Evaluating student writing using BERT: A Transformer Based Approach

Abdul Muqtadir I222849@nu.edu.pk

May 2023

1 Abstract

Writing skills are essential tool for success and different automated tools have been designed over the years for evaluating the writings, But many fail to identify the writing structure. Identifying the writing in to its key argumentative components is considered as difficult task,, such tasks require large, pretrained, state-of-the-art models. In this Paper we have presented BERT, a transformer based model to train on the task and provide the results. By first pre-processing and encoding the dataset using BERT preprocessor and BERT encoder, then training the model on the dataset we have obtained 61 percent accuracy. The study is limited to see results of BERT model on data but in future other transformer based models and longformers can be implemented to see their performance on the data.

2 Introduction

Effective writing is an essential skill for students to acquire and at the same time it is important for educators to assess students' writing skills and provide them with targeted feedback for improvements. With advancement in Natural language processing (NLP) technology, many automated writing evaluation tools are in market. With many AI based tools in market, many fails to identify the structure of writing. In this paper, we have evaluated the Bidirectional Encoder Representations from Transformers (BERT) in evaluating students' writing. BERT is a state-of-art model that has shown good results in many tasks such as text classification, sentimental analysis. we will examine the performance of BERT on Evaluating students' writing. In this study, A dataset of students' essays has been used that contains the disclosure text and argumentative components (as labels) for each text. A BERT state-of-art model will be implemented to classify argumentative components in writing. Additionally, several different plots will be plotted to analyze model performance. In summary, This study aims to evaluate the use of BERT model in evaluating students' writing and contribute to overall discussion of the topic.

3 Background

3.1 Problem statement

currently there are many NLP based automated models for writing evaluation but they fail to evaluate structure of writing. In this study we aim to analyze the BERT transformer model for evaluating students' writing.

3.2 Objective

The objective for the study are designed considering the knowledge and requirements of study.

1. To evaluate the performance of BERT (A state-of-art transformer model) in evaluating students' writing.

3.3 Significance of study

Writing is an important skill in academic and professional life for two purposes:

1. For communicating important information.
2. For documenting technical topics

For success in many fields, including academia, business, IT industries and in many other life aspects. Effective writing skills enables students to communicate their ideas clearly, demonstrate their knowledge and express themselves. Evaluation of students' writing can help educators identify areas of strength and weakness in their students' writing skills, and provide targeted feedback for improvement. This study will contribute to current research in evaluating the students' writing.

3.4 Limitations and future work

This study is limited to analyze BERT transformer model in evaluating students' performance and in future other transformer models and longformers will be analyzed in evaluating students' writing.

4 Methodology

The purposed methodology is divided into two parts:

1. Data collection and preprocessing
2. Using pretrained BERT model and training it on our Dataset

4.1 Dataset

The dataset used in this study is from competition sponsored by the Bill and Melinda Gates Foundation. The dataset consist 144290 labelled data samples. Approximately 68600 samples are used for training, 17100 for validation and 15100 for testing.

4.2 Preprocessing and Embeddings

For this study BERT pretrained Preprocesing and encoding from TensorFlow hub are used the 768 dense representation embedding is used from the hub.

4.3 Model and Evaluation

BERT model is implemented with pretrained preprocessing and encoders. With dropout layaer to reduce overfitting and last layer as dense layer softmax. Following are the details of model.

Table 1: Model Details

Optimizer	Loss	Metrics
Adam	Sparse categorical Entropy	Accuracy

The model is trained on 5 epochs and with last epoch it gives following training numbers.

Table 2: Training details

Epoch	Loss	Accuracy
5	1.25	58

5 Results

Very first we have results of testing accuracy against the training loss. that shows how accuracy is increasing with decreasing loss. but after 5 epoch the accuracy is not increasing.

