



Web Pages Classification Based on BERT

Presenter: Zebang Yang
Advisor: Professor Min Chi



Propose

- Realize Data Cleansing to improve data quality and utility.
- Analyze the dataset to find hidden features.
- Apply several traditional methods to serve as baselines.
- Implement the proposed BERT model.
- Fine-Tune BERT model to achieve better performance.

Dataset

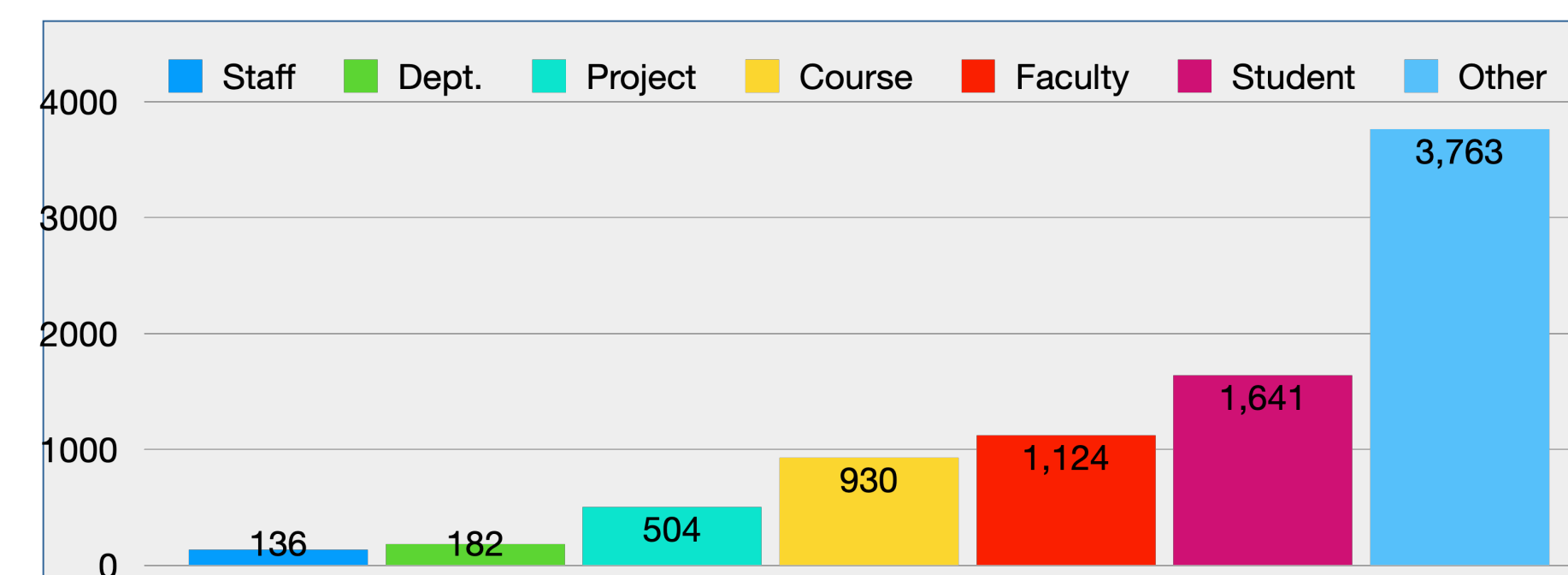


Figure 1: Statistics of Dataset

Table 1: Arrangement of Dataset

	Staff	Dept.	Project	Course	Faulty	Student	Other	Total
Train	95	128	353	651	786	1149	2634	5796
Valid	27	36	101	186	225	328	753	1656
Test	14	18	50	93	113	164	376	828
Total	136	182	504	930	1124	1641	3763	8280

