

基于Transformer-LSTM模型的股票预测

马昱晗*, 孙德山*, 代雨鑫

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2025年4月26日; 录用日期: 2025年5月19日; 发布日期: 2025年5月28日

摘要

近年来, 全球经济金融环境动荡不安, 充满挑战与风险, 在此情况下, 投资者在金融交易方面的抉择更加复杂, 股票交易市场也面临着更大的挑战, 传统的计量经济学模型难以充分适应此类变化。本文运用深度学习中的长短期记忆(LSTM)神经网络作为基本模型对股票数据的收盘价进行预测, 选用京粮控股(000505)、东北制药(000597)和中钢国际(000928)三组数据, 并在LSTM神经网络中引入Transformer模型。经实验, 该模型的预测精度有明显提升, 说明Transformer-LSTM网络模型在股票预测领域具有一定的可靠性。

关键词

神经网络, LSTM, Transformer, 股票预测

Stock Prediction Based on Transformer-LSTM Model

Yuhan Ma*, Deshan Sun*, Yuxin Dai

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: Apr. 26th, 2025; accepted: May 19th, 2025; published: May 28th, 2025

Abstract

In recent years, the global economic and financial environment has been volatile and full of challenges and risks. In this situation, investors' choices in financial trading have become more complex, and the stock trading market is also facing greater challenges. Traditional econometric models are difficult to fully adapt to such changes. This article uses the Long Short-Term Memory (LSTM) neural network in deep learning as the basic model to predict the closing price of stock data. Three sets of data from Jingliang Holdings (000505), Northeast Pharmaceutical (000597), and Zhonggang International

*通讯作者。

文章引用: 马昱晗, 孙德山, 代雨鑫. 基于 Transformer-LSTM 模型的股票预测[J]. 应用数学进展, 2025, 14(5): 387-393.
DOI: 10.12677/aam.2025.145268

(000928) are selected, and the Transformer model is introduced into the LSTM neural network. Through experiments, the prediction accuracy of the model has been significantly improved, indicating that the Transformer LSTM network model has certain reliability in the field of stock prediction.

Keywords

Neural Network, LSTM, Transformer, Stock Prediction

Copyright © 2025 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

从 1997 年, IBM 的深蓝(Deep Blue)超算程序战胜了世界象棋冠军加里·卡斯帕罗夫, 到 2012 年, 标志着深度学习兴起的 AlexNet 深度神经网络在 ImageNet 图像识别竞赛中获得巨大突破, 再到 2015 年, 谷歌的自动驾驶汽车首次成功通过完全无人驾驶测试, 机器学习的快速发展, 为人工智能领域的发展提供了坚实基础。因物联网、社交媒体、在线交易等各行行业的火爆, 时间序列类数据数量急剧增加, 从海量时序数据中挖掘潜在的有效信息并明晰数据之间的关系, 具有极大的学术价值与现实意义。深度学习因其强大的建模能力、对大数据的处理能力以及在各类金融任务中的应用潜力吸引了大量金融行业研究者的注意。从市场预测到算法交易、风险管理, 再到投资组合优化, 深度学习为金融领域提供了许多创新的方法和工具, 也极大地推动了金融科技的变革[1]。股票作为“经济健康的晴雨表”, 作为金融行业中的重要组成部分, 也间接影响到我们的日常生活。

在机器学习中, 循环神经网络对具有时间依赖性的序列数据有很好的处理效果, 但当序列长度增加时就会产生梯度消失或梯度爆炸的问题[2]。长短期记忆神经网络 LSTM 是专门针对这一问题而提出的改进模型, 在金融数据方面有广泛的应用[3]。杨瑞奇针对美股股指进行 BP、RNN、LSTM 三种神经网络模型训练, 分析对比得知模型效果逐步提升, LSTM 网络表现较显著[4]。赵红蕊和薛雷结合长短期神经网络、卷积神经网络和注意力机制模块(Convolutional Block Attention Module, CBAM)成为深度学习模型, 针对上证指数数据进行价格预测, 验证了结合模型的有效性 with 可行性, 进一步提高了股票的预测性能[5]。

本文将 Transformer 模型与长短期记忆网络相结合, 建立 Transformer-LSTM 模型, 选取 2018~2024 年间三只股票的收盘价作为实证分析, 并与 LSTM 神经网络进行对比, 得到了较好的结果。

