

基于 LSTM 神经网络的短期价格趋势预测^①



邓飞燕, 岑少琪, 钟凤琪, 潘家辉

(华南师范大学 软件学院, 佛山 528225)

通讯作者: 潘家辉, E-mail: panjh82@qq.com

摘要: 本文主要对 LSTM 模型结构改进及优化其参数, 使其预测股票涨跌走势准确率明显提高, 同时对美股周数据及日数据在 LSTM 神经网络预测效果展开研究. 一方面通过分析对比两者预测效果差别, 验证不同数据集对预测效果的影响; 另一方面为 LSTM 股票预测研究提供数据集的选择建议, 以提高股票预测准确率. 本研究通过改进后的 LSTM 神经网络模型使用多序列股票预测方法来进行股票价格的涨跌趋势预测. 实验结果证实, 与日数据相比, 周数据的预测效果表现更优, 其中日数据的平均准确率为 52.8%, 而周数据的平均准确率为 58%, 使用周数据训练 LSTM 模型, 股票预测准确率更高.

关键词: 长短期记忆网络; 股票价格预测; 时间序列; 短期价格; LSTM

引用格式: 邓飞燕, 岑少琪, 钟凤琪, 潘家辉. 基于 LSTM 神经网络的短期价格趋势预测. 计算机系统应用, 2021, 30(4): 187-192. <http://www.c-s-a.org.cn/1003-3254/7855.html>

Short-Term Price Trend Forecast Based on LSTM Neural Network

DENG Fei-Yan, CEN Shao-Qi, ZHONG Feng-Qi, PAN Jia-Hui

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: This study mainly optimizes the structure and parameters of the LSTM model, so that the accuracy of predicting the trend in stock prices is significantly improved. Besides, we investigate the weekly and daily data on US stock in terms of predicting the LSTM neural network. On one hand, we compare the difference to verify the impact of different data sets on the forecast. On the other hand, we provide selection suggestions for data sets so as to increase the accuracy of stock forecast. This study uses the multi-time-series stock forecast in the improved LSTM model to predict the trend in stock prices. The results demonstrate that the weekly data perform better in forecast than daily data. To be specific, the average accuracy of daily data and weekly data is 52.8% and 58%, respectively. In summary, the application of weekly data to training the LSTM model yields higher accuracy in stock forecast.

Key words: Long Short-Term Memory (LSTM); stock price forecast; time series; short-term price; LSTM

1 引言

1.1 研究背景

随着经济全球化的发展, 美股市场愈发受到我国投资者的青睐, 伴随着金融市场的发展和强烈需求, 股

票价格趋势预测倍受学术界、业界关注. 众所周知, 投资股票具有高收益的同时, 也伴随着高风险, 而影响股市波动的内外因素众多且难以量化, 处理海量、繁杂的股票数据工程量大, 所以传统的非人工智能的方法

① 基金项目: 广州市科技计划重点领域研发计划 (202007030005); 广东省自然科学基金面上项目 (2019A1515011375); 国家自然科学基金面上项目 (61876067)

Foundation item: Key R & D Plan of Science and Technology Plan of Guangzhou Municipality (202007030005); General Program of Natural Science Foundation of Guangdong Province (2019A1515011375); General Program of National Natural Science Foundation of China (61876067)

收稿时间: 2020-08-02; 修改时间: 2020-08-28; 采用时间: 2020-09-08; csa 在线出版时间: 2021-03-30

在股票价格预测上往往不尽如人意. 因此, 如何利用大数据技术去发掘股票中隐含的众多有价值的信息, 运用 LSTM 等神经网络技术进一步解决股票价格趋势预测的问题, 具有非常重要的意义.

1.2 研究现状

长短期记忆网络 (Long Short-Term Memory, LSTM) 在时间序列上具有优异的表现, LSTM 神经网络的时间序列和选择记忆性等特性非常适用于股票价格随机的时间序列预测, 可以从大量复杂的股票历史数据中选择性学习有用的隐藏信息, 进而辅助决策^[1].

1997年, Hochreiter 和 Schmidhuber^[2]通过对无法刻画时间序列长记忆性的 RNN 进行改造, 提出了 LSTM 模型. 2015年, Chen 等^[3]采用 LSTM 对中国股市的收益率进行预测. 2016年, Jia 等^[4]验证了 LSTM 在预测股票价格走势上具有有效性.

