

IMPROVING BERT FINE-TUNING VIA STABILIZING CROSS-LAYER MUTUAL INFORMATION

Jicun Li^{1,2}, Xingjian Li^{3,6}, Tianyang Wang⁴, Shi Wang^{1,2}, Yanan Cao⁵, Chengzhong Xu⁶, Dejing Dou³*

¹ Key Laboratory of Intelligent Information Processing
Institute of Computing Technology, Chinese Academy of Sciences

² University of Chinese Academy of Sciences

³ Big Data Lab, Baidu Research

⁴ University of Alabama at Birmingham

⁵ Institute of Information Engineering, Chinese Academy of Sciences

⁶ State Key Lab of IOTSC, University of Macau

ABSTRACT

Fine-tuning pre-trained language models, such as BERT, has shown enormous success among various NLP tasks. Though simple and effective, the process of fine-tuning has been found

severe instability, i.e. fine-tuning on the same dataset with different seeds often leads to quite different results (in terms of accuracy), some of which generalize very poorly.

To solve this problem, a major idea of existing studies is