Sentiment Analysis of IMDB Movie Reviews

Yasamin Hosseinzadeh Sani

Introduction

In the evolving field of natural language processing, sentiment analysis stands out as a key technique for deciphering the emotional undertones embedded within textual data. Our project harnesses advanced machine learning algorithms to analyze and categorize sentiment polarity—positive or negative—of movie reviews from the globally recognized IMDB website. This endeavor not only helps in understanding public sentiment towards films but also enhances the granularity of feedback analysis within digital media platforms.

Literature Review

Sentiment analysis has been extensively researched, with foundational studies highlighting its relevance in domains ranging from consumer behavior to political science. The seminal work by Pang et al., which explored the use of machine learning for classifying film reviews, provided a base for our methodology. Furthermore, advancements in text processing and machine learning techniques have allowed for more accurate and nuanced sentiment analysis, as demonstrated by the work of Liu and Zhang. Our study builds upon these insights, incorporating cutting-edge machine learning strategies to refine sentiment evaluation methods.

Preliminaries

Drawing on a dataset featured in Andrew L. Maas et al.'s influential paper, "Learning Word Vectors for Sentiment Analysis," we analyze 50,000 reviews equally divided between positive and negative sentiments. This dataset is meticulously split into sets for training, validation, and testing, consisting of 25,000, 12,500, and 12,500 reviews respectively. Our primary tool for classification is the multinomial Naïve Bayes classifier, chosen for its effectiveness in text classification tasks due to its simplicity and speed in handling large datasets.

Methodology

1. **Vocabulary Building** Starting with an initial set of 1000 words, our approach involves constructing a comprehensive vocabulary from the training data. This process identifies high-frequency words which are pivotal in representing textual features within the reviews.
2. **Feature Extraction** Employing a Bag of Words (BoW) model, we transform textual data into structured feature vectors. This representation quantifies the presence of vocabulary words in each review, alongside binary labels that indicate the

sentiment of the review, thus preparing the data for subsequent analytical processes.

3. **Classifier Training** We utilize the multinomial Naïve Bayes approach, training the model on prepared feature vectors. The classifier evaluates the probability of word features within the context of positive and negative sentiments, effectively categorizing the reviews.

Results and Comparative Analysis

Our experimental setup yielded the following results:

- Multinomial Naive Bayes Accuracy: 0.81
- Logistic Regression Accuracy: 0.83

The slight edge in performance by logistic regression suggests its superior capability in capturing complex feature relationships, an aspect somewhat overlooked by Naïve Bayes, which presumes independence between features.Logistic regression shows a slightly higher accuracy of 0.83 compared to 0.81 for the multinomial Naïve Bayes. This suggests that logistic regression may be better at capturing intricate relationships between features, possibly due to its ability to handle dependencies among input variables, which multinomial Naïve Bayes, with its assumption of feature independence, does not.

Vocabulary size :

changing the size of the vocabulary does influence the results. The report initially uses a small vocabulary of 1000 words. If this size is increased, it can potentially capture more nuanced language features in the reviews, which may improve the accuracy of the sentiment analysis. The extent of the influence would depend on the specific increase in vocabulary size and the nature of the text data. Typically, a larger vocabulary allows the model to better differentiate between sentiments but may also introduce noise if too many irrelevant words are included.

Conclusion

This detailed exploration into sentiment analysis underscores the significant impact of selecting appropriate classification models on the outcomes of automated sentiment assessment. Our investigation not only contributes to the broader discourse on machine learning applications in text analysis but also sets the stage for future research into optimizing classifier configurations for enhanced performance in diverse textual environments.

I affirm that this report is the result of my own work and that I did not share any part of it with anyone else except the teacher.