在国内, 2016年孙瑞奇^[5]使用 LSTM 对美股股指进行预测, 并与 BP 神经网络和传统 RNN 模型做对比, 证实了 LSTM 的预测准确度更高, 获得了较低的误差均值: 0.783% 股指. 邓凤欣等^[6]利用 LSTM 模型对美港股票中市值较大的 4 支股票进行预测, 得出 4 支个股价格趋势预测的 RMSE 均处于 (0, 0.1) 的区间内, 预测结果显著. 认为 LSTM 在美港股市中具有较好的适用性, 但从实验结果上看存在着部分的时滞效应, 主要是作为一种案例分享而并未进行深入探讨. 2019年, 方红等^[7]提出了一种基于 LSTM 的改进型股票预测方法, 多维输入预测的输出结果为 $acc=0$, $lag=12$, 提高了预测准确度, 并改善了预测的滞后性.

根据研究表明, LSTM 神经网络对未来股票的价格趋势有一定的预测能力, 但没有关于短期价格趋势预测的研究, 比如短期价格趋势预测中日数据、周数据两种不同数据集对 LSTM 神经网络的预测效果影响的研究. 面对这些问题, 基于 LSTM 神经网络对周数据与日数据集这两个短期趋势预测效果对比的实证研究具有一定重要的现实意义^[8].

2 股票价格趋势预测基本原理与实现

2.1 LSTM 模型结构及原理

LSTM 结构采用控制门的机制, 由记忆细胞、输入门、输出门、遗忘门组成. LSTM 模型的各个控制门的计算原理如下^[9]:

(1) 输入门: 记忆现在的某些信息. 计算输入门 i_t 的

值和 t 时刻输入细胞的候选状态值 a_t :

$$i_t = \sigma(W_i \times (h_{t-1}, X_t) + b_i) \quad (1)$$

$$a_t = \tanh(W_c \times (h_{t-1}, X_t) + b_c) \quad (2)$$

其中, W_i , W_c 代表相应的权重, b_i 与 b_c 代表相应的偏置.

(2) 遗忘门: 控制舍去哪些信息. 计算在 t 时刻遗忘门的激活值 f_t :

$$f_t = \sigma(W_f \times (h_{t-1}, X_t) + b_f) \quad (3)$$

其中 W_f , b_f 分别表示遗忘门的权重和偏置, σ 表示 Sigmoid 函数.

(3) 细胞状态更新: 根据输入门和遗忘门的计算结果, 对细胞状态进行更新, 从而得出 t 时刻的细胞状态更新值 C_t :

$$C_t = i_t \times a_t + f_t \times C_{t-1} \quad (4)$$

(4) 输出门: 控制决定哪些信息需要输出. 根据计算得到的细胞状态更新值 C_t , 可以得到输出门的计算公式:

$$h_t = \sigma(W_o \times (h_{t-1}, X_t) + b_o) \times \tanh(C_t) \quad (5)$$

其中, W_o 和 b_o 代表输出门的权重和偏置, h_t 为当前单元的输出值.

2.2 LSTM 股票价格趋势预测模型搭建

在 Windows 操作系统下搭建 CPU 版本的 Keras 框架. Keras 框架具有模块化、简单、易扩展等优点. 因此, 实验采用 Keras 进行模型搭建. 其模型训练步骤如下:

(1) 数据预处理: 为避免乱序, 需进行排序, 获得规整的股票数据集.

(2) 数据降噪: 因市场动态非常复杂, 这些数据含有不经常的噪声, 所以采用 Python 中的 Pywt 库进行小波变换去除数据噪声, 图 1 和图 2 是 AAON 股票小波变换前后对比图.

