Project Milestone Report

Sentiment Analysis on IMDB Movie Reviews Proposal: Developing a Model for Sentiment Classification

Lei Yang
Tulane University
lyang20@tulane.edu

Qinzheng Xu
Tulane University
qxu5@tulane.edu

1 Problem Overview

The issue at hand is using deep learning models to automatically figure out how people feel about movies in reviews. The goal is to correctly find, separate, and label the feelings behind movie reviews from the IMDB database as either positive, negative, or neutral. The subjective and often complicated nature of human language, which includes sarcasm, idiomatic expressions, and different emotional states, makes this job hard. To make things even more difficult, you have to understand the context and semantics of the texts in order to accurately classify the mood. The end goal is to use the information gathered from this research to help entertainment companies improve their services, make their marketing more effective, and make viewers happier. Achieving high accuracy in sentiment analysis can make these insights more open to everyone, so more people can understand and use the feelings and opinions shared in movie reviews.

2 Data

This sentiment study is based on information from the IMDB database, which has 50,000 movie reviews. This dataset was carefully chosen for tasks that involve natural language processing and text analytics. Its main focus is on binary mood classification, which means putting reviews into two groups: positive and negative. The data has a lot of it, which makes it a great resource for training and testing deep learning models that are meant to analyse mood. This dataset is unique because it is so big. Compared to other standard datasets in the same field, it lets you train and test models in more ways.

There are two parts to the dataset: a training set and a testing set. Each has 25,000 movie reviews. This fair split makes sure that both positive and negative feelings are represented equally, which is important for teaching computers to correctly sort feelings without any bias. The data comes from the IMDB library and can be found on Kaggle (Maas et al., 2011), a website for data science and machine learning projects. In the fields of mood analysis and natural language processing, this makes it easier to repeat studies and make sure that results are correct.

3 Method

3.1 Algorithms Used

We used three different types of neural networks to figure out how people felt about IMDB movie reviews: Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Transformer models. Each of these models takes a different approach to dealing with and making sense of the complex structure of textual data, and each has its own benefits when it comes to catching the subtleties of language and emotion.

3.2 Libraries and Variations

We used well-known Python tools for deep learning, like TensorFlow and Keras, to put these models into action. These libraries make the development process faster by giving you a lot of tools and features for creating and training neural network models.

- We used normal configurations from the Keras library for the ANN and RNN models, but changed the number of units, dropout rates, and sequence processing to make them work best for sentiment analysis.

- Because TensorFlow is so flexible, the Transformer model implementation added custom layers and attention mechanisms to make it work for our text classification job.

3.3 Evaluation Methods

A confusion matrix, which is a powerful way to see how well a model can classify things into different groups, was used to judge the models' success. Accuracy, precision, recall, and the F1 score are some of the key measures we got from the confusion matrix. These metrics give a full picture of how well the model can predict the future and tell the difference between happy and negative emotions.

3.4 Baselines for Comparison

A comparison with well-known benchmarks in sentiment analysis was part of our research to make sure that our models worked. By comparing our results to these standards, we wanted to show how our method led to progress and possible improvements. We specifically looked at standard metrics that regular machine learning methods and past neural network models got right in sentiment analysis tasks. We then used these as a starting point to see how much better our ANN, RNN, and Transformer models were.

4 Intermediate/Preliminary Experiment Results

4.1 Text Features for Extraction

We will extract the following text features, each offering unique insights into the text data:

1. **Polarity**: This is a measure ranging from -1.0 to 1.0. A score of 0 denotes neutral sentiment, -1 signifies extremely negative sentiment, and +1 reflects highly positive sentiment.

2. **Subjectivity**: Expressed as a float within the range of 0.0 to 1.0. A score of 0.0 indicates a highly objective stance, whereas 1.0 suggests a highly subjective viewpoint.

3. **Word Count**: The total number of words present in each review.

4. **Part-Of-Speech (POS) Tag Ratios**: We will calculate the ratio of specific POS tags to the total word count in a review. The tags include:

   - PROPN (Proper Noun), e.g., John, London, NATO.
   - PUNCT (Punctuation), e.g., ., (,), ?, !.
   - NOUN, e.g., woman, tree, beauty.
   - ADJ (Adjective), e.g., big, yellow, sparkling.

- VERB, e.g., eat, running.

5. **Uppercase Ratio**: The proportion of uppercase words to the total word count in the review.

6. **Digits Ratio**: The proportion of digits to the total word count in the review.

4.1.1 Polarity Distribution

The histogram shows the orientation scores of reviews that are both positive and negative. When reviews are good, the polarity scores are more concentrated on the right side of the scale, reaching their highest point in the range that shows strong positive sentiment. On the other hand, bad reviews have a polarity distribution that leans to the left and has a peak in the negative spectrum.
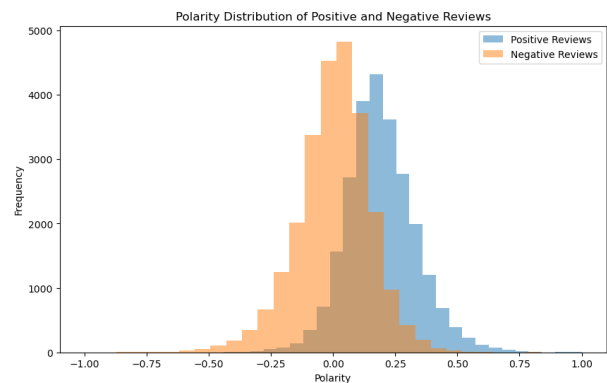

Figure 1: Polarity Distribution

This bimodal distribution highlights the clear demarcation between the sentiments of positive and negative reviews, essential for refining sentiment analysis algorithms.

