Transformers and Sentiment Analysis

Michael H. Shimanyula

University of San Diego

Masters in Applied Artificial Intelligence

Abstract

This paper explores the application of the BERT (Bidirectional Encoder Representations from Transformers) model in sentiment analysis on a dataset of movie reviews. We preprocessed the data, trained a BERT model, and evaluated its performance in distinguishing between positive and negative sentiments.

Methodology

Data Preprocessing

The dataset comprises reviews separated by "

". We preprocessed the data by removing stop words, punctuation, and converting the text to lowercase. The sentiments were labeled as 1 for 'positive' and 0 for 'negative'.

Evaluation and Interpretation:

The dataset was split into training (80%) and validation (20%) sets. We evaluated the model's performance on the validation set using metrics such as accuracy, precision, recall, and F1 score.

Discussion:

Our findings affirm the effectiveness of BERT in sentiment analysis tasks. However, the training time was considerably long, suggesting a need for optimization or the exploration of lighter models for similar tasks.

Conclusion

This study showcased the application and effectiveness of BERT in sentiment analysis on movie reviews. The insights gained from this work can be instrumental in refining sentiment analysis models for real-world applications.

References

Onat Topal, M., Bas, A., & van Heerden, I. (2021). Exploring Transformers in Natural Language Generation: GPT, BERT, and XLNet. *arXiv e-prints*, arXiv-2102.