Assessing Basic Natural Language Processing Approaches Towards Building Resume Screening Pipelines for Relevant Candidate Recruitment

Stephanie Henderson
shenderson62@gatech.edu

Hannah Nithilla
hpeter6@gatech.edu

Mahya Qorbani
mqorbani3@gatech.edu

Abstract

In this paper, we evaluate the effectiveness of basic natural language processing techniques in the creation of a simple pipeline, which takes as input a job description, and aims to output the five most relevant resumes. With the increasing use of automated candidate screening in the recruitment process, our team aimed to experiment with different NLP approaches in order to see if we could implement a successful pipeline which would retrieve ideal candidates for a given job. The most successful approach we attempted included embedding the resumes within our dataset as well as the job description, utilizing a BERT model to classify the job description to its closest matching profession, and utilizing cosine similarity to retrieve the most similar embedded resumes in high-dimensional vector space. However, this approach has its limitations in its effectiveness, which are further examined within our report.

1 Introduction

As Artificial Intelligence increases in popularity, one may observe its adoption within several tasks in life. For instance, current job seekers may have noticed the rapid nature of which their resumes are screened, forwarded to the next recruiter, or rejected, in the modern day recruitment process. With the increasing use of automated resume screening within the candidate recruitment process, our team aims to evaluate the question: how effective are basic Natural Language Processing techniques in the creation of a pipeline which returns the top resumes within a given dataset?

2 Related Works

There currently exists several different approaches that utilize natural language process in the task of automated resume scanning. One approach that is very similar to our team's experiment is by Ali et al. (2022), which also utilizes a BERT model to establish embeddings of resumes and job descriptions, and utilize similarity metrics to retrieve the top candidates. However, their approach included parsing resumes for skill sets. In our approach, we embed the entire resume and compare it's similarity to an embedding of an entire job description, as multiple items within a resume may match desired qualifications of a job description. We aim to experiment how effective evaluating semantic similarity is between an entire resume and entire job description in the selection of candidates. since not all resumes are formatted the same way, which could lead to failures in parsing.

Additionally, the approach proposed by Quynh Trinh (2021) is very similar to our team's, in that our approaches both simulate the same use case: a resume database stored by a recruitment system, which is evaluated to extract potential candidates for future job opportunities. However, our approach towards pruning candidates is slightly different. In the approach by Quynh Trinh (2021), the researchers first extract relevant information from resumes, and then utilize a Word2Vec model to create embeddings to help for later candidate retrieval. In our approach, we examine different classification methods (BERT and linear classifiers) to first limit our evaluation of resumes to those belonging to a classified, relevant profession, and then utilize a BERT model to create the embeddings for semantic search.

Lastly, Ali et al. (2022) explores an approach similar to the first phase of our experiment, which entails classifying a job description to its corresponding profession. In Ali et al. (2022), they do the reverse, and classify a resume to potential job openings, in order to best fill spots. They examined possible classification approaches, including K-Nearest Neighbors, SVM, etc. The experimenters of this approach were able to achieve 96% testing accuracy in classification. However, our team goes beyond just profession classification, and also supplies a recruiter with the top resumes for a given position, which would help increase efficiency in the screening process.

3 Data

We utilized a Kaggle dataset which contained more than 2400 resumes, labeled to their corresponding profession. Within this dataset are 24 professions, such as "accountant", "digital media", and "chef".
A limitation of this dataset is that the data is not balanced amongst the labels. For example, the "automobile" profession only has 36 resumes, while the "accountant" profession has 116 resumes. With an imbalanced dataset, our model may not learn meaningful patterns for classification within the profession with limited resumes.
Additionally, each resume was stored as a pdf file, so our team leveraged the PDFMiner python module to extract the text from each resume. With all of the resumes transformed into text format, we were able to utilize a bert-base-uncased BERT model, Devlin et al. (2018) to embed each resume within each profession. By embedding each resume, we effectively transformed the text into high-dimensional vector space, allowing for future comparisons of semantic similarity between a resume and a job description.

4 Methods

The approach our team took towards implementing a system which retrieves the top resumes has roots in basic natural language processing techniques. The steps outlining our approach are also visually depicted in Figure 1.

