

山东大学 计算机科学与技术 学院

大数据分析实践 课程实验报告

学号: 202300130235	姓名: 危弘毅	班级: 数据 23
实验题目: Bert 实践		
实验学时: 2	实验日期: 2025.11.7	
实验目的: 在预训练后的 Bert 模型上进行微调以适应下游任务, 掌握模型的训练方法和应用预训练模型的能力。		
实验环境: Windows 11, Edge 浏览器		
实验步骤与内容: 使用中文 Bert 模型在 ChnSentiCorp 情感分析数据集上进行微调, 解决情感分析问题		
<pre>===== 开始训练BERT模型 ===== 训练进度: 0% 1/300 [00:13<1:06:40, 13.38s/step]<IPython.core.display.HTML object> 训练进度: 100% ██████████ 300/300 [-792:50:11<00:00, 10.96s/step] 训练进度: 100% ██████████ 300/300 [-792:52:07<00:00, 10.96s/step] 开始评估验证集... 训练进度: 100% ██████████ 300/300 [-792:52:08<00:00, -0.00step/s]</pre>		
测试准确率和前 5 条样本		
<p>测试集准确率: 0.9350</p> <p>文本: 这个宾馆比较陈旧了, 特价的房间也很一般。总体来说一般... 真实标签: 正面 预测标签: 负面</p> <p>文本: 怀着十分激动的心情放映, 可是看着发现, 在放映完毕后, 出现一集米老鼠的动画片! 开始还怀疑是不是赠送... 真实标签: 负面 预测标签: 负面</p> <p>文本: 还稍微重了点, 可能是硬盘大的原故, 还要再轻半斤就好了。其他要进一步验证。贴的几种膜气泡较多, 用不了多... 真实标签: 负面 预测标签: 负面</p> <p>文本: 交通方便; 环境很好; 服务态度很好 房间较小... 真实标签: 正面 预测标签: 正面</p> <p>文本: 不错, 作者的观点很颠覆目前中国父母的教育方式, 其实古人们对于教育已经有了很系统的体系了, 可是现在的父... 真实标签: 正面 预测标签: 正面</p>		

结论分析与体会：

本次基于 ChnSentiCorp 数据集完成 BERT 中文情感分析微调实验。实验中，模型在测试集上取得了较好的分类准确率，能有效区分正负情感文本。通过实践我体会到，模型训练中数据集规模、epoch 数量与硬件适配直接影响训练效率与效果，而正确加载最优模型及分词器是实现精准预测的关键。同时，早停、混合精度训练等技巧可有效提升训练稳定性，为后续 NLP 文本分类任务提供了实践经验。