

# 基于LSTM算法与股票历史数据的 股票价格预测

郭沛颖

云南师范大学云南纪检监察学院, 云南 昆明

收稿日期: 2023年7月22日; 录用日期: 2023年8月22日; 发布日期: 2023年8月29日

## 摘要

股票价格的准确预测一直是金融领域关注的焦点之一。传统的股票预测方法通常依赖于k线图、十字线等方法, 然而仅依靠传统方法分析前一天的价格来预测股票走向存在较大的局限性。为了进一步提升股票价格预测的精度和可靠性, 本文提出了一种基于长短时记忆神经网络(Long Short-Term Memory Networks, LSTM)的股票价格预测模型。通过LSTM算法处理时序数据的优势捕捉价格数据中的复杂模式和非线性关系, 挖掘股票历史价格数据中隐藏的波动信息, 为股票价格预测领域的研究和实践提供了新的思路和方法。此外, 为进一步获取最优的模型参数量, 本文以平安银行2019年1月至12月的开盘价和收盘价数据为样本进行实验。实验结果表明, 基于LSTM算法的股票价格预测准确率可达78.2%; 历史股票价格序列数为5时, 模型的预测精度最高。

## 关键词

平安银行, 股票预测, 长短时记忆神经网络(LSTM), 神经网络

## Stock Price Prediction Based on LSTM Algorithm with Historical Stock Data

Peiying Guo

Yunnan Institute of Discipline Inspection and Supervision, Yunnan Normal University, Kunming Yunnan

Received: Jul. 22<sup>nd</sup>, 2023; accepted: Aug. 22<sup>nd</sup>, 2023; published: Aug. 29<sup>th</sup>, 2023

## Abstract

Accurate prediction of stock prices has always been a focal point in the financial sector. Traditional stock prediction methods often rely on techniques such as k-line charts and crosshairs. However,

relying solely on these traditional methods to analyze the previous day's price for predicting the direction of a stock is subject to significant limitations. To enhance the accuracy and reliability of stock price prediction, this study puts forth a model based on Long Short-Term Memory Networks (LSTM). By leveraging the capabilities of the LSTM algorithm for processing time-series data and capturing complex patterns and nonlinear relationships in price data, this study uncovers hidden fluctuation information in historical stock price data, offering novel insights and methodologies for research and practice in the field of stock price prediction. Additionally, to determine the optimal number of model parameters, this study utilizes the opening and closing price data of Ping An Bank for the period between January and December 2019 as experimental samples. The experimental results demonstrate that the LSTM algorithm achieves a stock price prediction accuracy of 78.2%. Moreover, the model exhibits the highest precision when utilizing 5 historical stock price sequences.

## Keywords

Ping An Bank, Stock Prediction, Long and Short Term Memory Neural Network (LSTM), Neural Network

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

自股票诞生以来, 由于股票市场的高回报性, 股票价格一直备受关注。研究者们希望通过科学的研究方法, 对股票的价格走势进行预测。但由于股票市场受多方面因素的影响, 其非线性、随机性的特点也给股票价格的预测带来了巨大的挑战。

传统的股票分析主要分为技术分析和价格趋势预测[1]。其中, 技术分析主要通过股票的历史价格对证券市场的变化因素进行分析, 通过股价、成交量等因素建立股价模型, 使用模型及图表来把握市场的整体情况, 从而对未来的价格趋势进行预测。价格趋势预测基于市场未来股息的贴现值。这种方法试图找到股票的基本价值, 并确定股票的价格是否合理。价格过低的股票触发买入信号, 价格过高的股票触发卖出信号, 价格合理的股票触发持有信号。ARIMA [2] [3]是价格趋势预测的重要方法, 许多学者对基于 ARIMA 算法的股票价格预测进行了优化工作, 如 Zheng [4], Rangel-Gonzalez [5], 宋刚[6], 丁文绢[7], 熊政[8], 徐幼恩[9], 陈登建[10]等。

