

A. Data Set

The researcher used the collection of job seekers' records from a Job Recruitment Agency for the period of one year. Specifically, the initial size of the data set is 2, 283 job seekers' records.

TABLE I. JOB SEEKER'S PROFILE SON VARIOUS ATTRIBUTES, ITS DESCRIPTION, AND POSSIBLE VALUES

Attribute	Description	Possible values
Age	Describes how old or young a particular job seekers	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,

		51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61
Sex	Gender of the job seekers	Male, Female
Civil Status	Refers if the Job seekers are single, married and the like.	Single, Married, Separated, Widowed
Educational Attainment	This refers to the highest level of education completed by the job seekers	e.g. Bachelor of Science In Business Administration, Bachelor Of Science In Information Technology, Bachelor of Science in Computer Engineering, Bachelor of Science in Computer Science, Bachelor of Science in Nursing, Bachelor of Science in Hotel and Restaurant Management, Bachelor of Science in Electronics and Communications Engineering, Bachelor of Science in Agriculture, Bachelor of Elementary Education, Bachelor of Science in Pharmacy, Others
Work experience	Denotes experience that a job seeker gains while working in a specific field or occupation	Relevant, Irrelevant
No of years Work experience	Number of years that job seekers work in a specific field	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

As demonstrated in Table 1, it described the attributes of the data and their possible values. The different attributes identified average, sex, Civil Status, Educational Attainment, Skills, Work experience, years of experience. These are corresponded by different descriptions and possible values for each entity.

TABLE II. COMPANY PROFILES ON VARIOUS ATTRIBUTES, ITS DESCRIPTION AND POSSIBLE VALUES

Attribute	Description	Possible values
Job Title	Describes the position of an applicant	e.g. Domestic Helper, Accounting Staff, Nurse, Chef, Restaurant Supervisor, Chef, House Nurse, House Driver
Educational Attainment	This refers to the highest level of education	e.g. Bachelor of Science In Business Administration, Bachelor Of Science In Information Technology, Bachelor of Science in Computer Engineering, Bachelor of Science in Computer Science, Bachelor of Science in Nursing, Bachelor of Science in Hotel and Restaurant Management, Bachelor of Science in Electronics and Communications Engineering, Bachelor of Science in Agriculture, Bachelor of Elementary Education, Bachelor of Science in Pharmacy, Others
Age	States how old or young, essential	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32,

	for the job	33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61
Sex	Gender of an applicant needed in a company	Male, Female
Civil Status	Signifies the state of being single, married and the like, required for the job	Single, Married, Widowed, Separated
Work Experience	Refers to the experience required for the job	Relevant, Irrelevant
No. of years work experience	Implies the number of years of experience necessary for the job	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

As shown in Table 2, it illustrated the attributes of the data and their possible values. The different attributes identified are Job Title, Age, Sex, Civil Status, Educational Attainment, Skills, Work experience, and years of experience. These are corresponded by different descriptions of possible values for each entity

B. Feature Selection

Feature selection is a method that chooses a subgroup of unique features. The best feature subgroup is calculated by an evaluation standard [10]. All profiles have been obtained from recruitment agency.

The features that were obtained significant in a job matching it to the two most important groups: Job Seekers and Company. In job seekers, the features that were present are age, sex, civil status, educational attainment, work experience, and no. of years of experience, then the features relating to the company requirements are a job title, sex, age, civil status, work experience, and no. of years of experience.

C. Data Categorization

As per the goal was to find out the criteria on which the job seekers, belonging to different age group, gender, education level, marital status, the focus for choosing the unfilled jobs, the whole categorization or generalization was done [11]. The job seekers information alongside company information together is grouped into different clusters intended for finding out the job seeker performance suitable to a particular group for choosing exact job upon the basis of six parameters as follows: Job title, Work Experience, Educational Attainment, Civil Status, Age and Sex-based on which matching methods are utilized.

