Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 175 (2020) 695–699

Procedia Computer Science

www.elsevier.com/locate/procedia

The International Workshop on Artificial Intelligence and Smart City Applications (IWAISCA)
August 9-12, 2020, Leuven, Belgium

Job Recommendation based on Job Profile Clustering and Job Seeker Behavior

D. Mhamdi*, R. Moulouki, M. Y. El Ghoumari, M. Azzouazi, L. Moussaid

Mathematics, Computing and Information Processing, Faculty of Sciences Ben M'Sik, Hassan II University, Casablanca 20670, Morocco

Abstract

This article presents a recommender system that aims to help job seekers to find suitable jobs. First, job offers are collected from job search websites then they are prepared to extract meaningful attributes such as job titles and technical skills. Job offers with common features are grouped into clusters. As job seeker like one job belonging to a cluster, he will probably find other jobs in that cluster that he will like as well. A list of top n recommendations is suggested after matching data from job clusters and job seeker behavior, which consists on user interactions such as applications, likes and rating.

Keywords: job recommendation; profile clustering; word2vec; k-means clustering.

1. Introduction

In the era of Big Data, both of employers and job seekers are confronted with the increasing data overload and the time-consuming process of recruitment. Candidate's profiles are so diverse that it is laborious for recruiter to find the suitable competencies. Consequently, it is important to identify the most important features of each job candidature.

Job recommendation system are machine learning solutions capable of suggesting pertinent jobs or candidates based on the behavior and needs of job seekers and on requirements of employers. Therefore, applicants could receive personalized online jobs and recruiters are supposed to find the most relevant candidatures with skills and qualifications that fit their needs.

* Corresponding author. Tel.: +212-69-833-5254.
E-mail address: drsmhamdi@gmail.com

10.1016/j.procs.2020.07.102

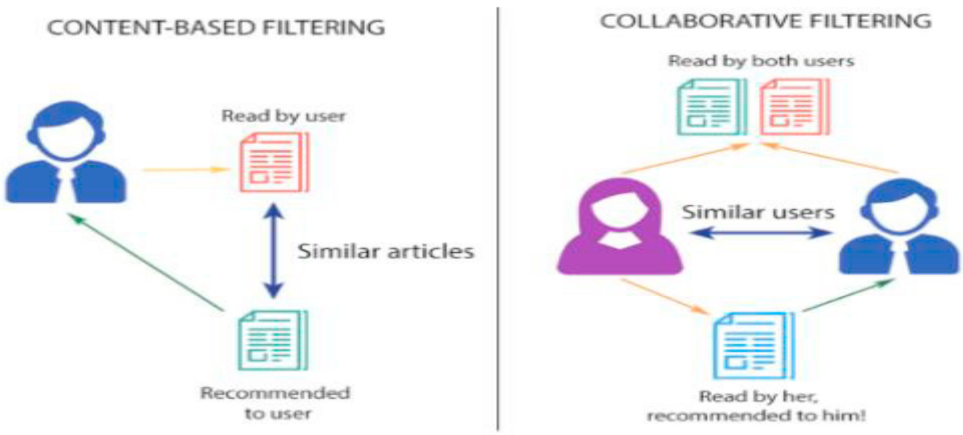

Fig. 1. Content based filtering and collaborative filtering recommendation [3].

Natural language processing as a branch at the intersection of computer science, artificial intelligence, and linguistics could be used to extract useful insights from job offers in order to match candidates to suitable offers. In addition, it offers the capabilities of processing and analyzing large quantities of unstructured job postings and job applications as to provide recruiters and job applicants the ability to understand the preferences and requirements of each other. As a result, they can save time and efforts.

In this paper, first, we have presented related works concerning automated recommendation and some text clustering methods, and then we have exposed the basis and rules of our proposed model.

2. Related works

2.1. Automated Recommendation

While conducting a search on the web, users are supported by automated recommendation to find and choose the right items that fit their needs, according to people they trust or sharing similar tastes. Automated recommendation is divided into content-based filtering and collaborative filtering [1].

As shown Fig. 1, Content-based methods suggest items similar to those a user has selected in the past. Collaborative filtering recommend objects based on the preferences of other users with tastes similar to those of the current user [2].

To overcome weaknesses of the two previous types of recommendation, we could use a hybrid approach that merge the two previous recommendation techniques to get the best advantage from both of them. Hybrid methods are achieved in different ways such as switching, mixing, weighting or using a cascade approach [3].

