

引用:王嵩巍,孙林. 融合时序卷积网络与多头自注意力的时间序列预测模型[J]. 聊城大学学报(自然科学版), 2026, 39(4): 475-485.

Citation: WANG Songwei, SUN Lin. Time series prediction model integrating temporal convolutional networks and multi-head self-attention[J]. Journal of Liaocheng University (Natural Science Edition), 2026, 39(4): 475-485.

文章编号 1672-6634(2026)04-0475-11

DOI 10.19728/j.issn1672-6634.2026010019

融合时序卷积网络与多头自注意力的 时间序列预测模型

王嵩巍, 孙林

(天津科技大学 人工智能学院, 天津 300457)

摘要 传统预测模型在捕捉长期依赖与提升多步预测精度方面存在明显不足。为进一步解决长序列训练中梯度消失、计算效率低及多步预测误差累积等问题,提出一种融合时序卷积网络、多头自注意力机制与序列到序列架构的时间序列预测模型。首先,使用时序卷积网络提取局部时序特征,利用其膨胀卷积结构与残差连接缓解梯度消失并支持并行计算;然后,引入多头自注意力机制,对时序卷积网络输出进行全局依赖建模与上下文语义增强,形成“局部-全局”协同的特征提取机制;最后,通过带注意力机制的序列到序列框架,将增强后的特征序列映射为未来多步价格序列。基于中国 A 股市场多只股票的日频历史数据集的实验结果表明,该模型在多项指标上均显著优于基线方法,其均方误差与平均绝对误差明显降低,验证了该模型在股票价格多步预测任务中的有效性与优越性。

关键词 时序卷积网络;多头自注意力;序列到序列;股票价格多步预测

中图分类号 TP391.41

文献标志码 A

开放科学(资源服务)标识码(OSID)



Time series prediction model integrating temporal convolutional networks and multi-head self-attention

WANG Songwei, SUN Lin

(Collage of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China)

Abstract Traditional forecasting models exhibit notable limitations in capturing long-term dependencies and enhancing the accuracy of multi-step predictions. To further address issues such as gradient vanishing during long-sequence training, low computational efficiency, and error accumulation in multi-step forecasting, this paper proposes a time series forecasting model that integrates temporal convolutional networks, a multi-head self-attention mechanism, and a sequence-to-sequence architecture. First, temporal convolutional networks are employed to extract local temporal features, leveraging their dilated convolution structures and residual connections to mitigate gradient vanishing and support parallel computation. Then, a multi-head self-attention mechanism is introduced to model global dependencies and enhance contextual semantics based on the outputs of the temporal convolutional networks, forming a “local-global” collabora-

收稿日期: 2026-01-31

基金项目: 国家自然科学基金项目(62076089); 天津市自然科学基金项目(24JCYBJC00890)资助

通信作者: 孙林, 男, 汉族, 博士, 教授, 河南省科技创新杰出青年, 研究方向: 深度学习、数据挖掘等, E-mail: sunlin@tust.edu.cn.

rative feature extraction mechanism. Finally, a sequence-to-sequence framework equipped with an attention mechanism maps the enhanced feature sequences into future multi-step price sequences. Experimental results based on daily historical datasets of multiple A-share stocks in the Chinese stock market demonstrate that the proposed model significantly outperforms baseline methods across multiple metrics, with notable reductions in both mean squared error and mean absolute error, validating its effectiveness and superiority in the task of multi-step stock price forecasting.

Keywords temporal convolutional network; multi-head self-attention; sequence to sequence; multi-step stock price forecasts

0 引言

时间序列预测与股票价格预测之间存在紧密关联,其根本原因在于股票价格本质上是按时间顺序排列的金融时间序列数据。早期的股价预测主要依赖基于统计学与计量经济学的传统模型,如自回归积分滑动平均模型^[1]、广义自回归条件异方差模型^[2]等。这类方法通常建立在线性假设基础上,而股票市场本质上是复杂、非线性、高噪声的动态系统,传统模型难以捕捉其复杂非线性关系,预测精度因而受到显著制约。

近年来,深度学习凭借其强大的表示学习与端到端特征提取能力,在股价预测研究中取得了显著进展,涌现出多种重要模型,如基于循环神经网络(Recurrent Neural Network, RNN)和长短时记忆网络(Long Short-Term Memory, LSTM)^[3]的模型。相关研究方面,Hiyam 等^[4]将 LSTM 应用于石油价格预测;Richiuti 等^[5]采用开盘价、最高价、最低价、收盘价数据结合 LSTM 进行股价建模。然而,上述模型只能提取序列数据的浅层特征,且对噪声与异常值较为敏感,影响了预测的稳定性。另外,LSTM 是通过递归方式循环处理输入信息,导致计算效率较低。双向长短时记忆网络通过同时考虑前向与后向信息,增强了上下文的建模能力,Kalra 等^[6]提出使用混合双向 LSTM 模型进行股票指数价格的实时预测。门控循环单元(Gated Recurrent Unit, GRU)作为 LSTM 的一种轻量化变体,可与 LSTM 构成混合网络^[7],利用 GRU 降低计算开销,同时借助 LSTM 保障建模能力。然而,这些方法在提取长序列的依赖关系方面仍存在不足,特别是在处理极长序列时面临梯度消失问题。此外,有研究者将卷积神经网络(Convolutional Neural Network, CNN)与 LSTM 结合^[8],利用 CNN 提取局部特征,再通过 LSTM 建模时序动态。但是 CNN 通常使用非因果卷积,在预测任务中会引入未来信息泄露,导致训练时效果虚高,但在真实预测时性能崩溃。为应对上述问题,本研究引入了带有残差连接的时序卷积网络(Temporal Convolutional Network, TCN)。该网络通过并行化结构同时处理输入序列,相较于 LSTM 的递归处理方式,提升了计算效率;其膨胀卷积机制通过设置不同层次的感受野,增强了对长程时序特征的提取能力;同时,其因果卷积机制也保证了在训练和推理时数据流的一致性;而残差连接则有效缓解了长距离梯度消失的问题。

由于股票数据本身具有较强的时序性,其预测方法亦可借鉴其他领域的时序建模成果,例如基于 CNN 及双向 LSTM 的锂电池充电量预测^[9]、地震事件预测^[10]等模型。然而,此类模型普遍存在误差累积、长期预测可靠性低等问题。在多步预测任务中,若仅使用最后一个时间步的输出进行迭代预测,误差累积问题尤为突出,易导致预测结果迅速偏离真实趋势。为此,本文进一步引入序列到序列(Sequence to Sequence, Seq2Seq)的端到端架构,以更系统地进行多步预测并提升建模的鲁棒性。

