

# 基于 LSTM 深度学习的股市预测研究

莫洁安

(广西民族师范学院 广西 崇左 532200)

【摘要】本文研究基于 LSTM 深度学习的股市预测问题,对股市运行状况进行研究,利用 LSTM 模型对时间序列数据的非线性关系处理的优势,运用 python3 基于 Keras 库进行实证研究,对深证指数进行了研究,其预测效果良好。

【关键词】LSTM;股市预测

对于当下比较热门的 LSTM 深度学习模型可以学习长期依赖信息,在很多问题中都取得相当巨大的成功并得到了广泛的使用,特别是当下比较热门的图像处理、文本挖掘等人工智能方面,因其善于发掘时间序列数据间的非线性关系,可考虑运用在股市时间序列预测问题,因此本文研究 LSTM 深度学习模型股市时间序列能否有效地进行预测。

## 1、LSTM 模型介绍

长短期记忆人工神经网络 (Long Short-Term Memory, LSTM)模型是递归神经网络 (recurrent neural network, RNN)特殊的一种变型类型,由 Hochreiter & Schmidhuber (1997)<sup>[1]</sup>提出,并在近期被 Alex Grave<sup>[2]</sup>进行了改良和推广。LSTM 模型通过巧妙设计来避免长期依赖问题记住长期的信息,经典的 LSTM 模型结构如下图 1 所示:

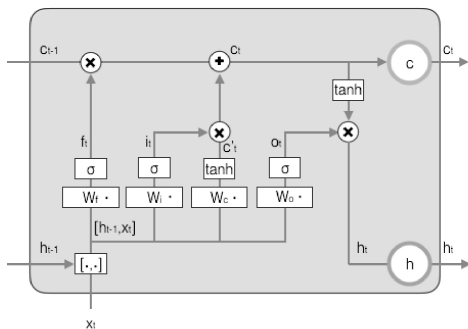


图 1 LSTM 模型结构详细图

LSTM 模型的结构是重复的模块拥有一个不同的结构。不同于单一神经网络层,以一种非常特殊的方式进行交互。下面简单介绍一下其工作原理

(1)忘记门层,是决定从细胞状态中丢弃什么信息。读取上一层的输出  $h_{t-1}$  和当前状态的输入  $x_t$ ,输出结果是在  $[0,1]$  区间的数值,每个在细胞状态  $C_{t-1}$  中的数字中 1 表示“完全保留”,0 表示“完全舍弃”。该层的计算过程如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

(2)输入门层,是决定什么信息将要更新。有两个部分,一是 sigmoid 函数层,决定输入什么信息,二是 tanh 函数层创建一个新的候选值向量,该向量会被加入到状态中。该层的计算过程如下所示:

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \end{aligned} \quad (2)$$

(3)更新门层,该层更新新旧细胞状态,由  $C_{t-1}$  更新到  $C_t$ 。该层计算过程如下所示:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3)$$

(4)输出层,决定需要确定输出什么值。这个输出将会基于细胞状态,先运行一个 sigmoid 层来确定细胞状态的哪个部分将输出,接着把细胞状态通过 tanh 进行处理,并将它和 sigmoid 门层的输出相乘,最终确定输出那部分。该层计算过程如下:

$$\begin{aligned} o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(C_t) \end{aligned} \quad (4)$$

根据以上对 LSTM 模型步骤分析,可知其特点就是在 RNN 结构以外添加了各层的阀门节点。阀门有 3 类:忘记门(forget gate),输入门(input gate)和输出门(output gate)。这些阀门可以打开或关闭,用于将判断模型网络的记忆态在该层输出的结果是否达到阈值从而加入到当前该层的计算中。

## 2、股市走势预测研究

本文对深证指数进行研究,获取数据的范围是从 1991 年 4 月到 2018 年 4 月共六千多个样本数据,该数据是从同花顺软件中导出来的,以收盘价为研究对象,数据如下:

表 1 深证指数收盘价数据情况

序号	时间	收盘
1	1991-04-03, 三	988.05
2	1991-04-04, 四	983.11
3	1991-04-05, 五	978.27
...	...	...
6643	2018-04-26, 四	10292.12
6644	2018-04-27, 五	10324.47
6642	2018-04-25, 三	10547.1

运用该数据样本探讨其预测规律情况,其总体的变化趋势如下图 3 所示。

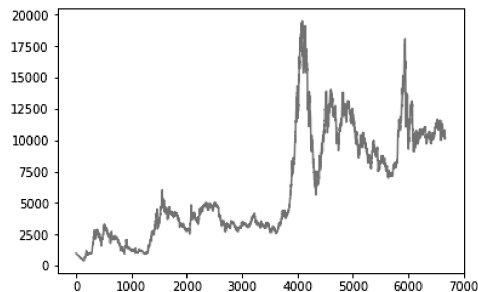


图 2 深证指数的收盘价变化趋势

观察图 3 可知深证指数总体情况是增长趋势,中间变动情况比较大。下面讨论运用 LSTM 深度学习模型对该数据集样本进行研究。具体实施步骤如下:

