沈立 2018000142

本项目主要是基于美团用户评论文本，进行情感细粒度的分析。

1. 本项目是基于美团用户评论文本，通过自然语言处理技术，从文本中生成情感标签，训练过程均在GPU服务器上；

2. 通过jieba、导入停止词等中文分词技术，将评论文本有效清洗，主要在“利用搜狗词库预处理”的文件中；

3. 对于评论样本过少的问题，通过爬虫抓取美团官网百万条评论以扩充样本数量；

4. 尝试通过gensim.word2vec技术主动生成中文词向量以及进行词嵌入；在遇到瓶颈的时候，通过导入搜狗新闻词嵌入矩阵（https://github.com/Embedding/Chinese-Word-Vectors），以提高词嵌入的质量；

5. 在训练过程中，尝试了简单的机器学习的训练技术，例如sklearn中的MLPClassifier，SVM，AdaBoostClassifier以及DecisionTreeClassifier等简单的训练函数，效果不佳；

6. 在训练过程中，最终尝试了复杂的深度学习模型，主要采用tensorflow进行搭建模型，例如有简单的CNN, RNN，以及复杂的双层LSTM+Attention，同时在训练过程中加入了正则损失以及对全连接层进行归一化操作，最后得到了较好的结果。