随着机器学习、大数据等技术的发展, 股票市场的海量数据引起了研究者的关注。基于机器学习的方法, 如神经网络[11] [12], 支持向量机[13]等方法在处理大数据时更具优越性, 可充分发掘股票市场中的隐藏信息, 解决了传统股票预测方法的局限性。基于机器学习的股票预测方法将股票市场中的大量历史数据结合机器学习算法进行建模并训练, 通过训练好的模型对股票的未来价格走势及具体价格进行预测[14]。与传统方法相比, 该方法可大幅度提高股票价格预测的准确率, 具有重要的研究价值和现实意义。研究表明, 基于机器学习的算法是股票价格预测领域最好的方法[15]。但目前基于机器学习的股票预测研究仍较少, 且方法过于单一。

LSTM 算法是为时间序列分析而设计的神经网络, 可克服梯度消失、梯度爆炸的问题, 通过记忆模块来学习时间序列的长期依赖关系。文献[16] [17]指出, 在众多的神经网络算法中, LSTM 算法更适用于股票价格预测。因此, 本文以股票历史价格研究为背景, 提出一种基于 LSTM 的股票价格预测方法。

## 2. 方法

### 2.1. LSTM 网络

LSTM 是一种基于循环神经网络(RNN)的特殊时间序列预测网络,可学习序列的依赖关系,被广泛的应用于语音信号处理和自然语言处理等领域。RNN 网络可循环利用模型中的权重,在一定程度上可以利用到模型的历史信息。然而,RNN 模型的反向传播算法较为简单,重复利用历史权重会导致梯度消失、梯度爆炸等问题,无法有效学习序列的历史信息。与 RNN 不同,该网络增加了输入门、遗忘门和输出门,可缓解时间序列预测中的梯度消失和梯度爆炸问题,有效的解决了 RNN 的不足。

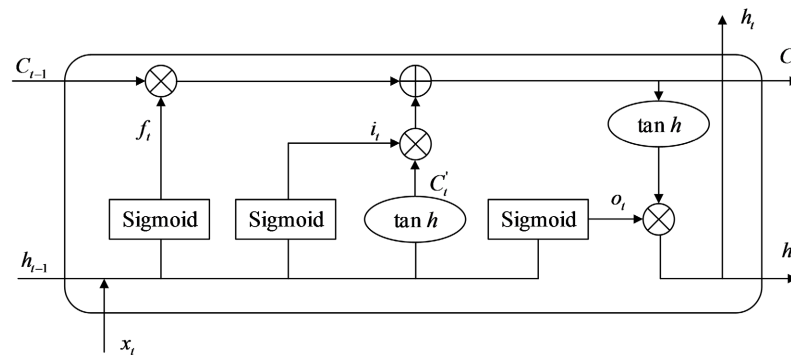


Figure 1. Structure of LSTM network  
图 1. LSTM 网络结构图

图 1 为 LSTM 的网络结构图。LSTM 网络一共有 4 层,包含一个输入序列  $x_t$ , 两个输出  $h_t$  和  $C_t$ 。其中,  $h_t$  为当前单元的输出,  $h_{t-1}$  为上一个单元的输出。sigmoid 和 tanh 为激活函数,  $\otimes$  为序列元素相乘,  $\oplus$  为序列元素相加,  $C_t$  为  $t$  时刻神经元的状态,  $f_t$  为遗忘门的遗忘阈值,可控制神经元的历史信息输入量。 $i_t$  为输入门的阈值,可控制当前输入神经元的时序信息量。 $o_t$  为输出阈值,可控制 sigmoid 输出神经元的状态信息。输入序列  $x_t$  与前一时刻的单元输出  $h_{t-1}$  进行融合,首先输入至遗忘门,通过 sigmoid 激活函数初步对特征进行提取,并在遗忘阈值  $f_t$  的调节下,与前一时刻神经元的状态  $C_{t-1}$  通过  $\otimes$  运算,得到遗忘门的输出状态。其次,输入至输入门,通过 sigmoid 和输入阈值  $i_t$  对输入的时间序列进行特征提取,并与 tanh 激活函数层的输出状态通过  $\otimes$  运算将序列元素进行对应相乘,最终与遗忘门的输出通过  $\oplus$  运算按序列元素相加,得到当前时刻的神经元状态  $C_t$ 。此外,输入序列在输出门通过 sigmoid 和 tanh 激活函数进行特征提取,通过输出阈值  $o_t$  控制输出信息,并通过  $\otimes$  运算与  $\tanh(C_t)$  按序列元素相乘,得到当前时刻的神经元输出  $h_t$ 。该过程可用公式(1)~(4)表示。

