Skill and Job Recommender

Literature Survey

K.S.R COLLEGE OF ENGINEERING

Team Member

Abishek M (73151913002)

Aathmiga M (73151913001)

Deepika V (73151913016)

Kanishkar R (73151913038)

Introduction

The Internet-based recruiting platforms become a primary recruitment channel in most companies. While such platforms decrease the recruitment time and advertisement cost, they suffer from an inappropriateness of traditional information retrieval techniques like the Boolean search methods. Consequently, a vast amount of candidates missed the opportunity of recruiting. The recommender system technology aims to help users in finding items that match their personnel interests; it has a successful usage in e-commerce applications to deal with problems related to information overload efficiently. In order to improve the e-recruiting functionality, many recommender system approaches have been proposed. This article will present a survey of e-recruiting process and existing recommendation approaches for building personalized recommender systems for candidates/job matching.

Job Recommendation based on Extracted Skill Embeddings

With the increasing popularity of online recruiting platforms in modern industry, most employers choose these platforms as a means of connecting with potential candidates for open positions. Developing job recommendation systems can significantly help both employers and job seekers in speeding up this process and finding the best matches. Using skill phrases

extracted from unformatted and unstructured CVs and Job Descriptions, we implement two approaches with different similarity metrics, namely Word Mover's Distance[1] and Cosine Similarity. We selected TF-IDF with Cosine Similarity as a baseline and evaluated our approaches on the real data from Kariyer.net, which is an employment-oriented online service located in Turkey.

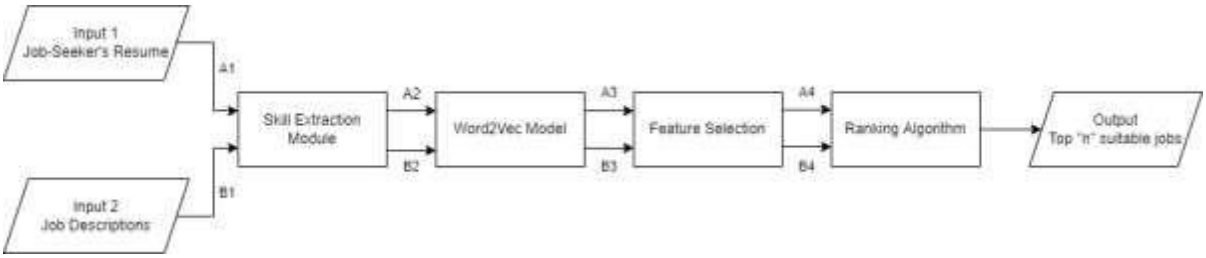
Datasets

We collected 500,000 Turkish and English Job Descriptions (JDs) related to Information Technologies (IT) posted on Kariyer.net. They are divided into two parts: 450,000 of them are used for training a word2vec model and the rest is used for experiments and evaluation. To represent job-seekers, we gathered 7700 CVs only consisting of people working in the IT industry.

Kariyer.net has collected a *Skill Dictionary* by human relations experts and from users' feedback over the years which consists of only English keywords. This dictionary is used in the skill extraction module to find existing skill keywords. We narrowed our focus and only gathered skills related to IT.

Methodology

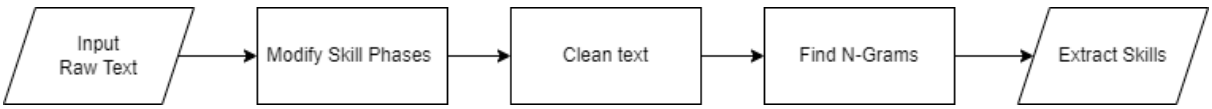
We propose two different approaches. The first one utilizes Word Mover's Distance [1], a similarity metric between documents, and the second one uses cosine similarity. Besides the choice of ranking algorithm part both of the approaches take the same input and have the same core phases



General Workflow

Skill Extraction

The skill extraction module aims to find all the skills, single-word or multi-word phrases. To benefit from the knowledge accumulated over the years at Kariyer.net and to be able to work with the unstructured CVs and JDs, the skill extraction module uses the Skill Dictionary to both identify and extract the skills. As shown below, this process takes raw text as input; JD or Job-Seeker's CV.



Workflow of the Skill Extraction Process

It first checks for the skill phrases that exist in the Skill Dictionary. For every skill phrase found in the raw text, every white space, punctuation, and stop-word is changed with a fixed character. For example, white space and the dot in "machine learning" and "asp.net" are changed with a fixed character. This ensures that those characters will not be removed in the cleaning phase. It also prevents multi-word skill phrases separated by white space

References

1. M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, "From Word Embeddings To Document Distances." Accessed: Jan. 22, 2022. [Online]. Available: http://proceedings.mlr.press/v37/kusnerb15.pdf.