

组会汇报

陈钊杰
专业:计算数学

October 10, 2023

目录

1

代码调试

- 序列分类任务比对
- k线图编码以后的进行降维可视化结果
- 下一步的计划

LSTM模型与自然语言模型进行序列分类对比实验

- ① 所选用的数据集:美股实时行情数据,经过筛选以后一共得到774支股票,将所有股票数据合并后再分成训练集和测试集.
- ② 关于预测的准确度,主要做了以下几个测试:
 - ① chatglm模型,历史节点:25
 - ② lstm模型,使用随机的词嵌入向量,历史节点:25,对每只股票使用一个模型
 - ③ lstm模型,使用word2vec方法,历史节点:25,对每只股票使用一个模型
 - ④ lstm模型,使用word2vec方法,历史节点:100,对每只股票使用一个模型
 - ⑤ lstm模型,使用word2vec方法,历史节点:25,对所有股票使用单个模型
- ③ 使用Word2Vec方法对所有数据集中的字符进行训练以后,构造新的一些字符组合,并对每个字符的高维向量进行降维,然后进行可视化。

Table: 准确率结果

模型	平均准确率
随机初始化词向量, 历史长度是25 (多个模型)	27.26%
Word2Vec方法, 历史数据点:25 (多个模型)	27.36%
Word2Vec方法, 历史数据点:100 (多个模型)	27.40%
Word2Vec方法, 历史数据点:25 (一个模型)	29.49%
chatglm模型, 历史数据点:25	32.34%



[C B D A]

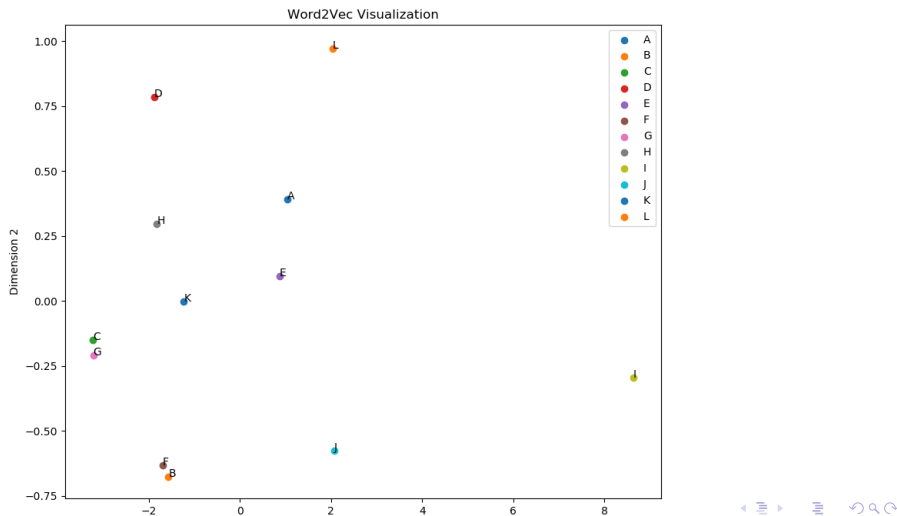


[G F H E]



[I L J K]

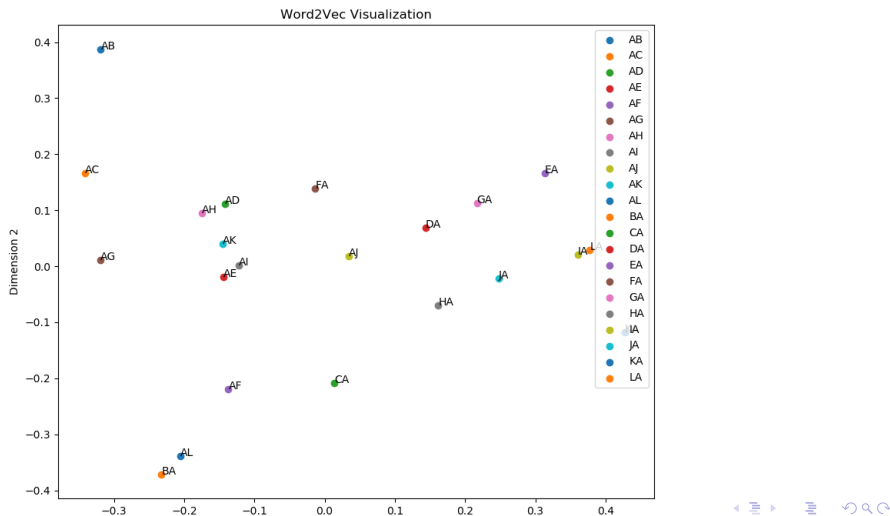
单个字符可视化



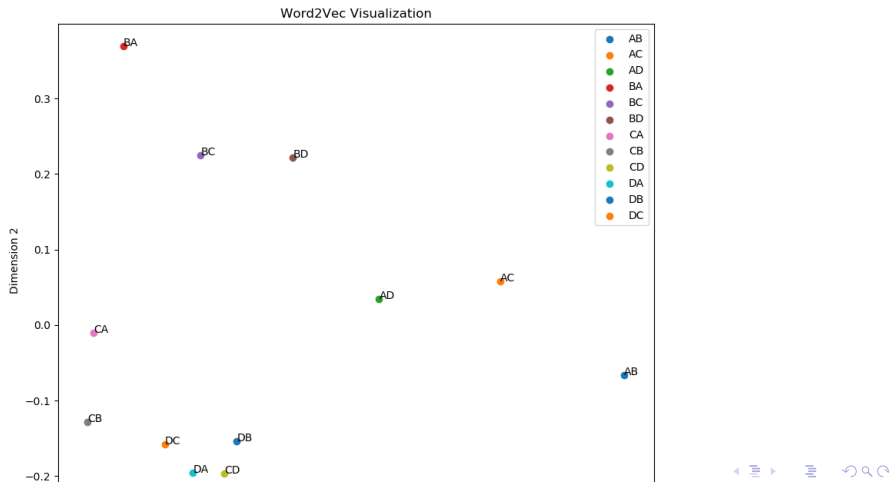
两个字符所有组合可视化



两个字符组合中所有包含A的可视化



A,B,C,D的两两组合可视化（即收盘价低于开盘价的组合）



- 经过比较，对lstm模型而言，使用单个模型训练的效果更好。
- chatglm模型相比lstm模型的平均准确率更高。
- 使用随机的词向量和使用word2vec方法的词向量最终结果相差不大，使用word2vec方法的准确度略高一点。
- 经过可视化以后，使用增加了基本词向量以后，依旧可以发现CG,FB,AE等之间有密切的关联。

下一步计划及相关问题

- 寻找其他相关模型进行比对?

谢谢老师和同学们的聆听!