RoBERTa

Facebook提出的a Robustly Optimized BERT Pretraining Approach（RoBERTa），模型结构与BERT相同，但预训练时采用了更多的训练数据，并使用了动态masking的策略，移除了Next Sentence Prediction任务的预训练。

在表现上，GLUE基准测试上效果很好

