Transfer Learning Approach For Sentiment Analysis

Shivani Gowda KS
Loyola Marymouny University
sks@lion.lmu.edu

Abstract

Previously, we did sentiment analysis of twitter data using SVM in our assignments, where we did hyperparameter selection and tried different kernels to improve the classification accuracy. During this class project, instead of using SVM, we will explore transfer learning approaches. Specifically, we finetune BERT language model on the same dataset as before and then compare its accuracy.

1 Transfer Learning

Transfer learning in machine learning is when the knowledge learned from previous training is used to help perform a new task. Training new ML models from scratch can be resource-intensive, so transfer learning saves both resources and time. Moreover, Transfer learning models are more generalized which means that are models are not being rigidly tied to a training data sets. Models developed in this way can be utilised in changing conditions and with different datasets. Specifically BERT(Bidirectional Encoder Representations from Transformers) because of its bidirectionally (Batra et al., 2021) (Hoang et al., 2019).

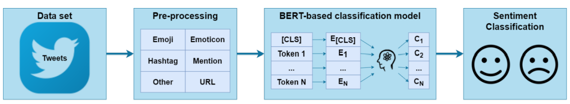

Figure 1: Overview of twitter sentiment classification using BERT model (Pota et al., 2021)

A multi layer transformer model which can understand the meaning of each word based on context both to the right and to the left of the word. Its two main techniques, Mask Language Model (MLM) and Next Sentence Prediction (NSP). Where MLM trains a model to predict a random sample of input tokens that have been replaced by a [MASK] placeholder in a multi-class setting over the entire vocabulary and in NSP the model is trained to predict whether the observed document segments come from the same or distinct documents.

2 Experimental setup

The datasest that we will be using for this experiment contains two Fields, text and labels. We used dataset from two different surces. The first data set[1] is the same data that we used for SVM which contains 620 texts and the second data is from Kaggle [2] It contains 16,00,000 texts, but we will be using only 1,000 texts from this.But in this report we will be concentrating mainly on the first data set.

We used Google Colab notebook along with GPU backend to perform the following steps 1. Installing Huggingface's transformers library

2. Familiarizing the use of Transformers Library

3. Loading the Dataset

4. Preprocessing to make it compatible with BERT as a parent model.

5. Finetune the pretrained model with suitable hyper parameters

6. Make predictions

7. Compare the results(accuracy)

3 Data prepossessing

The data set has text field of various length. To maintain equal length across the batch.

The length of these sentences is padded to 128 tokens, to make it of equal length. Furthermore, for the text sequence we add two special tokens, namely we prepend CLS and append SEP. We tokenize the sentence with Byte Pair Encoding (BPE) which is implemented in SentencePiece. As we

[1] https://raw.githubusercontent.com/sgowdaks/sentiment-analysis-using-BERT/master/Tscfilebert.txt

[2] https://www.kaggle.com/kazanova/sentiment140

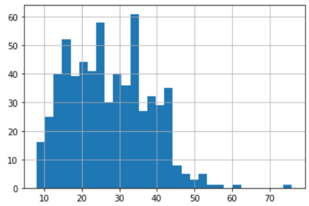

Figure 2: tweets before padding

don't need nor want the model to attend to the padded positions in sequences, we use attention mask to exclude padded positions. Finally, all the prepared sequences are converted to torch tensors. We divide the data set into training and validation set by 80:20 ratio.

4 Models and Results

BERT model is instantiated from Huggingface's transformers library's (Wolf et al., 2020) BertForSequenceClassification. We used `bert-base-uncased` model, where the encoder weights are copied from the Bert-base-uncased model. The number of labels is set to 2, as we have 2 classes in labels field. A newly added classifier head on top of encoder is randomly initialized. We use ADAM optimizer with learning rate = 2e-5 and trained with batch size of 32 sentences. The hidden_dropout_prob and attention_dropout_prob is set to 0.1. The accuracy obtained with change in epochs is given below.