为进一步提升模型对关键信息的捕捉能力,Uckan^[11]构建了基于主成分分析与 CNN 的混合模型,应用于土耳其股票市场预测。但该模型在处理极长序列时,长期依赖建模能力仍然有限,且模型在训练过程中容易出现过拟合,导致泛化性能下降。Sebastian 等^[12]引入注意力机制改进了 LSTM 在印度股票预测中的表现;Wang 等^[13]在 LSTM 模型基础上引入注意力机制,用于研究比特币与股票市场的关联性。然而,这些方法使用的是单头注意力机制,只关注到最主要的模式,而忽略了次要但同样重要的模式。Dai 等^[14]还提出了 TCN+Seq2Seq 预测模型用于股票价格预测。但该模型仅通过 TCN 提取了局部特征,无法提取全局特征,特别是对于长序列输入,该模型必须将所有信息挤压到最后一个时间步的隐藏状态,导致早期输入的信息容易丢失。为解决上述问题,本文模型引入多头自注意力机制(Multi-head Self Attention, MHSA)。多头能

够构建出远比单头注意力更全面的时序表征;同时可以充分提取时序数据中的全局有效特征,与 TCN 共同建立全局感受野。另外,本研究还在解码器中增加交叉注意力,使解码器在生成每一个预测点时,可以直接回看编码器保存的完整历史序列特征矩阵。

综上分析,本文提出一种融合时序卷积网络与多头自注意力的时间序列预测模型(Seq2Seq with TCN and MHSA, TCNMS)。该模型在架构层面融合了时序卷积网络的并行长程依赖建模能力、多头自注意力的全局特征筛选能力以及带交叉注意力的序列到序列框架的自回归生成能力。在国内 A 股市场多只股票历史数据构建的数据集上进行实验结果表明,该模型在多支股票价格预测中表现出良好的稳定性与预测精度,其性能相较于主流基线模型及当前常见的深度学习方法均有显著提升。

1 混合预测模型

1.1 总体网络结构

股票价格预测任务面临多噪声、长期跨度和复杂特征依赖的金融时序数据,在多步预测中误差易累积,导致现有模型难以充分提取有效时序特征,预测精度受限。为此,本文提出一种融合 TCN 与 MHSA 的 Seq2Seq 网络的时间序列模型 TCNMS。TCNMS 模型的整体网络架构如图 1 所示。该模型是可变序列(输入)经由编码器和解码器到可变序列(输出)的架构。编码器部分是由 TCN 及 MHSA 组合构成。TCN 通过堆叠多个具有不同膨胀率的一维卷积层构建多尺度时序特征提取架构。每个卷积模块均集成残差连接机制,有效缓解梯度消失问题并提升网络稳定性。在 TCN 提取的局部多尺度特征基础上,引入 MHSA,通过动态计算序列内所有时间步的关联权重,实现关键时间节点的自适应特征增强。最终,将融合后的特征输入至配备注意力机制的 Seq2Seq 解码器,该解码器采用教师强制训练策略,构成端到端的生成式预测框架,能够将学习到的时空特征映射为未来多时间步的连贯序列输出。

1.2 编码器特征提取模块

由于股票价格历史数据具有非线性、高噪声、长记忆性及波动率聚集等特性,传统预测模型在精度和泛化能力上往往受限,且在长序列预测中容易出现梯度消失的问题。为此本研究提出编码器特征提取模块。该子模块由时序卷积网络 TCN 与多头自注意力 MHSA 共同构成。TCN 通过膨胀卷积机制,有效捕捉序列中的长程依赖与多尺度时序模式,并能够高效提取非线性特征,从而为整个模型提供稳健的底层表征。在此基础上,MHSA 进一步对 TCN 输出的特征进行全局依赖建模,增强对远距离时间点间复杂因果关系的捕获能力,并实现动态关键时间点的注意力聚焦。以下将分别阐述 TCN 与 MHSA 两个子模块的工作原理。

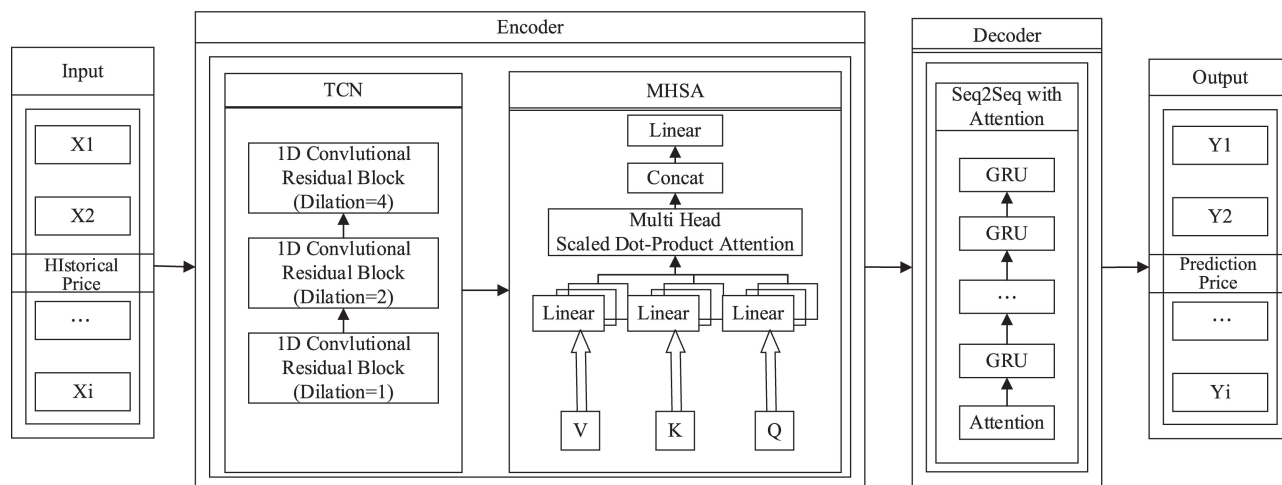


图 1 TCNMS 模型的整体网络架构

Fig. 1 Overall network architecture of the TCNMS model

1.2.1 时序卷积网络 TCN。使用 TCN 替代传统网络 RNN 或 LSTM 作为编码器,能够高效提取局部及短期的时序依赖关系(如日内波动、短期趋势),同时具备并行计算与梯度稳定的优势。TCN 接收一维时间序列作为输入,通过堆叠的残差块充分提取局部特征,生成包含丰富历史信息的输出序列,其中每个时间步表

