

基于注意力机制融合TCN和LSTM的 股价预测方法

——以中国银行(601988)为例

徐 赫, 宋瑾钰*

浙江理工大学计算机科学与技术学院(人工智能学院), 浙江 杭州

收稿日期: 2025年2月20日; 录用日期: 2025年3月13日; 发布日期: 2025年4月14日

摘 要

随着我国金融市场的不断深化与发展, 股票价格的波动预测成为投资者和研究者关注的焦点。本文以中国银行(601988)为研究对象, 使用Python爬取英为财经网站上的股价数据。在此基础上, 提出了一种基于注意力机制融合时间卷积网络(TCN)和长短期记忆网络(LSTM)的股价预测模型。通过对中国银行股票的开盘价格进行训练与预测, 本文所提出的模型在预测精度上优于传统的ARIMA模型和单一的LSTM模型。实验结果显示, 注意力机制融合TCN + LSTM模型在预测误差和损失函数MSE上均表现出较好的性能, 尤其在处理股票价格的极值点和长期趋势方面具有显著优势。本研究不仅为股票投资者提供了有效的预测工具, 也为金融时间序列分析领域提供了新的研究视角。

关键词

Python, 时间卷积网络, 长短期记忆神经网络, 注意力机制, 股价预测

Stock Price Prediction Based on Attention-Fused TCN and LSTM

—Taking Bank of China (601988) as an Example

He Xu, Jinyu Song*

School of Computer Science and Technology (School of Artificial Intelligence), Zhejiang Sci-Tech University,
Hangzhou Zhejiang

Received: Feb. 20th, 2025; accepted: Mar. 13th, 2025; published: Apr. 14th, 2025

*通讯作者。

文章引用: 徐赫, 宋瑾钰. 基于注意力机制融合 TCN 和 LSTM 的股价预测方法[J]. 电子商务评论, 2025, 14(4): 637-645.
DOI: 10.12677/ec.2025.144931

Abstract

As China's financial market continues to deepen and evolve, forecasting stock price fluctuations has become a focal point for both investors and researchers. In this study, China Bank (601988) is selected as the case study, and stock price data is collected using Python from the Investing.com website. Building on this foundation, we propose a novel stock price prediction model that integrates features from a Temporal Convolutional Network (TCN) and a Long Short-Term Memory (LSTM) network through an Attention Mechanism. By training and forecasting the opening prices of China Bank's stocks, our results demonstrate that the proposed model achieves higher prediction accuracy than both the traditional ARIMA model and the standalone LSTM model. Experimental outcomes further indicate that the attention-fused TCN + LSTM model outperforms in terms of prediction error and Mean Squared Error (MSE), particularly excelling in capturing extreme price fluctuations and long-term trends. This study not only provides a valuable predictive tool for investors but also offers a fresh perspective for financial time series analysis.

Keywords

Python, TCN, LSTM, Attention Mechanism, Stock Price 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

在当今经济全球化的大背景下,我国股票市场作为金融市场的重要组成部分,其波动性和复杂性日益凸显。根据东方财富数据网最新统计,截至2024年6月17日,我国A股账户数已达到2.3亿,其中活跃交易者数量在4000万至5000万之间。股票市场的高回报潜力吸引了众多投资者,但与之相伴的是高风险,一次失误的操作可能带来巨大的财产损失。因此,准确预测股票价格走势对于投资者来说至关重要。

深度学习模型在处理大规模历史数据方面具有显著优势。它们能够从复杂的数据中提取特征,准确捕捉到影响股价变化的关键技术指标。在众多机器学习模型中,长短期记忆网络(LSTM, Long short-term memory)因其能够处理和预测时间序列数据而备受瞩目。但是,LSTM模型在预测股票价格时也存在一定的局限性。例如,在基于ARIMA与LSTM对股票价格的预测研究中,LSTM模型的均方误差(MSE)指标为0.33125543,仍有提升空间[1]。