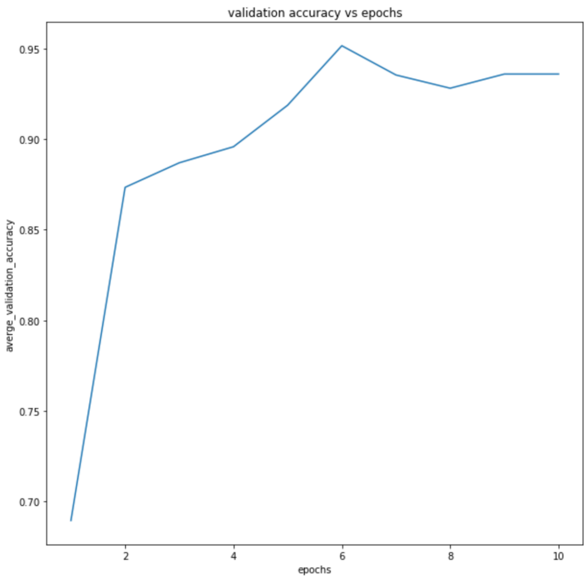

Figure 3: validation accuracy vs epochs

The maximum accuracy obtained is 94%, we cannot depend only on accuracy for overall performance as there may be chances for over-fitting. We will be calculating validation loss and training loss across the model using cross entropy loss. The training loss and validations loss is listed int the given table.

epochs	Validation Loss	Training Loss
1	0.12	0.61
2	0.08	0.44
3	0.06	0.25
4	0.05	0.11
5	0.06	0.06
6	0.04	0.03
7	0.05	0.03
8	0.06	0.01
9	0.06	0.0
10	0.06	0.00

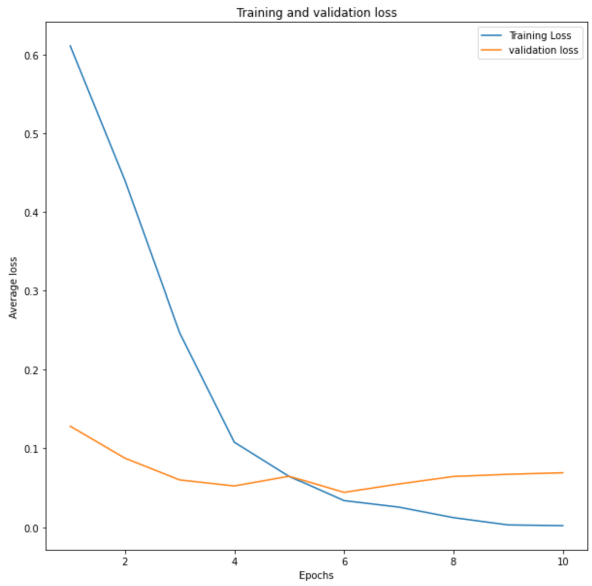

Figure 4: Training and validation loss

The results indicate that after 6th epoch the validation loss decreases, training loss decreases which indicates that it is overfiting and at the same epoch the maximum accuracy is obtained.

The accuracy obtained from Support vector machines, with RBF kernal was around 84%. Now with BERT we are able to reach upto 94% on the same datasets.

Acknowledgments

I am grateful to Prof. Mandy for Machine learning classes, feedbacks and opportunity to learn through projects and assignments. I thank Thamme Gowda for fruitful discussions, Google for providing free GPU's, HuggingFace for free API's with documentation.

References

Himanshu Batra, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal. 2021. Bert based sentiment analysis: A software engineering perspective. *arXiv preprint arXiv:2106.02581*.

Mickel Hoang, Oskar Alija Bihorac, and Jacobo Rouces. 2019. Aspect-based sentiment analysis using bert. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 187–196.

Marco Pota, Mirko Ventura, Rosario Catelli, and Massimo Esposito. 2021. An effective bert-based pipeline for twitter sentiment analysis: a case study in italian. *Sensors*, 21(1):133.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.