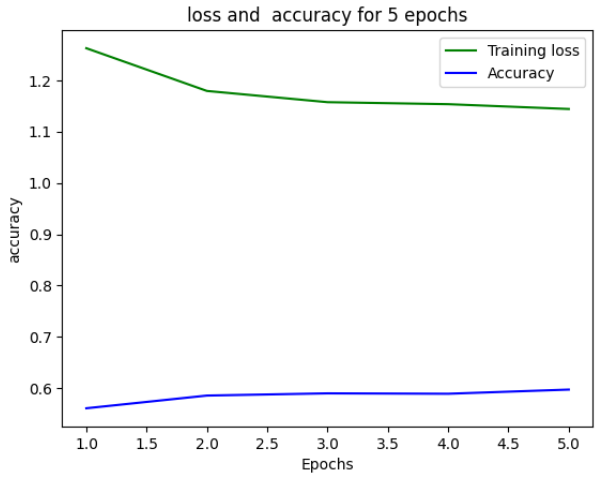

Figure 1: Example picture

Next we have validation loss against validation accuracy and this graph shows the relation validation loss and validation accuracy have.

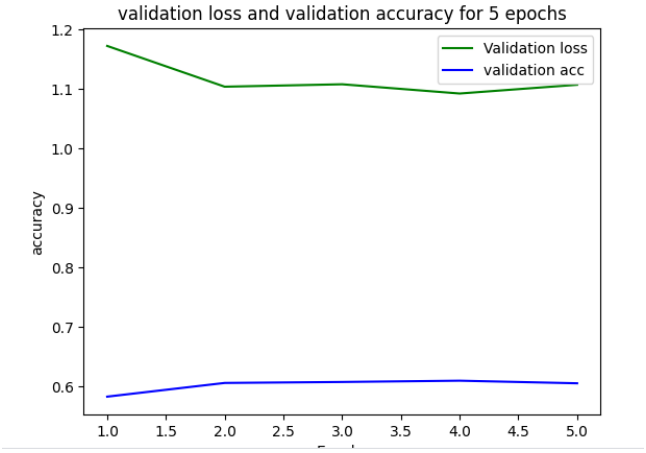

Figure 2: Example picture

Finally, we have plotted validation accuracy against testing accuracy for 5 epochs.

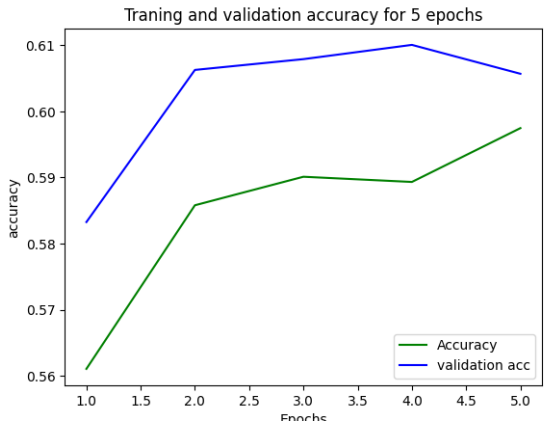

Figure 3: Example picture

6 Conclusion

In conclusion, Writing skills are essential tool for success, with different automated tools in market for automated evaluation of writing we have lot of options to select from. But they fail to identify the writing structure and Identifying the writing in to its key argumentative components. Therfore In this study, BERT, a transformer based model is implemented in evaluating the students' writing and in classifying argumentative components of essays. Accuracy of 61 percent is achieved. Future work includes other transformer based models and longformers can be implemented to see their performance on the data.

7 References

1. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova.

2. "Evaluating the Quality of Texts Written by Language Learners Using BERT" by Anastassia Loukina, Jiahong Yuan, Detmar Meurers, and Lei Chen.

3. "Evaluating and Improving the Use of BERT for Automated Essay Scoring" by Xiaofei Lu, Sowmya Vajjala, and Edward C. Council

4. "BERT: Fine-Tuning by Freezing Weights and Out-Of-Domain Data" by Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer