

A Vectorization Model for Job Matching Application of a Government Employment Service Office

Leah G. Rodriguez¹⁺, Enrico P. Chavez², Christopher A. Rodriguez³

¹Pangasinan State University

²Technological Institute of the Philippine – Manila

³Pangasinan State University

Abstract. The fast growth of the Internet caused a matching growth of the amount of available online information that increased the need to expand the ability of users to manage all this information. This encourages a substantial interest in specific research fields and technologies that could benefit the managing of this information overload. However, in the Philippines, it has been a challenge for most job agencies to find out and predict job matching intelligently due to lack of accurate models to be adopted. To address this problem, the researchers developed job matching application specifically for Pangasinan Employment Service Office (PESO). The aim is to know how these measures behave and whether they validate the idea that applicants' data have more in common with job profile. The process of matching an applicant with an offered job is performed through the use of vectorization model and cosine similarity. Large training sets of records of applicants and job profiles were used to define the weights of the parameters. The results show that the selection of cosine similarity measures and vector weights are significant in job matching application, especially in those where the applicants' information's likeness is measured.

Keywords: Vector space model, job matching application, cosine similarity, resumes

1. Introduction

Finding the right person for the right job takes never been an easy feat for companies, whose value is very often to a large degree derived from their manpower. With the increased mobility of job seekers in recent years, more and more jobs are seeing rapidly growing pools of potential candidates, requiring respective recruiters to wade through hundreds if not thousands of resume to find the perfect match [1]. In fact, the challenge of matching suitable candidates with job openings available at specified time is pervasive. Mainly, in times of economic stress where very large numbers of candidates may be seeking a small number of openings, the review process can tax even the most experienced human reviewer. On the other hand, an applicant may discover it very hard to find a genuinely appropriate position for themselves from among a huge number that are being promoted [2].

The job matching procedure typically yields over the information accessible in the resume and match against the information recorded in the list of open opportunities. A standout amongst the most difficult assignments of this sort of employment coordinating is that there are typically such a large number of information to coordinate against. Moreover, these information more often than not submitted in free structure, as every individual has their own inclination to set up the information. [3]. In actual scenario, an ordinary employment coordinating procedure more often than not begins with errand to recover information from both resume and opportunity. The information will be put away in an outside capacity in various organization. The job matching assignment is typically determined by predefined coordinating principles on removed information. These principles caught the vital criteria of resume and opportunity that is

+ Corresponding author. Tel.: +639151190208
E-mail address: mam_leah@yahoo.com.ph

characterized by a gathering of employment showcase area specialists. This great practice is to stay away from irreconcilable circumstances just as predisposition during job matching decision making by one person. Besides, the criteria got from the job market domain experts such as knowledge about the job and basic candidate’s requirement enhance the precision level for job matching. A weighting factor will be assigned to all the corresponding criteria in the rules for simplicity of applicants filtering and ranking. The weight decides the dimension of pertinence between the resume and employment opening or the other way around [4].

In view of this, this research highlights the incorporation of vector space model specifically in calculating cosine similarity to rank and match job applicants. Specifically, the paper is expected to give suggestions in 2 different ways: 1) recommending most qualified candidates for the predetermined job, to the recruiters and 2) endorsing occupations to the seeking applicants as per their corresponding profiles.

2. Background of the Study

2.1. Vectorization concept

Vectors have direction and magnitude. Because of this, we can calculate the angle between two vectors. A popular measure in data science is the cosine of this angle computed as follows:

$$\cos(x,y) = \text{dot}(x,y)/|x||y|$$

This measure will equal 1 when the vectors are parallel (they point in the same direction) and 0 when the vectors are orthogonal. Vectors that point in the same direction are more similar than vectors that are orthogonal [5].

2.2. Data collection

The data utilized in this article originate from job recruitment agency with a collection of resumes and company vacancies. There is no basic format for resume as it is a free format record. The authors utilized the agency’s Standard format in this article. Standard Format is the perfect layout that comprises of the considerable number of parameters that is fundamental in a resume. The more information the layout contains, the more matching and the better outcome can be expected. Information mined from candidate’s resume is automatically fill-in standard layout, so as to find missing parameters [6].

The variables that were obtained important in a job scenario be appropriate to two (2) major classifications: Applicant and Occupation. For applicant the variables that were taken into account for judging him/her are: Education, Age, Gender, Marital Status, and Experience. And the variables relating to the occupation are: Job Title, Required Qualification and Experience.

Table 1 exhibits the total number of files, number of vacancies, and number of applications in the corpus.

Table 1: Number of job offers, job applications and files in the corpus.