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

$$i_t = \text{sigmoid}(W_i \cdot [h_{t-1}, x_t] + c_i) \quad (2)$$

$$C'_t = \tanh(W_c \cdot [h_{t-1}, x_t] + c_c) \quad (3)$$

$$o_t = \text{sigmoid}(W_o \cdot [h_{t-1}, x_t] + c_o) \quad (4)$$

其中,  $W_f$ ,  $W_i$ ,  $W_c$ ,  $W_o$  分别为遗忘门,输入门,激活函数层和输出门的权重系数矩阵,  $c_f$ ,  $c_i$ ,  $c_c$ ,  $c_o$  分别为遗忘门,输入门,激活函数层和输出门的偏移系数矩阵。因此,神经元的状态  $C_t$  和输出  $h_t$  可表示为:

$$C_t = (f_t \otimes C_{t-1}) \oplus (i_t \otimes C'_t) \quad (5)$$

$$h_t = o_t \otimes (\tanh(C_t)) \quad (6)$$

与 RNN 不同, LSTM 通过三个门控制机制有效的发掘时间序列的历史信息, 并与当前序列信息相融合, 充分的利用了输入数据, 并对输入数据形成长期记忆。此外, LSTM 在通过反向传播对权重信息进行校正时, 一部分误差信息会通过输入门和激活函数层直接传递给下一个神经元, 另一部分则会通过遗忘门筛选掉部分信息, 有效的避免了梯度爆炸、梯度消失等问题, 在处理具有冗余信息的历史数据等问题时具有极大的优势。本文研究的股票价格预测是典型的具有冗余信息的时间序列问题, 当前时刻  $t$  的股价会受历史股价的影响。因此, 本文采用 LSTM 网络对股票价格进行预测。

## 2.2. 基于 LSTM 的股票价格预测模型

基于上述原理, 本文通过构建 LSTM 神经网络模型来对股票价格进行预测, 模型训练及预测的流程图如图 2 所示。

本文构建的股票价格预测系统首先对输入的历史股价数据进行预处理, 提取出“开盘价”和“收盘价”信息并将其转换为时间序列。为了减小计算量、加快模型的求解速度, 将不同的尺度的股价转换到特定的数据分布, 采用了数据归一化的方式对时间序列进行处理。然后对模型的参数进行设置, 包括训练次数、训练集和测试集的划分、单次训练的历史数据量、批次大小等。之后将预处理的数据输入至本文构建的 LSTM 模型进行训练, 采用 Adam 优化器使模型适应稀疏梯度, 动态更新模型的权重系数矩阵。此外, 还采用反向传播的方法将最优的历史权重系数矩阵传输至当前状态模型, 既使当前的权重信息保持最优, 又可包含训练数据的历史状态信息。训练完成后, 将测试数据输入至模型即可得到当前数据的评价指标, 通过预测数据可视化等操作即可完成模型的评估工作。

