Naren S. Chittem

Mind Brain Behavior Research Program

Dr. Ron Mallon

December 22th, 2022

Introduction

Social media has become a significant part of the average person's day, serving as their news source and their way of communication with the world. Social media sites like Twitter have amplified the layman's voice, allowing others to connect and respond to their content and messages. As social media has become more popular and used, it also becomes more important to be able to analyze said information. Due to the large number of users on these sites, a plethora of data and information is available which holds a sizable section of the public's opinion. Analyzing vast amounts of these messages to understand the overall feeling towards a specific event, person, or action is critical in developing strategies. This project used Bidirectional Encoder Representations from Transformers (BERT), specifically the Bert-cased pre-trained model that serves as the underlying neural network architecture. It was then tailored for sentiment analysis in tweets by adding additional dropout and linear layers. This creates a new neural network labeled the MBB model within the code, which utilizes the power of BERT for sentiment classification. This new neural network was trained on preprocessed Twitter sentiment analysis data in order to classify tweets as neutral, positive, or negative. This trained model was then applied to data surrounding tweets on the Unite the Right rally in Charlottesville and tweets about 2020 presidential candidates Donald Trump and Joe Biden.

The Internet has become the primary form of communication for various fringe or counter-public groups; specifically, it benefits radical right-wing groups and other various members of extreme ideologies (Xu, 2020). These types of hateful groups often use popular social media sites to recruit and spread their message, making what was once radical more mainstream (Törnberg & Törnberg, 2021). Not only that but new nationalist propaganda and ideology have been spreading like wildfire on the Internet. A connected militant diaspora can now directly contribute to nationalist sentiments or to "long-distance nationalism," where members of these social media groups can form a national consciousness without even meeting face-to-face (Lamensch, 2021). The flexibility afforded by the internet has led to increased diversification of nationalist propaganda and polarization. The internet connects so many different individuals that it creates different nationalisms and radical ideologies, thereby creating more unpredictability. All this increased diversity also leads to more polarization of national images allowing for the mobilization of more extreme ideologies (Mihej, 2020). The influence of the internet on global and national politics makes the importance of a sophisticated sentiment analysis program clear: detecting negative sentiments against certain groups or events could go a long way in helping preemptively ban or monitor toxic accounts.

Model

Before the introduction of the Transformer model, there was a heavy reliance on recurrent neural networks (RNNs) or convolutional neural networks (CNNs). Slow training speeds plague these models and struggle to capture long-range dependencies; in stark contrast, the Transformer models' attention-based architecture eliminated the need for recurrence or convulsions, which in turn allows for a much better ability to capture diverse dependencies and allows for efficient computations. The model has a self-attention mechanism that allows the model to weigh the importance of independent positions in a given input sequence while generating the output. This model consists of encoders and decoders in stacks of identical layers

(Vaswani et al., 2017). Based on this new transformer technology, the BERT model is comprised of stacked self-attention and feed-forward layers. Self-attention mechanisms enable the model to assign different weights to different words in a sentence, capturing their contextual dependencies effectively. BERT is designed to pre-train deep bidirectional representations from unlabeled text by joint conditioning on both left and right context in all layers, analyzing the data from both sides and allowing it to develop deeper contextual understandings. The training for BERT is unique as there was masked language modeling where some words were masked, so the algorithm had to predict the missing word, helping develop contextual learning. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create advanced natural language processing models (Devlin et al., 2018).

Tokenization is a crucial step in NLP that involves splitting the input text into individual tokens or subwords. In BERT, tokenization is performed using WordPiece tokenization, which splits words into subword units. Each token is then mapped to a corresponding index in the model's vocabulary. Additionally, special tokens, such as [CLS] and [SEP], are added to the input to provide context and structure. Attention masks are used to indicate which tokens should be attended to by the model and which ones should be ignored. In BERT, attention masks are created by assigning a value of 1 to input tokens and a value of 0 to padding tokens. This allows the model to focus only on the relevant tokens during computation and ignore the padded tokens (Peters et al., 2018). Tokenization and attention masks allow the neural network to 'read in' natural language, which is why these steps are included in the code. Our MBB model is a variant of BERT designed explicitly for sentiment analysis on tweets which also utilizes the AdamW optimizer to help with computations (Loshchilov & Hutter, 2019). It was trained on 3,000 preprocessed tweets classified as neutral, positive, or negative. After ten epochs of training, the training accuracy was nearly 100% (see Figure 1). The model also boasts a 93.69% accuracy on the test data indicating that it may be able to reveal some exciting patterns in other datasets. Figure 2 shows just how accurate this model is through a confusion matrix which also shows that the mode seems to have the most trouble with classifying tweets as neutral (see Table 1 for further evidence).