征了从序列起始至当前时刻的累积上下文。每个残差块由主分支与捷径分支组成。主分支依次包含膨胀因果卷积、权重归一化、激活函数及随机丢弃操作。捷径分支用于处理输入与主分支输出通道数不一致的情况:若通道数相同,则直接传递输入;否则通过 1×1 卷积(点卷积)进行通道对齐。最终,最后一个残差块的输出经由全连接层或点卷积生成预测结果。TCN 模块的数学原理及相关公式可以参考文献[15]。

1.2.2 多头自注意力 MHSA。多头自注意力 MHSA 作为模型的“洞察力增强器”,它能捕捉序列内部任意两个时间点之间的长程依赖和全局关联。例如,发现“一年前的某个事件”与“当前股价”的潜在联系,或者识别出重复出现的周期性模式。多头机制通常会自发分化出不同的功能。比如周期性头,权重呈现规律的周期性分布,表明该头在捕捉季节模式;趋势性头,权重集中在最近的几个时间步(局部性)或分布在序列两端(全局性);滞后性头,权重集中在特定的滞后步长上。MHSA 模块中使用的数学原理主要有以下三部分。

(1) 查询(Query)矩阵 Q 、键(Key)矩阵 K 和值(Value)矩阵 V 的表达式为: $Q = W^q h, K = W^k h$ 和 $V = W^v h$,其中,每个时间步的特征向量 h 通过线性变换 W^q, W^k 以及 W^v 投射到三个向量空间。

(2) 计算 Q 和所有 K 的点积(相似度)衡量重要性,并用 Softmax 将其转化为权重,对 V 进行加权求和,则缩放点积注意力 Attention 的表达式为: $\text{Attention}(Q, K, V) = \text{Softmax}(QK^T / \sqrt{d_k})V$,其中, QK^T 计算了所有查询-键对之间的相似度分数; $\sqrt{d_k}$ 是缩放因子,防止点积结果过大导致 Softmax 梯度消失。

(3) 允许模型在不同的表示子空间中并行学习不同的依赖关系,例如,一个头关注短期波动,另一个头关注周期性模式,则多头机制 MutiHead 的表达式为: $\text{MutiHead}(Q, K, V) = \text{Concat}(h_1, h_2, \dots, h_n)W^o$,其中, $h_i = \text{Attention}(QW_i^q, KW_i^k, VW_i^v)$ ($i = 1, 2, \dots, n$) 为第 i 个头 h_i 学习到的特征, $\text{Concat}(h_1, h_2, \dots, h_n)$ 为合并多头学习到的不同特征, W^o 为线性变换,表示将特征投影到新的表示空间。

1.3 解码器预测序列生成模块

序列到序列的 Seq2Seq 框架为“变长序列到变长序列”的映射任务提供了通用建模范式。该框架首先将经由 TCN 与 MHSA 编码得到的历史时序特征压缩为一个固定维度的上下文向量,进而通过基于 GRU 的解码器以自回归方式逐步生成未来多日的价格预测序列。该结构天然适配“输入 N 日历史数据,输出未来 m 日价格”的时序预测场景。该子模块主要由交叉注意力机制与 GRU 解码器两部分构成。传统的编码器-解码器结构依赖固定的上下文向量传递信息,限制了长序列建模的能力。本模型引入交叉注意力机制,使解码器在每一时间步生成预测时,能够动态计算并聚焦于编码器输出序列中相关性最高的部分,而非仅依赖最终状态。该机制通过加权聚合编码器的全部隐藏状态,生成随时间步变化的上下文向量,显著增强了对长序列关键信息的捕捉能力。解码器部分采用基于 GRU 的循环神经网络,以自回归方式生成的预测序列如下: $X = (x_1, x_2, \dots, x_t) \rightarrow Y = (y_1, y_2, \dots, y_t)$,表示将输入序列 X 映射到输出序列 Y ; $s_t = \text{fdec}(y_{t-1}, s_{t-1}, c)$; $P(y_t | y_{<t}, X) = g(s_t, y_{t-1}, c)$,其中, s_t 是解码器的隐藏状态, g 是 Softmax 层输出概率分布; $\text{Loss} = \sum_{t=1}^{T'} \log P(y_t^* | y_{<t}^*, X)$, $y_t = \text{argmax} P(y_t | y_{<t}, X)$ 。

当训练时,采用最大似然估计来计算 Loss, y_t^* 是真实标签,每一步的预测作为下一步的输入计算得出 y_t 。该设计结合注意力机制的动态信息选择与 GRU 的序列建模,提升多步预测的准确性与鲁棒性。

1.4 业务流程及算法描述

本文的预测方法主要包含以下三个阶段:数据准备阶段,负责采集并预处理股票历史价格数据;模型训练阶段,将处理后的数据输入模型进行训练,并输出相应的股价预测值;模型评估阶段,基于预测值与真实股价计算多项性能评价指标,以量化模型预测效果。具体流程如图 2 所示。首先,数据准备阶段从现有数据源采集原始时序数据,随后按比例划分为训练集、验证集与测试集。为提升模型收敛稳定性,对数据进行标准化处理。实验中对比了最小-最大归一化与零均值归一化两种方法,结果表明零均值归一化能够取得更优的标准化效果,使模型训练过程更为稳定。模型训练阶段首先完成模型构建与参数初始化。模型依次通过 TCN 提取局部时序特征,经 MHSA 增强全局依赖表征,最终输入至带有注意力机制的 Seq2Seq 结构中。解码器以自回归方式逐步生成预测序列。训练过程中采用反向传播算法更新权重,并引入早停机制,迭代至模型性能最优或达到预设轮次为止。

该方法通过上述系统化流程,实现了基于 TCNMS 混合模型的股票价格预测。以下是该算法的详细步骤概述,具体实现细节请参见算法 1 的完整描述。

算法 1. TCNMS 股票价格预测

输入:股票历史价格数据集,共 10 个特征。

输出:未来 5 日股票价格预测值。

1. 通过 akshare 接口获取股票历史数据
2. 创建数据集类 StockDataset(data)
3. 构造输入为 15 和输出为 5 的滑动窗口 build_sliding_window(data, window=15, horizon=5)
4. 创建训练(0.7)、验证(0.1)和测试(0.2)数据集,split_data(data, ratios=[0.7, 0.1, 0.2])
5. 零均值归一化训练、验证以及测试数据集 standard_scaler(数据集)
6. 创建 TCN 模块类 TCN_Block(nn. Module)
7. 分别构建膨胀因子为 1、2 和 4 的卷积层 build_convolution_layer(dilation=[1, 2, 4])
8. 通过残差连接作为该模块的输出 $output = \text{Residual connection} + \text{LayerNorm}$
9. 创建多头自注意力类 MHead_SAttention(nn. Module)
10. 分出 8 头分别计算 8 组权重,通过 bmm 运算后进行合并
11. 创建端到端解码器类 Seq2Seq_Decoder(nn. Module)
12. 建立 GRU_Layer + 交叉注意力 Attention 层
13. Function train()
14. 构建模型、优化器和损失函数。
15. For epoch = 1: max_epochs
16. 训练模型 model.train()并评估模型 model.eval()
17. Function predict()
18. 用最优模型进行预测 best_model(latest_data)

