专业综合训练报告

学号： 姓名：

学号： 姓名：

学号： 姓名：

学号： 姓名：

指 导 教 师： 聂栋栋 王刚

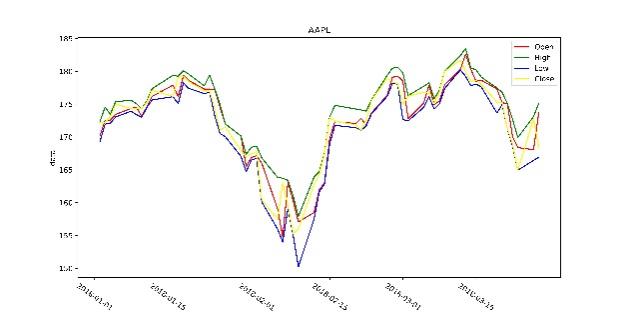
2022年11月

利用LSTM模型进行股票数据预测分析

1. 本课题主要进行股票数据分析预测，通过下载官方提供的交易数据对其进行数据分析，并使用LSTM长短期记忆模型对当日成交信息进行预测，并利用模型预测结果进行模拟操作，最终根据模拟操作结果反向评价LSTM模型表现。

# 交易数据汇总

本文中所使用交易数据主要来自于维基百科的美股交易接口，通过下载苹果、谷歌、微软、Facebook以及波音公司的2018年1月1日至2018年6月1日的交易数据进行分析。受限于本文篇幅，此处仅展示各家公司2018年1月1日至2018年3月27日的交易数据折线图，导入结果如下:

图表, 直方图

描述已自动生成

图 1苹果交易数据 图 2微软交易数据

图表, 直方图

描述已自动生成图表, 折线图

描述已自动生成

图 3谷歌交易数据 图 4 Facebook交易数据

图表, 直方图

描述已自动生成

图 5波音交易数据

其中，红线表示当日开盘价，绿线表示当日最高成交价，蓝线表示当日最低成交价，黄线表示当日收盘价。各曲线反映了对应交易信息在2018年1月1日至2018年3月27日的交易价格变化情况，为后文分析结果提供有力支撑。

本课题同时还考虑了各公司市值变化趋势，交易信息汇总折线图如下：

图表

描述已自动生成图表

描述已自动生成

图 6苹果市值变化趋势 图 7微软市值变化趋势

图表

描述已自动生成图片包含 图表

描述已自动生成

图 8谷歌市值变化趋势 图 9 Facebook变化趋势

图表

描述已自动生成

图 10波音市值变化趋势

# 建立预测模型

## 2.1模型介绍

正常情况下，前馈神经网络默认当前输入信息与前一刻信息完全无关，在处理无前后关系问题时，该类型神经网络有较好表现，但是在处理金融市场交易数据时无法忽视交易信息变化趋势对预测结果的影响。因此，为了运用到时间信息，人们构建了递归神经网络（RNN），该神经网络模型的特点时单个神经元会递归调用自身并将时刻的信息传递给时刻。而递归神经网络在利用时间序列信息的目标下运行良好，尤其时在短时间序列情况下表现优秀，但是该模型在处理长期以来问题时有较为明显的不足。

本课题使用LSTM神经网络模型进行预测，该方法是一种时间循环神经网络，全称为长短期记忆网络（Long Short Term Memory），主要目的是为了解决一般的循环神经网络存在的长期依赖问题而专门设计的，通过记忆不定时间长度的数值，LSTM适合处理和预测时间序列中间隔和延迟非常长的事件。同时，LSTM模型还可以解决RNN模型中的梯度消失以及梯度爆炸问题。

## 2.2模型原理

LSTM神经元结构如下：

图示

描述已自动生成

图 11 LSTM神经元结构图

图11中，每个红色圆圈代表对向量做出的操作，而黄色矩形代表一个神经网络层。

LSTM的关键内容就在于神经元的隐藏态，也成为神经网络的单元状态。在该模型中，可以简单理解成递归神经网络对于输入信息的存储记忆，该向量覆盖了对于时刻前神经网络对于所有输入信息的概括总结。

每个神经元一般包含四个函数层：

第一层通常为遗忘门，该层函数通常作用为判断是否选择遗忘在之前单元状态中传递下来的记忆信息，并将该神经元输入结果存储在记忆向量中，作为时刻新的记忆向量向下传递。

第二层/第三层通常为记忆门，用于控制是否将在时刻的数据并入单元状态中的控制单位。主要目的便是从当前输入信息中提取有效信息，并将提取的有效信息进行筛选，为每个分量做出评级，评级越高的信息便会在该单元存储的信息中占据更多容量。

第四层通常为输出门，该层函数就是LSTM单元用于计算当前时刻的输出值的神经曾。输出层会先将当前输入值与上一时刻输出值整合后得到的向量提取其中得到的信息，并将其整合后可以得到LSTM在时刻的输出值。

# 模型实战

利用LSTM模型，对所获数据进行处理。鉴于本课题篇幅有限且该模型单次运行时长过长，因此在本章节仅提供单次模型运行结果作为实例。

## 3.1构建训练集

构建训练集关键代码如下：

图形用户界面, 文本, 应用程序

描述已自动生成

图 12构建训练集代码展示

构建验证集代码展示如下：

图片包含 图表

描述已自动生成

图 13构建训练集代码展示

构建测试集代码展示如下：

图片包含 图形用户界面

描述已自动生成

图 14 构建验证集代码展示

模型中单个神经元构建代码如下：

图片包含 文本

描述已自动生成

图 15构建单个神经元代码展示

建立损失函数并进行可视化处理：

文本

描述已自动生成

表格

描述已自动生成

图 16损失函数代码及部分结果展示

预测效果趋势图展示：

图表, 直方图

描述已自动生成

图表, 折线图

描述已自动生成

图 17模型预测结果（缩略/详细）

# 预测结果

根据预测结果中所示交易数据对所选择股票进行交易操作，本金为10000美元，利用贪心算法原理，对每日交易数据进行预测，使得每日总交易利润最高，构建交易操作代码如下：

形状, 矩形

描述已自动生成

形状

描述已自动生成

图形用户界面, 文本

描述已自动生成

图形用户界面

中度可信度描述已自动生成

图形用户界面, 文本

描述已自动生成

图形用户界面

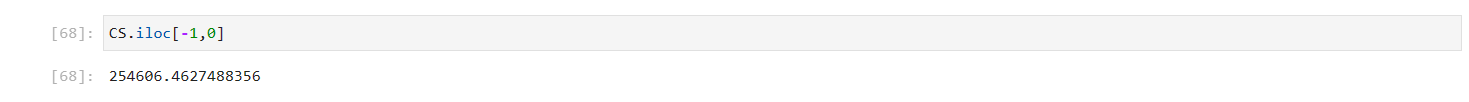
中度可信度描述已自动生成

模拟展示结果如下：

图形用户界面, 应用程序

描述已自动生成

该模拟结果仅展示部分时间内的交易操作，其中C表示当前持有现金总量，shares1~5分别表示持有苹果、微软、谷歌、Facebook和波音的股份，最终模拟操作结果为：



本课题利用LSTM算法，分别对苹果、微软、谷歌、Facebook和波音在2018年1月1日至2018年6月1日的交易数据进行训练，并通过模拟进行交易操作以验证模型可行性，本金为10000美元，在完成模拟交易周期后共获利254606.4627488356美元，模型可行性验证成功。