Figure 1

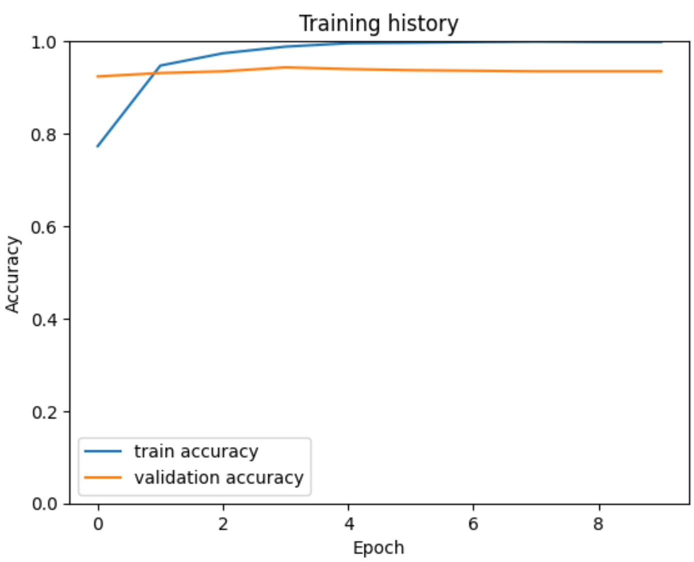


Figure 2

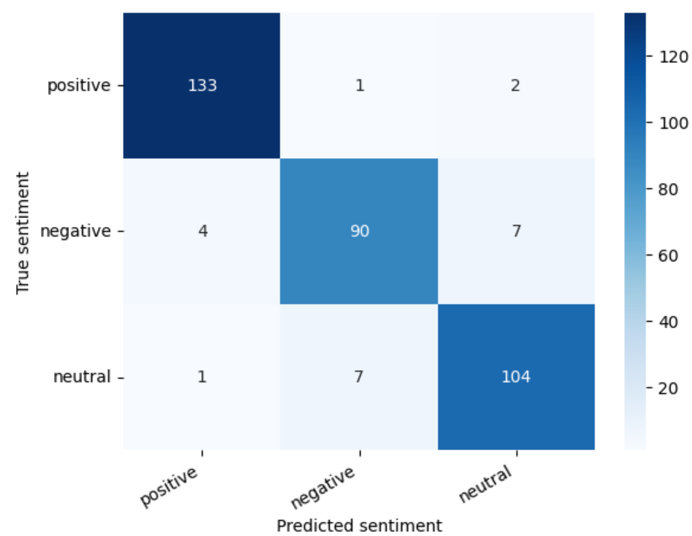


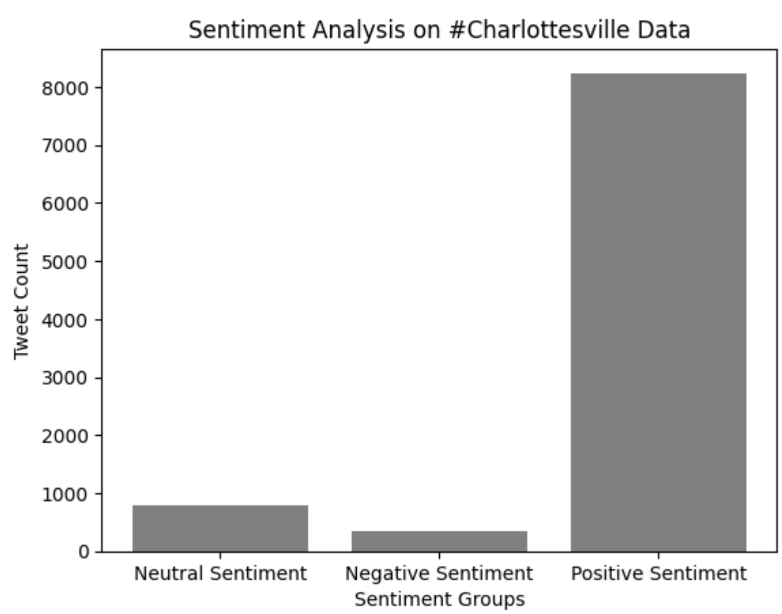
Table 1

	Precision	Recall	F1-Score	Support
Positive	0.96	0.98	0.97	136
Negative	0.92	0.89	0.90	101
Neutral	0.92	0.93	0.92	112
Accuracy			0.94	349
Macro Average	0.93	0.93	0.93	349
Weighted Average	0.94	0.94	0.94	349

Application: Charlottesville

The first application scenario of the MBB model was to look at the Unite the Right rally in Charlottesville. The model was applied to a dataset that consisted of reactions to the rally days after the event after the tokenization of the tweet data and the application of attention masks. Analysis shows that an overwhelming number of tweets surrounding the event had a positive sentiment (see Figure 3). The positive sentiment here was not in support of the rally but rather a positive sentiment in terms of the need for change and condemning ‘nazism’ and ‘white supremacy’ when looking at the classified positive tweets. Here we see why the context that the BERT model can achieve is clear; it is able to identify negative words such as ‘nazi’ and, through its fine turning, was able to realize the condemnation of that is a positive tweet. Looking through this dataset, it is clear that most tweets condemned the