Chengbin huang   shichen gyan

Overview

In this project, we aim to use the BERT (Bidirectional Encoder Representations from Transformers) model to predict the sentiment of movie reviews. By analyzing the text of the reviews, the goal is to predict the sentiment (such as positive, negative, or neutral). Sentiment analysis of movie reviews is a widely studied problem in natural language processing (NLP). With the advancement of deep learning and pre-trained models like BERT, the accuracy and effectiveness of sentiment prediction have improved significantly. The purpose of this project is to build a sentiment prediction system based on BERT that can quickly and automatically evaluate audience feedback on movies.

What is the problem?

The problem we are solving is sentiment classification of movie reviews. Specifically, we aim to classify the sentiment of a review into different categories (e.g., positive, negative, or neutral). This task involves processing text data and extracting meaningful information to predict the sentiment of the review. The motivation behind solving this problem is to automate the analysis of public sentiment toward movies, which is traditionally done manually and is often subjective. By using natural language processing, we can build an efficient and objective system to handle large-scale sentiment analysis tasks.

Why is this problem interesting?

This problem is interesting because sentiment analysis has wide applications in various industries such as entertainment, marketing, and customer service. In the movie industry, understanding audience sentiment is crucial for filmmakers, marketers, and production companies. It helps them gauge the popularity of movies. Additionally, sentiment analysis can improve content recommendation systems, enhance movie ratings, and predict the success of a movie based on early reviews. More broadly, sentiment analysis helps understand public opinion, which is valuable for both consumers and producers. It allows filmmakers to improve content based on audience preferences and provides clearer emotional feedback for viewers.

What approach do you propose to tackle the problem?

To solve the problem of movie review sentiment prediction, we propose using the BERT model for sentiment analysis. BERT is a popular choice for NLP tasks because of its attention mechanism, which captures context from both sides of a sentence. This makes BERT ideal for understanding complex emotional expressions in movie reviews, as they often require deep contextual understanding.

Our approach is to fine-tune a pre-trained BERT model on a labeled movie review dataset. This fine-tuning allows the model to learn how to predict the sentiment based on the content of the review. Specifically, the process involves tokenizing the review text, passing it through the BERT model, and fine-tuning the model to predict sentiment.

What is the rationale behind the proposed approach?

BERT has already demonstrated state-of-the-art results in many NLP tasks, including sentiment analysis. BERT understands the context in a sentence and has been pre-trained on large amounts of text data, making it an ideal model for this task. In previous research, other models like logistic regression, support vector machines (SVM), and recurrent neural networks (RNN) were used for sentiment analysis. However, these methods often fail to capture long-range dependencies and complex language structures as effectively as BERT.
The key difference in our approach is the use of BERT's transfer learning capabilities. Instead of training the model from scratch, we leverage BERT's pre-trained knowledge and fine-tune it to suit the task of movie review sentiment analysis.

Key components and results

The key components of the approach include:

BERT-based model: We use a pre-trained BERT model and add a classification head (fully connected layer) on top for sentiment prediction.
Tokenization: The text data is preprocessed and tokenized using BERT's tokenization tools.
Fine-tuning: We fine-tune the pre-trained BERT model using a labeled movie review dataset.
Evaluation: We evaluate the model using accuracy, precision, recall, and F1 score metrics.
The main limitations of this approach include the high computational cost of training the BERT model and the reliance on a large labeled dataset. Additionally, the model might struggle with sarcasm or sentiment that heavily depends on context.

Experiment setup

Dataset: We will use a movie review dataset that includes labeled text reviews for sentiment classification. The dataset consists of tens of thousands of reviews to ensure a balanced distribution of sentiment categories.

Basic statistics:
Number of reviews: 50,000
Sentiment categories: 0 - negative   1 - somewhat negative   2 - neutral   3 - somewhat positive   4 - positive

Implementation:

Model: We use the BERT model (e.g., bert-base-uncased or bert-large-uncased).
Hyperparameters: Batch size is set to 16, learning rate is 2e-5, and maximum sequence length is set to 128 words.
Environment: The experiments are run on machines with GPUs (e.g., NVIDIA Tesla V100) to handle the high computational demands of training the BERT model.

Model architecture:

We start with a pre-trained BERT model and add a classification head (fully connected layer) to predict the sentiment of the reviews. During training, both the BERT layers and the classification head are fine-tuned.

Experiment results

Main results: The fine-tuned BERT model achieves an accuracy of 63% on the test set. Precision, recall, and F1 score are also high, indicating that the model performs well in sentiment classification tasks.

Model	Dataset	Accuracy	Weighted F1-Score
BERT	Cleaned	63.78%	64.00%
BERT	Original	70.03%	70.00%

Supplementary results: During the experiment, we made the following parameter choices:

Learning rate: Set to 2e-5 to ensure stable training without overfitting.

Batch size: Set to 16 due to GPU memory limitations.

Maximum sequence length: Set to 128 words to balance performance and computational efficiency.

Discussion

The results show that BERT performs exceptionally well in the sentiment analysis task for movie reviews, achieving much higher accuracy than traditional machine learning models (such as SVM and logistic regression). However, the model still faces challenges in handling sarcasm or sentiment that heavily depends on context. Future work could explore using other BERT variants (e.g., RoBERTa or DistilBERT) to improve model performance, especially in terms of inference speed.

Conclusion

We have demonstrated that the BERT model is highly effective in movie review sentiment prediction. By fine-tuning BERT, we achieved excellent results in sentiment classification, outperforming traditional methods. However, there is still room for improvement, especially in handling sarcasm and context-dependent sentiment. Future work could focus on using domain-specific models or experimenting with larger datasets.

References

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin et al., 2019): https://arxiv.org/abs/1810.04805

Transformers Library (Hugging Face): https://huggingface.co/transformers

IMDB Movie Reviews Dataset: https://ai.stanford.edu/~amaas/data/sentiment/