2. 长短期记忆网络与 Transformer 模型

2.1. 长短期记忆网络

长短期记忆网络(Long Short-Term Memory, LSTM)是一种特殊类型的循环神经网络(Recurrent Neural Network, RNN), 具有一定的记忆功能, 能够更好地处理序列中的长期依赖关系。LSTM 由多个组件组成, 其中最核心的部分是三个门: 输入门(Input Gate)、遗忘门(Forget Gate)和输出门(Output Gate)。其中, 输入门决定当前输入的信息有多少会传递到细胞状态; 遗忘门决定当前细胞状态中的信息有多少需要被“遗忘”; 输出门决定当前细胞状态中有多少信息将输出。每个门都由一个激活函数控制, 它的输出在 0 到 1 之间, 表示信息的保留程度[6]。神经元具体结构见图 1, 其中 Cell 是细胞单元。

在该模型中, 每个记忆单元都要从输入开始训练完整的三套权重, 包括前一时间步中的所有隐藏状

态。信息进入单元后，会同时流经遗忘门、输入门、输出门，这些门通过不同的计算和更新机制，决定了细胞状态和隐含状态。每个门都独立计算，再反馈到本身。具体公式如下：

$$\text{输入门: } i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (1)$$

$$\text{遗忘门: } f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (2)$$

$$\text{输出门: } o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (3)$$

$$\text{循环层状态的输出值: } h_t = o_t \odot \tanh(c_t) \quad (4)$$

其中， W 与 b 分别为三个门的权重矩阵与偏置项，通过训练得到； c_t 为细胞单元在 t 时刻维持的记忆值，更新公式为 $c_t = f \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$ 。

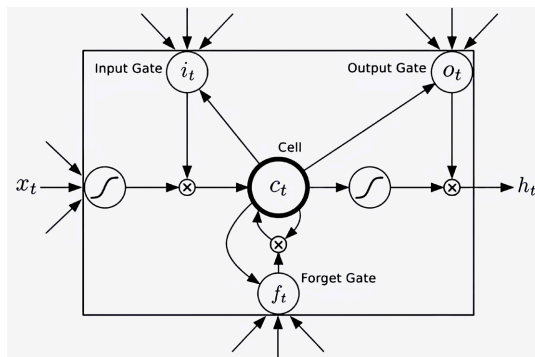


Figure 1. LSTM neural network structure
图 1. LSTM 神经网络结构

2.2. Transformer 模型

Transformer 模型是一种用于自然语言处理和其他序列任务的深度学习模型架构，核心创新在于引入了自注意力机制，使得每个输入的元素在处理时可以考虑整个输入序列的所有其他元素，而无需像传统的 RNN 或 LSTM 那样逐步处理序列。这使得 Transformer 在训练时能并行处理整个序列，大大提高了训练速度和效率[7]。Transformer 模型分为两部分：编码器(Encoder)和解码器(Decoder)，运行原理结构如图 2 所示。

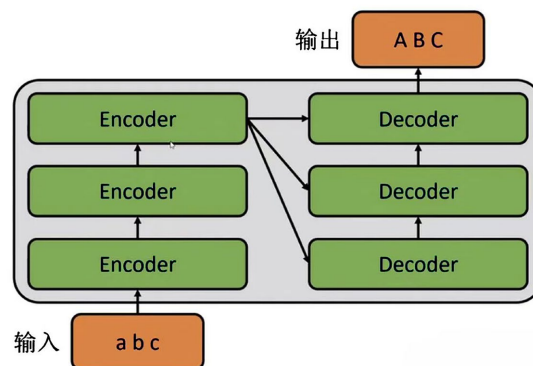


Figure 2. Simple structure of Transformer model
图 2. Transformer 模型简单结构

在金融股票数据分析中，Transformer 模型通过其强大的自注意力机制能够有效地捕捉情绪与市场数据之间的复杂关系。通过集成情绪数据(如社交媒体评论、新闻情绪等)，Transformer 模型不仅能基于历

史股市数据进行预测，还能根据市场情绪的变化动态调整预测结果，提升模型的准确性。情绪调控在股市预测中尤为重要，因为市场情绪对价格波动有深远影响，Transformer 通过长程依赖和多头注意力机制能够深入理解这些复杂的情绪驱动因素，从而在动态市场环境中提供更为精准的预测[8]。