event. A fundamental problem with the model that this dataset exposed was that the model was not trained to look at retweets, meaning it would look at retweets as original tweets without the context of what they were responding to, allowing it to be misled.

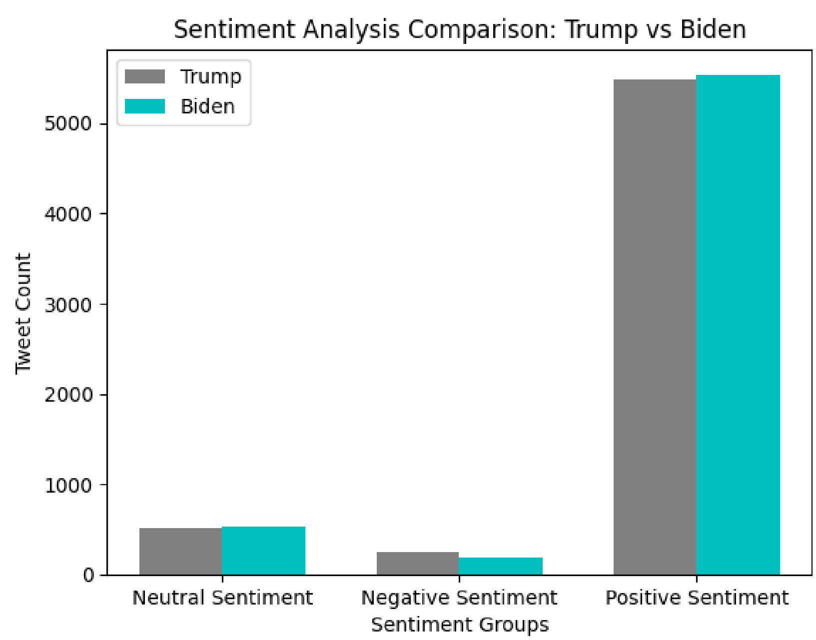
Figure 3



Application: 2020 Election

The second application of the MBB model looked at data surrounding the 2020 US presidential election, which was split between the candidate by #Trump and #Biden. The model was applied to the dataset after tokenizing the tweet data and applying attention masks. Analysis shows that Biden had slightly more Positive and Neutral tweets about him while having fewer Negative tweets while looking at the same number of overall random tweets (see Figure 4). This future application of something like this is evident where a more sophisticated model trained on more area or platform-specific data may help predict election outcomes. BERT again proved its usefulness here by seeing condemnation of one candidate, for example, using #VoteThemAllOut or #NotMyPresident, as a positive for the other. Also, because the model uses tokenization, it is able to make connections and patterns on these hashtags as obviously, they are consistent, and so is their ID within its library.