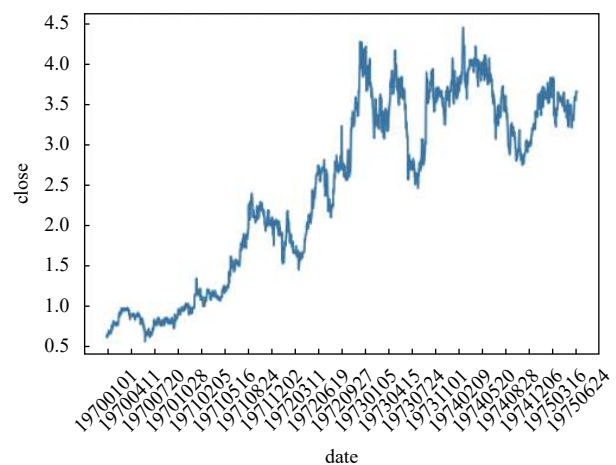


图 1 小波变换前

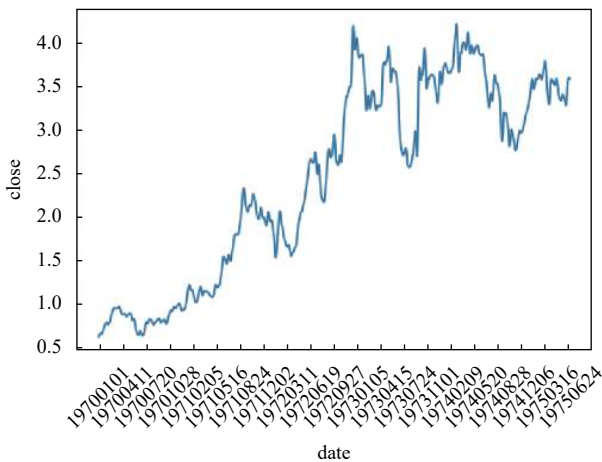


图2 小波变换三层分解后

(3) 数据归一化: 由于收盘价和交易量等参数同时作为特征值输入, 取值范围相差较大, 需避免数值过大的交易量对预测结果的影响过大, 需要进行归一化处理, 将数据的范围处理并控制在 0 到 1 之间, 数据归一化的公式如下:

$$y = \frac{x - MinValue}{MaxValue - MinValue} \quad (6)$$

其中, x 和 y 为转换前与转换后的值, $MaxValue$ 是样本的最大值, $MinValue$ 是样本的最小值。

(4) 数据分类: 开始训练前, 将数据集划分为训练集和测试集, 训练集占 85%, 测试集占 15%。

(5) 训练模型: 选取最低价 (low)、最高价 (high)、收盘价 (open)、开盘价 (close) 以及交易量 (volume) 作为特征数据输入, 配置 LSTM 模型的训练参数。配置完毕后, 调用模型训练函数。训练完后, 将训练好的模型存入 saved_models 文件夹中, 使用时可直接加载文件夹中保存的模型文件, 节省训练模型的时间。

(6) 股票预测: 采用多序列预测方法进行股票预测, 并将预测结果进行存储。

(7) 模型评估: 采用后述第 2.4 章节的模型评估方法, 计算评估指标。

(8) 参数优化: 运用控制变量法, 对不同参数选取适当的取值范围, 在训练的过程中不断调整参数, 直至模型的预测效果最佳。

经过多次调参及优化, 使 AAON 股票预测准确率从 55% 提高到 69%, 其改进前后预测结果如图 3、图 4 所示, 得到的最终模型结构及各项参数为: 测试窗口序列长度为 55, Sequential 模型包括 3 层 LSTM 层, 各层

的神经元个数分别为 100、100、150, 为避免过拟合, 增加 2 层 Dropout 层, Dropout 为 0.2, 输入数据维度为 5, 加入 Dense 层将其维度聚合为 1, 激活函数为 linear, 损失函数定为均方差 (Mean Squared Error, MSE), 优化算法采用 Adam, 模型采用 2 个 epochs, 每个 batch 的大小为 32, 得到的最终模型结构及参数设置详情分别如图 5、表 1、表 2 所示。

