Twitter Sentiment Analysis Report on NLP Model Performances

- **Name:** Sarthak Vajpayee
- **NetID:** SXV220020

Overview

This report summarizes the findings from three distinct natural language processing experiments conducted by Sarthak Vajpayee. Each experiment involved different model architectures and configurations, focusing on evaluating the performance of various transformer-based models on a specific classification task. The results and setups were meticulously documented using Weights & Biases (W&B), providing a robust platform for tracking and comparing model performances.

Experiment 1: DistilBERT Model

Model Architecture: The model used in this experiment is `DistilBERTClass`, a distilled version of the BERT model which is optimized for faster performance while retaining most of BERT's accuracy. The model includes layers for self-attention, dropout, and linear transformations geared towards a final classification layer with 11 outputs.

Tokenizer: `DistilBertTokenizer` from the 'distilbert-base-uncased' model was used, facilitating effective tokenization suited for DistilBERT.

Training Parameters:

- Epochs: 1
- Train Batch Size: 4
- Eval Batch Size: 4
- Learning Rate: 0.00001
- Weight Decay: 0.1

Results:

- Eval Loss: 0.3303

- F1 Score: 0.6713
- Eval Runtime: 13.10 seconds
- Samples per Second: 117.896

W&B Project Link: https://api.wandb.ai/links/sarthak-vajpayee/05pylktq

Experiment 2: AlbertRobertClass Model

Model Architecture: This unique setup combines layers from both RoBERTa and ALBERT models, named `AlbertRobertClass`. It includes dual token type embeddings, a mixture of self-attention mechanisms, and multiple dropout layers, followed by separate classifiers for combining features extracted via RoBERTa and ALBERT pathways.

Tokenizer: Two tokenizers, `RobertaTokenizer` and `AlbertTokenizer`, were used corresponding to each model's architecture component.

Training Parameters:

- Epochs: 1
- Train Batch Size: 4
- Eval Batch Size: 4
- Learning Rate: 0.0001
- Weight Decay: 0.1

Results:

- Eval Loss: 0.3115
- F1 Score: 0.6613
- Eval Runtime: 61.67 seconds
- Samples per Second: 25.053

W&B Project Link: https://api.wandb.ai/links/sarthak-vajpayee/edkqhg7m

Experiment 3: FLAN-T5 Model

Model Architecture: `FLAN-T5`, a variant of T5 adapted for few-shot learning, was used targeting multi-label classification. It simplifies adapting the T5 architecture to various NLP tasks without extensive task-specific tuning.

Tokenizer: `AutoTokenizer` for `google/flan-t5-base`.

Training Parameters:

- Epochs: 3
- Train Batch Size: 8
- Eval Batch Size: 8
- Learning Rate: Dynamic adjustment
- Evaluation Strategy: Every 200 steps

Results:

- Eval Loss: 0.3037
- Accuracy: 0.8767
- F1 Score: 0.6833
- Eval Runtime: 34.22 seconds
- Samples per Second: 45.145

W&B Project Link: https://api.wandb.ai/links/sarthak-vajpayee/94fn3ev6

Comparative Analysis

- **Performance:** FLAN-T5 outperformed the other models in terms of accuracy, showcasing its efficiency in handling multi-label classification tasks. DistilBERT provided a good balance between speed and accuracy, while AlbertRobertClass demonstrated robustness but with slower evaluation times.
- **Efficiency:** DistilBERT's faster evaluation times make it suitable for applications requiring quick responses. In contrast, AlbertRobertClass might be better suited for scenarios where model complexity and depth yield better performance accuracy.
- **Scalability:** FLAN-T5, with its few-shot learning capabilities, represents a scalable model for diverse NLP tasks without needing extensive re-training.

Conclusions

The experiments highlight the effectiveness of transformer-based models in handling complex NLP tasks, with each model offering unique advantages depending on the specific requirements of the task and the operational constraints. FLAN-T5's superior accuracy and flexibility make it an excellent choice for future NLP challenges, especially in multi-label classification scenarios.

However, the best test set results were achieved using DistilBERT, followed by FLAN-T5, and then the hybrid Albert+RoBERTa model.