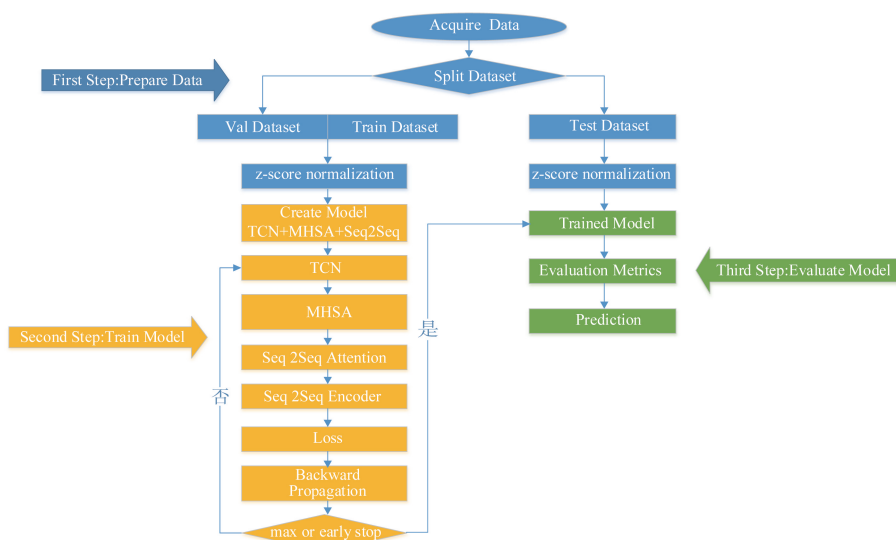


图 2 整体业务流程图

Fig. 2 Overall business process flowchart

2 实验与结果分析

2.1 数据来源

本实验选取国内股票市场中的上证指数、上证央企、上证公用及上证电信四大板块中的多只股票作为研究对象。股票历史数据通过 AKshare 接口从雅虎财经获取,时间范围为 2010 年 1 月 1 日至 2025 年 9 月 30 日,单只股票交易日数约 300 天。原始数据共包含 12 个字段:日期、股票代码、开盘价、收盘价、最高价、最低价、成交量、成交额、振幅、涨跌幅、涨跌额及换手率。在特征选择中,提取其中除日期及股票代码以外的 10 项作为模型输入特征。本文以股票代码为 600004 的股票作为示例,原始股票数据集如表 1 所示。

2.2 数据集划分与预处理

本文研究任务为基于过去 t 个交易日的股票历史数据,预测未来 n 个交易日的收盘价序列,即依据历史窗口 $[1, t]$ 内的特征预测目标窗口 $[t+1, t+n]$ 的收盘价格。为防止未来信息泄露,数据按严格时间顺序划分为训练集、验证集和测试集,比例为 $7:1:2$ 。为提高时序建模的有效性,本文采用滑动窗口方法划分样本:设置固定长度为 t 的滑动窗口,按交易日顺序逐日向前滑动,每次取窗口内 t 天的数据作为输入特征,紧随其后的 n 天收盘价作为输出标签。该数据划分与样本生成方式如图 3 所示。

表 1 原始数据格式(部分)
Tab. 1 Raw data format (partial)

Date	Open	Close	High	Low	Volume	Turnover	Amplitude	Percentage change	Price change	Turnover rate
2010-01-04	10.15	10.15	10.28	10.14	154 789	157 747 809	1.38	0.00	0.00	1.53
2010-01-05	10.15	10.19	10.24	10.14	110 709	112 683 563	0.99	0.39	0.04	1.09
2010-01-06	10.17	10.04	10.19	10.02	94 107	952 058 63	1.67	-1.47	-0.15	0.93
2010-01-07	10.0	9.78	10.05	9.71	89 618	88 611 912	3.39	-2.59	-0.26	0.88
2010-01-08	9.73	10.18	10.25	9.72	149 238	149 969 456	5.32	4.09	0.40	1.47
2010-01-11	10.26	10.39	10.46	10.18	213 822	22 1095 349	2.75	2.06	0.21	2.11
2010-01-12	10.33	10.46	10.47	10.25	128 943	133 964 253	2.12	0.67	0.07	1.27
2010-01-13	10.26	10.31	10.50	10.21	144 030	149 155 601	2.77	-1.43	-0.15	1.42
2010-01-14	10.30	10.68	10.73	10.28	185 264	196 244 561	4.36	3.59	0.37	1.83
2010-01-15	10.67	10.60	10.73	10.53	100 716	106 958 605	1.87	-0.75	-0.08	0.99
2010-01-18	10.60	10.96	11.08	10.50	191 836	208 241 848	5.47	3.40	0.36	1.89
2010-01-19	10.90	11.10	11.20	10.85	202 184	224 091 042	3.19	1.28	0.14	1.99
2010-01-20	11.12	10.99	11.16	10.89	198 384	218 072 462	2.43	-0.99	-0.11	1.95
2010-01-21	11.25	10.84	11.35	10.80	194 460	215 431 588	5.00	-1.36	-0.15	1.92
2010-01-22	10.90	11.08	11.58	10.81	406 182	454 945 617	7.10	2.21	0.24	4.00

