Tales Araujo Leonidas
Professor Patricia McManus
ITAI 2376: Deep Learning
April 14, 2024

Fine-Tuning BERT: Reflective Journal

Introduction

In this journal, I report my experience performing Lab 5 of the AWS Academy course “*Application of Deep Learning to Text and Image Data*”, Module 2. The Lab introduces BERT, “which stands for Bidirectional Encoder Representations from Transformers, [...] is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers” (Devlin, Jacob et al), and proposes a practical fine-tuning on the model to classify product reviews. The goal is to understand this revolutionary model and acquire hands-on experience in fine-tuning it.

Experience Description

After loading the dataset, I performed feature engineering by dropping all the empty fields of the “*reviewText*” column and selecting the first 2000 rows to work with to reduce training time. Working with the columns “*reviewText*” and “*isPositive*”, I split the dataset into train and validation sets with a proportion of 90/10, before tokenizing (breaking sentences into small blocks) and encoding (converting tokens into numbers) the train and validation splits, as pictured on the code snippet below.

```
tokenizer = DistilBertTokenizerFast.from_pretrained('distilbert-base-uncased')

train_encodings = tokenizer(train_texts,
                            truncation=True,
                            padding=True)
val_encodings = tokenizer(val_texts,
                          truncation=True,
                          padding=True)
```

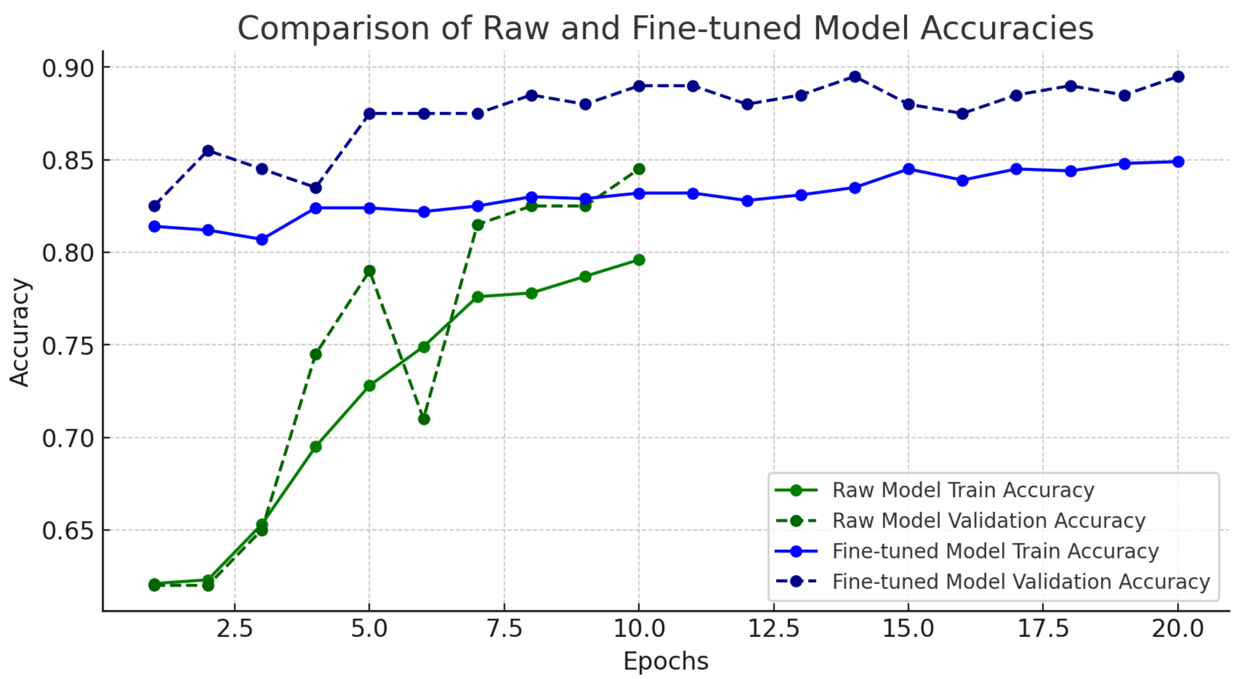
Since BERT is a computationally heavy model, the Lab used its light version called DistilBert, which was designed to retain most of the original model's performance while being smaller and faster. Additionally, I created a new “*ReviewDataset*” class to use the training and validation encoding-label pairs with it, which can later be used to create a data loader.

To help accelerate the training process, I froze all weights until the last classification layer, as suggested by AWS Academy.

Upon defining the accuracy function, I created the training and validation loop to train and test the model, and after some experimentation, I was satisfied with my hyperparameter setting for fine-tuning the model on the proposed dataset.

Personal Reflection

This Lab was extremely insightful and provided a great opportunity to fine-tune a large pre-trained model by following a simple script of steps and continuous testing. The graph below, generated by Chat GPT4, showcases the performance of the BERT model on the uploaded dataset in green, and the performance of the fine-tuned model in blue.



The graph underscores the practical benefits of fine-tuning large pre-trained models like BERT, providing enhanced predictive performance on specialized tasks.

Improvements and Learning

The performance improvement from the raw to the fine-tuned model demonstrates the effectiveness of the fine-tuning process. The fine-tuned model achieves higher and more stable accuracies both in the training and validation phases, indicating better generalization and understanding of the specific characteristics of the uploaded dataset. Moreover, another great learning is that DistilBert is a light version of BERT, which is great since it enables DistilBert implementation in various scenarios.

Conclusion

In conclusion, this lab deepened my understanding of BERT and its capabilities but also highlighted the practical implications and steps of fine-tuning such models on specific tasks. The ability to improve a model's accuracy and efficiency through fine-tuning proves invaluable, particularly in fields requiring nuanced text interpretation. This hands-on experience has equipped me with the knowledge to leverage state-of-the-art AI models effectively, ensuring their applicability across varied scenarios and datasets.

Works Cited

AWS Academy. “Application of Deep Learning to Text and Image Data: Module 2.” AWS Academy, 2024, https://awsacademy.instructure.com/login/canvas

Devlin, Jacob et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” North American Chapter of the Association for Computational Linguistics (2019).