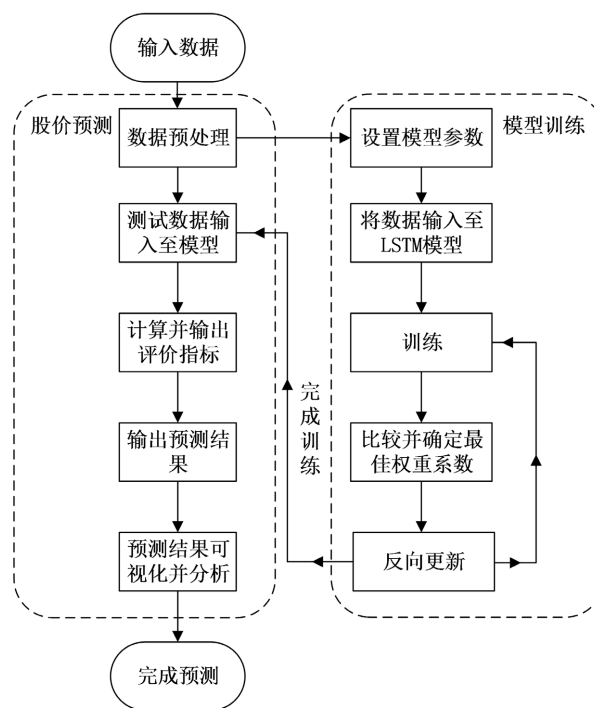


Figure 2. LSTM stock price prediction flowchart

图 2. LSTM 股票价格预测流程图

图 3 为本文构建的股票价格预测模型，包括四个全连接层(fully connected layer)，两个 LSTM 层(LSTM layer 1, LSTM layer 2)，两个 Dropout 层。该模型的核心思想是通过 LSTM 层的记忆单元(memory cell)和门控机制(gate mechanism)，帮助网络有效地处理长期和短期依赖关系。其中，记忆单元是网络中的核心组件，可以存储和传递信息。门控机制由三个关键的门控单元组成，即遗忘门(forget gate)、输入门(input gate)和输出门(output gate)。在预测股票价格时，LSTM 模型将历史的股票价格数据序列作为输入，通过记忆单元和门控机制来学习和捕捉输入序列中的模式和规律。首先，遗忘门通过一个 sigmoid 函数来决定是否从记忆单元中忘记之前的记忆。接下来，输入门根据当前输入数据和先前的记忆状态来更新和存储新的记忆。最后，输出门根据当前输入数据和先前的记忆状态来决定输出的值。

通过 LSTM 模型，可以有效地捕捉和利用股票价格数据中的非线性关系和复杂模式，提高股票价格预测的准确性和可靠性。该模型可以处理不同时间尺度上的数据，并能够有效地处理长期和短期的影响因素，从而更好地捕捉市场的动态变化。

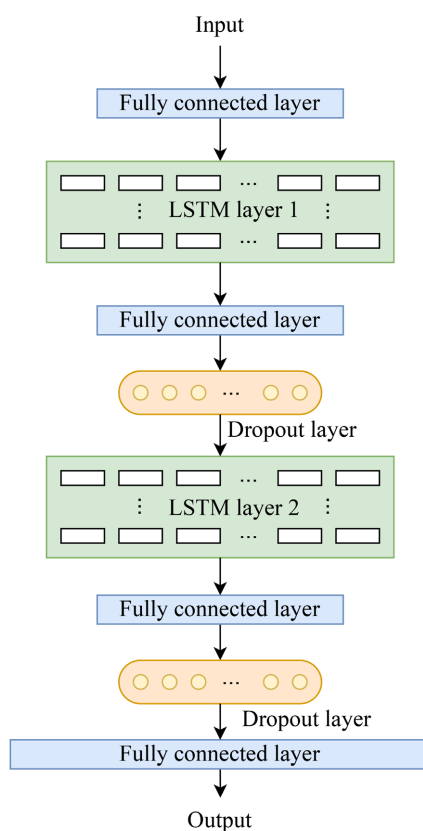


Figure 3. Stock price forecasting model

图 3. 股票价格预测模型

### 3. 实验及结果分析

#### 3.1. 数据来源及评价指标

本文采用的数据为平安银行(0000001) 2019 年 1 月~12 月的最高价、最低价、开盘价、收盘价、成交量等数据(数据来源于通达信金融终端)，一共有 1952 条。选取开盘价、收盘价作为输入时间序列数据，80%作为训练集来对 LSTM 模型进行训练，20%作为测试集来测试模型的预测准确度和模型的泛化能力。

目前在股票价格预测领域，并没有通用且统一的评估标准。为了显示模型的预测效果，本文采用均方误差(Mean Squared Error, MSE)、均方对数误差(MSLE)、平均绝对误差(MAE)、预测准确率(Accuracy)来对模型进行数值评估；采用图表的方式来对预测结果进行可视化评估。公式定义如下：

$$\text{MSE} = \frac{1}{n} \sum_{n=1}^N (y'_n - y_n)^2 \quad (7)$$

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^N |y'_n - y_n| \quad (8)$$

$$\text{MSLE} = \frac{1}{n} \sum_{n=1}^N ((\log y'_n + 1) - (\log y_n + 1))^2 \quad (9)$$

其中， $N$  为输入样本数， $y'_n$  为预测值， $y_n$  为真实值。

### 3.2. 实验及结果分析

本实验的环境配置如表 1 所示。所有实验均在此平台上进行。

**Table 1.** Experimental environment configuration  
**表 1.** 实验环境配置

操作系统	Ubuntu18.04
CPU	I7-9700K
内存	32G
显卡	RTX 2080Ti
深度学习框架	Tensorflow
编译语言	python 3.7.8

