

Natural Language Processing Project Report

Fuming Zhang, 118033910025

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.

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.

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