Sentiment Analysis

Mohamad Rasmy*

1 Introduction

Named Entity Recognition (NER) is a crucial task in Natural Language Processing (NLP) that involves identifying and classifying named entities in text data into predefined categories such as person names, organizations, locations, etc. The objective of this project is to develop a pipeline for NER using Large Language Models (LLMs), specifically fine-tuning BERT and RoBERTa models for token classification, and to compare their performance.

2 Data Description

The CoNLL-2003 dataset, a well-known dataset for NER, was used in this project. It includes four types of named entities: persons (B-PER/I-PER), organizations (B-ORG/I-ORG), locations (B-LOC/I-LOC), and miscellaneous names (B-MISC/B-MISC). For tokens that don't correspond to any entity, it has the code (O). The dataset is divided into three subsets: training, validation, and testing.

- **Training Set:** Contains 14,041 sentences.
- **Validation Set:** Contains 3,250 sentences.
- **Test Set:** Contains 3,453 sentences.

3 Experiments

This section describes the preprocessing steps, the training arguments used, and the performance of the BERT and RoBERTa models.

3.1 Preprocessing

The text data underwent tokenization and label alignment to prepare it for modeling. This included:

- Tokenization of text data using the respective tokenizers for BERT and RoBERTa.
- Alignment of NER tags with the tokenized inputs to ensure correct label associations.

3.2 Training Arguments

The models were trained with the following parameters:

- Learning rate: 2e-5
- Batch size: 16
- Epochs: 3
- Weight decay: 0.01

*mohamad.hassan.rasmy@gmail.com

3.3 BERT Model

Model: `distilbert-base-cased` **Results:**

- Precision: 91.2%
- Recall: 90.1%
- F1-Score: 90.6%

Table 1: BERT Model Classification Report

	Precision	Recall	F1-score
LOC	0.93	0.92	0.92
MISC	0.78	0.78	0.78
ORG	0.86	0.88	0.87
PER	0.96	0.94	0.95
Micro avg	0.90	0.89	0.90
Macro avg	0.88	0.88	0.88
Weighted avg	0.90	0.89	0.90

3.4 RoBERTa Model

Model: `distilroberta-base` with `add_prefix_space=True` **Results:**

- Precision: 92.3%
- Recall: 91.5%
- F1-Score: 91.9%

Table 2: RoBERTa Model Classification Report

	Precision	Recall	F1-score
LOC	0.91	0.94	0.92
MISC	0.78	0.80	0.79
ORG	0.88	0.90	0.89
PER	0.95	0.95	0.95
Micro avg	0.90	0.91	0.91
Macro avg	0.88	0.90	0.89
Weighted avg	0.90	0.91	0.91

4 Overall Conclusion

Both BERT and RoBERTa models performed well on the NER task with high precision, recall, and F1-scores. However, the RoBERTa model slightly outperformed the BERT model, indicating its stronger ability to capture context and entity relationships in text. This is due to that RoBERTa was pre-trained on larger data, while having larger model.

5 Tools and Libraries Used

- **Programming Language:** Python
- **Libraries:** PyTorch, Transformers, Datasets, seqeval, Numpy