Natural Language Processing Project Report

Fuming Zhang, 118033910025
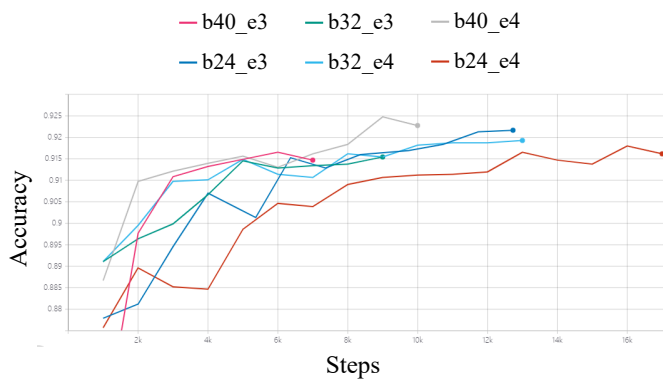
✦

Fig. 1. Evaluation Accuracy
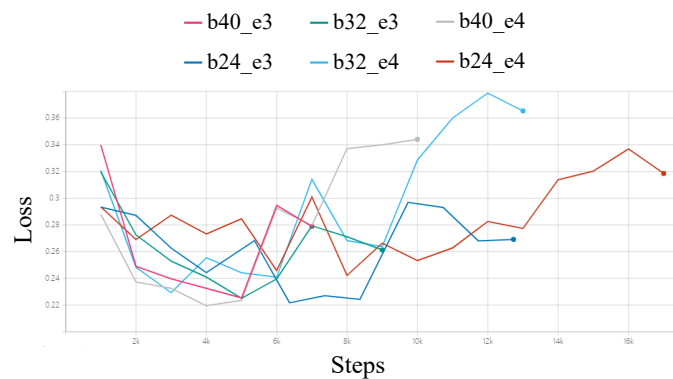
1 INTRODUCTION

2 BACKGROUND

2.1 Transformer

Transformer [1] [2]

2.2 GPT

GPT [3]

2.3 BERT

BERT [4]

3 IMPLEMENTATION

Considering the limited gpu resources we have, it is almost impossible to retrain the bert model. So we completed this project based on the existing open source implementation and its pre-trained model. To be more specific, we use the PyTorch implementation provided by Hugging Face (https://github.com/huggingface/pytorch-pretrained-BERT).

4 EXPERIMENTAL RESULTS

5 CONCLUSION

ACKNOWLEDGMENTS

The author would like to thank Prof. Hai Zhao and TA Shu Jiang and Zhuosheng Zhang for their instructions on this work.

- *Project link: https://github.com/AlexChang/NLU-QNLI*
- *E-mail: zhangfuming-alex@sjtu.edu.cn*
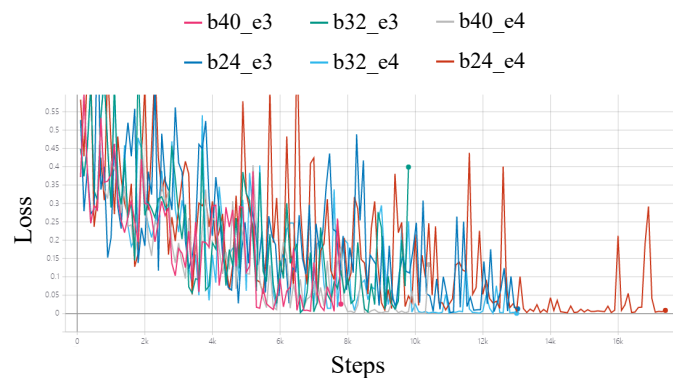

Fig. 2. Evaluation Loss

Fig. 3. Train Loss

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[2] Z. Dai, Z. Yang, Y. Yang, W. W. Cohen, J. Carbonell, Q. V. Le, and R. Salakhutdinov, "Transformer-xl: Attentive language models beyond a fixed-length context," *arXiv preprint arXiv:1901.02860*, 2019.

[3] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," *URL https://s3-us-west-2. amazonaws. com/openai-assets/research-covers/languageunsupervised/language understanding paper. pdf*, 2018.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

Results for submission qnli_ensemble_1

Score: 60.9

Task	Metric	Score
The Corpus of Linguistic Acceptability	Matthew's Corr	0.0
The Stanford Sentiment Treebank	Accuracy	80.0
Microsoft Research Paraphrase Corpus	F1 / Accuracy	81.5/73.4
Semantic Textual Similarity Benchmark	Pearson-Spear...	61.2/58.7
Quora Question Pairs	F1 / Accuracy	51.4/79.1
MultiNLI Matched	Accuracy	56.0
MultiNLI Mismatched	Accuracy	56.4
Question NLI	Accuracy	92.7
Recognizing Textual Entailment	Accuracy	54.1
Winograd NLI	Accuracy	62.3
Diagnostics Main	Matthew's Corr	9.2

Fig. 4. Best Prediction Result(ensemble)

Results for submission qnli_4

Score: 60.8

Task	Metric	Score
The Corpus of Linguistic Acceptability	Matthew's Corr	0.0
The Stanford Sentiment Treebank	Accuracy	80.0
Microsoft Research Paraphrase Corpus	F1 / Accuracy	81.5/73.4
Semantic Textual Similarity Benchmark	Pearson-Spear...	61.2/58.7
Quora Question Pairs	F1 / Accuracy	51.4/79.1
MultiNLI Matched	Accuracy	56.0
MultiNLI Mismatched	Accuracy	56.4
Question NLI	Accuracy	92.2
Recognizing Textual Entailment	Accuracy	54.1
Winograd NLI	Accuracy	62.3
Diagnostics Main	Matthew's Corr	9.2

Fig. 5. Best Prediction Result(single)