2.3. Transformer-LSTM 模型

部分数据由于其特性需同时捕捉局部特征和全局特征，但 LSTM 擅长捕捉局部时序依赖，难以建模长距离全局关系；Transformer 模型可以依靠自注意力机制，通过计算任意两个时间步的关系权重，直接建模全局依赖。所以提出使用 Transformer 与 LSTM 的协同机制，由 Transformer 提供“全局上下文特征”，LSTM 细化“局部时序动态”，两者通过注意力机制进行融合。模型结构分为编码器和解码器两部分：

1) Transformer 编码器：将输入序列 $X = \{x_1, x_2, \dots, x_T\}$ ($x_t \in \mathbb{R}^d$ 表示 t 时刻的股票特征)映射为全局上下文表示 $H_{trans} \in \mathbb{R}^{T \times d}$ ，核心公式为：

$$H_{trans} = \text{Soft max} \left(\frac{QK^\top}{\sqrt{d_k}} \right) V \quad (5)$$

$$Q = W^Q X, K = W^K X, V = W^V X \quad (6)$$

其中， W^Q 、 W^K 、 W^V 为可学习参数矩阵， d_k 为缩放因子。

2) LSTM 解码器：基于 Transformer 的全局特征和自回归机制逐步生成预测序列 $\hat{Y} = \{\hat{y}_1, \dots, \hat{y}_N\}$ 。LSTM 通过门控机制为每一步 t 更新隐藏状态 h_t 和细胞状态 c_t ，并融合 Transformer 的上下文向量 c_t ：

$$i_t, f_t, o_t = \sigma(W \cdot [h_{t-1}; \hat{y}_{t-1}; c_t]) \quad (7)$$

$$c_t = \sum_{i=1}^T \alpha_{t,i} H_{trans,i} \quad (8)$$

$$\hat{y}_t = W_o h_t + b_o \quad (9)$$

其中， $\alpha_{t,i}$ 为解码器对编码器第 i 时间步的注意力权重，通过 h_{t-1} 与 H_{trans} 的相似度计算得到。

3. 实证分析

3.1. 数据选取

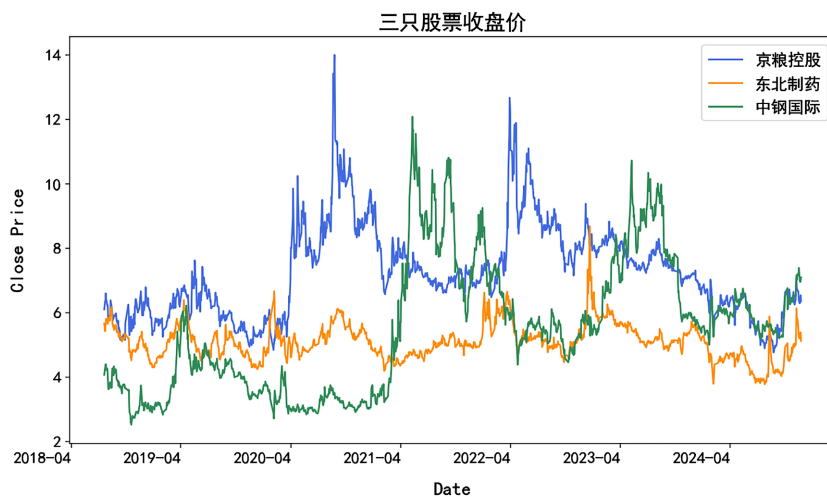


Figure 3. Closing prices of three stocks

图 3. 三只股票的收盘价

本文获取的原始数据来自 AKShare, 选取了京粮控股(000505)、东北制药(000597)和中钢国际(000928)三只 A 股股票来预测收盘价, 包含 2018 年 7 月 20 日到 2024 年 11 月 22 日共 1544 个交易日的数据, 前 1244 个数据作为训练集, 后 300 个数据作为测试集。考虑到对股票价格影响比较重要的一些指标, 选取了 21 个指标数据。其中包含 5 个基本交易数据, 分别是开盘价、最高价、最低价、收盘价、成交量, 也包含了基于基本交易数据通过数学公式计算得到的能够反映股价多方面信息的技术指标, 分别是移动平均线(MA)、指数平滑异同移动平均线(MACD)、趋向指标(DMI)、能量潮指标(OBV)、随机指标(KDJ)、均量线(VOL)、心理线(Psy)、相对强度指标(RSI)和布林带(Bollinger Bands)共九个指标, 特征个数分别为 3、1、2、1、3、1、1、1、3。三只股票的收盘价如图 3 所示。