D. Profile Matching

To limit the scope of profile matching and give more exact matching outcomes, first clean the incorrect profiles corresponding to the significance of features (age, civil status, sex, educational attainment), and afterward match up the profile fitting to the representation of other features [12]. Since the distinct description of features must be matched up

in various ways, in this research only one kind of matching method which is proposed for clustering. Job Title/Position, age, gender, and educational attainment are the four essential features that companies matter, yet, those features are customarily accorded physically.

In this study, the ideas of those attributes offer complete suitability to employ models similarity to match it.

E. Proposed Profile Matching Model

Figure 1 shows the proposed framework for profile matching. The input of the study includes the Job seekers data from NSRP Form1 of PESO Employment Information System Registration Form such as age, sex, Civil Status, Educational Attainment, Skills, Work experience, years of experience; and Company/job Data such as Job requirements.

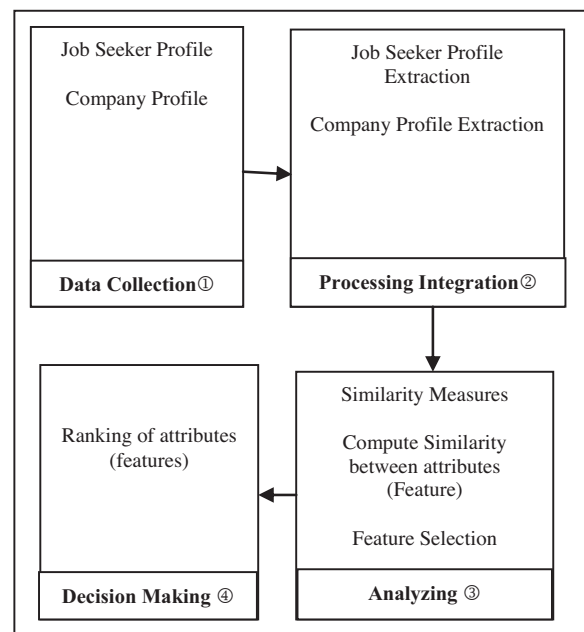


Figure 1. Proposed Profile Matching Framework.

In Data Collection, the data from the job seeker and company requirements were collected. The gathered data requires to be preprocessed. Preprocessing include changing information to plain text, clean up the impurities, for example, blank areas and extra boisterous characters; eliminating HTML labels, accentuation marks; changing the data training to widespread configuration to make it prepared used for the preparation of the framework, and incorporating the information into a unified format [13].

On the other hand, Processing and Integration involve information extraction and data categorization wherein all data collected were cleaned and integrated to develop a profile matching model.

For analyzing, the proposed framework will compute similarity scores between two profiles. The values of common attributes in both profiles are extracted and their similarity scores are computed. Then, the obtained similarity scores are tuned in order to have more realistic scores that take into consideration the importance assigned to each

attribute. By doing so, the new similarity value will tend to increase or decrease depending on the importance of each attribute. This tuning is an attribute based operation that outputs a new similarity score to each attribute by applying a weight to the computed similarity scores. [14].

In decision making, it returns whether the two compared profiles are the same or not. This decision, is computed using the weighted similarity scores. Matching is an essential stage in the recruitment procedure [15]. The matching algorithm will be able to select relevant clusters and only match against all vacancies contained within these clusters. As soon as the matching is completed, the recommendation of jobs to the job seeker is executed. Lastly, with the job match output at hand, the system recommends the job to the job seeker and job seeker to the company.

IV. RESULTS AND DISCUSSION

The first step to determine the best fit between candidates and job description is building the job profile and the prospective company profiles. We extract some features from job seekers resumes and company requirements to build both profiles.