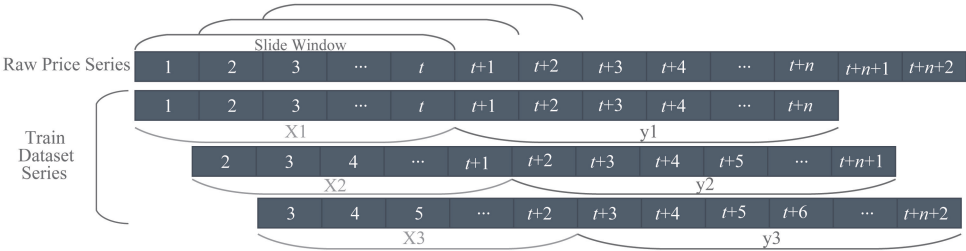


图 3 滑动窗口示意图
Fig. 3 Sliding window illustration

为消除不同特征间量纲差异对模型训练的影响,避免因输入尺度不一致导致神经网络收敛困难及预测性能下降,本文对原始数据进行了标准化处理。具体采用零均值归一化方法,将各特征数据转换为均值为 0、标准差为 1 的分布,使其数值范围主要集中于 $[-1, 1]$ 区间,计算公式为 $x^* = (x - \mu) / (\sigma + \epsilon)$,其中, x 为原始数据值, μ 为 x 的均值, σ 为 x 的标准方差, ϵ 为很小的量 (1×10^{-9}),保证公式中的分母不为零。经过零均值归一化后的数据如表 2 所示。

2.3 实验环境与评估指标

本文实验在 Windows 10 操作系统下进行,开发语言为 Python 3.12。硬件配置为:第 11 代 Intel Core i7-1165G7 处理器(主频 2.80 GHz),内存 8.00 GB。模型训练阶段采用 Adam 优化器对网络权重与偏置进行迭代更新,损失函数选用均方误差,初始学习率设为 1×10^{-4} 。训练过程中,最大迭代轮数设为 100,并引入早停机制监控验证集损失,若连续多轮性能无提升则提前终止。批处理大小设置为 32,以平衡内存占用与训练效率。网络隐藏层维度统一设为 64,随机丢弃率设定为 0.1。模型评估阶段使用训练完成的模型对测试集进行预测,评估指标采用均方误差(Mean Squared Error, MSE)、平均绝对误差(Mean Absolute Error, MAE)及拟合优度(R -squared, R^2),通过对比预测结果与真实值,综合评估模型在股票价格多步预测

任务中的性能。

2.4 实验结果与分析

2.4.1 对比模型。为系统评估本文提出的 TCNMS 模型,选取了当前广泛使用的深度学习时序预测模型进行对比实验,以验证本文模型在股票指数预测任务中的有效性与优越性。所选对比模型如下:(1) LSTM^[16]:经典循环神经网络变体,擅长建模序列长期依赖关系,是时序预测常用基准模型;(2) CNN^[17]:通过卷积核自动提取序列局部模式与短期依赖特征;(3) CNN-LSTM^[18]:结合 CNN 的局部特征提取能力与 LSTM 的长期依赖建模,实现端到端的时序预测;(4) CNN-LSTM-AM^[19]:在 CNN-LSTM 基础上引入注意力机制,动态强化关键特征;(5) LSTM-RNN^[20]:双通道结构,通过不同循环单元并行提取时序特征,增强特征融合能力;(6) ResNet-LSTM^[21]:融合残差卷积模块与 LSTM,构建深层网络以充分提取序列相关性;(7) TCN^[22]:基于膨胀因果卷积的时序网络,支持并行计算,具有明确的长程依赖捕获能力;(8) Seq2Seq^[23]:编码器-解码器框架,专用于处理变长序列的多步预测问题;(9) TCN-Seq2Seq^[14]:结合 TCN 的长程建模与 Seq2Seq 的序列生成能力,实现高效多步预测。其中,前 6 种对比模型作为基线对比模型,用于与本文提出的 TCNMS 模型进行性能对比;后 3 种模型作为消融实验参照模型,用于分析 TCNMS 模型中 3 个组件(TCN、MHSA 和 Seq2Seq)的贡献与必要性。

表 2 归一化处理后的数据(部分)

Tab. 2 Data after normalization(partial)

Date	Open	Close	High	Low	Volume	Turnover	Amplitude	Percentage change	Price change	Turnover rate
2010-01-04	-0.35	-0.35	-0.35	-0.32	0.52	0.12	-0.74	-0.02	-0.00	1.31
2010-01-05	-0.35	-0.34	-0.36	-0.32	0.05	-0.19	-0.97	0.16	0.13	0.59
2010-01-06	-0.34	-0.38	-0.38	-0.35	-0.13	-0.31	-0.58	-0.69	-0.49	0.33
2010-01-07	-0.39	-0.45	-0.41	-0.43	-0.18	-0.35	0.41	-1.20	-0.84	0.25
2010-01-08	-0.46	-0.34	-0.36	-0.43	0.46	0.07	1.52	1.85	1.29	1.21
2010-01-11	-0.32	-0.29	-0.31	-0.31	1.15	0.55	0.04	0.93	0.67	2.26
2010-01-12	-0.30	-0.27	-0.30	-0.29	0.24	-0.04	-0.32	0.29	0.22	0.89
2010-01-13	-0.32	-0.31	-0.30	-0.30	0.40	0.06	0.05	-0.67	-0.49	1.13
2010-01-14	-0.31	-0.21	-0.24	-0.28	0.84	0.38	0.97	1.62	1.19	1.80
2010-01-15	-0.21	-0.24	-0.24	-0.24	-0.21	-0.23	-0.46	-0.36	-0.26	0.43
2010-01-18	-0.23	-0.14	-0.15	-0.22	0.91	0.46	1.60	1.54	1.16	1.90
2010-01-19	-0.15	-0.11	-0.12	-0.13	1.02	0.57	0.29	0.57	0.45	2.07
2010-01-20	-0.10	-0.13	-0.13	-0.12	0.98	0.53	-0.14	-0.47	-0.36	2.00
2010-01-21	-0.06	-0.17	-0.08	-0.14	0.94	0.51	1.33	-0.64	-0.49	1.95
2010-01-22	-0.15	-0.11	-0.02	-0.14	3.20	2.14	2.54	0.99	0.77	5.36

2.4.2 对比实验结果分析。为评估模型在预测任务中的性能,本文将所提模型与基线模型在同一数据集上(股票代码:600004)进行对比实验,并采用均方误差 MSE、平均绝对误差 MAE 及拟合优度 R^2 三个指标进行量化评估。其中,MSE 反映预测误差平方的均值,对较大误差更为敏感,其值越小表明模型整体预测偏差越小;MAE 为预测误差绝对值的均值,用于衡量典型误差水平; R^2 则用于评估模型预测方向的有效性,正向值代表具备预测能力,负值则表明模型无效。各基线对比模型与本文模型的预测结果如表 3 所示。实验结果表明,本文模型在三个指标上均取得最优性能:拟合优度最高,MSE 与 MAE 均为最低。具体而言,前 6 个基线对比模型的平均拟合优度为 0.767 7,而本文模型达到 0.973 8,相对提升约 26.8%。进一步分析发现,基线对比模型中包含 CNN 模块的 3 个模型表现较优,平均拟合优度为 0.874 7,本文模型仍高出约 11.3 个百分点。

