Application of BERT variations in sentiment analysis

## Introduction

BERT is a language model that has been widely applied in a variety of NLP tasks since its first release. Its architecture was developed based on transformer. It achieved high performance in common NLU tasks. Due to high computation cost, it was pre-trained using a huge unlabeled dataset and can be finetuned with fewer resources to fit specific downstream language tasks.

Sentiment analysis is regarded as a contextual mining task with the purpose of identify and extract subjective information in texts. One possible use case of sentiment analysis is to analyze customer feedbacks about certain product in an online retail scenario. We can use AI tools to automate the process of obtaining user preference statistics.

BERT and its variants can be used to conduct sentiment analysis tasks. This review will introduce several variants of BERT and their application in sentiment analysis. Their performance in general sentiment analysis tasks will also be compared.

## Body

Vanilla BERT

The vanilla BERT, which is the base for all subsequent variants, has two model configurations. The base version contains 12 transformer encoders while the large version doubles the number of encoders. The two tasks that were used in pre-training process are language modeling and next sentence prediction. By making no or only slight changes to the architecture, BERT can be used for different NLU tasks such as sentiment analysis, question answering and named entity recognition.

During pre-training, a CLS token is used for the task of next sentence prediction. In the case of sentiment analysis, one can simply add a classification layer on top of the CLS token output. Sentiment analysis is a supervised learning task and one should follow the suggested pipeline to train the model. To begin with, one should prepare a labelled dataset with text statement as features and their corresponding sentiment direction as label. Depending on the required precision of text polarity, one may need to adjust the number of label categories. The next main step is usually text data preprocessing. This may include a sequence of substeps such as stopwords removal, stemming, lemmatization, tokenization, and word embedding. During inference, it is important that data points should go through the same processing steps before fed into the model. As the final step in this pipeline, the processed data is used for model parameter optimization. The above discussed pipeline is a high-level solution procedure for all variants of BERT.

It has been shown that vanilla BERT can achieve a fairly high accuracy in sentiment analysis tasks with only a few epochs of training. Using TensorFlow, an accuracy of 80%-85% on the IMDB reviews dataset with 3-5 epochs of training was reported (GeeksforGeeks, 2021).

RoBERTa

RoBERTa, with nearly the same model architecture as vanilla BERT, is one well-known variants of BERT. However, its pre-training process was changed. The new pre-training phase only includes a developed language modeling task with dynamic masking technique applied. Furthermore, a much more complex dataset is used for pretraining which significantly increased the total training time. Similar to vanilla BERT, RoBERTa can be applied to solve a wide range of NLU tasks.

To fit the task of sentiment analysis, one need to add one more linear layer on top of the whole

architecture. The afore-mentioned pipeline can then be used to train a RoBERTa model.

In terms of model performance, RoBERTa is similar to vanilla BERT in most cases. In an experiment of analyzing emotions of tweets, both models got a 0.9 F1-score with only 4 epochs of training (Ludovico, 2021). RoBERTa achieved a higher precision while vanilla BERT was better in recall rate.

**Conclusion**

BERT as well as its variants have been shown to be powerful in sentiment analysis. The pre-training technique has made it possible to train a well-performing model in fairly short amount of time. BERT should be considered as an option when the task of sentiment analysis is in need of automation.

References

GeeksforGeeks. (2021). *Sentiment Classification Using BERT*. https://www.geeksforgeeks.org/sentiment-classification-using-bert

Ludovico, C. (2021). *Twitter Sentiment Analysis with BERT + RoBERTa*. https://www.kaggle.com/code/ludovicocuoghi/twitter-sentiment-analysis-with-bert-roberta