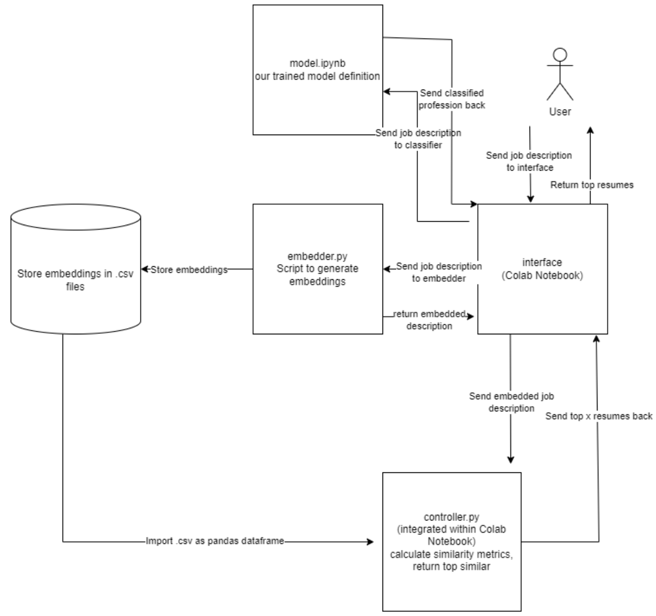

Figure 1: Diagram of our implementation design, illustrating the embedding and comparison process for retrieving the top resumes.

4.1 Profession Classification

To prune the possible resumes to a limited quantity, we first use a classification approach to classify a given job description to its most suitable profession. That is, the first step towards resume generation is to first provide the system with a job description. We then classify the job description to its most suitable profession. For example, if a job description sought candidates with higher education relating to finance, previous experience relating to tax returns and tax accounts, a successful model classification could be "accountant". To perform this classification, we tested a few basic NLP approaches, including the use of a BERT model, SVM classifier, and logistic classifier.

4.1.1 Profession Classification: BERT

First, our team employed the approach of training a BERT (Bidiectional Encoder Representations from Transformers) Model. Our team hypothesized that this NLP approach would work best, due to BERT models ability to leverage context from surrounding words within the resume to create meaningful embeddings, allowing for successful tests of semantic similarity between embeddings, as entailed in Hashemi-Pour and Lutkevich (2024). We trained the BERT model for 10 epochs, and it was able to obtain 80.72% testing accuracy, the highest of the three tested approaches. We utilized the AdamW optimizer, an alternate form of the Adam optimizer which decouples weight decay, as explored in Loshchilov and Hutter (2019). We utilized a learning rate of 0.00005 to achieve this accuracy.

4.1.2 Profession Classification: Linear Classifiers

As a baseline, we compared the BERT model approach towards profession classification with a linear classifiers, including a SVM (Support Vector Machine) and multinomial logistic classifiers. As hypothesized, the accuracy of both of the linear classifiers are lower than the BERT model. Linear classifiers generally are not as effective as transformers for capturing complex (nonlinear) textual data. The BERT model is able to capture meaningful information within sequential text. Additionally, the relationship between a profession and text extracted from a resume is nonlinear; this can be because resumes can be formatted within several different ways, candidates may not have relevant experience pertaining to the profession, etc. Ultimately, we found that the linear classification methods underperformed in comparison to the BERT model for classification.

4.1.3 Profession Classification: Results

As anticipated, the BERT classification method worked best for classifying a job description to its most relevant profession. These results are seen in Table 1.

Table 1: Testing accuracy for BERT, SVM, and Logistic classification of a job description to a relevant profession

Approach	Accuracy (%)
BERT Model	80.72
SVM Classifier	66.26
Logistic Classifier	67.06

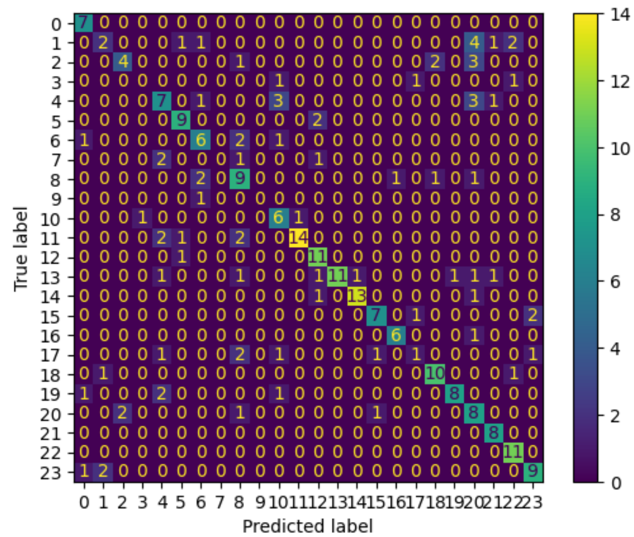

Figure 2: Confusion matrix for the Logistic Regression classifier.