2.2. Text clustering methods

Text clustering consist on an unsupervised learning approach that aims to group a given text document set into clusters our groups in a way that documents in a same cluster are more similar between each other [4]. Many techniques are used to accomplish textual content clustering of documents. Here are two main methods :

- Word2vec: a neural network based model considering words from a corpus and representing them as vectors with contextual comprehension. Two words are considered similar if the distance between their respectively vectors is lower [5].

Word2vec is a model used to produce word embedding, which is an embedded representation of documents that consists on mapping words or phrases from a vocabulary to corresponding vectors of real numbers. Words that appear together in the text will also be very close in vector space [6].

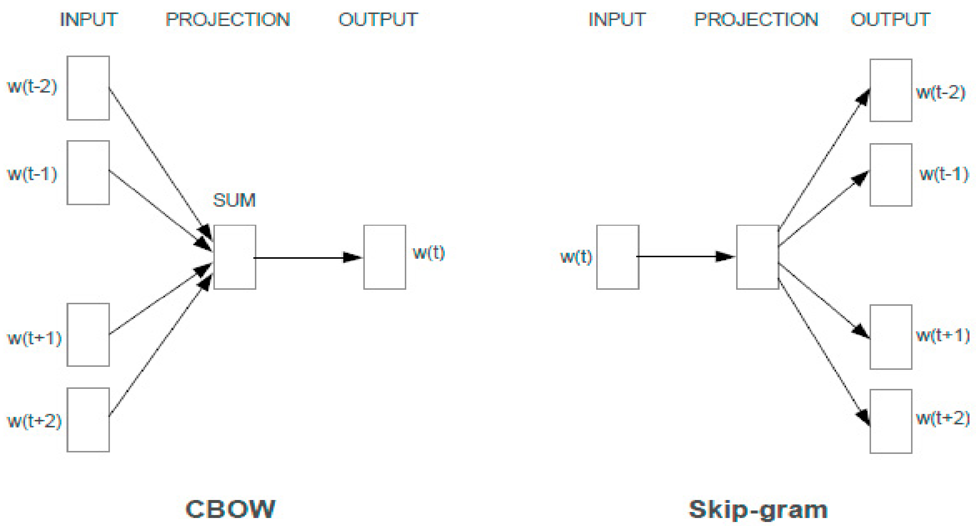

Fig. 2. Difference between SkipGram and CBOW [7].

As shown in Fig. 2, Word2vec includes two architectures of performing distributed presentation of words: continuous bag-of-words (CBOW) and skip-gram.

While the CBOW architecture predicts the current word based on the context, the Skip-gram predicts surrounding words given the current word.

- K-means clustering: K-means is a popular and simple to implement algorithm. It groups N data points into k clusters by minimizing the sum of squared distances between every point and its nearest cluster mean called centroid. It starts by selecting k random data points as the initial set of centroids of clusters. Then the centroid of every cluster is recalculated as the mean of all data points assigned to this cluster [8].

There is a variant of k-means algorithm called Spherical k-means considering all vectors as normalized and using cosine dissimilarity based on the angle between the vectors as calculated in formulae (1) presented below [9]:

$$\mathrm{d(x,p)} = 1 - \cos(\mathrm{x,p}) = 1 - \frac{< x, p >}{||x||\ ||p||} \tag{1}$$

The spherical K-means has as input a set of unit vectors that lie on the surface of the unit hypersphere about the origin, and uses the cosine similarity as its proximity measure [10].

3. Proposed recommender system

Our Dataset includes job offers and job seeker interactions such as rating, likes and reviews. The objective is matching job offers to the right candidates.

A job posting is a document made up of structured data such as the position title and unstructured data such as the map of a location. Many features describe a job offer including the company name, the job title, the job description, the location, the post time, the salary, and the job requirements.

The job title and the job description are main attributes that hold meaningful data. The title is a short and succinct text that designates the job position. The description of the job contains all the skills, qualifications and conditions related to the job.