### 2.1 数据初始化(生成模型训练数据集)

由于 LSTM 模型是神经网络结构类型,需要将样本的列数组转换为数据集矩阵,并将数据集进行正则化处理,再把处理后的数据分为两部分,训练数据集和验证测试数据集,本文按照样本的 90%为训练数据集,10%为验证测试数据集,随机拆分为训练数据集和验证测试数据集,这是为了防止过度拟合。

### 2.2 LSTM 模型结构的确定与调整

本文采用 Python 的 Keras 库对数据样本进行处理,同时构建 LSTM 模型的网络进行设定与调整,涉及到非常多的参数调整,经过不断的运行调整,得到本问主要的参数设置:

A. 确定 LSTM 模型的激活函数 (Keras 默认是用 tanh 函数), 确定接收 LSTM 输出的完全连接人工神经网络中的激活函数(Keras 默认是 linear 函数);

B. 确定每一次网络节点的遗忘率,本文使用默认值 0.2;

C. 确定计算误差计算公式,本文采用均方根误差(RMSE).

D. 确定权重参数的迭代更新方式, 本文采用 RMSprop 算法。

E. 确定模型训练的 epoch 和 batch size,本文采用 100 和 1.

F. LSTM 的隐含层数为 20,输出层为 1.

### 2.3. 模型训练和结果预测

```
trainScore = math.sqrt(mean_squared_error(trainY[0], trainPredict[:,0]))
print('Train Score: %.2f RMSE' % (trainScore))
testScore = math.sqrt(mean_squared_error(testY[0], testPredict[:,0]))
print('Test Score: %.2f RMSE' % (testScore))
```

```
Train Score: 0.01 RMSE
Test Score: 0.01 RMSE
```

图 3 结果的均方根误差

将本文采集的数据集以及训练好的参数导入模型,便可得到本模型的预测情况,与实际数据集相比便可得到该模型的优劣。

通过对数据的预测值与实际值的对比,可以清晰的看到其预测情况的优劣。如下如图 4 所示,蓝色为预测值,可以看出其效果比较好。

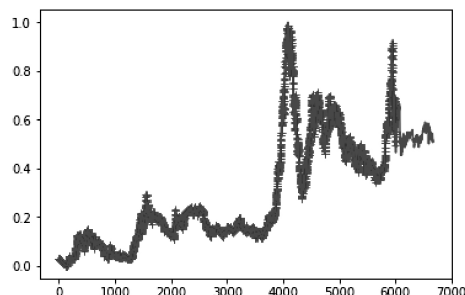


图 4 预测值与实际值的对比

### 3 结论

本文运用当下比较热门的人工智能模型 LSTM 深度学习对深证指数数据进行建模预测研究,建一个基于 LSTM 深度神经网络的股市时间序列预测模型应用在股票市场的预测中,其效果良好,具有一定的实践参考意义。

### 参考文献:

- [1]Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural computation, 12(10), 2451-2471
- [2]Graves, Alex. Supervised sequence labelling with recurrent neural networks. Vol. 385. Springer, 2012.

(上接第 21 页)

### 3.2 系统实现关键技术

#### 1. 基于位置的服务(LBS)

签到系统采用百度地图的定位 SDK, 利用手机的 GPS 信息、基站信息和 WiFi 信息完成学生位置的定位。在系统中集成百度地图 SDK, 通过定位 SDK 接口发送定位服务到百度定位服务<sup>[4]</sup>。

(1)在 application 中声明 service,用以定位。

(2)声明使用权限。

(3)设置 AccessKey。

(4)注册监听函数。

(5)开启/关闭服务。

(6)定位回调,位置信息在 onReceiveLocation 回调参数 location 中。

#### 2. 采用 ASP.NET 进行后台数据管理。

能够实现根据签到时间或姓名进行查询和排序,可以导出 Excel 文件,并打印。可以综合查询签到记录并生成报表数据,方便教师进行成绩的考核,摆脱手工管理模式。

### 4 总结

针对目前高校签到管理上存在的缺点,本文设计和实现一

个基于 Android 平台的高校签到管理系统,系统采用手机操作的方式,利用 GPS 定位技术实现高校学生签到的高效率管理,使高校学生考勤管理更加科学、有效。系统所采用的 Android 平台和 ASP.NET 技术实现签到系统的全部功能,且系统签到数据更加精准,维护更加方便,可以在各高校进行推广应用。

### 参考文献:

- [1]陈庆勤. 基于 Android 的学生考勤管理系统[D].南昌航空大学,2017.
- [2]基于 Android 系统的掌上班级课表与学生考勤管理系统[J]. 周紫瑞. 电子技术与软件工程. 2016(18) .
- [3]基于物联网技术的学生考勤管理系统应用[J]. 申玉宏. 科技创新与应用. 2015(30) .
- [4]基于 Android 平台下手机定位程序的设计及实现[J]. 代敏. 计算机与数字工程. 2012(04) .

### 作者简介:

林兴(1989-),男,福建永泰人,在读研究生,主要从事移动多媒体技术方面的研究,E-mail:995894393@qq.com。