影响股价预测模型的主要因素有两个：

1) LSTM 层输入不同数量的历史数据。LSTM 层可通过历史数据的权重系数矩阵来学习序列的历史信息。通常情况下，在一定的范围内，历史数据的数量越多，模型的性能越准确。但增加历史数据的数量，并不会无限制的提升模型的性能。因此，本文对不同数量的历史数据对模型的影响进行了研究。

2) LSTM 层的神经元模块数。如图 3 所示，模型有两个 LSTM 层，每个 LSTM 层有一定模块数。在训练数据有限的情况下，提升 LSTM 的模块数会造成梯度不饱和等问题，影响模型对股价的评估。因此，本文对 LSTM 层的最优模块数进行了研究。

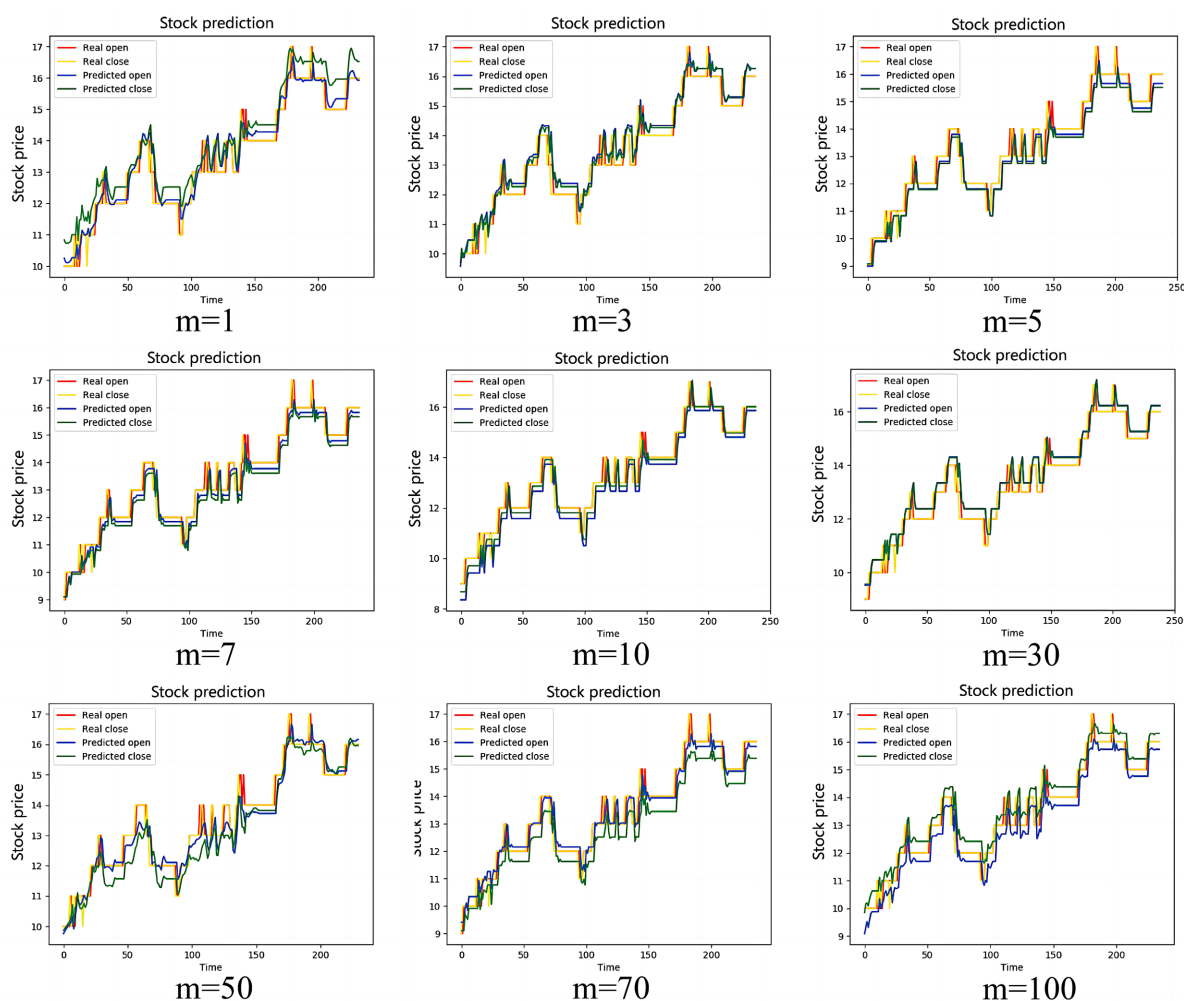
#### 3.2.1. 不同数量的历史数据对模型的影响

为了比较不同数量的历史数据(m)对模型的影响，该节分别选取 1、3、5、7、10、30、50、70、100 组数据进行实验，迭代次数为 1000 次，分别选取 MSE、MAE、MSLE、Accuracy 这四个指标对模型进行评估。表 2 为实验性能参数。

由表 2 可知，当输入的历史数据量小于 5 时，模型的准确率随着输入数据量的增加而增加，在输入数据量为 5 时达到最大。在输入的历史数据量大于 10 以后，模型的准确率趋于稳定，并随着输入数据量的增加而减少。模型的 MSE、MAE、MSLE 可反映模型在训练过程中的收敛速度。在输入历史数据量为 100 时，MSE、MAE 达到最小值，这说明输入更多的历史数据量可加快模型的收敛速度。但过快的收敛并未使模型性能达到最优。