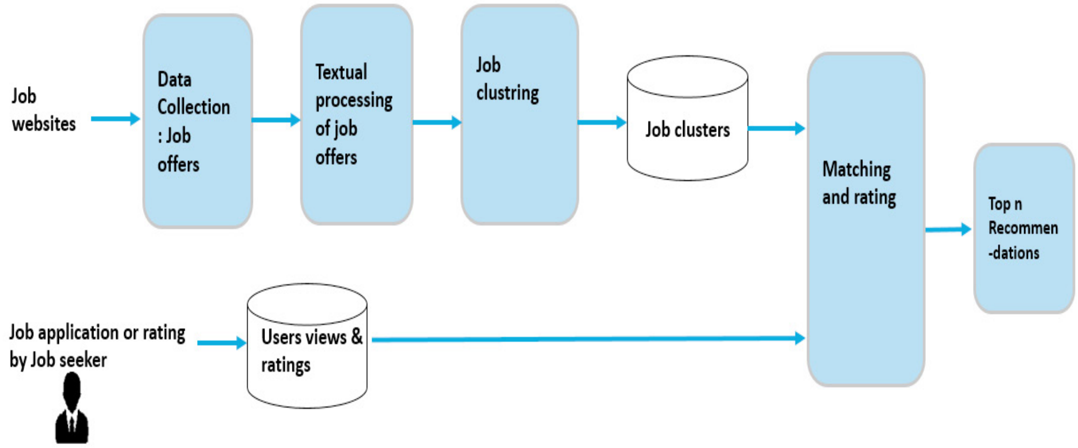

Fig. 3. Matching job offers to job seekers.

The data that will be processed and analyzed is collected from different job search websites using web-scraping techniques.

As shown in Fig. 3, the output of our model is a list of sorted recommendations. Many steps are to perform including data collection, textual processing and matching profiles. In a first step, the unstructured data from job offers is transformed and validated to make it more easily understood. Then, the formatted and cleaned data is processed and analyzed in order to be divided into job clusters based on common features. The attributes from the data contained in these clusters are matched with behavior attributes of the job seekers and a list of n recommendations is suggested to the user. When a job offer is liked or rated by a candidate, all relevant job offers belonging to the same cluster are suggested to the same candidate.

This model is based on cluster analysis approach, which is a self-organized learning that helps to identify groups of job offers according to the degree of similarity, or dissimilarity between their features.

4. Conclusion and future work

In this paper, we presented a job recommender model aiming to extract meaningful data from job postings using text-clustering methods. As a result, job offers are divided into job clusters based on their common features and job offers are matched to job seekers according to their interactions.

Our future Work will focus on training and evaluating our model using Word2vec method and k-means clustering algorithms used to capture and represent the context of job profiles. Subsequently, it will be easy to match set of job offers to a given job seeker based on its past interactions toward specific job offers. The dataset that will be used is built from scraping job search websites.

References

[1] Frank Faeber, Tim Weitzel, and Tobias Keim. (2003) “An Automated Recommendation Approach to Selection in Personnel Recruitment.” *Americas Conference on Information Systems (AMCIS)*.

[2] Marko Balabanovic, and Yoav Shoham. (1997) “Content-based, Collaborative Recommendation.” *Communications of the Association for Computing Machinery* 40(3), 66-72.

[3] Marwa Hussien Mohamed, Mohamed Helmy Khafagy, and Mohamed Hassan Ibrahim. (2019) “Recommender Systems Challenges and Solutions Survey.” *International Conference on Innovative Trends in Computer Engineering*, Aswan, Egypt.

[4] Hoajun SUN, Zhihui LIU, and Lingjun KONG. (2008) “A Document Clustering Method based on Hierarchical Algorithm with Model Clustering.” *22nd International Conference on Advanced Information Networking and Applications*.

[5] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. (2013) “Efficient Estimation of Word Representations in Vector Space.” *Proceedings of the International Conference on Learning Representations*.

[6] D. Mhamdi, R. Moulouki, M. Y. El Ghoumari, and M. Azzouazi. (2020) "Job Recommendation System based on Text Analysis." *Jour of Adv Research in Dynamical & Control Systems*, Vol. 12, 04-Special Issue.

[7] Kalyan Katikapalli Subramanyam, and Sangeetha Sivanesan. (2019) "SECNLP: A Survey of Embeddings in Clinical Natural Language Processing." *Journal of Biomedical Informatics*.

[8] Pasi Franti, and Sami Sieranoja. (2019) "How much can k-means be improved by using better initialization and repeats?" *Pattern Recognition* 93 (2019) 95-12, Machine Learning Group, School of Computing, University of Eastern Finland.

[9] Kurt Hornik, Ingo Feinerer, Martin Kober, and Christian Buchta. (2012) "Spherical k-Means Clustering." *Journal of Statistical Software*, Vol. 50, Issue 10.

[10] Rehab Duwairi, and Mohammed Abu-Rahmeh. (2015) "A novel approach for initializing the spherical k-means clustering algorithm." *Simulation Modelling Practice and Theory* 54 (2015) 49-63.