Methods

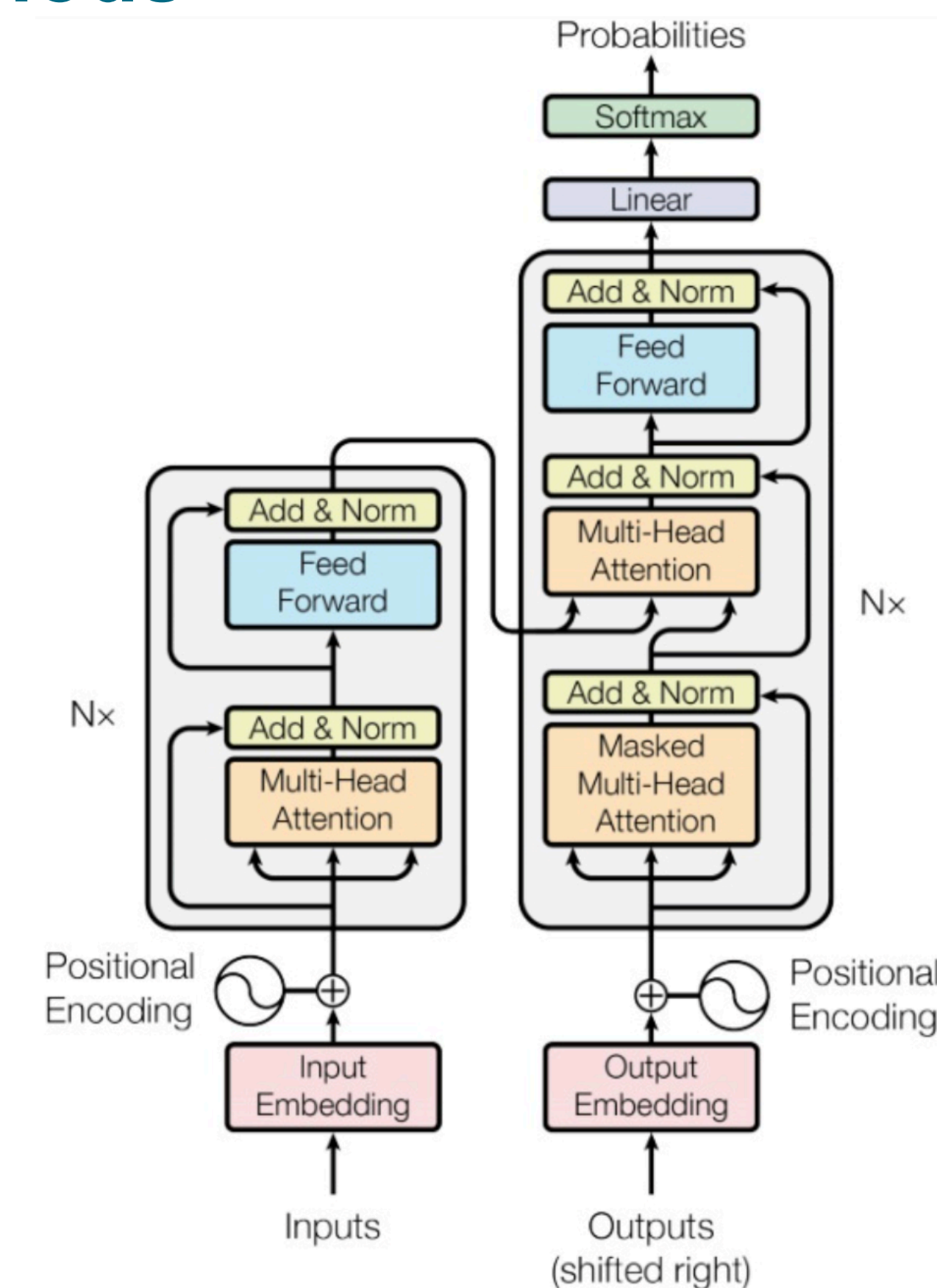


Figure 2: Basic Transformer Architecture

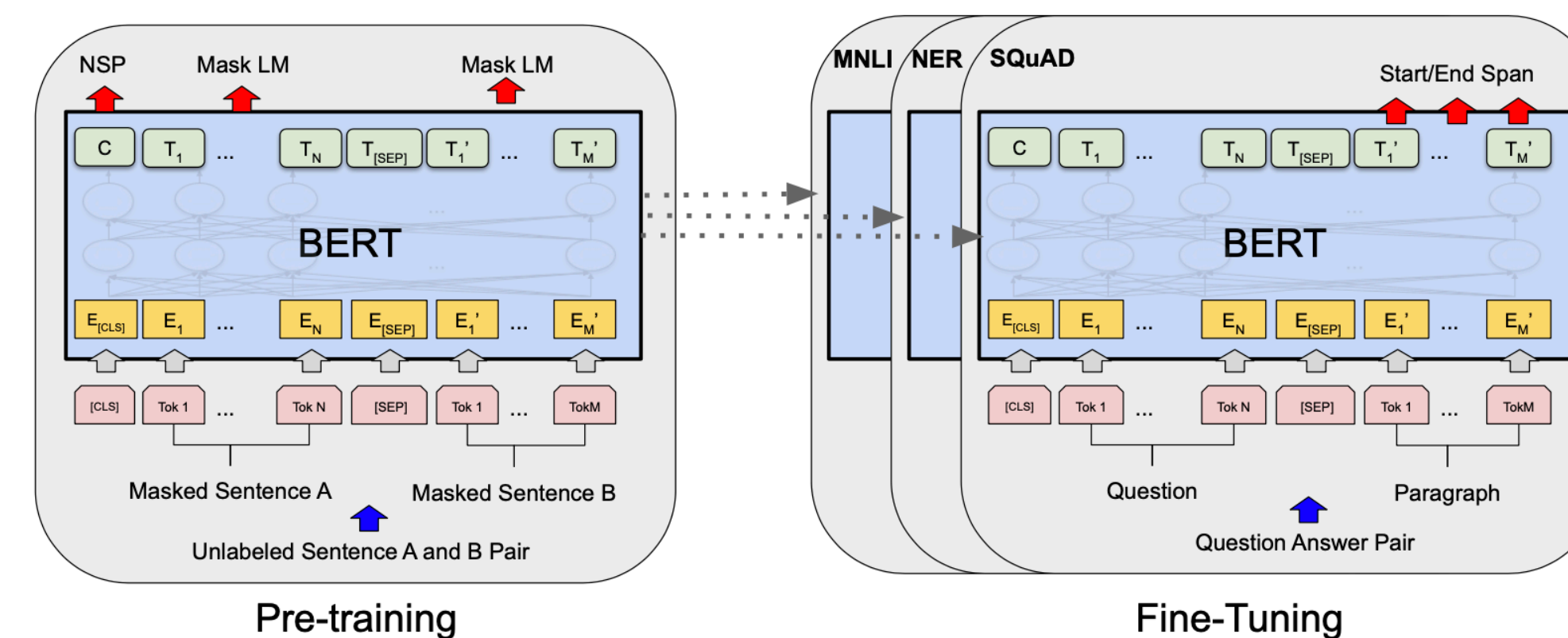


Figure 3: Pretrained Model and Specific Fine-Tuning

Results

Table 2: Performance Between Different Methods

Method	Accuracy
Naïve Bayes	0.6032
Random Forest	0.8108
Logistic Regression	0.8350
Bidirectional LSTM	0.9118
BERT + Fine-Tune	0.9457