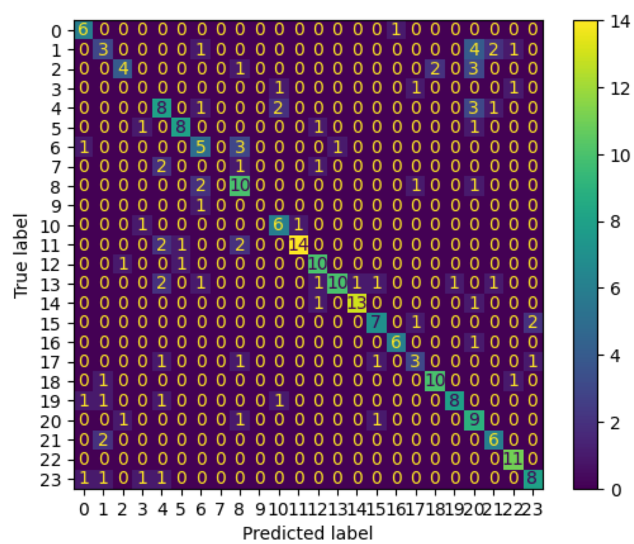

Figure 3: Confusion matrix for the SVM classifier.

4.2 Retrieval of Top Resumes: Semantic Search in High-Dimensional Vector Space

To retrieve the top resumes, our team utilized an approach relating to semantic search. First, we utilized a bert-base-uncased model to embed each resume found within the Kaggle dataset. We stored these embeddings within a csv, and later imported a single csv corresponding to the resumes of the classified profession as a pandas dataframe.

After the job description was classified to its closest matching profession, we also utilized a bert-base-uncased model to embed the job description. With an embedding representing the job description, and embeddings of all of the resumes, we were able to conduct semantic search, by comparing the cosine similarity, jaccard distance, and euclidean distance between the embedded job description and each embedded resume.

4.2.1 Semantic Search: Cosine Similarity

Our team hypothesized that the use of cosine similarity would yield the most relevant resumes. Cosine similarity is a renowned metric that can be used to compare datapoints within high-dimensional vector space. With embeddings of job descriptions and resumes, we utilized cosine similarity to find the closest matching embeddings in terms of the cosine measure of the angle between the high-dimensional vectors (Medium). To test, we created several sample job descriptions which closely matched resumes within our dataset. Upon human inspection of the retrieved resumes, we found that the cosine similarity metric was often able to return a desired resume if the job description was semantically similar to an existing resume.

4.2.2 Semantic Search: Jaccard Distance

To examine other possible similarity metrics, our team also looked at how well Jaccard distance could capture similarity between a resume and a job description. The jaccard distance metric evaluates similarity between vectors by looking at their intersection opposed to their union (Medium). We found that, even when including skills on a test job description found verbatim within resumes of our existing dataset, the jaccard distance metric was not effective in yielding the desired resume.

4.2.3 Semantic Search: Euclidean Distance

With the popular use of cosine similarity as a metric to compare embeddings in natural language projects, our team had hypothesized that it would yield markedly better results than other similarity metrics, such as jaccard and euclidean distance. However, we found that euclidean distance was nearly as effective in our human inspection of the returned resumes. Often, the euclidean distance metric would yield the same top candidate, and even the same top 3-4 candidates as the cosine similarity metric. With this result, it may be helpful to include several similarity metrics (cosine similarity and euclidean distance) in the retrieval of candidate resumes in the automated screening process for job opportunities, in order to help reinforce relevant candidate retrieval.

5 Limitations and the Ethical Debate of Automated Resume Screenings

As with any automated screening of a candidate, we anticipated a multitude of limitations with our approach. First, our approach only highlights the semantic similarity between a job description and a resume. While our model may be able to find relevant candidates due to returning resumes with skills that are very similar to skills listed on the job description, our model will fail to capture important information regarding unique qualifications of individuals.

For instance, two candidate resumes may both have education relating to Accounting, but one individual may have attended a school with a top program relating to accounting, while the other may have not. Our model will not prioritize the top program candidate, unless the job description specifically mentions the name of the top program. This is a big limitation within our approach, and illuminates a possible downside of automated resume screenings.

References

Ali, I. et al. (2022). "Resume classification system using natural language processing and machine learning techniques". In: *Mehran University Research Journal Of Engineering & Technology* 41.1, pp. 65–79. DOI: 10.3316/informit.263278216314684. URL: https://search.informit.org/doi/10.3316/informit.263278216314684.

Devlin, Jacob et al. (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *CoRR* abs/1810.04805. arXiv: 1810.04805. URL: http://arxiv.org/abs/1810.04805.

Hashemi-Pour, C. and B. Lutkevich (Feb. 15, 2024). *What is the BERT language model?* Enterprise AI. URL: https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model.

Loshchilov, I and F Hutter (2019). "Decopupled Weight Decay Regularization". In: DOI: 10.48550/arXiv.1711.05101. URL: https://arxiv.org/pdf/1711.05101.

Quynh Trinh, Thanh-Tuan Dang (Jan. 2021). "Automatic Process Resume in Talent Pool by Applying Natural Language Processing". In: *International Conference on Logistics and Industrial Engineering*. Vol. ISBN: 978-604-308-723-9; ISSN: 2734-956X. Van Lang University. Ho Chi Minh City, Vietnam.

Acknowledgements

We would like to thank the instructors of this course for their helpful comments and feedback.