鉴于此,本文提出了一种基于注意力机制融合时间卷积网络(TCN)和LSTM的股价预测模型。该模型旨在通过结合TCN在处理时间序列数据方面的优势以及LSTM在捕捉长期依赖关系的能力,进一步提高股价预测的准确性。本文以中国银行(601988)为例,通过实证研究验证了所提出模型的有效性,为股票市场预测提供了新的研究视角和方法论。

2. 国内外研究现状

关于资产价格预测的研究起源较早,国外学者在早期便提出了许多成熟的理论模型。资本资产定价模型(CAPM)以及Fama与French提出的三因子、五因子模型为后续的预测工作奠定了理论基础[2]。随着

对时间序列内在规律认识的不断深入以及计算机数据处理能力的提升, 资产价格预测方法逐步从单纯基于统计与概率理论的传统模型(如 AR、MA、ARMA、ARIMA、ARCH 及 GARCH 模型)演变到融合新型机器学习与深度学习方向。Mat Yusoff & Abdullah (1999)利用吉隆坡证券交易所的日股票指数建立了 GARCH 模型及其衍生形式[3], 而 Minjun Kim & Hiroki Sayama (2017)则通过构建标准普尔 500 指数基础公司的时间序列复杂网络, 将网络指标引入 ARIMA 模型, 从而提高了预测准确性[4]。

随着机器学习理论的兴起, Cortes 和 Vapnik 提出的支持向量机(SVM)在小样本、高维非线性模式识别中显示出独特优势, 成为股价预测领域的重要工具[5]。同时, 近年来深度学习方法的广泛应用也为股价预测带来了新的思路。已有研究表明, 基于 LSTM 的模型在捕捉时间序列的长期依赖关系方面具有明显优势; 此外, 部分学者还提出了将 LSTM 与其他模型(如 GARCH、SVR、Adaboost 等)进行组合的混合模型[6]-[9], 以期在不同数据特征与噪声条件下取得更优预测效果。

近年来, 新型深度学习模型如 GNN[10], GCN[11]出现, 这些创新的网络结构为股价预测提供了全新的思路, 尤其是在利用图结构数据建模金融市场中的复杂关联性方面, 展现了其独特的优势[12]-[14], 通过对节点(代表个股或市场指标)之间的关系进行建模, 实现了对市场结构的深入挖掘。同时, 强化学习方法优化的模型在股价预测与交易策略优化中展现出潜在优势[15], 其通过与环境的持续交互, 动态调整决策策略, 为投资者提供实时的风险控制与收益优化方案。

总体来看, 无论是传统统计方法还是新兴的机器学习和深度学习模型, 都存在一定局限性, 单一模型往往难以全面捕捉股票价格这一既不平稳又受多因素影响的复杂时序信号。针对这一问题, 本文致力于深入挖掘股价序列中的局部特征和长期依赖信息。考虑到 TCN 在提取局部时序特征方面表现出高效性[16], 以及 LSTM 在捕捉长时依赖性上的优势[17], 本文从注意力机制融合多特征思路入手[18], 提出了一种融合注意力机制的 TCN + LSTM 股价预测模型, 并以中国银行(601988)为例开展实证研究。注意力机制的引入使得模型能够根据不同分支输出的重要性进行动态调整, 从而显著提升了预测精度和整体模型性能。

3. 研究方法 with 模型构建

3.1. 时间卷积网络

时间卷积网络 TCN 是一种基于卷积神经网络(CNN)改进而成的模型, 专门用于处理时间序列数据。与传统 CNN 不同, TCN 利用因果卷积结构确保模型输出仅依赖于历史信息, 从而有效避免未来信息的泄露。为进一步扩大模型的感受野, TCN 引入了扩张卷积(Dilated Convolution), 该方法通过在卷积核元素之间插入零(即空洞)来增加卷积间隔, 在不显著增加参数数量的情况下实现感受野的扩展。其中, 扩张因子 d 定义了相邻卷积核元素之间的间隔, 其具体计算公式可表示为:

$$y(t) = \sum_{k=0}^{K-1} f(k) \cdot x(t-d \cdot k) \quad (1)$$

其中 K 表示卷积核的大小, $f(k)$ 为卷积核的权重, $x(t)$ 为输入信号。

此外, 为了解决深层网络训练过程中常见的梯度消失问题, TCN 还引入了残差连接(Residual Connections)和批量归一化(Batch Normalization)技术。其核心思想可以描述为:

$$o = \text{Activation}(x + F(x)) \quad (2)$$

其中 $F(x)$ 表示经过卷积操作得到的特征映射, x 为输入信号, Activation 表示激活函数。

3.2. 长短期记忆神经网络

LSTM 是一种特殊的循环神经网络(RNN), 在处理和预测序列数据方面展现出显著优势, 尤其适用于

时间序列预测任务。与传统 RNN 在长序列学习过程中容易遭遇梯度消失和梯度爆炸问题不同, LSTM 通过引入门控机制, 有效地调控了信息在网络中的传递和更新, 从而大大改善了模型的学习效果。

具体来说, LSTM 的核心在于其多个门控单元:

- 输入门: 该门通常由一个 sigmoid 函数和一个 tanh 函数组成, 用以决定哪些新信息需要被写入单元状态。形式上, 可以表示为

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \tilde{C}_t = \sigma(W_c[h_{t-1}, x_t] + b_c) \quad (3)$$

其中 i_t 为输入门的激活值, \tilde{C}_t 为候选状态。

- 遗忘门: 该门决定了哪些信息应从单元状态中舍弃, 其机制有助于清除不再需要的历史信息, 从而保持状态信息的精简和有效性。遗忘门的计算通常表示为

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (4)$$

其中 f_t 表示遗忘门的激活值。

- 输出门: 用于控制单元状态中哪些信息将传递到下一时刻。

上述门控机制共同作用, 保证了 LSTM 在进行如股价预测等时间序列信息推理任务时, 能够更好地捕捉长期依赖关系, 同时防止信息在传递过程中过度衰减。由此, LSTM 成为金融数据、语音识别等领域中广泛应用的重要模型。

3.3. 注意力机制

注意力机制是一种能够使模型更加关注输入序列中关键部分的技术。在股价预测模型中, 注意力机制可以帮助模型识别出对预测结果影响最大的历史信息, 从而提高预测准确性。本文提出的模型通过融合来自 LSTM 与 TCN 两个分支的多特征信息, 实现了对时序数据中长期依赖关系与局部特征的有效捕捉。

对于注意力机制, 其基本计算过程可描述为:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

其中, Q (查询向量)、 K (键向量) 和 V (值向量) 分别由 LSTM 和 TCN 分支提取的特征经过线性变换获得, d_k 为键向量的维度。该公式使得模型能够根据不同时间步之间的相关性自适应地分配注意力权重, 从而突出对预测最有价值的信息。

为了实现多分支特征的有效融合, 本文设计了如下融合操作:

$$F = \alpha F_{\text{LSTM}} + (1 - \alpha) F_{\text{TCN}} \quad (6)$$

其中, F_{LSTM} 和 F_{TCN} 分别表示来自 LSTM 与 TCN 分支的特征表示, $\alpha \in [0, 1]$ 为融合权重。通过这种加权融合, 模型既能利用 LSTM 捕捉长序列中的依赖关系, 又能借助 TCN 提取局部时序特征, 进而通过注意力机制自适应地调整对各特征的关程度, 最终实现对股价走势的高精度预测。

3.4. 模型搭建

结合 TCN 和 LSTM 两者的优势, 构建了一种混合模型结构, 以提升时间序列数据的预测精度。模型主要由两个并行分支组成, 分别是基于卷积网络的 TCN 分支和基于递归神经网络的 LSTM 分支, 两者各自对时间序列数据进行特征提取, 在特征融合阶段, 借鉴 ACMix 模型的设计思路[19], 通过注意力机制动态地为每个分支分配权重, 使其能够根据各自的特征贡献进行加权融合, 从而增强关键时间步的表征