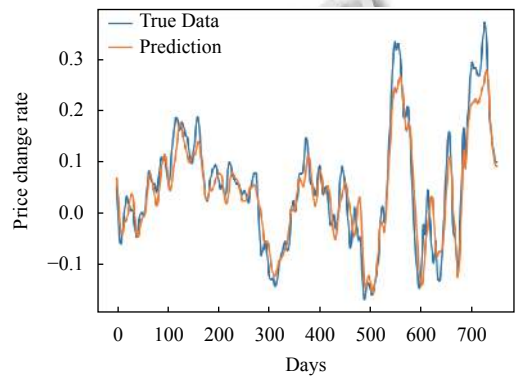


图3 模型改进前预测结果

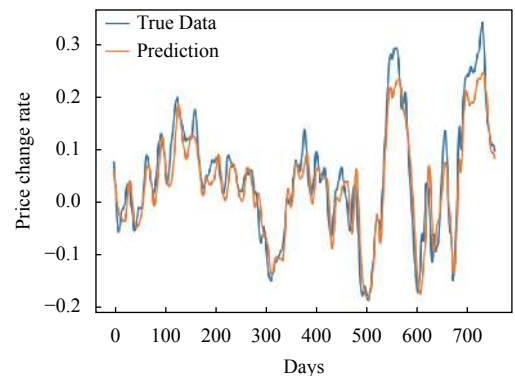


图4 模型改进后预测结果

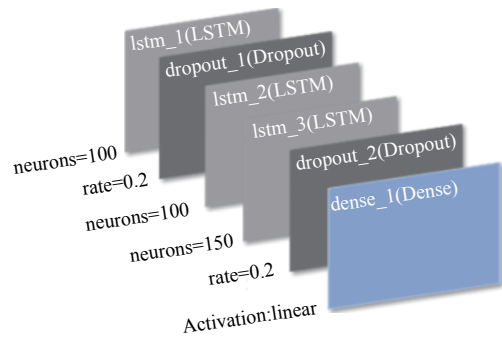


图5 基于 LSTM 的短期价格趋势预测模型层次结构

表1 模型参数设置详情

参数名	参数值
sequence_length	55
train_test_split	0.85
epochs	2
batch_size	32

表2 模型结构及其参数

模型结构	参数
loss	MSE
optimizer	Adam
LSTM一层	neurous = 100input_dim = 5return_seq = trueinput_timesteps = 54
Dropout层	rate = 0.2
LSTM二层	neurous = 100return_seq = true
LSTM三层	neurous = 150return_seq = false
Dropout层	rate = 0.2
Dense层	neurous = 1activation = linear

2.3 股票预测方法

本文搭建的股票价格预测模型采用多序列预测的方式. 首先将测试数据初始化为一定序列长度的测试窗口, 预测下一个点收盘价; 将预测得到的这个点收盘价加入到窗口中, 生成相同序列长度的新窗口, 循环往复. 当窗口中的数据全部由过去预测点组成时停止, 再以一个完整窗口长度向前移动, 用真实测试数据重置窗口, 再重新启动该过程.

相比完全序列预测, 多序列预测避免了一直使用过去的错误预测结果进行下一步预测, 可以有效提高预测准确率.

2.4 模型评估方法

模型的预测性能评价指标采用准确率 (*Accuracy*)、精确率 (*Precision*)、召回率 (*Recall*)、F1 值. 其计算流程如下:

首先, 计算收盘价涨幅, 判断涨跌, 公式如下:

$$\text{今日收盘价涨幅} = \frac{\text{今日收盘价} - \text{昨日收盘价}}{\text{昨日收盘价}} \quad (7)$$

根据涨幅正负值信息判断股票收盘价涨跌, 若涨幅大于 0, 则为股票上涨并设置标签为 1, 否则为 0. 通过这种方法可以实现股价涨跌预测^[10].

根据样本的真实结果和模型预测结果, 将样本分为 *TP* (样本与模型预测结果均为正类的个数)、*FN* (样本为正类, 预测结果为负类的个数)、*FP* (样本为负类, 预测结果为正类的个数)、*TN* (样本与模型预测结果均为负类的个数)4 类.

具体计算公式如下:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

模型评估的具体流程如图 6 所示.

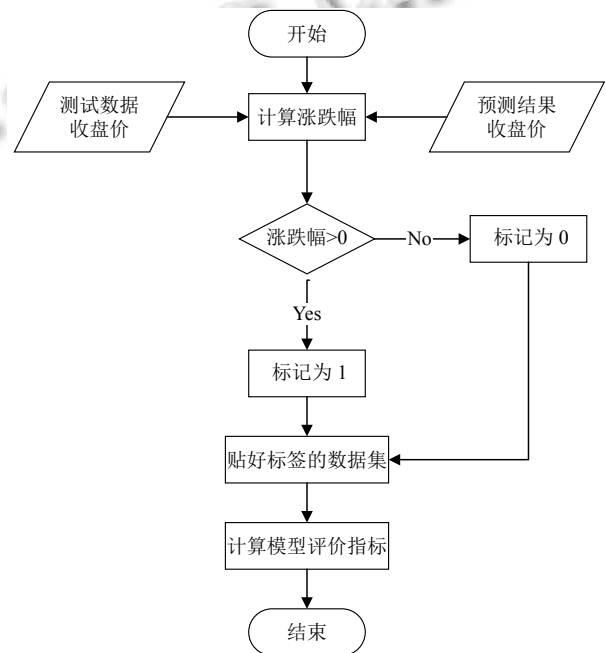


图 6 模型评估流程图

3 实验和结果

3.1 不同算法对比实验

实验选取 1998 年 1 月 9 日至 2019 年 5 月 31 日的 AAON 股票日数据为例进行研究, 预测第 2 日收盘价涨跌趋势. 数据特征属性为收盘价、开盘价、最高价、最低价、交易量, 对数据进行插值排序、降噪、归一化操作, 建立前述 LSTM 模型、并与支持向量机 (SVM)、随机森林 (random forest) 模型进行对比实验. 股票预测方法选取逐点预测方法.