3.2. LSTM 神经网络模型

针对实际问题, 数据集集中的样本特征往往存在差异, 故在建模前对数据进行规范化处理至关重要。首先对所选数据清洗后得到完整样本, 然后采用“max-min 标准化”方法来消除指标间的量纲差异, 公式为:

$$x_i' = \frac{x_i - \min x_i}{\max x_i - \min x_i}, \quad (10)$$

其中, x_i 为原始数据, x_i' 为标准化后的数据, $\max x_i$ 为收盘价中的最大值, $\min x_i$ 为收盘价中的最小值, 数据处理后不改变原始意义。在建模预测后, 再将收盘价数据作反归一化处理, 使其在检验数据建模效果时能够更加直观。

数据处理好之后, 输入到 LSTM 网络当中, 采用前 60 日的特征信息来预测下一日的股票收盘价格, 并采用滑动窗口预测法进行实验。设置不同的参数进行测试, 设置不同层数进行超参优化后发现选用 Adam 作为优化算法, 隐含层个数设置为 2 时预测精度最高, 对应的隐含层节点数分别为 128、192、128, 学习率为 0.0001、0.0008 和 0.0005, 迭代次数为 200 次。选用决定系数(R-Square, R^2)来检验实验精度, 反映因变量的全部变异能通过回归关系被变量解释的比例, 计算公式为:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}, \quad (11)$$

其中, y_i 为原始收盘价, $\bar{y}_i$ 为平均收盘价, $\hat{y}_i$ 为预测收盘价。

LSTM 神经网络对京粮控股、东北制药和中钢国际的模型拟合优度分别为 0.8394、0.8261 和 0.8404, 均超过 80%, 说明 LSTM 神经网络模型可以较好地预测股票的收盘价, 但对数据的波动不敏感, 模型存在优化空间。

3.3. Transformer-LSTM 神经网络模型

与文献[9]中先降维进行特征选择再训练模型的方法不同, 由于股票数据的高复杂性, 本次实验使用全部指标进行训练。实验先将处理好的数据输入到 Transformer 编码器(Encoder)中, 数据会被映射到向量空间进行位置编码, 并利用自注意力机制来捕捉序列中不同位置之间的关系, 提取全局特征, 然后将提取的特征输入到 LSTM 神经网络中, 利用 LSTM 来捕捉时间序列中的局部时间依赖性。通过 Keras-Tuner 的 3 种不同的搜索调试策略(即 RandomSearch、贝叶斯优化和 HyperBand), 针对不同的层数组合调参后发现, 使用 HyperBand 方法, 三只股票均在 Transformer 编码层数为 1, 使用两个多头注意力组合, 且 LSTM 隐藏层层数也为 1, 对应的隐含层节点数分别为 64、256 和 192。迭代次数为 200 时, 模型运行效

果最佳，预测结果如图 4 所示。

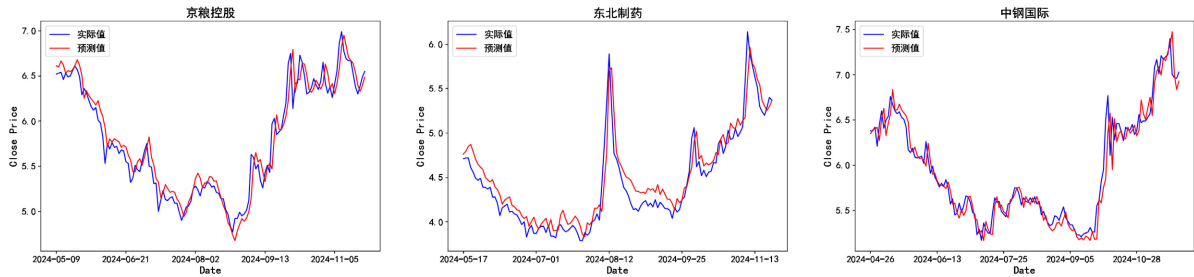


Figure 4. Prediction of stock closing price

图 4. 股票的收盘价预测

实验仍选择决定系数来检验实验精度。Transformer-LSTM 网络对京粮控股、东北制药和中钢国际的模型拟合优度为 0.9329、0.8903 和 0.9330，其中对京粮控股和中钢国际两只股票的模型拟合优度均达到 93%，效果提升 10%左右，显著优于 LSTM；而对东北制药的模型拟合优度为 89%，未达到 90%但也有 6.42%的提升。实验结果表明，Transformer 的引入能够有效提高模型的拟合能力，尤其是在全局特征提取方面，Transformer 帮助 LSTM 更好地捕捉股票价格波动趋势，使模型的预测精度大幅提升。