4.1.2 Word Count Distribution Analysis

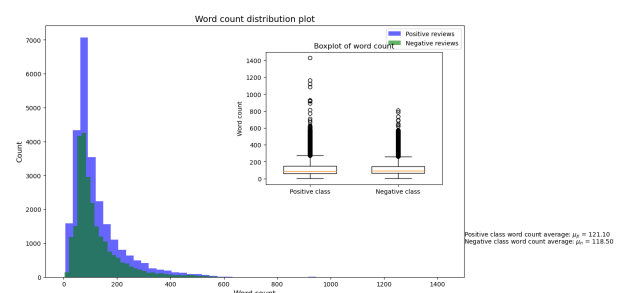The histogram and boxplot visualization indicate the frequency of word counts for both positive and

Figure 2: Polarity Distribution

5 Related Work

When researchers use sentiment dictionaries to study sentiment analysis, the first thing they need to do is label the expression trend of sentiments. This means that they need to make the emotional meanings of words clearer. The study is based on predefined semantic rules and the sentiment scoring of the lexical expressions in the sentiment dictionary. A series of text processing is used to get the general sentiment tendency of the text. Some researchers, including Tureny et al. (Turney, 2007), look at and figure out how similar words are by adding mutual information intent. They also rebuild the dictionary by adding polar semantics, which makes the dictionary's lexical construction more complex. Yang et al. (Chen and Yang, 2013) start with a sentiment dictionary that already exists and use the LDA model tool to find the topic words in that dictionary. This makes the sentiment tendencies more expressive and gives topic words with multiple meanings a lot of extra information. Pang et al. (Pang et al., 2002) were the first to use machine learning to solve the sentiment analysis binary classification problem. They combined standard bag-of-words methods with machine learning to make the classification work better. Wikarsa et al. (Wikarsa and Thahir, 2015) gathered a huge number of Twitter comments as a dataset and then put emotions into groups based on their meanings. Once the groups were made, the authors used standard machine learning algorithms to figure out how people felt about the tweets, taking into account the emojis and words in the dataset to improve the algorithm's performance and accuracy. Govindarajan et al. (Govindarajan, 2013) made a big step forward in combining genetic algorithms and traditional machine learning algorithms in a natural way. They used the combined algorithm as a new way to solve the problem of sentiment binary classification and got better results in binary classification effects. Chikersal et al. (Chikersal et al., 2015) changed the emotional rules of emojis and polysemous words in comment texts by mixing rule processing and SVM and processing Twitter corpora. This made text semantic prediction more accurate. But the text opinion research based on machine learning has hit a wall. This is because the research field is getting too big. When it comes to current long text jobs that involve semantics, the combination of dictionary rules and machine learning often doesn't work as planned, which slows down research and development.

6 Division of Labor and Timeline

6.1 Division of Labor

The project tasks have been divided into two main categories: the coding part and the report part.

- **Coding Part:**
  - **[name1]** will focus on developing the backend algorithms for sentiment analysis, including the implementation of the neural network models and data preprocessing.
  - **[name2]** is responsible for the frontend development, creating the user interface for the web application and integrating the backend algorithms.

- **Report Part:**
  - **[Qinzheng Xu]** will contribute to writing the methodology section of the report, detailing the algorithms used and the rationale behind model choices.
  - **[Lei Yang]** will handle the results and discussion sections, analyzing the performance of the implemented models and discussing the implications of the findings.

6.2 Timeline

1. **Frontend Development and Integration:** Completion by [March 02, 2024].

   - Task involves developing the web application frontend and integrating it with the backend algorithms.

2. **Testing and Refinement:** Completion by [March 09, 2024].

   - This phase focuses on testing the complete system, refining the models based on performance, and ensuring usability of the web interface.

3. **Final Report Writing:** Completion by [March 15, 2024].

   - The final step includes compiling the project report, incorporating all sections, and reviewing the document for completeness and accuracy.

References

L Chen and Y Yang. 2013. Emotional speaker recognition based on i-vector through atom aligned sparse representation. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7760–7764. IEEE.

P. Chikersal, S. Poria, and E. Cambria. 2015. Sentu: sentiment analysis of tweets by combining a rule-based classifier with supervised learning. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 647–651.

M. Govindarajan. 2013. Sentiment analysis of movie reviews using hybrid method of naive bayes and genetic algorithm. *International Journal of Advanced Computer Research*, 3(4):139.

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*.

B Pang, L Lee, and S Vaithyanathan. 2002. Thumbs up: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. Association for Computational Linguistics.

B. W. Turney. 2007. Anatomy in a modern medical curriculum. *The Annals of The Royal College of Surgeons of England*, 89(2):104–107.

L Wikarsa and S. N. Thahir. 2015. A text mining application of emotion classifications of twitter's users using naive bayes method. In *2015 1st International Conference on Wireless and Telematics (ICWT)*, pages 1–6. IEEE.