Vacancies	Applications	Total Number of Files
1367	2, 283	3000

2.3. Parameters used in data collection

The standard parameters are data that hold distinctive values of the job seeker. It is the main criteria used to determine the level of suitability of a job seeker to the vacancies. This was determined by the phase one of the project through the research entitled, “Feature Selection for Job Matching Application using Profile Matching Model”. Below is the list of standard features defined by the system as a result of the said research:

Job title. States to the name given to a position held by an employee in an organization. It is a short depiction of the position held by the worker and gives a thought regarding the position in the organizational hierarchy. Typically, it helps employers classify positions in their organization and help job seekers get a proper line of work dependent on their experience and requirement [7].

Educational qualification. Talk about the official confirmation, more often in the form of a certificate or degree, ensuring the effective fulfilment of an education program or a phase of a program. Effective completion of an education program refers to the accomplishment of determined learning goals, regularly

approved through the appraisal of acquired knowledge, skills and competencies. The term 'qualification' is synonymous with 'credential' [8].

Work Experience. Alludes the experience that an individual by now has of working. It is the jobs that somebody has had, or the kind of work they have done, previously. This bit of information permits the employer to know the candidates better in the event that they are suitable for the opening [9].

Age. The period of time someone has been alive or something has existed

Gender. It is the state of being either male or female

Civil Status. It is the fact of someone being married or not.

2.4. Cosine similarity

Similarity in a data mining context is typically portrayed as a distance with measurements denoting highlights of the objects. If this distance is little, there will be high level of similarity; if a distance is big, there will be low level of similarity. Similarity is particular and is very much reliant on the field and application [10].

Cosine similarity metric finds the standardized dot product of the two features. By clarifying the cosine similarity, we will successfully endeavouring to discover cosine of the angle between the two objects. Cosine similarity is mainly utilized in positive space, where the result is precisely limited in (0, 1). One justification designed for the popularity of cosine similarity is that it is extremely well-organized to assess, particularly for sparse vectors [11].

3. Related Works

On job matching, a variety of matching rules use the analysis results to automatically construct and adapt an underlying matching process for a given match task. However, one of the most difficult tasks a business owner faces today is finding the right person for the job.

3.1. Proposed System Model Architecture

The key purpose of the proposed model is to discover an applicant who matches to the requirements of occupation profile and recommend jobs to the aspiring applicants as per to their matching profiles.

Figure 1 shows the proposed system model for Job Matching Application of a Government Service Office. It is illustrated as follow:

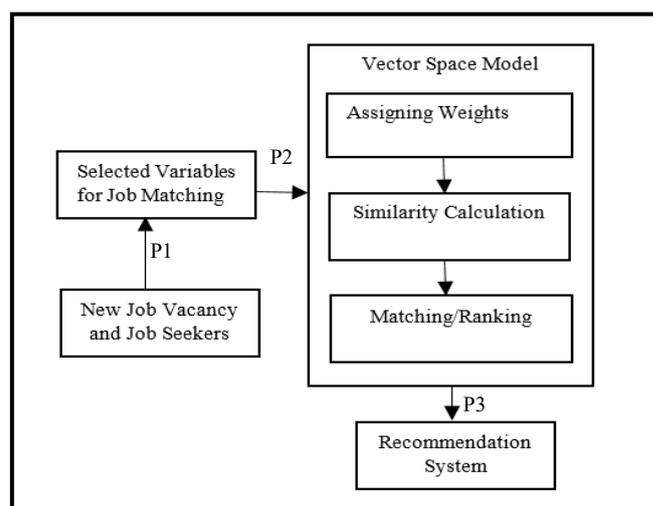


Fig1: Proposed System Model for Job Matching Application of Government Service Office.

In the system architecture, a number of processes have been identified to perform the job matching analysis and recommendation.

Process 1 (P1) indicates the extracted features (variables) from applicants' resumes and job description to build Job profiles and Job applicants profiles. In job case, the applicant's profile consists of some features

that required for a specific job. Thus, the job's profile comprises of the job requirements that ought to be possess by applicants.

Process 2 (P2) specifies the vectorization process that includes assigning weight, similarity calculation, and matching/ranking, where the job profile along with the applicants' profile were denoted as vectors (Assigning weight). The model parameters by creating a rating matrix $R_{x,y}$, where x denotes the job and y denotes the Job seekers resumes [12] explain as follows:

$$R_{x,y} \begin{cases} 1 \text{ (TRUE = "Exist")} \text{ „If the target attribute is existed} \\ 0 \text{ (FALSE = "not Exist")} \text{ „If the target attribute isn't existed} \end{cases}$$

The ranking matrix $R_{x,y}$ converted by considering the values of job seekers' features as ratings of all the features mined from the resumes utilizing any similarity measures.