3.4. 预测比较分析

将 LSTM 神经网络与 Transformer-LSTM 模型进行比较，针对测试集对收盘价进行预测，并将预测结果还原与真实值进行比较，采用均方误差 (MSE) 作为标准进行评价，公式为：

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{12}$$

对比结果如表 1 所示。

Table 1. Closing price prediction MSE

表 1. 收盘价预测 MSE

	LSTM	Transformer-LSTM
京粮控股	0.0613	0.0245
东北制药	0.0553	0.0302
中钢国际	0.0517	0.0222

由图表对比结果发现，京粮控股、东北制药和中钢国际在 Transformer-LSTM 模型下的预测误差分别为 0.0245、0.0302 和 0.0222，而在 LSTM 神经网络模型下的误差为 0.0613、0.0453 和 0.0517，说明 Transformer-LSTM 模型对股票收盘价预测更有效。

训练过程中发现，Transformer-LSTM 模型通过扩展了一层 Transformer 架构，使用了更多的参数，数据得到了深度挖掘。而由于其并行化的特点，可以在一次计算中处理整个序列，所以模型最终训练时间与单一 LSTM 模型训练时间相差不大，这说明在保证高预测精度的情况下，Transformer-LSTM 的混合结构通过全局 - 局部特征融合，平衡了长短期依赖关系，验证了其设计合理性，仍具备较好的优势。

4. 总结

本文通过实证分析比较了 LSTM 神经网络和 Transformer-LSTM 混合模型在股票收盘价预测中的应

用与效果。通过对京粮控股(000505)、东北制药(000597)和中钢国际(000928)三只 A 股股票的历史数据进行建模与预测, Transformer-LSTM 混合模型在股票收盘价预测中表现出了更强的拟合能力和预测精度,特别是在捕捉股价的长期趋势方面具有较好的优势。实验也存在不足之处,一是在模型网络参数设置中, Keras-Tuner 超参调优的过程耗时耗力;二是虽然模型提高了预测精度,但对股票数据的近期剧烈波动仍不敏感。未来的研究可以对所选指标进行筛选并评估模型贡献值,以提高模型效率;其次,进一步优化 Transformer 和 LSTM 的结合方式,探索更多的特征选择与模型调优策略,以提升预测的精度和稳定性;与此同时,考虑到股市本身的高度复杂性和不确定性,之后还可以结合更多的外部因素,如宏观经济数据、投资者情绪数据等,来进一步增强模型的预测能力。

参考文献

- [1] 苏治, 卢曼, 李德轩. 深度学习的金融实证应用: 动态、贡献与展望[J]. 金融研究, 2017(5): 111-126.
- [2] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 735-780.
<https://doi.org/10.1162/neco.1997.9.8.1735>
- [3] 赵薇. 基于 LSTM 神经网络的金融数据预测分析[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2020.
- [4] 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[D]: [硕士学位论文]. 北京: 首都经济贸易大学, 2016.
- [5] 赵红蕊, 薛雷. 基于 LSTM-CNN-CBAM 模型的股票预测研究[J]. 计算机工程与应用, 2021, 57(3): 203-207.
- [6] 陈治颖. LSTM 模型优化及其在中国股票指数预测中的对比研究[D]: [硕士学位论文]. 济南: 山东财经大学, 2024.
- [7] 孟祥福, 石皓源. 基于 Transformer 模型的时序数据预测方法综述[J]. 计算机科学与探索, 2025, 19(1): 45-64.
- [8] 郑国杰. 基于互联网投资者情绪的股票时间序列分析与预测[D]: [硕士学位论文]. 杭州: 浙江工业大学, 2019.
- [9] 危冰淋, 刘春雨, 刘家鹏. 基于 Transformer-LSTM 模型的多因素碳排放权交易价格预测[J]. 价格月刊, 2024(5): 49-57.