Sentiment Analysis of RateMyProfessors Reviews: Investigating Evolving Perspectives on "Good" Professors Across Institutions

Finn Clark (Junior)
Anushil Jha (Senior)
Department of Computer Science and Engineering
{clarka58, jhaanush} @msu.edu

Dan Ka Le (Senior)
Emma Fountain (Junior)
Department of Computational Mathematics, Science, and Engineering
{ledan3, founta57} @msu.edu

Abstract

Institutions have been largely rated over two criteria – one of them being their ability to support and nourish talent. The second method is analyzing the data provided by students of an institution. This data appears robust and reflects the current situation of an institution in its flesh. RateMyProfessor has been an excellent tool for aggregating this data and assisting people rather than relying on the institution's legacy. Professors make up a huge chunk of the academic experience, and a knowledgeable professor is as useful as one who is willing to help their students. Sentiment analysis is a tool that can assist an individual in overcoming the ancient method of trial and error. This method classifies the given data set into positive, negative, or neutral, also known as polarity detection. In this analysis, we will be tackling the issue of identifying the sentiments of students (namely, positive and negative) towards any professor through Sentiment mining [2]. This machine-learning experiment is aimed at analyzing the sentiment polarity of students through Optimized Sentiment Analysis Framework.

Keywords – Machine learning, Sentiment analysis, Polarity detection, Optimized Sentiment Analysis Framework

Introduction

RateMyProfessors is a popular review site that allows anyone to anonymously rate professors from American, Canadian, and United Kingdom institutions. The site currently has more than 19 million evaluations and ratings of over 1.7 million professors at over 7,500 schools, making it the largest online destination for professor ratings. RateMyProfessors reviews have profound impact on students worldwide: students seem to have confidence in reviews on the site and there have been studies showing they use the site to make academic decisions.

Considering the popularity and academic influence of RateMyProfessors, we have decided to look at the language used in professor reviews on the site. Specifically, we aim to predict the rating of a professor based on the contents of reviews. This will ultimately help us answer the following questions: What makes a professor good or bad? What do students look for in a professor, and how has that changed over time, especially as a result of COVID-19 and a shift to online learning? This analysis will not only show how educational dynamics have evolved to their present state—it will also provide insight on potential trends and the future of education.

Related Works

Some work has already been done on predicting and analyzing professor reviews. A group at the University of Michigan investigated reviews from 33 different institutions, creating the first ever study of RateMyProfessors reviews in the field of natural language processing. Much like we are intending to do, the group used the text of professor reviews to predict whether the review was reflective of a "good" or

"poor" evaluation among other tasks. To represent the text, they found that features consisting of a mix of unigrams and bigrams worked the best. We plan on experimenting with a variety of more preprocessing methods to extract features from the text, outlined in the 3.2 Modeling section. Moreover, our analysis aims to investigate all institutions on the RateMyProfessors (RMP) site instead of just 33. We also intend to investigate the trends of RMP professor reviews amidst the COVID-19 pandemic, which the original study was unable to do as it was published in 2016 using 2015 data.

There is also a large body of work related to sentiment analysis using SVMs. For example, the work by Ahmad, et. al. reviews several other papers and proposes a basic framework for sentiment analysis using SVMs which they call the Optimized Sentiment Analysis Framework (OSAF). This framework, which we draw heavily from, involves four main phases. First, the Dataset phase, which involves loading and cleaning the dataset. Second, the Pre-Processing phase, which involves applying transformations to the data to better extract information from it. Third, Classification, which is the training of the model and optimization of its hyper-parameters using grid search and k-fold cross validation. Finally, the Results phase which involves measuring and reporting metrics including precision, recall, and f-score.

Methodology

1. Dataset

The dataset will be collected from ratemyprofessors.com via a webscraping tool designed specifically for the purpose. The tool can be found on github, and provides a utility to get all the professors for a particular school as well as data for those professors. The dataset will initially be composed of professors from a small subset of schools, and if time permits this will be expanded using other web-scraping utilities such as beautifulsoup to get a large amount of data from as many schools as possible.

The dataset features will be the comments/reviews left by students and the class will be "positive" or "negative" depending on whether the corresponding score is above or below 3. Two team-members will work together to gather, format, and clean the data. Specific team members have not yet been assigned to particular tasks.

2. Modeling

The model will follow the Optimized Sentiment Analysis Framework (OSAF) framework outlined by Ahmad, et. al. This involves four main steps; gathering the dataset, preprocessing the data, classification using an SVM, and results.

The dataset will be stored and read in JSON format, as that is the output of the webscraper we are using.

Preprocessing will be done incrementally. We will start with the basic necessity of converting individual words into vectors. After ensuring that the full pipeline is functional, further preprocessing steps will be applied including stemming, stop-words, term frequency–inverse document frequency weighting, and generating word n-grams. All of these preprocessing steps will be implemented using either the nltk library or scikit-learn. Since there are 4 additional pre-processing steps, each team member will research and implement one of the steps, with a group review and collaboration during meetings.

Classification will be performed using an SVM with a Radial Basis Function kernel. The hyperparameters will be adjusted using grid search and k-fold cross validation.

Results will be measured in terms of precision, recall, and f-score. Training and testing sets will be generated using a 70:30, set aside, and only used for final metric reporting. The training set will be used for hyperparameter tuning. A group of two team-members will be responsible for running the SVM, adjusting the hyperparameters, and reporting results.

References

Azab, Mahmoud, et al. "Analysing ratemyprofessors evaluations across institutions, disciplines, and cultures: The tell-tale signs of a good professor." *Lecture Notes in Computer Science*, 2016, pp. 438–453, https://doi.org/10.1007/978-3-319-47880-7_27.

Ahmad, Munir, et al. "SVM optimization for sentiment analysis." *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 4, 2018, https://doi.org/10.14569/ijacsa.2018.090455.

Brown, M.J., Baillie, M., Fraser, S.: Rating ratemyprofessors.com: A comparison of online and official student evaluations of teaching. College Teaching 57(2), 89–92 (2009)