Cosine Similarity measures is computed as follows:

$$\text{sim}(x, y) = \frac{\text{dot}(x, y)}{\sqrt{|x| \cdot |y|}}$$

Moreover, in matching process, the result of the matching instrument is a level of entities by their match rates. The instrument likewise gives a complete details of the matching process (i.e. similarity values for all clusters and for each single property within a cluster together with the information about matchers used, property values, and weights) which can be used to generate explanations for the calculated object similarity.

Process 3 (P3) recommends the most qualified applicants to the recruiters and/or endorse occupations to the seeking applicants as per their matching profiles.

3.2. Vectorization Results and Discussion

We conduct the example using one job description and list of five prospective applicants. The profile vectors are constructed as follows: 3(BS IT required), 2(if the candidate's required experience more than two years), 2(if the candidates' required age is above 22 years old), 1(if the status required is single), and 1(if the gender required is male). The result of job's vector is [3 2 2 1 1] and the result of candidates' vectors are: 1st person: [2 1 2 1 1], 2nd person: [2 1 2 1 0], 3rd person: [2 2 0 1 1], 4th person: [2 2 2 1 1], and 5th person: [3 2 2 1 1]. The applicants' ranking after applying the Cosine Similarity, is presented in table 2.

Table 2: Ranking of Candidates for the job position using cosine similarity measures.

<i>Job Seekers (Candidates)</i>	<i>Cosine Similarity Result</i>	<i>Rank</i>
5 th applicant	1	1 st
4 th applicant	0.98	2 nd
1 st applicant	0.97	3 rd
2 nd applicant	0.94	4 th
3 th applicant	0.87	5 th

This example aims to find a candidate who fits best to the requirements of job profile [3 2 2 1 1]. Based on the similarity measures, 5th applicant is the best candidate who fits job requirements, followed by 4th and 1st applicant. The 2nd and 3rd applicant are the least appropriate candidates for the job requirements.

4. Conclusion

The proposed vectorization model for job matching and recommendation system for Public Employment Service Office (PESO) is developed to aid the matching process of applicants in Pangasinan. This system evaluates whether applicants could match the requirements using identified variables as extracted from the PESO Guidelines through a utility software that incorporates text analysis. Specifically, the study considers selected variables such as job title, educational qualification, age, gender, and civil status that are considered important indicators to be considered in matching. The proposed system seeks to help PESO implements its objective to create sound decisions on matching applicant to intensify employability rate and eliminate job mismatch. Thus, ensures that government produce the best fit between jobs and applicants. Furthermore, this research can be made further by considering more factors for evaluating the matching risk for applicants. Also, this system could also be reflected in other organization in recruitment environment to improve the matching quality.

5. Acknowledgements

The authors thank members of the Public Employment Service Office for their support and comments on this work.

6. References

- [1] Knight, J., & Yueh, L. (2004). Job mobility of residents and migrants in urban China. *Journal of comparative economics*, 32(4), 637-660.
- [2] Furtmueller, E., Wilderom, C., & Tate, M. (2011). Managing recruitment and selection in the digital age: e-HRM and resumes. *Human Systems Management*, 30(4), 243-259.
- [3] Fujita, S., & Ramey, G. (2007). Job matching and propagation. *Journal of Economic dynamics and control*, 31(11), 3671-3698.
- [4] Muralidharan, C., Anantharaman, N., & Deshmukh, S. G. (2002). A multi-criteria group decisionmaking model for supplier rating. *Journal of supply chain management*, 38(3), 22-33.
- [5] Widdows, D. (2003, July). Orthogonal negation in vector spaces for modelling word-meanings and document retrieval. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1* (pp. 136-143). Association for Computational Linguistics.
- [6] Thomas, R. (2002). *U.S. Patent Application No. 09/877,474*.
- [7] Noe, R. A., Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (2017). *Human resource management: Gaining a competitive advantage*. New York, NY: McGraw-Hill Education.
- [8] Isopahkala-Bouret, U. (2015). „It's considered a second class thing.“The differences in status between traditional and newly established higher education credentials. *Studies in Higher Education*, 40(7), 1291-1306.
- [9] Granovetter, M. (2018). *Getting a job: A study of contacts and careers*. University of Chicago press.
- [10] Das, G., & Mannila, H. (2000, September). Context-based similarity measures for categorical databases. In *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 201-210). Springer, Berlin, Heidelberg.
- [11] Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261-273.
- [12] Al-Otaibi, S. T., & Ykhlef, M. (2012). A survey of job recommender systems. *International Journal of Physical Sciences*, 7(29), 5127-5142.