该模型训练过程中损失变化情况及最终股价预测如图 4 所示。可以看出,在训练过程中,随着迭代次数的增加,训练损失和验证损失在下降后趋于稳定,且稳定后未出现大幅度地下降或上升状况。总体而言,传统预测模型在处理股票数据非线性、非平稳及高噪声特性时存在明显局限。本文提出的 TCNMS 模型通过引入 TCN 模块有效提取局部时序依赖,结合 MHSA 捕捉全局关键信息,从而更好地识别序列中的重要模

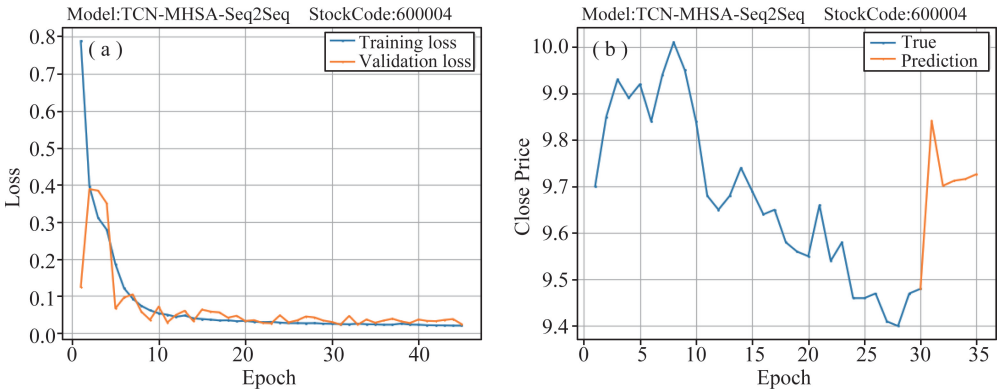
式与特征,提升了对股票价格趋势预测的科学性与准确性。

表 3 各模型股指预测实验结果

Tab. 3 Experimental results of stock index prediction by different models

Model	MSE	MAE	R^2
LSTM	0.186 8	0.385 8	0.498 2
CNN	0.030 3	0.146 7	0.918 5
CNN-LSTM	0.073 3	0.224 6	0.802 9
CNN-LSTM-AM	0.036 2	0.136 4	0.902 8
LSTM-RNN	0.045 2	0.170 4	0.878 5
ResNet-LSTM	0.146 8	0.320 9	0.605 5
TCNMS	0.009 7	0.072 9	0.973 8

2.4.3 模型性能检验。为验证 TCNMS 模型的有效性,本文首先在上证指数成分股 600004 上进行了实验,结果表明其性能优于基线对比模型。为进一步检验模型的普适性与稳健性,扩展实验至不同主题指数及个股,选取以下四只股票作为样本:上证指数—600000(浦发银行)、上证公用—600023(浙能电力)、上证央企—600019(宝钢股份)、上证电信—600498(烽火通信)。使用相同实验设置,分别采用 TCNMS 及各基线对比模型进行预测,结果如表 4 所示。



注:(a) 训练损失变化图;(b) 股价预测。

图 4 训练损失变化图及股价预测

Fig. 4 Training loss variation and stock price prediction of multiple stocks

表 4 不同股票数据下各模型性能对比

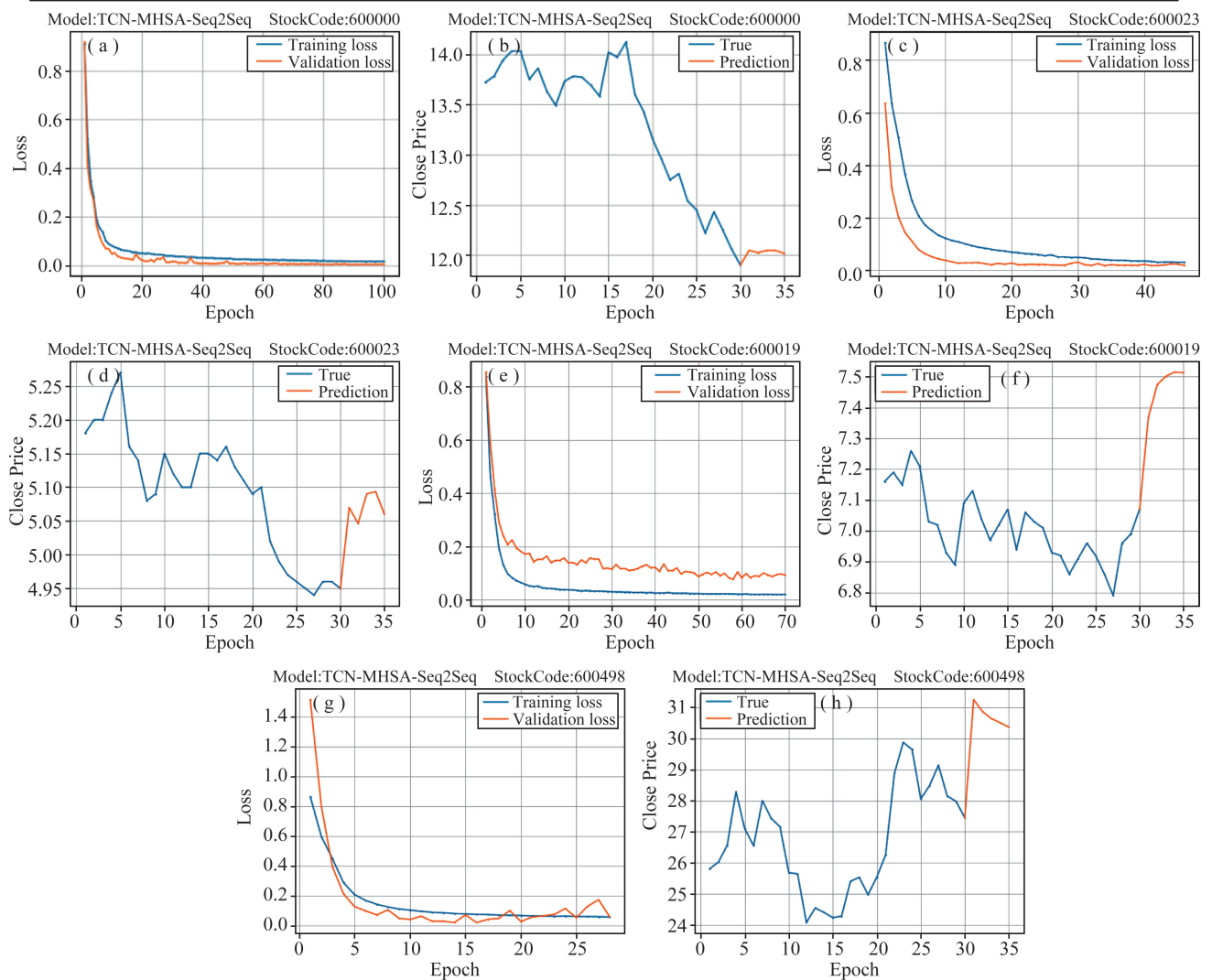
Tab. 4 Performance comparison of model on various stock datasets