**Table 2.** Performance parameters of the prediction model with input of different amount of historical data (m)  
**表 2.** 预测模型输入不同数量历史数据(m)的性能参数

m	MSE	MAE	MSLE	Accuracy
1	0.0073	0.067	0.0048	0.54
3	0.0077	0.07	0.0041	0.67
5	0.0083	0.0643	0.0038	<b>0.7532</b>
7	0.0039	0.0435	0.0039	0.6481
10	0.0034	0.039	0.0011	0.5739
30	0.0032	0.0469	0.0018	0.6311
50	0.002	0.0339	0.0012	0.6243
70	0.0024	0.0394	<b>0.001</b>	0.5706
100	<b>0.0015</b>	<b>0.0291</b>	0.0026	0.5714



**Figure 4.** Results of stock price prediction using different amount of historical data (m)  
**图 4.** 采用不同数量历史数据(m)的股价预测结果

图 4 为采用不同数量历史数据(m)时, 模型对股价预测结果可视图。当  $m = 1$  时, 模型相当于 RNN, 无法充分学习序列的历史信息, 预测结果与实际结果有较大的误差。增大  $m$  之后, 模型的预测能力有了

明显提升。当  $m = 5$  时，预测曲线与实际曲线几乎重合，达到最优。继续增大  $m$  值，虽然模型的正均方差 (MSE)、平均绝对误差 (MAE)、均方根误差 (MSLE) 均有下降的趋势，但预测效果越来越差。综合所有的评价指标，当输入历史数据量为 5 时，模型达到最优性能。

