Is it Required? Ranking the Skills Required for a Job-Title

Sarthak Anand, Jens-Joris Decorte, Niels Lowie

1 Utrecht University
2 TechWolf
3 Ghent University

Abstract

In this paper, we describe our method for ranking the skills required for a given job title. Our analysis shows that important/relevant skills appear more frequently in similar job titles. We train a Language-agnostic BERT Sentence Encoder (LaBSE) model to predict the importance of the skills using weak supervision. We show the model can learn the importance of skills and perform well in other languages. Furthermore, we show how the Inverse Document Frequency factor of skill boosts the specialised skills.

Introduction & Related Work

Hiring the right candidate can be a challenging and time-consuming task. Companies are constantly trying to optimize this process by building systems to engage with more relevant candidates. Knowledge about skills and job titles associated with a job vacancy helps in automating the matching process to a suitable candidate. Recent work in this direction to explore the capabilities of BERT to learn the representations of the job titles was done by (Decorte, Jens-Joris and Van Haute, Jeroen and Demeester, Thomas and Develder, Chris 2021). Following this work, (Bhola et al. 2020) worked on identifying the skills from the job descriptions through an eXtreme Multi-Classification (XLMC) objective. One drawback of treating the task as a classification problem is that some skills are more important or relevant for a job vacancy than others. For instance for a job of Python Web Developer knowledge of Python would be more relevant than Problem Solving or Communication Skills. To solve this, we frame the problem to learn the importances of skills given a job title. We only take job input into account in the , since the complete description can be noisy and contain irrelevant information as well.

Dataset & Methodology

Dataset We use an in-house taxonomy of around 37K skills, which was collected using a mix of sources and data mining techniques. We collect job postings from various job portals, and using the skills taxonomy we collect job titles and the skills associated with them. We also do some preprocessing to clean the job titles. In the end, we have around 170K unique job titles in English with a list of skills associated with them, which we later split into training, development and test in 70:10:20 ratio respectively.

Weak Supervision Similar job positions can be represented in multiple ways, for instance Python Developer can also be expressed as Python Programmer, Python Software Engineer etc. In our analysis, the skills that appear more frequently in similar job titles tend to be more important and relevant for a given title. Therefore, for each job title we first find similar job titles using the sentence embedding of the job titles using Sentence BERT (Reimers and Gurevych 2019). The threshold for the cosine similarity is set to 0.75 after empirical analysis. We then find the frequency of all skills present within the similar job titles and normalise it on a scale of 0-1, which are then learned by the model to predict for the input job title. This process is illustrated in the Figure 1.

Training The training objective of the model is to predict the importance of a skill given a job title. We use the pre-trained Language-agnostic BERT sentence embedding (LaBSE) model (Feng et al. 2022) to extract the sentence encoding of the job title. The encoding is then passed to a linear layer, and sigmoid activation is applied to map the output values between 0 and 1, as illustrated in Figure 2. Finally, we rank the skills for the job titles using the computed
importance.

**Results & Discussion**

**Results** In order to evaluate the models during training we use Mean Average Precision for top 20 skills. The table [1] shows the improvement on fine tuning the pre-trained LaBSE model. The Table 2 shows the results for different job titles. The model also performs fairly on other languages (as seen in the table 3), but by fine tuning we lose the multilingual capabilities of the ranking system.

**Boosting specialised skills** If we look at the top skills for Stock Broker in the Table 2 we find some skills which are very generic and therefore have high occurrence in most of the job titles like Excel, Communication Skills, Sales, Finance. In order to find specialised skills we apply a technique similar to TF-IDF, using the training data we compute the $IDF_s$ score of each skill $s$ as

$$IDF_s = \log\left(\frac{N}{f_s}\right)$$

where $N$ equals the total number of titles and $f_s$ is the number of titles associated with skill $s$. On comparing the IDF scores for Excel (0.92) and Tableau (4.72), we get an idea how it helps specialised skills. For finally ranking the skills we compute the importance of skill from the model and multiply with the IDF score of the skill.

The improvement in the rankings can be easily seen in table 4.

**Conclusion & Future Work**

We address the problem of ranking of skills for a given job title. Importantly, we propose a novel method that does not require manually annotating the importance of skills for different job titles. Further, we show how IDF factor of a skill can improve the rank of specialised skills. In future work can be done to rank skills required using the entire job description.