



PowerAI 人工智能马拉松编程大赛

Hackathon for Finance Industry

作品展示-第十五组

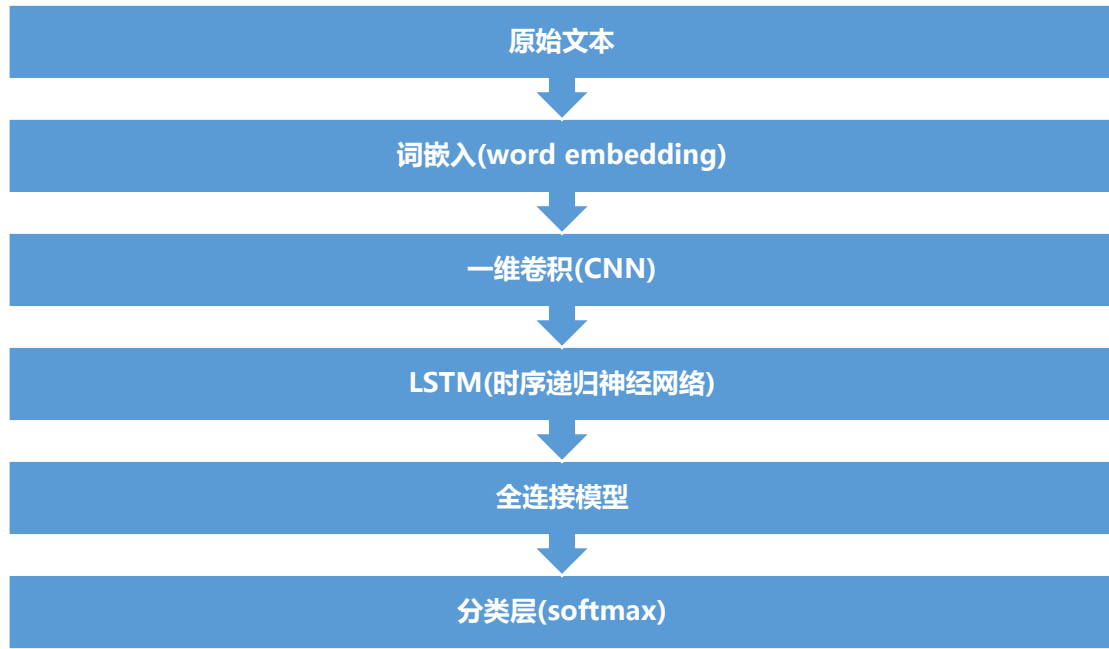
(大赛三等奖)

CSDN

Power Systems



作品展示



● 作品展示 (算法/代码等)

主要思路是先使用结巴分词对对话进行分词，然后使用词嵌入、LSTM、softmax进分类，加入CNN后发现效果更好，计算框架使用keras2.0+tensorflow1.0

作品展示

文本分类属于有监督学习过程，在分类中需要做的有：

- ◆ 定义类别，本例中类别分为两种，一种为情感分类，一种为产品倾向分类
- ◆ 划分训练集与测试集，本例划分比例为75%的训练集，25%的测试集合，首次划分采用自然划分法，后期调整为shuffle划分，避免原始数据集初始状态的影响。
- ◆ 向量降维度，选取分词后前128个词语作为特征向量，降低特征向量维度。
- ◆ 分类模型选择，考虑到本例是多分类问题，因而选择softmax模型进行分类。
- ◆ 测试集模型性能测试，在测试集合上首次迭代，情感分类准确率为64%，产品分类准确率76%
- ◆ 模型调优，考虑到实际情况中，客服角色的发言多为正向引导，而最终产品购买结果和情绪的载体是客户本身，因此考虑，仅使用客户发言作为训练集合重新进行训练，调整后，情感分类准确率为：72%，产品选择准确率为：86%。
- ◆ 继续调优，初始迭代次数15次，后期对其进行调整，考虑到过拟合的问题，每次调整迭代次数增加幅度为10，最终的迭代次数为45次，最终情感分类准确率为：77%，产品选择准确率为：92%，模型的准去率进一步提升。
- ◆ 实际数据测试，在最终的实际数据中，应用上述模型得到的最终情感分类准确率为：77%，产品选择准确率为：92%，模型表现稳定性。

- **作品描述：**诸如研发背景、实现思路、算法实现、框架的设计、对Caffe, Tensorflow技术的应用的经验，等等。

个人评价



大赛题目的选择还算是一个非常典型的金融业应用场景，赛前我们也准备了一点文本情感分析方面的模型，比赛过程中发现大量的时间放在数据的处理中了，最后用一个比较标准和文本处理模型来做的，而aspect level的情感分析模型没有来得及实现。

比赛的评委非常专业，在过程中也给了很多指导，学习到了很多知识。

团队风采



左起：徐作新、贺楠、于琛、裴勇泉
来自兴业数金的研究开发总部与算法实验室团队