Rating Reviews Based on Sentiment Analysis

Saeed Al-Qarni
School of Computing and Engineering
University of Missouri - Kansas City
saacfb@mail.umkc.edu

Kahlid Dhabbah
School of Computing and Engineering
University of Missouri - Kansas City
kmdk2t@mail.umkc.edu

Anshit Saxena
School of Computing and Engineering
University of Missouri - Kansas City
asm5y@mail.umkc.edu

Abstract—Nowadays, Social media has become popular places for sharing the best description of a movie through texts. Analyzing reviews or text helps in recognizing the sentiment stated by people. IMDb movie has a lot of reviews that have been expressed by reviewers towards a movie. A positive review and a negative review have taken place for many researchers to apply sentiment analysis to them. In the proposed work, we show how to give a rating based on reviews and compare them with IMDb rating. Getting positive and negative reviews based on our rating plays a significant role in text classification using sentiment analysis. As a result, we will see how the difference will be between our rating and IMDb. Moreover, our model will help people find the easiest way to read the top words for a movie instead of spending more time to just read a few completed reviews.

I. INTRODUCTION

Text classification is a valuable task in Natural Language Processing (NLP), such as sentiment analysis, web search, and reviews classification. sentiment analysis is one of the hot topics in the area of Natural Language Processing (NLP). Nowadays, people watch a lot of movies after they can see the rating for that movie. However, people can only view 10-20 reviews at a time. Through sentiment analysis, people can know the words that are described for a movie and the review ratings from above 5000 reviews. It is an easy way to describe and find the best movie instead of reading a lot of text in a review. In this paper, we explore ways to scale these reviews and compare them with IMDB rating, in the context of text classification. We will grab the reviews for a movie and do the sentiment analysis. After that, we will give the rating based on that sentiment and compare it with IMDB rating. Once we get our rating, we can find the top words which describe a movie.

II. RELATED WORK

In [1], they have used IMDb reviews to do sentiment analysis. This paper is about enhancing the accuracy and improving the overfitting by utilizing two networks, CNN and LSTM. Also, by using the Dropout layer with max Pooling and Batch Normalization, it helped them overfitting issues. They achieved the best accuracy of 89.5

The author's objective in [2] is to improve the vanishing gradient issue. They applied sentiment analysis on IMDb reviews. Their model was a mix of CNN and residual connection, and they tested it on four different models. They achieved the best accuracy of 90.02

In [3], two models were proposed. The first model is a hybrid LSTM-CNN, which performs better than the second hybrid CNN-LSTM model. The accuracy achieved was 79%. They have used preprocessing techniques to enhance the model's performance. Not only that but also, they used an embedded layer to embeds the words into a vector.

[4] used (TF, TF-IDF) beside Lexicon features to result in a better output of the model. It is also used the IMDb dataset. The pr-eprocessing and feature extraction were applied to the dataset to give a better result. They have used different algorithms, such as SVM, KNN, Maximum Entropy, and Naïve Bayes.

[5] applied a new way of sentiment analysis using Bayesian Rough Decision Tree, which uses the Decision Tree and Bayesian Rough set combined. They have done the text pre-processing, such as tokenization and conversion to lowercase, as well as feature extraction, such as POS. They achieved around 95
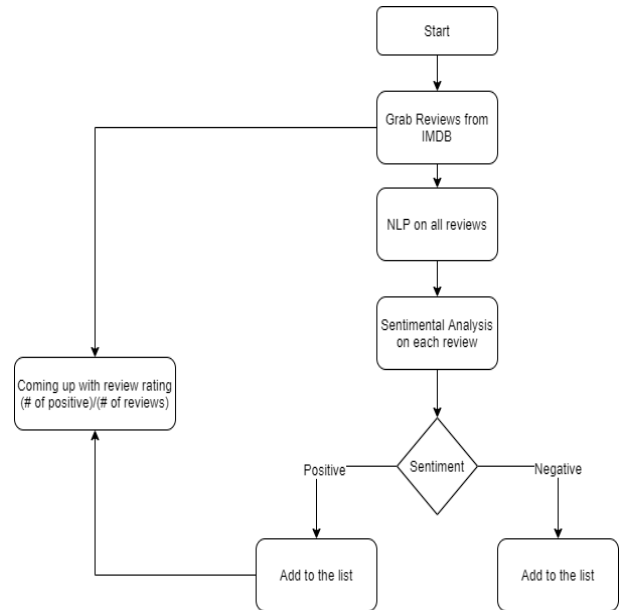
III. ARCHITECTURE

Figure 1: Flowchart for the proposed architecture

We have followed several steps to implement the sentiment analysis as shown in Figure 1.

1) Collecting the Dataset

For the purpose of collecting the data from the IMDB for a particular movie, we are using the Chrome Browser. We use a driver dedicated for chrome for opening up the IMDb reviews

page for that particular movie and wait for it to load all the reviews. After the page has loaded all the reviews we get all the reviews using BeautifulSoup and HTTP Parser.

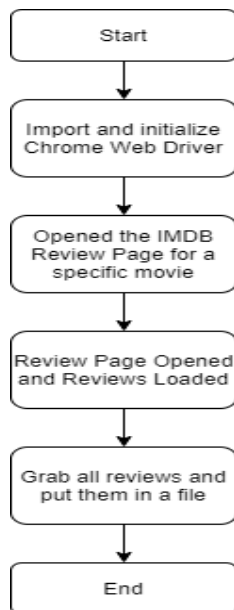Figure 2 shows our approach for collecting the IMDb Dataset.

Figure 2: Flowchart for getting the IMDb dataset

2) Performing NLP

Removing stop words Encoding the data in proper format (UTF-8) Tokenization and Capitalization/De-Capitalization Breaking attached words Lemmatization/Stemming

3) Performing Sentiment Analysis

After running sentiment analysis on each review, we classify them as positive and negative and separate them.

4) Calculating the Rating

We calculate our rating using the total number of reviews and the number of positive reviews by:

Total number of Positive reviews / Total number of reviews

IV. RESULT AND DISCUSSION

We will reach to the final result by the end of this project.

V. CONCLUSION

The conclusion of our work will be reached at the end of the project.

REFERENCES

[1] A. Yenter and A. Verma, "Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis," 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, 2017, pp. 540-546.

[2] N. K. Thinh, C. H. Nga, Y. Lee, M. Wu, P. Chang and J. Wang, "Sentiment Analysis Using Residual Learning with Simplified CNN Extractor," 2019 IEEE International Symposium on Multimedia (ISM), San Diego, CA, USA, 2019, pp. 335-3353.

[3] A. Sajeevan and L. K. S, "An enhanced approach for movie review analysis using deep learning techniques," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1788-1794.

[4] Kumar, Himanshu, B. S. Harish and H. K. Darshan. "Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method." IJIMAI 5 (2019): 109-114.

[5] İ.Tarımer, A. Çoban, A.E. Kocaman. "Sentiment Analysis on IMDB Movie Comments and Twitter Data by Machine Learning and Vector Space Techniques" 2019.