**Fine-Tuning BERT for News Topic Classification using AG News Dataset**

**Name:** Aishwarya Sukumaran (22058088)  
**Module:** 7PAM2015 – Large Language Models (Github - [link](https://github.com/17031910/Research-Assignment/blob/main/Ass_2_final.ipynb))

**Introduction**

Large Language Models (LLMs) like BERT are trained to understand language in a smart way. Unlike old models, BERT looks at the whole sentence instead of reading one word at a time. This helps it understand the meaning better. In this project, I fine-tuned a BERT model to do a classification task. The aim was to predict the topic of a news article using its title and description. I used the AG News dataset, which has four topics: World, Sports, Business, and Science/Technology.

**Literature Review**

BERT, introduced by Devlin et al. (2019), changed natural language processing by understanding the context of words better through bidirectional training. Many researchers have since fine-tuned BERT for tasks like text classification, achieving strong results. Recent studies (Yang et al., 2024; Biswas & Gil, 2023; Wu, Ye & Manoharan, 2023) have applied BERT to multi-class classification with improvements in accuracy and efficiency.

**Dataset Used**

I used the balanced AG News dataset with 4 topic classes (World, Sports, Business, Science/Tech), containing about 120,000 training and 7,600 test samples. Each sample includes a news title and short description.

**Exploratory Data Analysis (EDA)**

To understand the data better, I checked how many samples each class had. The dataset is balanced, with about 30,000 samples in each topic.

|  |  |  |
| --- | --- | --- |
| **Class Label** | **Topic** | **Count** |
| 0 | World | ~30,000 |
| 1 | Sports | ~30,000 |
| 2 | Business | ~30,000 |
| 3 | Science/Tech | ~30,000 |

**Table 1**: The AG News dataset has four balanced classes: World, Sports, Business, and Science/Tech — each with about 30,000 samples.

A graph of a training set label distribution

AI-generated content may be incorrect.

**Figure 1: Class Distribution of AG News Training Data**

This Figure 1 shows the number of training samples in each of the four classes: World, Sports, Business, and Science/Technology. Each class has around 30,000 samples, indicating the dataset is balanced and suitable for training a classification model without introducing bias.

**Methodology**

**Data Preparation:**

I combined the title and description into one text. I used BertTokenizerFast to tokenize the text and padded or trimmed all inputs to 128 tokens. Then, I converted the dataset into PyTorch format for training.

**Model Training:**

I used the pre-trained bert-base-uncased model with a classification layer for 4 labels. The model was trained for 3 epochs with a batch size of 16 and a learning rate of 2e-5. Training and evaluation were done using Hugging Face’s Trainer API. Accuracy was used to measure the model’s performance.

**Results**

After training for 3 epochs, the model gave the following results:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | **Epoch** | **Training Loss** | **Validation Loss** | **Accuracy** | | 1 | 0.1946 | 0.1764 | 94.36% | | 2 | 0.1262 | 0.1929 | 94.86% | | 3 | 0.0786 | 0.2389 | 94.61% | |  |  |  |

**Table 2:** The model reached over 94% accuracy across all epochs, showing strong and consistent performance.

The above table 2 shows the model's performance over 3 training epochs. Accuracy improved after the first epoch and stayed above 94%, indicating the model learned the task well. Validation loss slightly increased in the last epoch, suggesting possible overfitting. Final test accuracy was **94.36%**, which shows the model learned well and made accurate predictions.

**Test Predictions**

Here are some test examples where the model predicted the correct topic:

A screenshot of a computer

AI-generated content may be incorrect.

**Figure 2:** Example test news articles with true and predicted topics. The model correctly classified all samples, showing good accuracy.

The above figure 2 shows some example news articles from the test set with their true topics and the topics predicted by the fine-tuned BERT model. The model correctly classified all examples, demonstrating its ability to understand and categorize news content accurately

**Improvements**

The model did a good job, but here are some ways I could improve it:

* Train for more epochs using early stopping.
* Try other models like RoBERTa or DistilBERT.
* Tune the learning rate and batch size.
* Add more evaluation metrics like precision, recall, and F1-score.
* Add more training data or use techniques to slightly change the input (data augmentation).

**Conclusion**

This project shows how to fine-tune a pre-trained BERT model for a text classification task. Using the AG News dataset, I trained the model to classify news articles into four topics. The final accuracy was over 94%, which is very good.

It proves that large language models like BERT can be fine-tuned easily and used for real-world text classification problems with good results.

**References**

* Hugging Face Transformers: <https://huggingface.co/transformers/>
* AG News Dataset: <https://huggingface.co/datasets/fancyzhx/ag_news>
* Yang, B., Zhang, B., Cutsforth, K., Yu, S. & Yu, X. (2024) ‘Emerging industry classification based on BERT model’, Science of The Total Environment. Available at: <https://www.sciencedirect.com/science/article/pii/S030643792400142X>.
* Biswas, D. & Gil, J.-M. (2023) ‘Research Paper Classification and Recommendation System based-on Fine-Tuning BERT’, IEEE. Available at: <https://ieeexplore.ieee.org/document/10229381/authors#authors>.
* Wu, H., Ye, X. & Manoharan, S. (2023) ‘Enhancing Multi-Class Text Classification with BERT-Based Models’, IEEE. Available at: <https://ieeexplore.ieee.org/document/10487760/authors#authors>.