Job Fair Hunt

Sreevani Suvarna
ssuvarna@ucsc.edu

Abstract

We present a CNN model that predicts which parts of a resume are subjective and which parts are objective. Based on these results, the interviewer can further ask a potential candidate more questions to get to know their skillset better.

Introduction

Background

Job hunting is a common process that goes on in everyone's life. Everyone wants a job and enters the job market with a customized ticket, a resume. From there onwards, the job process expands into recruiting and interviews.

Problem Statement

However, it is getting hard for hiring managers to get to know a candidate as it is difficult to see how accurate the candidate's resume is due to the amount of subjective phrases present in their resume. As a result, it gets hard for that candidate to get hired because the managers may not see them as viable. In order to solve this problem, our goal was to use machine learning to create a model that predicts which parts of a resume are subjective and which parts are objective so it would help make the job hunting process easier for the hiring managers.

Dataset

Source Data

Our first step was gathering the resumes that served as our primary source to create the dataset. Once we gathered all the resumes we could find, we decided to use an Excel document that would contain two columns: one for all the sentences from each resume and one for the labels for sentences that would have the label subjective or objective. We coded up a parser using Python to parse all the sentences from each resume onto the Excel document and mapped each sentence to a label : {objective, subjective}. Overall we annotated 2,753 points.

Labeling Schema

In order to label each sentence as objective or subjective, we first defined what subjective and objective meant. Subjective is when a person feels influenced and puts some of their perspective in their words. Objective relates to words that are more factual. With this definition, we labelled sentences in the way such that if a sentence seemed like it was more towards the candidate's personal feelings as subjective. On the other hand, if the sentences seemed factual, we labelled it as objective. Here are a few examples of some of the sentences we labelled:

Subjective (more personalized):

- An experienced finance professional specializing in fixed income and related products **(may be experienced in their point of view, but not to others)**
- Involved in selection of various web based applications for book building of new issues **(the various keyword makes this subjective as the candidate may feel like 3-5 tasks is a lot where as another may not think so)**
- I have worked on a wide range of multi-million dollar master planned real estate projects and developments in the United States **(the wide keyword makes this subjective as they themselves considered that they worked on a lot of projects whereas another may not think so)**

Objective(more factual)

- Industry generalist advising regional and international Private Equity Funds **(stating a fact about what the candidate's occupation is)**
- Member of a cross regional & functional internal Microfinance consulting team **(stating a fact of what they work on)**
- University of Western Ontario - Richard Ivey School of Business - Canada 2004 - 2006 **(stating their college which is a fact)**

Overall there were 747 objective phrases and 253 subjective phrases.

Model Description

We experimented with a CNN model as it is known to be used for text classification which was the main goal of this project. We used Pytorch to build our model which had a word embedding layer, 2 convolution layers, one dropout layer, and one linear layer. We used the word2vec word embeddings to map each sentence to a vector for the model to perform this classification task.

Results

# of Epochs	# Batch Size	# of CNN layers	Learning Rate	Accuracy
20	8	2	0.001	75%
4	2	2	0.001	67%
16	4	2	0.01	69%
30	16	2	0.01	79%

Conclusion

Overall, the accuracy seems to be around with it being 70% regardless of the parameters changing. This tells us that the CNN model is pretty accurate in classifying sections of the resume as objective or subjective. We would like to further improve our accuracy by annotating more data (we only have 2,000 data points right now) and having more than one person annotate the data because this may lead to a bias if only one person sees the data from their perspective. We would also like to experiment more with our model and add more dimensions or layers to it to see how it would affect the accuracy metric.