Sentiment Analysis for Movie Reviews

Jiachen Ning
Linhong Li
Shiqi Xiao
Xiangting Chen

March 2, 2019

1 Introduction

Sentiment Analysis is the task of classifying text documents by their sentiment, or overall opinion towards the subject matter [3]. Such analysis can not only provide concise summaries of texts to readers, but also has extended applications in security, medical, finance, and entertainment domains [2].

In this project, the primary task is to classify the sentiment of a collection of 50,000 reviews posted on the Internet Movie Database. The dataset is balanced, containing the same number of positive and negative reviews, where positive reviews had a score greater than or equal to 7 and negative reviews had a score less than or equal to 4 [1].

2 Literature Review

A lot of work has been done on sentiment analysis for movie reviews. There are broadly three types of approaches towards polarity detection in the previous works : (1) Using a machine learning based text classifier - Random Forests, Support Vector Machine (SVM) or Naïve Bayes (NB) - with chosen feature space (2) Using the unsupervised semantic orientation scheme of using frequent n-grams as features and labeling them (3) Using SentiWordNet, a lexical resource especially created for assisting sentiment classification.

Some challenges in sentiment analysis includes separating irrelevant information from the topic of interest, determining the sequence of words impacted by negation, and finding the weights of words with different part of speech.

3 Methods

4 Preliminary Results

5 Evaluation of Preliminary Work

References

[1] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., Potts, C. (2011, June). Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1* (pp. 142-150). Association for Computational Linguistics.

[2] Mäntylä, M. V., Graziotin, D., Kuutila, M. (2018). The evolution of sentiment analysis?A review of research topics, venues, and top cited papers. *Computer Science Review, 27*, 16-32.

[3] Pang, B., Lee, L., Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Association for Computational Linguistics.