Figure 4



Limitations

The most significant limitation of this model and its application, besides its simplicity and lack of extensive training data, is the lack of memory and advanced GPUs. These limitations had a direct impact on the training process and the ability to handle more extensive and more complex datasets. Memory constraints refer to the limited amount of memory available to store and process the data during training. Neural network models, especially those based on transformers like BERT, require significant memory to accommodate the model parameters, intermediate computations, and data batches. Insufficient memory can lead to out-of-memory errors or significantly slower training due to frequent data transfers between GPU and system memory. Additionally, GPU constraints also played a role in limiting the model's performance and the amount of data it could handle. GPUs are essential for accelerating deep learning computations and are particularly useful for training large models like BERT. However, GPUs have their own memory limitations, and training complex models with large datasets can quickly exhaust the available GPU memory. These memory and GPU constraints resulted in several challenges and trade-offs during the model development process, for example, I had to opt for the 12 Transformer Bert Model rather than the 24 Transformer version. It often became necessary to reduce the dataset size, simplify the data, or use alternative strategies to overcome memory limitations; the most explicit example of this in the code is the use of subsets for the Biden and Trump data, as the original dataset was too large to process. In some cases, datasets with more significant complexities had to be abandoned in favor of smaller or simpler datasets that could fit within the memory and GPU constraints. These limitations have implications for the model's performance and generalizability. With limited training data and the inability to handle more complex datasets, the model may struggle to capture the full range of linguistic patterns and nuances present in real-world data. This can lead to reduced accuracy, reliability, and overall effectiveness in sentiment analysis tasks.

References

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Google AI Language. *ArXiv*. /abs/1810.04805

Lamensch, M. (2021). As Technology Evolves, So Does the Nature of Nationalism. Centre for International Governance Innovation. https://www.cigionline.org/articles/as-technology-evolves-so-does-the-nature-of-nationalism/

Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization. University of Freiburg. *ArXiv*. /abs/1711.05101

Mihej, S. & Jiménez-Martínez, C. (2020). Digital nationalism: Understanding the role of digital media in the rise of ‘new’ nationalism. Journal of the Associations for the Study of Ethnicity and Nationalism. Volume 27, Issues. Pp. 331-346. https://doi.org/10.1111/nana.12685

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. The University of Washington. *ArXiv*. /abs/1802.05365

Törnberg, P. & Törnberg, A. (2022). Inside a White Power echo chamber: Why fringe digital spaces are polarizing politics. New Media & Society. 24(9), 1311-1327. https://doi.org/10.1080/1369118X.2021.1880704

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2020). How to Fin-Tune BERT for Text Classification? Fudan University School of Computer Science. https://arxiv.org/pdf/1905.05583.pdf

Valkov, V. (2020). Sentiment Analysis with BERT and Transformers by Hugging Face using PyTorch and Python. Curiously. https://curiousily.com/posts/sentiment-analysis-with-bert-and-hugging-face-using-pytorch-and-python/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. Google Brain, Google Research, & The University of Toronto. *ArXiv*. /abs/1706.03762

Xu, W. W. (2020). Mapping Connective Actions in the Global Alt-Right and Antifa Counterpublics. International Journal of Communication. 14, 1070-1091. https://ijoc.org/index.php/ijoc/article/view/11978

Appendix

GitHub Repository with the Code:

Chittem, N. (2023). GitHub. https://github.com/NarenChittem/MBBResearch

Data and Code Sources:

Hui, M. (2020). US Election 2020 Tweets Oct 15th 2020 - Nov 8th 2020, 1.72M Tweets. Kaggle. https://www.kaggle.com/datasets/manchunhui/us-election-2020-tweets

La, V. (2017). #Charlottesville on Twitter. Kaggle. https://www.kaggle.com/datasets/vincela9/charlottesville-on-twitter/versions/1

Valkov, V. (2020). Sentiment Analysis with BERT and Transformers by Hugging Face using PyTorch and Python. Curiously. https://curiousily.com/posts/sentiment-analysis-with-bert-and-hugging-face-using-pytorch-and-python/

Yadav, S. (2017). Pre-processed Twitter tweets - The dataset is divided into positive, negative and neutral pre-processed tweets. Kaggle. https://www.kaggle.com/datasets/shashank1558/preprocessed-twitter-tweets?resource=download