

NCUSCC 2025 秋季考核深度学习赛道实验报告

目录

1. 环境搭建
2. 数据集准备
3. 模型训练
4. 特别感谢

1. 环境搭建

1. 安装wsl2并在其中搭建Ubuntu 24.04 LTS
2. 安装Miniconda并创建虚拟环境
3. 更新NVIDIA驱动并安装cuda toolkit
4. 安装python3.10
5. 在官网根据cuda和python的版本下载对应的Pytorch
6. **安装Hugging Face Ecosystem(其中的transformers库包含了bert-base-chinese模型)**
7. 用vscode远程连接wsl2

2. 数据集准备

1. **寻找并下载数据集,我用的数据集是百度飞浆中的商品评价数据集,数据量较大共7万条,对评价的情感倾向分为5分,其中1—2分表示消极,3分表示中性,4-5分表示积极。**
2. 进行数据清洗:去除停用词,处理缺失值,去除部分不必要的字符,删除过长评论(减少一些无用的信息,提高模型训练的效率)(clean_data.py)
3. 进行数据强化:进行同义词替换,随机插入,随机删除(使用情感词库避免情感词被删除)(提高数据的多样性以提高模型的准确度,避免模型死记硬背)(test.py)
4. 划分训练集、验证集(用一段Python代码进行划分)(classify.py)

3. 模型训练(train.py)

1. 加载数据集
2. **数据预处理,用Transformer中的分词器对文本进行分词、填充和截断(我的理解是分词之后模型更容易抓住情感判断的关键词,减少无用的信息造成的负面影响)**
3. 定义评估指标:使用f1-score作为主要评估指标,额外计算Accuracy
4. 加载模型

4. 特别感谢

1. Kimi大模型
2. Gemini
3. 雨木木对于模型三分类的提醒