能力, 最后经过 MLP (Multi-Layer Perceptron)进行推理, 得到最终预测结果。模型结构如图 1 所示。

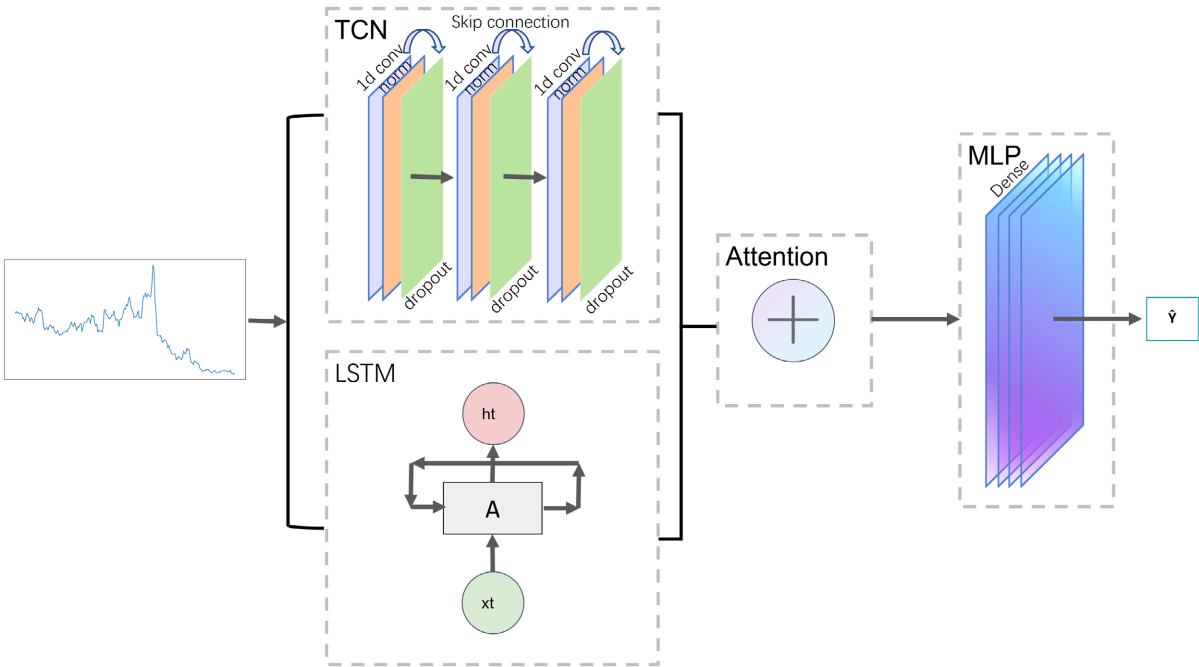


Figure 1. Model structure diagram
图 1. 模型结构图

4. 实验与结果分析

4.1. 数据获取

数据来自 investing (英为财经)的数据内容, 采用 python 构建的爬虫工具获取中国银行(601988)的数据, 覆盖了从 2010 年至 2024 年的交易记录。数据主要内容分为收盘价格、开盘价格、最高值、最低值、交易值和涨跌幅, 每周记录一次。交易数据样例如表 1 所示。

Table 1. Sample data
表 1. 样本数据

| 日期 | 收盘 | 开盘 | 高 | 低 | 交易量 | 涨跌幅 |
|------------|------|------|------|------|----------|-------|
| 2024-12-8 | 5.20 | 5.08 | 5.25 | 5.07 | 806.35 M | 1.96% |
| 2024-12-1 | 5.10 | 5.01 | 5.15 | 4.96 | 930.30 M | 1.80% |
| 2024-11-24 | 5.01 | 4.92 | 5.10 | 4.88 | 940.06 M | 2.04% |
| 2024-11-17 | 4.91 | 4.88 | 5.05 | 4.88 | 1.12 B | 1.03% |
| 2024-11-10 | 4.86 | 4.84 | 4.90 | 4.78 | 1.04 B | 0.00% |

4.2. 模型超参数选择

在模型训练阶段, 对关键超参数进行了精心选择和调整来优化模型性能。模型的编译采用了 Adam 优化器, 并使用均方误差(MSE)作为损失函数。并且引入了早停(Early Stopping)回调机制, 以防止模型在训练过程中过拟合。该机制监控验证集损失(val_loss), 当连续 10 个训练周期内验证集损失未改善时, 自