Stock Code	Model	MSE	MAE	R^2
600000	LSTM	0.132 1	0.316 7	0.707 9
	CNN	0.008 6	0.067 6	0.980 9
	CNN-LSTM	0.031 9	0.136 7	0.929 6
	CNN-LSTM-AM	0.034 0	0.159 7	0.924 7
	LSTM-RNN	0.028 4	0.141 9	0.937 1
	ResNet-LSTM	0.192 7	0.387 5	0.573 8
	TCNMS	0.008 8	0.070 4	0.980 6
600023	LSTM	0.135 8	0.294 1	0.440 7
	CNN	0.019 9	0.105 3	0.918 3
	CNN-LSTM	0.098 5	0.246 9	0.594 4
	CNN-LSTM-AM	0.084 4	0.234 1	0.652 6
	LSTM-RNN	0.043 6	0.164 0	0.820 4
	ResNet-LSTM	0.111 6	0.273 7	0.540 6
	TCNMS	0.020 6	0.107 5	0.915 1

(续)表4 不同股票数据下各模型性能对比

(Continued)Tab. 4 Performance comparison of model on various stock datasets

Stock Code	Model	MSE	MAE	R^2
600019	LSTM	0.095 7	0.259 3	0.287 2
	CNN	0.020 8	0.108 9	0.844 7
	CNN-LSTM	0.081 0	0.226 1	0.396 5
	CNN-LSTM-AM	0.080 4	0.220 4	0.401 2
	LSTM-RNN	0.046 5	0.170 9	0.653 5
	ResNet-LSTM	0.125 6	0.287 8	0.064 7
	TCNMS	0.014 9	0.090 3	0.888 3
600498	LSTM	0.264 1	0.348 0	0.169 1
	CNN	0.072 6	0.163 9	0.772 3
	CNN-LSTM	0.151 3	0.278 7	0.524 4
	CNN-LSTM-AM	0.170 5	0.264 4	0.464 1
	LSTM-RNN	0.100 0	0.230 4	0.685 8
	ResNet-LSTM	0.232 2	0.364 7	0.269 4
	TCNMS	0.058 3	0.177 9	0.817 6



注:(a) 600000 训练损失;(b) 600000 股价预测;(c) 600023 训练损失;(d) 600023 股价预测;(e) 600019 训练损失;(f) 600019 股价预测;(g) 600498 训练损失;(h) 600498 股价预测。

图5 多支股票训练损失变化图及股价预测

Fig. 5 Training loss variation and stock price prediction of multiple stocks

由表 4 可知,从整体上看,在不同股指和个股样本下,TCNMS 模型在各项指标上的表现普遍优于基线对比模型,说明本文模型能够较好地适应不同的股票数据集,验证了模型的普适性与鲁棒性。此外,可以注意到各基线对比模型在个股数据上的拟合优度相比于股票 600004 均有一定程度的下滑。主要归因于个股价格波动具有更高的随机性和噪声,增加了建模难度,导致预测结果的不稳定。而本文所提的 TCNMS 模型的预测拟合优度能够稳定保持在 80%以上,由此说明了模型在处理较复杂的个股数据时也具有独特的建模能力和优势,进一步验证了其在实际应用中的潜力。本文所提模型在训练过程中的损失变化及股价预测如图 5 所示。可以看出模型在训练时的收敛速度很快,远没有到 100 轮次即实现早停。这也验证了本模型具有较高的训练效率。

2.4.4 消融实验。为了验证本文提出的 TCNMS 预测模型中关键组件的必要性与有效性,本文进行了消融实验,在相同的数据集上(上证指数的 600004 和 600011、上证央企的 600030 和 600048)设计了 4 种消融策略,由表 5 所示。其中策略 4 为本文模型,通过拟合优度评估各个策略的模型性能。

从表 5 的实验结果可知,策略 1 和策略 2 分别为单组件模型,这两个模型的拟合优度相当,其中由 TCN 构成的策略 1 略高于 GRU 构成的策略 2。说明 TCN 的大视野特征提取能力优于 GRU。策略 3 为由策略 1 与策略 2 中组件构成的混合模型,策略 3 的拟合优度高于策略 1 和策略 2。说明编码器采用 TCN,解码器采用 GRU 的混合模型的预测能力要好于单组件模型。与策略 3 相比,本文模型(策略 4)在策略 3 模型的基础上又增加了多头自注意力机制,从全局把握序列数据的依赖关系,其拟合优度更高,比策略 3 平均提升了 0.75 个百分点。说明本文模型综合了各个模块的优势,有利于整体模型提取高级、深度的特征,从而提高模型对不同时间戳之间复杂依赖关系的建模能力。

表 5 不同消融策略对模型预测性能的影响

Tab. 5 Ablation study on model performance under different strategies

Stock Code	Strategy	MSE	MAE	R^2
600011	TCN	0.045 8	0.158 1	0.899 9
	Seq2Seq	0.064 9	0.194 9	0.858 1
	TCN-Seq2Seq	0.049 5	0.160 6	0.891 7
	TCNMS	0.042 2	0.153 0	0.907 7
600048	TCN	0.026 1	0.118 0	0.972 8
	Seq2Seq	0.045 1	0.161 2	0.953 2
	TCN-Seq2Seq	0.025 7	0.119 1	0.973 2
	TCNMS	0.025 6	0.119 4	0.973 3
600004	TCN	0.013 3	0.087 6	0.964 3
	Seq2Seq	0.017 4	0.100 5	0.953 3
	TCN-Seq2Seq	0.009 3	0.069 2	0.975 0
	TCNMS	0.008 3	0.068 4	0.977 6
600030	TCN	0.023 6	0.093 6	0.946 9
	Seq2Seq	0.028 6	0.114 4	0.935 5
	TCN-Seq2Seq	0.021 1	0.089 7	0.952 4
	TCNMS	0.017 2	0.086 6	0.961 3