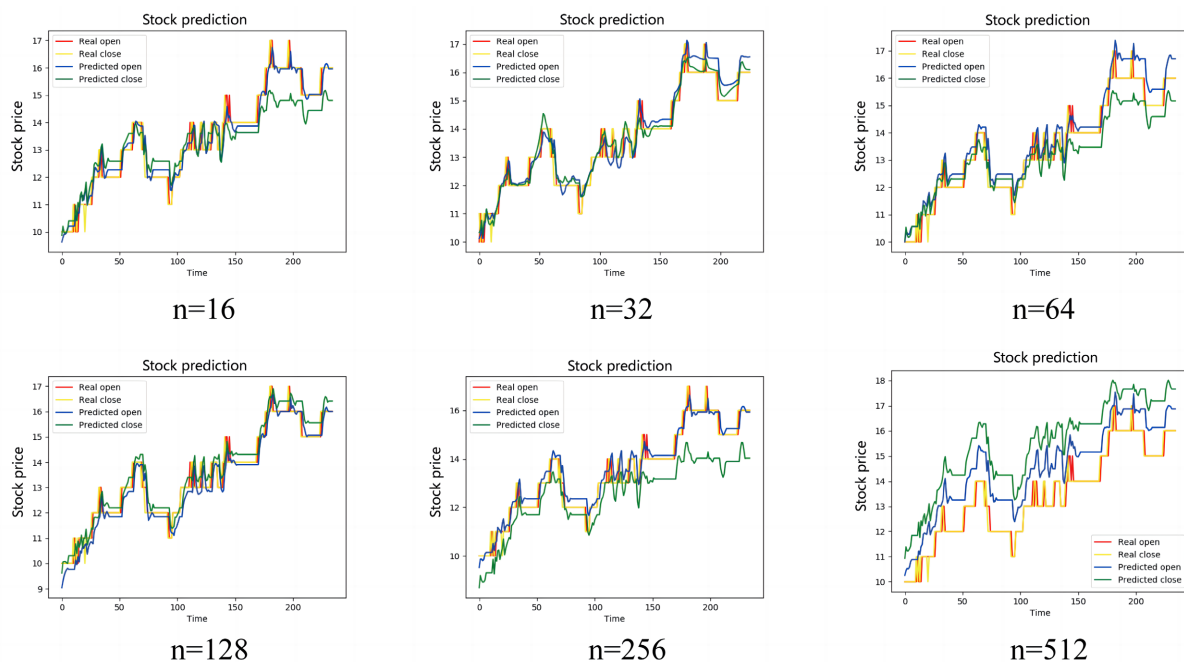
### 3.2.2. 不同的模型参数对模型的影响

为了比较 LSTM 中不同数量的神经元模块 ( $n$ ) 对模型性能的影响，该节分别选取 16、32、64、128、256、512 个模块，进一步研究使模型达到最优的神经元模块数量。采用 3.2.1 的实验结果，将输入历史数据量设置为 5，模型的其他参数与 3.2.1 保持一致，表 3 为实验结果。

**Table 3.** Model performance parameters using different number of neuron modules ( $n$ )

**表 3.** 采用不同数量神经元模块 ( $n$ ) 的模型性能参数

$n$	MSE	MAE	MSLE	Accuracy
16	0.0033	0.0473	0.0014	0.5106
32	0.0024	0.0411	0.0016	0.5787
64	0.0025	0.0358	0.0020	0.6332
128	0.0026	0.0352	0.0024	0.7817
256	0.0034	0.0503	0.0021	0.5404
512	0.0036	0.0543	0.0024	0.4340



**Figure 5.** Stock price prediction results using different number of neuron modules ( $n$ )

**图 5.** 采用不同数量神经元模块 ( $n$ ) 的股价预测结果

实验结果显示，当神经元模块数量由 16 增加至 128 时，模型的正均方差 (MSE)、平均绝对误差 (MAE) 呈下降趋势，准确率在 128 时达到最优值。当模型的神经元由 128 增长至 512 时，MSE、MAE 呈上升趋势，准确率也在急剧下降。当神经元数量为 32~128 时，模型的正均方差 (MSE) 和平均绝对误差 (MAE) 值趋于稳定状态。这说明模型在该阶段达到了最优。此外，模型的均方根误差 (MSLE) 随着神经元的增加呈上升趋势。将表 3 的实验数据与图 5 结合分析，神经元数



量并不是越多越好。当神经元的数量为 128 时, 模型的性能达到最优。进一步增加神经元的数量, 会造成模型过拟合, 导致模型性能急剧下降。

#### 4. 结论

本文提出了一种基于 LSTM 算法的股票价格预测方法。与传统的股票预测方法不同, 基于深度学习的股价预测可发掘更多历史股票数据的信息, 仅依靠历史股票价格数据即可预测未来股价。针对股票价格预测研究较少、并没有统一的评价指标等问题, 本文选取了均方误差(Mean Squared Error, MSE)、均方对数误差(MSLE)、平均绝对误差(MAE)、预测准确率(Accuracy)来对模型进行数值评估。此外, 本文还对 LSTM 层输入不同数量的历史数据、LSTM 层的神经元模块数两个方面对股票价格预测模型进行进一步研究, 实验结果证明了本文提出方法的有效性。

但仅依靠历史股价来对未来的股价进行预测仍存在较大的局限性, 这也导致模型的准确率并不太高。因此, 在下一步的研究工作中, 可将市场上多种影响股票价格因素(如大户资金流向, 买入卖出信号, 股价拉升拉回信号等)结合情感分析算法进行建模, 获取更精准的股票价格预测模型。

