Jade Sanchez

2376 ITAI

2/26/2025

Reflective Journal

Lab 05

Learning Insights

The experiments in this lab focused on modifying BERT (Bidirectional Encoder Representations from Transformers) through its distilled version named DistilBERT for performing sentiment analysis on Amazon product reviews. Computing transformer-based models proved to be the main focus of this research lab which concentrated on text information processing. Among contemporary machine learning concepts I learned about were tokenization and attention mechanisms together with model fine-tuning. Tokenization remains vital in model processing since the developed text must first transform raw text into a format that the model understands.

I found it most significant to learn that BERT along with other large language models can receive specialized training for carrying out sentiment analysis tasks. Working with large pre-trained models requires freezing model layers because it creates crucial performance benefits for the process. The freezing process speeds up training parameters because it reduces the number of values which need changes. The practical transformer implementation together with direct exploration of its structure helped me grasp the core principles of transfer learning better.

During this exercise I deepened my knowledge about accuracy metrics within deep learning models and their calculation methods. I experienced direct observation of how different learning rate parameters together with batch size values influence performance from my model. The process heavily relied on Hugging Face transformers library which gave users an easy-to-use API interface to run DistilBERT and other state-of-the-art models.

Challenges and Struggles

The main issue I encountered during this laboratory work was trying to work with memory-limited BERT models. The memory capacity of my Colab instance reached its maximum at first so I needed to decrease the batch size followed by kernel restarts to proceed. The process emphasized how transformer models require substantial computing power especially during operations involving extensive datasets or extended sequences.

The training and validation loops that work specifically within transformer models presented themselves as a difficult concept to master. Special attention_mask input treatment was needed during the loop because it specifies which tokens the model should focus on. The model learning required the correct handling of this information since it played a vital role in its proper functioning.

My solution to address these issues involved keeping the progress slow and steady. I lowered the number of examples to 2000 to reduce memory requirements so I could concentrate on mastering the fine-tuning methods. Through my study of library documentation I gained better familiarity with the training procedure within the Hugging Face library. The assessment of loss and accuracy during training intensified through monitoring the print statements which resulted in smoother diagnosis of potential problems.

Personal Growth

My knowledge about deep learning has dramatically improved through this practical research work. Traditional machine learning models constituted most of my work experience before this project and deep learning covered only a small fragment of my knowledge base. Studying DistilBERT's fine-tuning process has enhanced my knowledge about transfer learning techniques and their practical usages. Pretrained models enable researchers to perform most necessary work prior to model integration since dataset adaptation remains the key task.

I found it most unexpected that transformer architecture with BERT could perform sentiment analysis with basic training when existing models are optimized properly. The ease of working with pre-trained models combined with their accessibility made fine-tuning technology for new problems turn out to be simple.

The laboratory work led me to become more interested in NLP tasks while watching how BERT-type models evolve the understanding of natural language. These models demonstrate practical applicability in different disciplines including customer service operations while also performing social media assessment and automated content screening. The possibility of these models to advance computational interactions between humans and computers excites me because it matches my interest in AI and machine learning.

Critical Reflection

The major change I would implement while repeating this lab would be devoting more attention to optimal model hyperparameter settings. I believe the model could benefit from testing multiple optimizers including Adam and AdamW along with varying epochs between adjustments of batch size and learning rate to prevent memory problems.

I plan to research approaches that manage class imbalance because such techniques will help improve model accuracy for unbalanced data sets. Knowledge about oversampling and class weights techniques would help in situations where particular classes like negative reviews become underrepresented.

The lab experience has made me reflect on the ethical issues surrounding the deployment of large language models when they operate in decision systems that analyze sentiment such as review evaluations. My research will focus on studying inherited biases in large models from their training datasets to establish methods that build ethical AI systems.

The training experience demonstrated the need to analyze both technological and moral aspects when deploying AI models in actual deployment scenarios. The optimization of BERT models demands both robust expertise in machine learning principles together with knowledge about potential effects from deploying such systems across different programs. The practical experience linked fundamental theoretical concepts to real-world industrial needs and this knowledge base will boost my career in machine learning and AI.