3 结论

为解决传统时间序列预测模型在捕捉长期依赖、全局动态特征以及多步预测精度方面的不足的问题,本文提出了 TCNMS 模型,验证了融合时序卷积网络 TCN、多头自注意力机制 MHSA 与序列到序列架构 Seq2Seq 的混合深度学习框架,用于实现股票价格的多步预测。通过理论分析、模型构建、对比实验与结果讨论,证明了该模型结合 TCN 的局部特征提取与长程依赖建模能力、MHSA 的全局动态聚焦机制,以及 Seq2Seq 的序列生成特性,有效形成了协同互补的时序建模体系;预测性能显著优于基准模型,在多个评估指标上,TCNMS 模型均优于单一的 LSTM、GRU 及标准 Seq2Seq 等基准模型,显示其在精度上的明显提升。综上所述,本研究通过集成 TCN 的高效并行特征提取、MHSA 的重点信息聚焦与 Seq2Seq 的序列生成

机制,为复杂金融时序数据的多步预测提供了一种新颖的解决方案。但是,该模型仍存在一定的局限性,未来工作尝试采用图神经网络构建股票特征间关联矩阵,以提升模型特征提取能力与抗噪声鲁棒性。

参 考 文 献

- [1] HSU M L, LI H H, LI S T. ARIMA-driven memory market insights: forecasting DRAM spot price[J]. *Asia Pacific Management Review*, 2025, 30(2): 100351.
- [2] MUSHTAQ H, ISHTIAQ M, JAMAL S, et al. Regime-Switching volatility and risk quantification in South Asian and developed stock markets: a comparative perspective using Markov-Switching GARCH with MLE and MCMC estimations[J]. *The North American Journal of Economics and Finance*, 2026, 82: 102576.
- [3] OSMAN E G A, OTAIBI F A. Integrating deep learning and econometrics for stock price prediction: a comprehensive comparison of LSTM, transformers, and traditional time series models[J]. *Machine Learning with Applications*, 2025, 22: 100730.
- [4] ABDULRAHIM H, ALSHIBANI S M, IBRAHIM O, et al. Prediction OPEC oil price utilizing long short-term memory and multi-layer perceptron models[J]. *Alexandria Engineering Journal*, 2025, 110: 607-612.
- [5] RICCHIUTI F, SPERLI G. An advisor neural network framework using LSTM-based informative stock analysis[J]. *Expert Systems with Applications*, 2025, 259: 125299.
- [6] KALRA R, SINGH T, MISHRA S, et al. An efficient hybrid approach for forecasting real-time stock market indices[J]. *Journal of King Saud University-Computer and Information Sciences*, 2024, 36(8): 102180.
- [7] HADAM L, AMRANE M, HANDEL N, et al. Deep sequence learning architectures for predicting cyclic soil stress-strain behavior: comparative evaluation of LSTM, GRU, TCN, and hybrid models[J]. *Transportation Geotechnics*, 2026, 56: 101822.
- [8] WANG Y L, SU Y K, ZHENG Z Z, et al. Bayesian-optimized CNN-LSTM neural network for predicting road construction dust concentrations[J]. *Developments in the Built Environment*, 2026, 25: 100843.
- [9] SHERKATGHANAD Z, GHAZANFARI A, MAKARENKOV V. A self-attention-based CNN-Bi-LSTM model for accurate state-of-charge estimation of lithium-ion batteries[J]. *Journal of Energy Storage*, 2024, 88: 111524.
- [10] KAVIANPOUR P, KAVIANPOUR M, JAHANI E, et al. A CNN-BiLSTM model with attention mechanism for earthquake prediction [J]. *The Journal of Supercomputing*, 2023, 79(17): 19194-19226.
- [11] UÇKAN T. Integrating PCA with deep learning models for stock market forecasting: an analysis of Turkish stocks markets[J]. *Journal of King Saud University-Computer and Information Sciences*, 2024, 36(8): 102162.
- [12] SEBASTIAN A, TANTIA D V. Multi-variate LSTM with attention mechanism for the Indian stock market[J]. *International Journal of Information Management Data Insights*, 2025, 5(2): 100350.
- [13] WANG Y X, GHAZALI M F, AB RAZAK R, et al. Complex system and PS-LSTM prediction of cryptocurrencies, stocks, bonds, exchange rates and commodities[J]. *Physica A: Statistical Mechanics and its Applications*, 2025, 679: 130976.
- [14] DAI Y M, YU W J. Short-term power load forecasting based on Seq2Seq model integrating Bayesian optimization, temporal convolutional network and attention[J]. *Applied Soft Computing*, 2024, 166: 112248.
- [15] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. *ArXiv Preprint: arXiv:1803.01271*, 2018.
- [16] SIAMI-NAMINI S, NAMIN A S. Predicting stock market prices using LSTM networks: a time series forecasting approach[J]. *Intelligent Systems in Accounting, Finance and Management*, 2019, 26(2): 55-64.
- [17] BOROVYKH A, BOHTE S, OOSTERLEE C W. Conditional time series forecasting using a deep convolutional neural network[EB/OL]. *ArXiv Preprint: arXiv:1703.04691*.
- [18] FANG Y, WANG J, OUYANG D. A hybrid CNN-LSTM model for stock price prediction[C]//*Proceedings of 2018 International Conference on Big Data and Artificial Intelligence (BDAI)*. London, UK: IEEE, 2018: 54-58.
- [19] BORRÉ A, SEMAN L O, CAMPOGARA E, et al. Machine fault detection using a hybrid CNN-LSTM attention-based model[J]. *Sensors*, 2023, 23(9): 4512.
- [20] HARBAOUI H, ELHADJAMOR E A. Enhancing stock price prediction: LSTM-RNN fusion model[J]. *Procedia Computer Science*, 2024, 246: 920-929.
- [21] DUBEY A, DIXIT I. Stock price prediction using ResNLS technique[C]//*Seventh International Conference on Recent Trends in Image Processing and Pattern Recognition (RTIP2R-2024)*. *Procedia Computer Science*, 2025, 260: 752-760.
- [22] GUO W C, LI Z G, GAO C, et al. Stock price forecasting based on improved time convolution network[J]. *Computational Intelligence*, 2022, 38(4): 1474-1491.
- [23] CHEN L, CHEN J. Stock price prediction based on Seq2Seq with attention mechanism[C]//*Proceedings of 2018 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*. Chengdu, China: IEEE, 2018: 157-160.

(责任编辑:刘庆松)