#### 参考文献

- [1] Harel, A. and Harpaz, G. (2021) Forecasting Stock Prices. *International Review of Economics & Finance*, **73**, 249-256. <https://doi.org/10.1016/j.iref.2020.12.033>
- [2] Mondal, P., Shit, L. and Goswami, S. (2014) Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications*, **4**, 13-29. <https://doi.org/10.5121/ijcsea.2014.4202>
- [3] Rao, T.S. and Gabr, M.M. (2012) *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer, New York.
- [4] Zheng, T.S., Farrish, J. and Kitterlin, M. (2016) Performance Trends of Hotels and Casino Hotels through the Recession: An ARIMA with Intervention Analysis of Stock Indices. *Journal of Hospitality Marketing & Management*, **25**, 49-68. <https://doi.org/10.1080/19368623.2015.970725>
- [5] Rangel-Gonzalez, J.A., Frausto-Solis, J., González-Barbosa, J.J., Pazos-Rangel, R.A. and Fraire-Huacuja, H.J. (2018) Comparative Study of ARIMA Methods for Forecasting Time Series of the Mexican Stock Exchange. In: Castillo, O., Melin, P. and Kacprzyk, J., Eds., *Fuzzy Logic Augmentation of Neural and Optimization Algorithms: Theoretical Aspects and Real Applications*, Springer, Cham, 475-485. [https://doi.org/10.1007/978-3-319-71008-2\\_34](https://doi.org/10.1007/978-3-319-71008-2_34)
- [6] 宋刚, 张云峰, 包芳勋, 等. 基于粒子群优化 LSTM 的股票预测模型[J]. 北京航空航天大学学报, 2019, 45(12): 2533-2542.
- [7] 丁文娟. 基于股票预测的 ARIMA 模型、LSTM 模型比较[J]. 工业控制计算机, 2021, 34(7): 109-112, 116.
- [8] 熊政, 车文刚. ARIMA-GARCH-M 模型在短期股票预测中的应用[J]. 陕西理工大学学报(自然科学版), 2022, 38(4): 69-74.
- [9] 徐幼恩, 罗扬. 基于组合模型的股票预测研究综述[J]. 科技和产业, 2022, 22(11): 214-220.
- [10] 陈登建, 杜飞霞, 夏换. 基于 ARIMA 与 SVR 滚动残差模型组合的股票预测[J]. 计算机时代, 2022(5): 76-81. <https://doi.org/10.16644/j.cnki.cn33-1094/tp.2022.05.019>
- [11] 黄明星, 姚晨. PSO-BP 神经网络模型在股票预测中的应用[J]. 福建电脑, 2023, 39(3): 36-40. <https://doi.org/10.16707/j.cnki.fjpc.2023.03.007>
- [12] 张如梦, 张华美. BP 神经网络与 ARMA-GARCH 模型在股票预测中的对比分析[J]. 高师理科学刊, 2021, 41(5): 14-20.
- [13] Bao, Y., Lu, Y. and Zhang, J. (2004) Forecasting Stock Price by SVMs Regression. In: Bussler, C. and Fensel, D., Eds., *Artificial Intelligence: Methodology, Systems, and Applications*, Springer, Berlin, 295-303. [https://doi.org/10.1007/978-3-540-30106-6\\_30](https://doi.org/10.1007/978-3-540-30106-6_30)
- [14] 徐浩然, 许波, 徐可文. 机器学习在股票预测中的应用综述[J]. 计算机工程与应用, 2020, 56(12): 19-24.
- [15] 孔宇萍. 基于时间序列与机器算法的工商银行股票预测分析[J]. 中国管理信息化, 2023, 26(6): 146-148.
- [16] Chen, K., Zhou, Y. and Dai, F. (2015) A LSTM-Based Method for Stock Returns Prediction: A Case Study of China

Stock Market. 2015 *IEEE International Conference on Big Data (Big Data)*, Santa Clara, 29 October-1 November 2015, 2823-2824. <https://doi.org/10.1109/BigData.2015.7364089>

- [17] Fischer, T. and Krauss, C. (2017) Deep learning with Long Short Term Memory Networks for Financial Market Predictions. *European Journal of Operational Research*, **270**, 654-669. <https://doi.org/10.1016/j.ejor.2017.11.054>