Module 5: Portfolio Milestone

NLP Chatbot

Ryan Thompson

Colorado State University - Global

CSC 525

Dr. Banerjee

13 May 2025

Model Training Milestone Report

For this milestone, I trained a text classification model designed to determine the sentiment of a sentence by classifying it as positive, negative, or neutral. The dataset used contained approximately 20,000 labeled sentences sourced from online reviews. To expand the dataset and improve model robustness, I applied several data augmentation techniques, including synonym replacement, random word insertion, and random deletions. These methods helped increase linguistic variety without changing the underlying meaning of the sentences, ultimately expanding the dataset to around 80,000 examples.

The model I used was a fine-tuned version of BERT, specifically the bert-base-uncased model available through Hugging Face's Transformers library. BERT, a transformer-based architecture developed by Google, has been shown to achieve state of the art results in various NLP tasks due to its ability to understand context in a bidirectional manner. I fine tuned the base model by adding a classification head consisting of a fully connected layer followed by a softmax activation function to output probabilities over the three sentiment classes. The implementation was done using PyTorch and the Hugging Face Trainer API, which helped streamline the training and evaluation processes.

The dataset combined public product reviews with sentiment labeled tweets, particularly from the Sentiment140 dataset. After data augmentation, the expanded dataset was split into training and validation subsets, with 80% of the data used for training and 20% reserved for validation. Each sentence was tokenized using BERT’s tokenizer. The sentiment labels were mapped to integer classes for training compatibility.

For the training process, I used a batch size of 16 and a learning rate of 2e-5. The optimizer selected was AdamW, with a weight decay of 0.01 to help prevent overfitting. I also included 500 warmup steps to allow the model to stabilize early in training. Training was conducted for three epochs. These settings were chosen based on common defaults recommended for fine-tuning transformer models.

The initial results were promising. The model achieved a training accuracy of ~89% and a validation accuracy of ~81%. The cross-entropy loss decreased steadily across epochs, indicating effective learning. Although the validation performance was solid, I noticed a slight gap between the training and validation accuracy, suggesting some overfitting. In future iterations, I plan to address this by incorporating early stopping mechanisms and dropout regularization. I also intend to experiment with different learning rate schedules and longer training durations to further improve the model's performance.

In summary, the training experience was productive and provided valuable insights into fine-tuning pre-trained models for sentiment classification. The model's current performance shows it generalizes well enough for practical use, and the next steps will involve more detailed evaluation using precision, recall, and F1 score metrics on a separate test set.