MultiNERD Named Entity Recognition Report

Yung-Ching Yang

Introduction

In this assignment, a series of Transformer-based models were fine-tuned for the Named Entity Recognition task, utilizing the MultiNERD dataset. The main experiment is to compare the performance of utilising different encoder, classifier layers and multitask. Table 1 shows the testing result of the system A and system B.

Model Type	**System A(ALL)**				**System B(FILTERED)**			
	Accuracy	Macro F1	Precision	Recall	Accuracy	Macro F1	Precision	Recall
Bert+Linear	0.983	0.822	0.818	0.830	0.988	0.905	0.892	0.919
Bert+LSTM	0.983	0.816	0.815	0.826	0.989	0.904	0.894	0.916
Roberta+Linear	0.983	0.801	-	-	0.989	0.902	-	-
MultiTaskBert	0.983	0.886	-	-	0.988	0.923	-	-

Table 1: Comparison of Metrics Between System A and System B

1 Findings

Due to the imbalanced data distribution, System B, which focuses on the top five NER tags, demonstrates better performance over System A. This can be attributed to the concentrated representation of major tags, which inherently possess higher accuracy due to their prevalence in the dataset. The result analysis, which can be found under the directory named notebook, also proves that those major tags have higher accuracy than the unrepresentative data. The implementation of the MultiTaskBert model tried to enhance the representation of minor tags by adapting multiple binary classification tasks targeted at O, B, and I tags. This approach improved the performance of System A, underscoring the effectiveness of multitasking in addressing class imbalance. Furthermore, a pre-analysis of token lengths has informed optimal padding strategies, striking a balance between training efficiency and model performance. The study reveals that the inclusion or exclusion of special tokens in optimization (padding as 0 or -100) impacts the outcomes in token-wise classification tasks, suggesting that such tasks has little degree of context sensitivity.

2 Limitation

Fine-tuning encoders with one layer or single task is less effective especially when the data is imbalanced. MultiTaskBert model tries to solve this issue by adding sub tasks, however, the sub classification tasks in the model are independent as the weighted losses are added together instead of being hierarchically dependent on each other. This method can be further improved by designing more complicated sub tasks as a joint learning framework where the interdependencies between different entity types are considered. For example, the learning from one task, say distinguishing between 'O' (Outside) and 'Non-O' (B or I) tags, informs and enhances the performance of other tasks, like distinguishing 'B' (Beginning) and 'I' (Inside) tags. Another limitation is that the current methodology lacks in data preprocessing techniques, particularly in data augmentation or balancing strategies such as synthetic data generation or oversampling methods.