1998 年 1 月 9 日至 2019 年 5 月 31 日的 AAON 股票日数据共计 5383 条, 其中 85% 作为训练数据, 15% 作为测试数据, LSTM 模型以前述 2.2 章节参数设置进行实验. 实验结果如表 3 所示: LSTM 模型各项评估指标均高于 SVM、随机森林. 可见, 本文改进后的

LSTM 模型预测效果优于其他模型。

3.2 周数据与日数据预测效果对比实验

(1) 实验对象: 美国纳斯达克股票市场 10 支股票, 其代码分别为 AAON、ABMD、ACHC、ACHV、ACIW、ACNB、AXAS、EGHT、SRCE、AXDX, 数据特征属性为收盘价、开盘价、最高价、最低价、交易量。

(2) 数据下载: 利用 pandas_datareader 中的 data 程序包下载上述 10 支股票的日数据, 周数据下载使用 Alpha Vantage API key 的 URL 地址进行爬取. 时间范围统一为 1998 年 1 月 9 日至 2019 年 5 月 31 日, 得到日数据每支股票的数据总量为 5383 条, 周数据每支股票数据总量为 1117 条。

(3) 实验流程: 按照数据下载、数据预处理、数据降噪、数据归一化、数据分类、LSTM 模型训练、股票预测、模型评估、记录实验结果这几个步骤完成。

具体实验流程如图 7 所示。

表 3 不同模型评估指标对比表

评估指标	LSTM	SVM	随机森林
准确率	0.67	0.56	0.53
精确率	0.67	0.61	0.29
召回率	0.67	0.56	0.53
F1	0.67	0.41	0.37

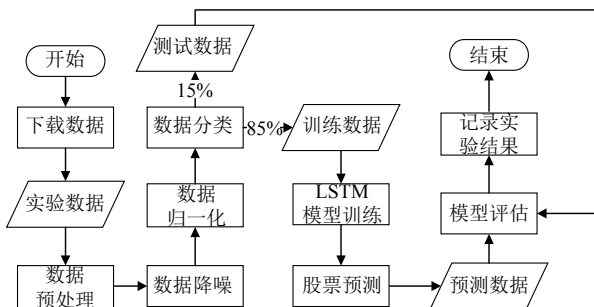


图 7 实验流程图

(4) 预测效果对比分析: 周数据与日数据均以前述第 2.2 节参数设置进行实验, 其中 85% 的数据划分为训练集, 剩余 15% 划分为测试集, 按照前述实验流程进行实验, 股票预测结果如表 4 和表 5 所示。

2019 年, 谢琪等^[11]使用多支指数的日数据进行测试多层 LSTM 神经网络的预测性能, 实验结果显示平均准确率为 47.33%, 平均精确率为 49.83%, 平均召回率为 63.5%, 平均 F1 值为 55.5%。与本实验相比, 其采用的是指数日数据集进行预测, 但预测效果表现不佳。通过表 6 和图 8 可发现, 与日数据相比, 周数据的各项指标均高于日数据, 其中日数据的平均准确率为 52.8%,

而周数据的平均准确率为 58%, 实证说明在短期价格趋势预测中, 与日数据相比, 周数据的预测效果表现更优, 使用周数据训练 LSTM 模型, 股票预测准确率有望进一步提高^[10]。

表 4 周数据预测结果表

股票代码	准确率	精确率	召回率	F1
AAON	0.70	0.73	0.69	0.66
ABMD	0.58	0.72	0.58	0.49
ACHC	0.50	0.51	0.50	0.48
ACHV	0.63	0.63	0.63	0.63
ACIW	0.60	0.76	0.60	0.46
ACNB	0.61	0.64	0.61	0.61
AXAS	0.61	0.61	0.61	0.57
EGHT	0.59	0.59	0.59	0.59
SRCE	0.50	0.50	0.50	0.49
AXDX	0.48	0.41	0.48	0.33
平均值	0.58	0.61	0.579	0.531

表 5 日数据预测结果表

股票代码	准确率	精确率	召回率	F1
AAON	0.53	0.52	0.53	0.50
ABMD	0.54	0.56	0.54	0.48
ACHC	0.53	0.58	0.53	0.43
ACHV	0.50	0.52	0.50	0.44
ACIW	0.55	0.55	0.55	0.55
ACNB	0.51	0.50	0.51	0.44
AXAS	0.49	0.53	0.49	0.34
EGHT	0.52	0.50	0.52	0.48
SRCE	0.56	0.55	0.56	0.54
AXDX	0.55	0.55	0.55	0.55
平均值	0.528	0.536	0.528	0.475

