

KWABENA MENSAH

Assignment 5.1: Sentiment Analysis with BERT on IMDb Movie Reviews

"IMDB Dataset of 50K Movie Reviews" Dataset

Content

The dataset consists of legal sentences for your analysis.

Text preprocessing

This is a dataset for binary sentiment classification. It provides a set of 25,000 highly polar movie reviews for training and 25,000 for testing.

Observation:

Training took about 2 hours. To reduce the time I trained on a 10% sample size. See summary below

Notebook 1

Sample Size 10%

Epochs 3

Accuracy: 0.8740

Precision: 0.8880

Recall: 0.8560

F1-Score: 0.8717

Review: The movie was fantastic! The storyline was gripping and the actors did an excellent job.

Predicted Sentiment: positive

Review: I didn't enjoy the film. It felt too long and the plot was hard to follow.

Predicted Sentiment: negative

Review: It's an okay movie. Not the best I've seen, but not the worst either.

Predicted Sentiment: positive

Review: Absolutely loved it! Would definitely recommend watching.

Predicted Sentiment: positive

Conclusion on use of BERT

Deep Understanding of Context:

Review: "It's an okay movie. Not the best I've seen, but not the worst either."

Predicted Sentiment: positive

This review is nuanced. While the movie isn't portrayed as outstanding, it is also not being negatively criticized. Traditional models might struggle with such ambivalent sentiments, but BERT, due to its bidirectional attention mechanism, can understand the context better and predict that the sentiment is more positive than negative.

Handling of Complex Sentence Structures:

Review: "The movie was fantastic! The storyline was gripping and the actors did an excellent job."

Predicted Sentiment: positive

BERT's ability to analyze the relationships between different parts of a sentence allows it to capture the overall positive sentiment of this review, even though multiple positive aspects are mentioned in different parts of the sentence.

Pre-trained on a Massive Corpus:

BERT has been pre-trained on a massive amount of text (like the entire Wikipedia). This means it has seen various sentence structures, words, and contexts. Therefore, even without much fine-tuning, it can often provide reasonably accurate predictions.

Bidirectional Context:

Traditional models, like LSTM or GRU, read the text either left-to-right or right-to-left. In contrast, BERT reads text bidirectionally, meaning it understands the context from both sides of each word in a sentence. This allows it to capture the sentiment more accurately.

Less Need for Domain-specific Data:

While domain-specific data always helps, BERT's extensive pre-training means it requires less of it. You can fine-tune BERT on a relatively smaller dataset (like the IMDb dataset) and still achieve high accuracy, as seen in the provided results.

Generalization Across Various NLP Tasks:

The same BERT model can be used for multiple NLP tasks, such as sentiment analysis, question answering, and named entity recognition, with just fine-tuning. This makes it versatile and adaptable.

In summary, the ability of BERT to understand the context deeply, its extensive pre-training, and its bidirectional nature give it an edge in capturing sentiments accurately, even in complex or nuanced reviews.

Note:

In notebook 2 we increased sample size to 20% and used 5 epochs and saw some improvement.

Accuracy: 0.9130

Precision: 0.8852

Recall: 0.9529

F1-Score: 0.9178