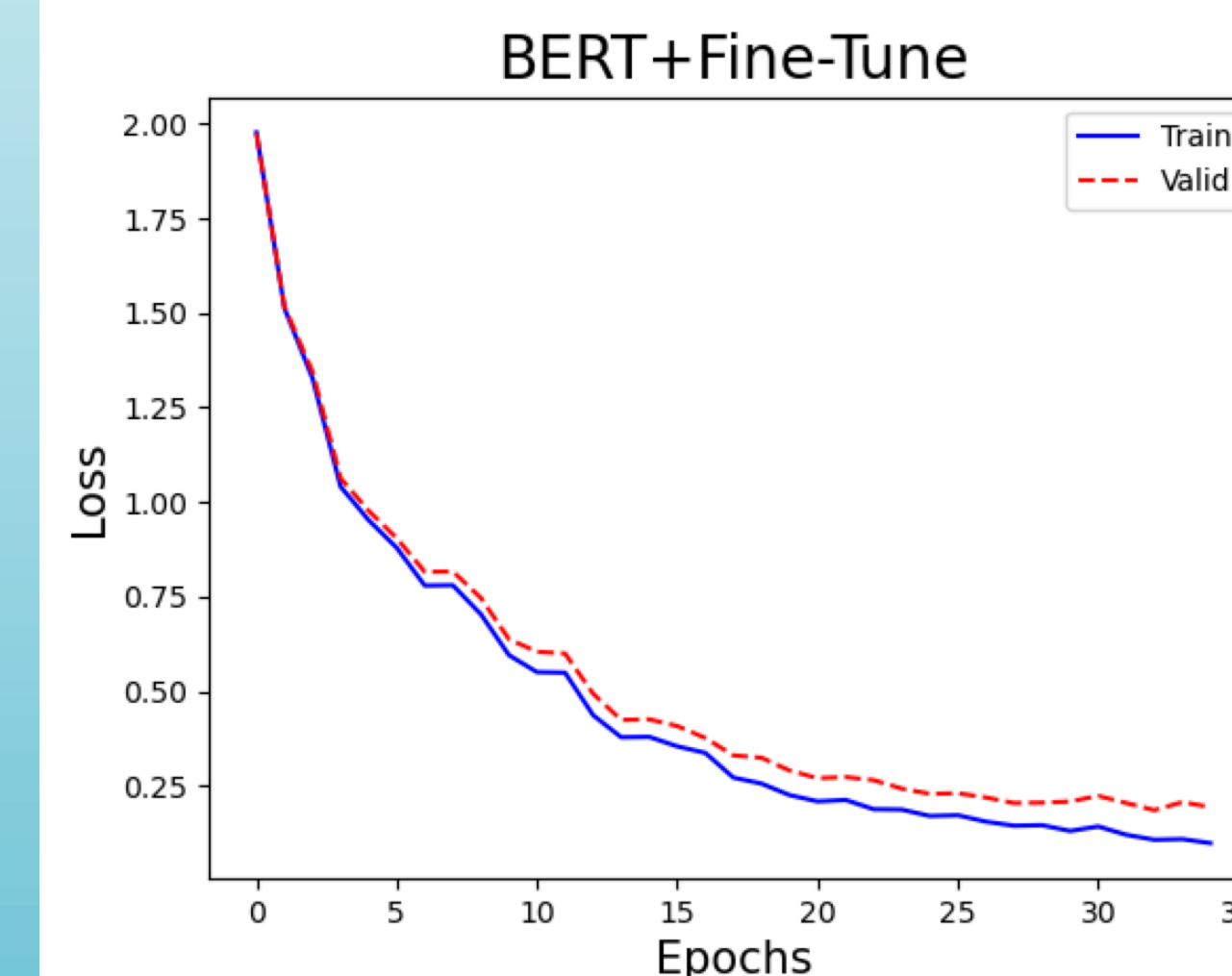


Figure 4: Loss of Proposed Method

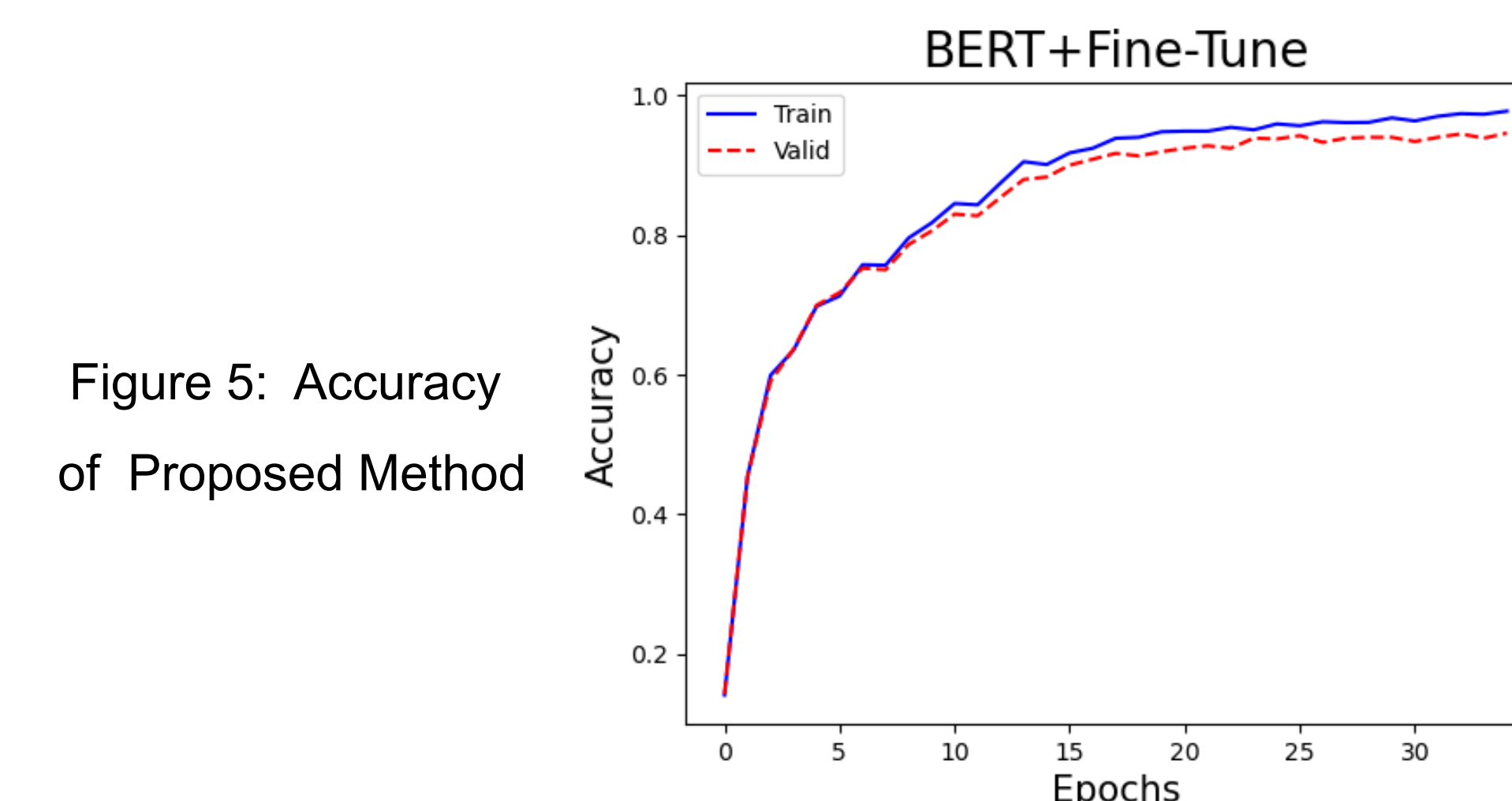


Figure 5: Accuracy of Proposed Method

Conclusions

- The pre-trained BERT model with some specific fine-tuning can easily deal with the target task.
- Due to self-attention, the fine-tuned pre-trained BERT model has better performance than any other traditional methods.
- Despite its excellent performance, the proposed model is easy to be over-fitting. Therefore, well-designed hyper-parameters and training process should be implemented.

Reference

- Bert: Pre-training of deep bidirectional transformers for language understanding. In NAACL, 2018.
- Contextual string embeddings for sequence labeling. In ICCL, 2018.
- Deep contextualized word representations, In NAACL, 2018.
- Attention is all you need. In NIPS, 2017.