动停止训练。这有助于节省计算资源，并提高模型的泛化性能。通过上述方式在一定程度上提升模型的预测精度和鲁棒性。

4.3. 激活函数对模型性能的影响分析

为检验不同激活函数对模型性能的影响，本文设计了一组对比实验，构建三种 TCN-LSTM-Attention 模型变体，分别采用 sigmoid、relu 和 tanh 作为激活函数，并在模型训练过程中记录了损失函数(loss)的变化情况，同时在测试集上计算了均方误差(MSE)、均方根误差(RMSE)、平均绝对误差(MAE)以及最大误差(Max_error)等评价指标。

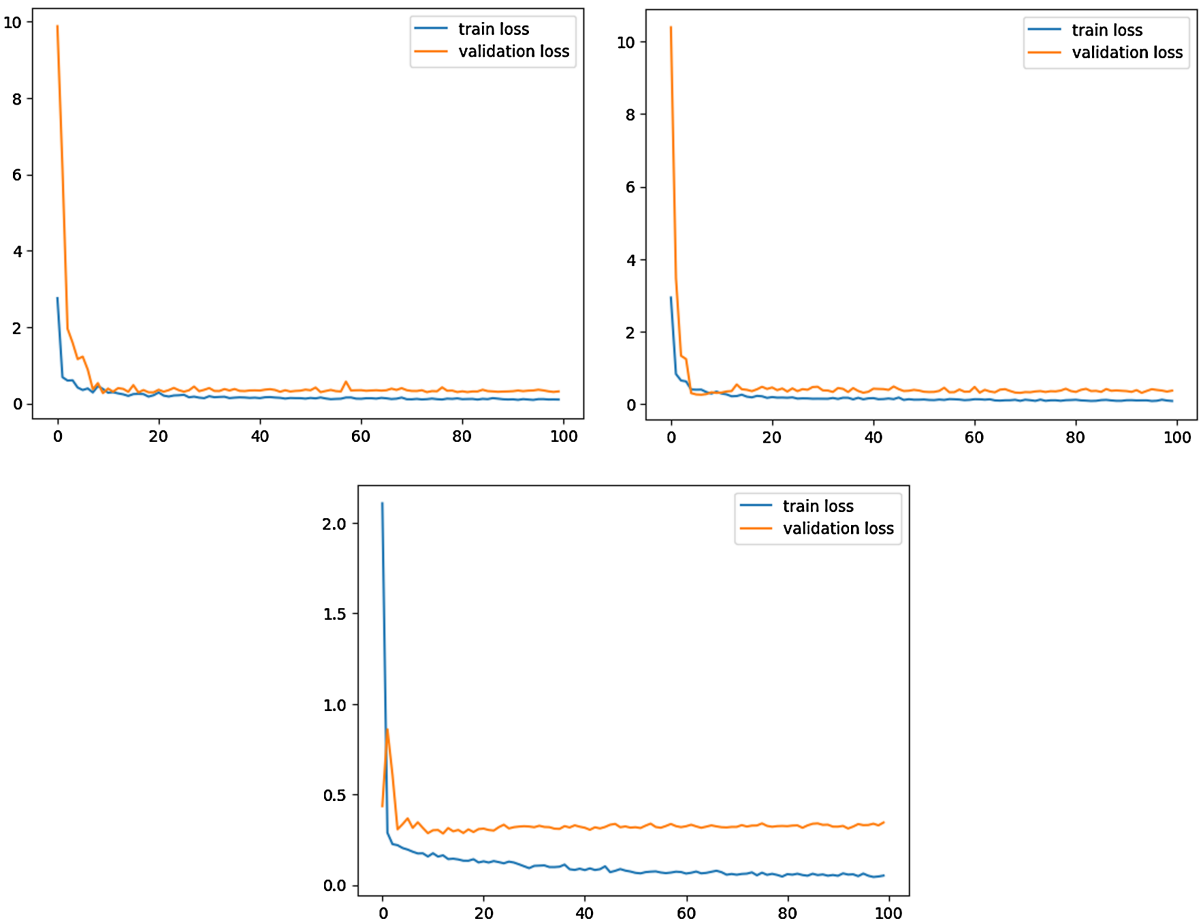


Figure 2. The impact of different activation functions on loss (from left to right are relu, sigmoid, tanh)
图 2. 不同激活函数对 loss 的影响(从左到右分别是 relu, sigmoid, tanh)

