Comparing Deep Learning Transformers To Naive Bayes and Random Forest Models in the Context of Emotion Detection

Graduate Project Proposal

Richard Hoehn*

Middle Tennessee State University

CSCI 6350

Prof. Dr. Sal Barbosa

February 29, 2024

*Electronic address: rhoehn@mtmail.mtsu.edu; corresponding author

1 Introduction

In the CSCI 6350 graduate project[1] the focus of research and student learning is on implementing a Deep Learning Transformer model to perform Emotion Detection (ED) and comparing it's prediction accuracy to previous work by Hoehn's[3] paper that employed Naive Bayes and Random Forest ML models for ED analysis.

2 Background

Emotion detection (ED) has gained significant interest and investment[5] in recent years as a means to understand human behavior and improve communication effectiveness[3]. By analyzing text, speech or facial expressions, these detection processes and consequently ML models try to accurately identify and classify emotions expressed by individuals or groups, often even in real-time. By use of ED ML models, which include natural language processing (NLP) that employ deep learning, applications with the right amount of data can detect emotions with high precision[1]. These ED systems find applications in diverse fields, such as sentiment analysis, customer feedback analysis, voting and news stance prediction, and mental health monitoring, which all aim at contributing to a deeper understanding of human emotions and their impact on various aspects of life.

3 Motivation & Specific Aims

The motivation for the this CSCI 6350 graduate project is two fold. **Firstly**, building upon previous work that Hoehn[3] completed for the qualifying exam of the COMS[2] PhD program is a superb expansion of that research focused on ML (Naive Bayes & Random Forest) models. **Secondly**, building on previous projects in CSCI 6350 allows for continuation of NLP[3] branching into an interesting area of the use of Deep Learning for NLP tasks.

In order to complete the project, data from Hoehn's[3] paper that uses a labeled emotions will be the basis for comparison. The dataset[4] itself is from 2021 Twitter, which is basically a collection of tweets annotated with the emotions. There are 13 different labeled emotion and corresponding text with over 40,000 rows in a `csv` file.

The aim for this Graduate Project using a Transformer for Emotion Detection has three (3) main goals.

- Build, Train, and Validate a Transformer for Emotion Detection
- Using Plots display the learning rates and validation results in graphical forms
- Evaluate the Transformer's results with that of Hoehn's[3] Naive Bayes & Random Forest

3.1 Build, Train & Validate Transformer for Emotion Detection

More details here...

[1]The purpose of the graduate project is to afford students the opportunity to research modern technologies in use in the natural language processing (NLP) field.
[2]COMS - Computational Data & Science Department
[3]NLP - Natural Language Processing

3.2 Plotting of Learning Rates, Validation, and Accuracy

More details here...

3.3 Transformer vs. Naive Bayes & Random Forest Prediction Accuracy

More details here...

4 Proposed Methods

There are three main components to achieve my aims from above. The first will be to pre-process the Dataset in order to clean the tweets and remove any irrelevant characters or words. Next will be to tokenize the data, which is to convert the tweets into a format that the model can understand. I envision I will be using the BERT-Tokenization[2].

Once tokenization has taking place the transformer, most likely BERT[2] since it is known for its effectiveness in understanding the context of human language.

The the transform in place I place to use 80 / 20% split of the 40,000 labeled records into training, validation, and test sets.

Train the model on the training set, where it learns to map the input text to the correct emotion labels. Use back propagation and gradient descent to optimize the model's parameters.

Assess the model's accuracy and F1 score on the validation set to gauge its effectiveness in emotion prediction. Make necessary adjustments to the model's configuration or training approach based on these performance metrics.

And finally compare the transformer's prediction results to that of the Naive Bayes & Random Forest results from Hoehn's[3] paper.

5 Expected Results

In the context of this Graduate Project in Emotion Detection from text data, it is anticipated that a Deep Learning Transformer model will perform better in predicting emotions compared to a Naive Bayes & Random Forest Machine Learning (ML) approaches from Hoehn's[3] paper.

This expectation is based on the notion that transformers, which are designed to capture contextual nuances of human language through their attention layers, are more adept in recognizing intricacies of spoken and written language such as sarcasm and idioms. Consequently, when the Transformer is applied to a dataset of 40,000 labeled records for emotion detection, it is likely to achieve higher accuracy rates that the Naive Bayes & Random Forest approaches, demonstrating its robustness and effectiveness in capturing the emotions conveyed in text, beyond the capabilities of traditional ML classifiers.

References

[1] Mengnan Chen et al. "Research on Emotion Recognition for Online Learning in a Novel Computing Model". In: *Applied Sciences* 12.9 (2022). ISSN: 2076-3417. DOI: 10.3390/app12094236. URL: https://www.mdpi.com/2076-3417/12/9/4236.

[2] Jacob Devlin et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *CoRR* abs/1810.04805 (2018). arXiv: 1810.04805. URL: http://arxiv.org/abs/1810.04805.

[3] Richard Hoehn and Jaishree Ranganathan. *Improving Emotion Detection through Translation of Text to ML Models Trained in Different Languages*. Version 1.0.1. https://github.com/richardhoehn/mtsu.coms.qualifying-exam. Aug. 2023. URL: https://github.com/richardhoehn/mtsu.coms.qualifying-exam.

[4] Kaggle and data.world. *Emotion Detection from Text*. 2021. URL: https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text.

[5] Iqbal H. Sarker. "Machine Learning: Algorithms, Real-World Applications and Research Directions". In: *SN Computer Science* 2.3 (Mar. 2021), p. 160. ISSN: 2661-8907. DOI: 10.1007/s42979-021-00592-x. URL: https://doi.org/10.1007/s42979-021-00592-x.