表 6 周数据与日数据评估指标对比表

评估指标	周数据	日数据
平均准确率	0.58	0.528
平均精确率	0.61	0.536
平均召回率	0.579	0.528
平均F1	0.531	0.475

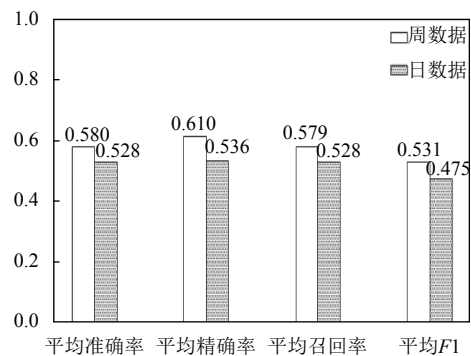


图 8 周数据与日数据评估指标对比图

4 总结

本文在 LSTM 模型的基础上进行了模型结构改进及参数优化,使得模型预测准确率提升 10% 以上,预测效果表现优于 SVM、随机森林。同时本文还随机选取了 10 支美股数据,进行了短期价格趋势预测中的周数据与日数据对 LSTM 模型预测效果对比实证研究,研究表明周数据预测效果更佳。对比其他文献研究,大部分研究者使用日数据或者更高频的数据,而较少使用周数据对股票进行预测。

在本次股票价格趋势预测模型上,不同的股票数据会对预测效果产生不同影响,因此数据集的选择也非常重要。根据本实验结果,说明周数据更适合用于股票预测。下一步的工作:本文只从收盘价、开盘价、最高价、最低价、交易量 5 个特征数据进行股票预测,股票预测的特征数据还可以增加其他数据,比如股市资金情绪特征、技术指标等,使用更多的特征数据进行股票预测,有望进一步提高股票预测的准确率。

参考文献

- 1 李春伟. 基于神经网络与遗传算法的股票智能预测 [硕士学位论文]. 西安: 西北工业大学, 2005. [doi: [10.7666/d.y711828](https://doi.org/10.7666/d.y711828)]
- 2 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)]
- 3 Chen K, Zhou Y, Dai FY. A LSTM-based method for stock returns prediction: A case study of China stock market. *Proceedings of 2015 IEEE International Conference on Big Data*. Santa Clara, CA, USA. 2015. 2823–2824. [doi: [10.1109/BigData.2015.7364089](https://doi.org/10.1109/BigData.2015.7364089)]
- 4 Jia H J. Investigation into the effectiveness of long short term memory networks for stock price prediction. arXiv: 1603.07893, 2016.
- 5 孙瑞奇. 基于 LSTM 神经网络的美股股指价格趋势预测模型的研究 [硕士学位论文]. 北京: 首都经济贸易大学, 2016.
- 6 邓凤欣, 王洪良. LSTM 神经网络在股票价格趋势预测中的应用——基于美港股票市场个股数据的研究. *金融经济*, 2018, (14): 96–98.
- 7 方红, 韩星煜, 徐涛. 改进型基于 LSTM 的股票预测方法. *安徽大学学报(自然科学版)*, 2019, 43(6): 36–42.
- 8 彭燕, 刘宇红, 张荣芬. 基于 LSTM 的股票价格预测建模与分析. *计算机工程与应用*, 2019, 55(11): 209–212. [doi: [10.3778/j.issn.1002-8331.1811-0239](https://doi.org/10.3778/j.issn.1002-8331.1811-0239)]
- 9 陈卫华. 基于深度学习的上证综指波动率预测效果比较研究. *统计与信息论坛*, 2018, 33(5): 99–106. [doi: [10.3969/j.issn.1007-3116.2018.05.015](https://doi.org/10.3969/j.issn.1007-3116.2018.05.015)]
- 10 肖琪. 人工神经网络在股票预测中的应用研究 [硕士学位论文]. 广州: 华南理工大学, 2017.
- 11 谢琪, 程耕国, 徐旭. 基于神经网络集成学习股票预测模型的研究. *计算机工程与应用*, 2019, 55(8): 238–243. [doi: [10.3778/j.issn.1002-8331.1807-0294](https://doi.org/10.3778/j.issn.1002-8331.1807-0294)]