Table 2. Experimental results
表 2. 实验结果

| 模型 | MSE | RMSE | MAE | Max_error |
|----------------------------|----------|----------|----------|-----------|
| TCN-LSTM-Attention_sigmoid | 0.047082 | 0.216984 | 0.159349 | 0.897534 |
| TCN-LSTM-Attention_relu | 0.084269 | 0.29029 | 0.21881 | 0.942492 |
| TCN-LSTM-Attention_tanh | 0.055687 | 0.235981 | 0.187316 | 0.586331 |

如图 2 所示, 不同激活函数下的 loss 曲线展示了各模型在训练过程中的收敛情况, 可以看出各变体在初期和后期的下降速度及稳定性存在差异; 而表 2 则汇总了各模型在测试集上的关键性能指标。从数据上看, 采用 sigmoid 激活函数的模型取得了最低 MSE (0.047082)和较低的 MAE (0.159349), 表明其在平滑非线性映射方面具有一定优势; 相比之下, relu 变体的误差指标均较高, 而 tanh 模型则在最大误差方面表现相对优异(0.586331)。

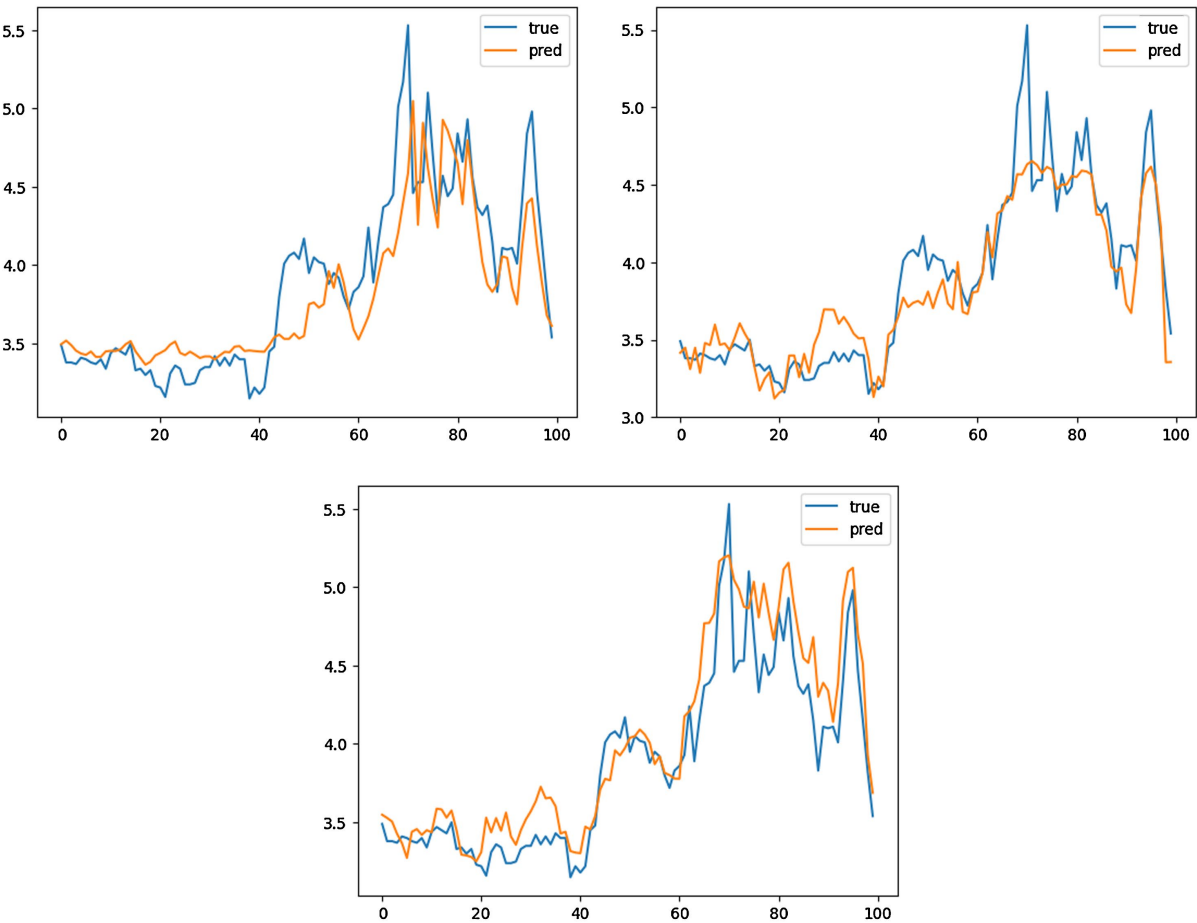


Figure 3. The impact of different activation functions on prediction results (from left to right are relu, sigmoid, tanh)
图 3. 不同激活函数对预测结果的影响(从左到右分别是 relu, sigmoid, tanh)

图 3 展示了采用不同激活函数(relu、sigmoid、tanh)的 TCN-LSTM-Attention 模型在相同数据集上的推理结果。由图中可见, 不同激活函数下模型的预测曲线存在明显差异。其中, sigmoid 激活函数的模型难以预测峰值数据的出现, tanh 激活函数对峰值数值预测比较准确, 而 relu 激活函数整体拟合效果好于前两者。

4.4. 多时段样本预测分析

