Abstract

Managing employees is perhaps the most critical task within an organization. Recent work in economics and finance has shown that employee perception of culture, managerial integrity, and other intangible factors is correlated with financial performance. Thus, it is critical that managers develop and curate skills to effectively manage employees. In this paper, we explore models that predict employee sentiment based on text in employee review data from a popular career website, Glassdoor.com. By developing a high-fidelity model that can accurately predict employee sentiment from text, we provide organizations with a new avenue to examine unstructured text data generated by their employees, e.g. internal quarterly employee reviews.

Introduction

Reviews are a very common way of giving feedback to entities and helps the reader to get a sense of the quality. Such reviews are used in various contexts: For companies, books, films, etc. However, the qualitative nature of a review oftentimes leave an unsatisfying feeling of not being able to identify how to compare different entities just based on the review. A translation of a qualitative review into a quantitative metric can close this gap. Websites like Glassdoor.com already offer the option to also add a quantitative (star ratings out of 5) rating. The goal of this project is to predict those quantitative ratings by applying an advanced sentiment analysis to the qualitative reviews based on reviews and ratings from Glassdoor. Such a model can then potentially also be applied to review platforms where there is no quantitative component and therefore enrich the review and enable comparisons between different reviews. We approach the problem of review classification by exploring the fidelity of deep-learning models such as long-short term memory recurrent neural networks (LSTM RNNs) with models from more traditional machine learning methods such as Naïve Bayes classifiers.

Background

Understanding employee perception of the firm is a critical task for managers (cite). Research in strategy and finance has started to look at the importance that employee satisfaction can have on firm level outcomes. Erhard, Jensen, and Zaffron (2007) show that maintaining a culture of integrity can have short-term costs, but long-term benefits to a firm. In another study, Edmans (2011) looks at whether the stock market fully values intangibles. Using the “100 Companies to Work for in America”, Edmans finds that that a portfolio of these companies earned an annual four-factor alpha of 3.5% from 1984 to 2009, 2.1% above the industry benchmark. Lastly, Guiso, Sapienza, and Zingales (2015) study which elements of corporate culture are related to a firm’s performance. They use data from the Greatest Places to Work Institute and find that managerial integrity and managerial ethics as rated by employees both have a positive correlation with Tobin’s Q.

A common problem in Natural Language Processing (NLP) is predicting the sentiment of a paragraph. In this report, we will look at employee reviews from Glassdoor.com. Employees write about their employer and then provide a star rating from one to five. This star rating will serve as a good proxy for the sentiment of the review. Glassdoor stands to learn about relevant employee satisfaction metrics after analyzing sentiment across many firms. Employers can also keep a finger on the pulse of the firm by monitoring the sentiment that appears in reviews. By first training models on Glassdoor data, employers could then apply the same models to unstructured reviews and feedback provided by employees. This would allow firms to quantitatively analyze and evaluate text in internal reviews provided by employees.

In order to tackle this problem, we take an iterative approach to try and build a model that can correctly classify sentiment. We begin with a Naïve Bayes classifier that analyzes one-hot encoded reviews. Next, we implement simple 1-ReLU and 2-ReLU networks and classify reviews that are encoded using bag-of-words integer encoding, word2vec, and GloVe. Lastly, we use a Long-Short Term Memory (LSTM) Recurrent Neural Network to build a model based on both the integer-encoded and GloVe encoded reviews.

Approach

Dataset

Glassdoor is a jobs and recruiting site that holds a database of millions of company reviews, CEO approval ratings, salary reports, interview reviews, and more. Before reviews are posted, employees must verify that they currently or previously worked at the listed employer. Along with this, review postings are completely anonymous and either unsolicited or contributed by employees searching for jobs in return for unlimited site access. These aspects of Glassdoor help ensure the authenticity of the reviews and reduce the potential for reviewer bias.

We obtained the dataset for this project by scraping the reviews of all public companies on Glassdoor.com. After extensive work developing the script to scrape Glassdoor, we ran our script for one week to obtain 1,015,163 reviews. Each of these reviews contains text sections for employees to fill in pros, cons, and advice to management. For this paper, we choose to focus only on the pros and cons when predicting the star rating. This is because employees tend to directly describe the state of the firm in the pros and cons, whereas in the advice to management they write how they would like the firm to be. Thus, we remove this potentially confounding data from our analyses.

Data Pre-processing

As we said, we extract only the pros and cons from the Glassdoor reviews. For our analysis, we further limit the pros and cons to 100 words. The average number of words in each section is 67, and the median is 49. Thus, this cutoff will capture the true sentiment of the majority of reviews while keeping the data at a manageable size. We use the Natural Language Toolkit (nltk) package in python to word tokenize the reviews and make this cut. If the combined text of the pros and cons is less than 200 words, we pad the reviews with a special word, ‘PAAAAAD’, until they reach this length. We then integer tokenize the words using the top 10,000 words contained in the reviews used for training. Any words that are outside of the top 10,000 words are replaced with a predetermined integer that represents unknown characters. This provides our first encoding of reviews: integer encoding using bag of words. The integer that represents a word corresponds to its position in the overall count of words. For instance, ‘the’ is the 5th most common word and is thus represented by a 5 in a review vector.

We split our reviews into three subsections of training (100k), dev (10k), and test (10k). We select this random subset of the reviews to allow for faster training times, thus allowing us to explore more model configurations.