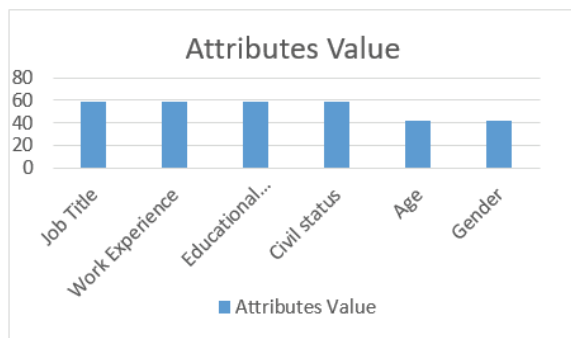


Figure 2. Ranking of identified attributes for profile matching.

Figure 2 shows the ranking of identified attributes for profile matching through the use of Weka. It was found out that the attributes Job title, work experience, Educational Attainment and Civil Status has the same rank with an attribute value of 58.333 while Age and Gender attribute ranks 5 and 6 with the attribute value of 41.66.

V. CONCLUSION

This presents a paper in progress work aims to identify the main attributes needed for the profile matching model in the proposed job matching system. This attributes are extracted from the 2,283 job seekers resume and job requirements information provided by Public Employment Service Office (PESO) of Pangasina, Philippines. The data is treated using data mining in an analytic software known as WEKA and is statistically presented using ranking.

With the identification of the attributes, the proposed system is directed to adopt a clustering algorithm to match the profile of the job seekers against the requirements of the job posted by the prospect employers.

To progress, the study aims to (a) add more training data set from job seekers and company, (b) conducting tests of the clustering model to verify reliability and performance of the job matching system, and (c) perform an end-user software evaluation.

ACKNOWLEDGMENT

The researchers would like to extend their gratitude to the Public Employment Service Office for approving the conduct of this research and to PESO Manager Jocelyn Z. Siapno for sharing vital information and assistance in the conduct of this study.

REFERENCES

- [1] McGovern, R. J., Winchester Jr, J. A., Evans, A. B., Farmer, B. E., Koffman, J. A., & Walker, A. P. (2002). *U.S. Patent No. 6,370,510*. Washington, DC: U.S. Patent and Trademark Office.
- [2] Pellizzari, M. (2010). Do friends and relatives really help in getting a good job?. *ILR Review*, 63(3), 494-510.
- [3] Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4), 491-502.
- [4] HIERS, M. 2014. What 'Job Matching' Really Means. Retrieved from <https://www.pandologic.com/job-boards/job-matching-really-means/>
- [5] Nasiriyah, R., Arham, Z., & Aini, Q. (2016). Profile matching and competency-based human resources management approach for employee placement decision support system (case study). *Asian J. Appl. Sci*, 9(2), 75-86.
- [6] Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4), 491-502.
- [7] Belsare, R. G., PRMIT&R, B., & Deshmukh, V. M. Employment Recommendation System using Matching, Collaborative Filtering and Content Based Recommendation.
- [8] Raad, E., Chbeir, R., & Dipanda, A. (2010, September). User profile matching in social networks. In *Network-Based Information Systems (NBIS), 2010 13th International Conference on* (pp. 297-304). IEEE.
- [9] Susilowati, T., Anggraeni, E. Y., Andewi, W., Handayani, Y., & Maselena, A. (2018). Using Profile Matching Method to Employee Position Movement. *Int. J. Pure Appl. Math*, 118(7), 415-423.
- [10] Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4), 491-502
- [11] Gupta, A., & Garg, D. (2014, September). Applying data mining techniques in job recommender system for considering candidate job preferences. In *Advances in Computing, Communications and Informatics (ICACCI), 2014 International Conference on* (pp. 1458-1465). IEEE.
- [12] Raad, E., Chbeir, R., & Dipanda, A. (2010, September). User profile matching in social networks. In *Network-Based Information Systems (NBIS), 2010 13th International Conference on* (pp. 297-304). IEEE.
- [13] Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *Acm Sigkdd Explorations Newsletter*, 1(2), 12-23.
- [14] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), 651-666.
- [15] Calvó-Armengol, A., & Zenou, Y. (2005). Job matching, social network and word-of-mouth communication. *Journal of urban economics*, 57(3), 500-522.