如图 4 所示, 对来自不同时间段的多个样本进行预测结果分析可以看出, 整体推理趋势较为一致, 模型能够准确捕捉数据的整体走势, 展现出较高的全局拟合能力。这些样本覆盖了不同时间段, 确保了验证的全面性和丰富性。但是细致观察各子图中的局部细节会发现, 部分时间段的预测结果与实际情况存在一定偏差。例如, 在局部波动剧烈的时刻, 模型预测值与真实值之间出现明显偏离, 这可能与数据

噪声或短期市场波动有关。总体上, 本文提出的模型在长时间尺度上表现出良好的鲁棒性和趋势捕捉能力, 其均方误差(MSE)仅为 0.047082, 相较于基于 ARIMA 与 LSTM 构建的模型误差指标 0.33125543 有显著提升。



Figure 4. Model prediction results for different time periods
图 4. 不同时段样本预测结果

影响模型预测结果的因素主要分为市场的短期波动、数据噪声以及不同时间段的波动幅度差异。短期市场波动如突发新闻、政策变化等, 可能导致股价的剧烈波动, 进而影响模型在这些时刻的预测准确性。模型的训练过程中未能充分考虑这些突发因素, 因此在短期快速变化的时段, 预测误差较为显著。市场在不同时间段的波动幅度也存在差异, 对于波动较小的时间段, 模型能够较好地捕捉到股价的整体趋势, 预测结果较为精确; 而在波动较大的时间段, 模型的预测结果可能受到更多噪声的干扰, 导致误差增大。

5. 结论

本文提出了一种基于注意力机制融合 TCN 与 LSTM 的混合预测模型。该模型利用 TCN 分支提取局部时序特征, 结合 LSTM 分支捕捉长期依赖信息, 通过注意力机制对两者提取的特征进行自适应加权融合, 并最终由 MLP 进行非线性映射, 输出预测结果。通过对比实验, 本文详细探讨了不同激活函数(sigmoid、relu、tanh)对模型训练收敛性及预测性能的影响, 并通过实验数据证明了该模型在 MSE 误差上优于基于

ARIMA 与 LSTM 构建的模型。

本文的工作不仅为金融领域中基于深度学习的股价预测提供了一种有效的解决方案, 也为多分支模型中注意力机制的设计与应用探索了新思路。尽管所提模型在实验数据上取得了较好的预测效果, 但在实际应用中仍存在诸如数据噪声、市场突变等不确定性因素的影响, 因此, 未来工作将进一步结合更多宏观经济变量和低频数据, 对模型进行改进和扩展, 以期提升模型在更复杂环境下的鲁棒性和泛化能力。

参考文献

- [1] 肖杰. 基于 ARIMA 与 LSTM 对股票价格的预测研究——以中信证券(600030)为例[J]. 电子商务评论, 2024, 13(4): 6072-6083. <https://doi.org/10.12677/ecl.2024.1341847>
- [2] Fama, E.F. and French, K.R. (1992) The Cross-Section of Expected Stock Returns. *The Journal of Finance*, **47**, 427-465. <https://doi.org/10.1111/j.1540-6261.1992.tb04398.x>
- [3] Chong, C.W., Ahmad, M.I. and Abdullah, M.Y. (1999) Performance of GARCH Models in Forecasting Stock Market Volatility. *Journal of Forecasting*, **18**, 333-343. [https://doi.org/10.1002/\(sici\)1099-131x\(199909\)18:5<333::aid-for742>3.0.co;2-k](https://doi.org/10.1002/(sici)1099-131x(199909)18:5<333::aid-for742>3.0.co;2-k)
- [4] Kim, M. and Sayama, H. (2017) Predicting Stock Market Movements Using Network Science: An Information Theoretic Approach. *Applied Network Science*, **2**, Article No. 35. <https://doi.org/10.1007/s41109-017-0055-y>
- [5] Vapnik, V. (2013) *The Nature of Statistical Learning Theory*. Springer Science & Business Media.
- [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] 史建楠, 邹俊忠, 张见, 等. 基于 DMD-LSTM 模型的股票价格时间序列预测研究[J]. 计算机应用研究, 2020, 37(3): 662-666.
- [8] 韩金磊, 熊萍萍, 孙继红. 基于 LSTM 和灰色模型的股价时间序列预测研究[J]. 南京信息工程大学学报, 2023, 15(6): 631-642.
- [9] 王文姣, 张娜. 基于 LSTM 变权组合模型的股价预测[J]. 商展经济, 2023(7): 95-97.
- [10] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020) Graph Neural Networks: A Review of Methods and Applications. *AI Open*, **1**, 57-81. <https://doi.org/10.1016/j.aiopen.2021.01.001>
- [11] Kipf, T.N. and Welling, M. (2016) Semi-Supervised Classification with Graph Convolutional Networks.
- [12] Sonani, M.S., Badii, A. and Moin, A. (2025) Stock Price Prediction Using a Hybrid LSTM-GNN Model: Integrating Time-Series and Graph-Based Analysis.
- [13] Tian, H., Zheng, X., Zhao, K., Liu, M.W. and Zeng, D.D. (2022) Inductive Representation Learning on Dynamic Stock Co-Movement Graphs for Stock Predictions. *INFORMS Journal on Computing*, **34**, 1940-1957. <https://doi.org/10.1287/ijoc.2022.1172>
- [14] Shi, Y., Wang, Y., Qu, Y. and Chen, Z. (2023) Integrated GCN-LSTM Stock Prices Movement Prediction Based on Knowledge-Incorporated Graphs Construction. *International Journal of Machine Learning and Cybernetics*, **15**, 161-176. <https://doi.org/10.1007/s13042-023-01817-6>
- [15] Zhang, J. and Lei, Y. (2022) Deep Reinforcement Learning for Stock Prediction. *Scientific Programming*, **2022**, 1-9. <https://doi.org/10.1155/2022/5812546>
- [16] Bai, S., Kolter, J.Z. and Koltun, V. (2018) An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling.
- [17] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [18] Li, Y., Zeng, J., Shan, S. and Chen, X. (2019) Occlusion Aware Facial Expression Recognition Using CNN with Attention Mechanism. *IEEE Transactions on Image Processing*, **28**, 2439-2450. <https://doi.org/10.1109/tip.2018.2886767>
- [19] Pan, X., Ge, C., Lu, R., Song, S., Chen, G., Huang, Z., et al. (2022) On the Integration of Self-Attention and Convolution. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 805-815. <https://doi.org/10.1109/cvpr52688.2022.00089>