



基于 BiLSTM-SA-TCN 时间序列模型 在股票预测中的应用

摘要

针对股票预测模型存在时效性和预测功能单一化的问题,本文在长短期记忆网络(LSTM)的基础上,提出了融合自注意力机制(SA)和时间卷积网络(TCN)的双向长短期记忆(BiLSTM)神经网络(BiLSTM-SA-TCN)股票预测模型。BiLSTM-SA-TCN模型中的学习单元和预测单元可以有效学习重要的股票数据,同时能够抓取长时间的依赖信息,输出次日股票收盘价预测值。实验结果表明,BiLSTM-SA-TCN模型在多个数据集上的预测结果更加稳定,模型泛化能力较高,在对比实验中,BiLSTM-SA-TCN模型在大部分数据集上均方根误差最小,平均绝对值误差最小,拟合度 R^2 最优。

关键词

股票价格预测;长短期记忆网络;注意力机制;时间卷积网络

中图分类号 TP183

文献标志码 A

收稿日期 2022-10-31

资助项目 重庆市自然科学基金(cstc2021ycjh-bgzxm0088);重庆市教育委员会科学技术研究计划项目(KJQN201903402)

作者简介

杨智勇,男,博士,教授,主要研究领域为人工智能、大数据处理与分析。zyy@cqvie.edu.cn

1 重庆工程职业技术学院 大数据与物联网学院,重庆,402260

2 重庆师范大学 计算机与信息科学学院,重庆,401331

3 重庆工程职业技术学院 财经与旅游学院,重庆,402260

0 引言

20世纪90年代,中国证券市场重新开放并正式改革,中国股市随之进入新的发展阶段。随着市场经济改革的深化,股票市场不仅扩大了经济活动的规模,同时也提高了人们的投资意识^[1-2]。中国股市总市值从1990年的3.3亿美元增长到2022年的12万亿美元,中国股市已经成为世界第二大股市。股票因其交易便利、高风险和高回报吸引了广大投资者,如何在股票市场中获取高收益则成为投资者们津津乐道的话题。股票预测是一种预测股价的方法和有效的投资方式,它通过分析股票市场的历史数据,来预测未来股票价格的运动趋势。

股票预测研究早期主要通过技术分析和统计学模型来预测股票走势,如道氏理论、自回归移动平均模型(Autoregressive Integrated Moving Average, ARIMA)、量价理论等,但面对日益壮大的股票市场及其存在的海量数据,仅仅依赖传统的技术分析不能满足市场的需求。随着机器学习的兴起,很多新兴的技术都广泛地应用在股票数据研究中,如决策树(Decision Tree, DT)、支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest, RF)等^[3-5]。然而,股票价格数据具有非线性、高噪声、强时变性等特点^[6],机器学习模型通常会因过拟合而降低预测模型的泛化能力,以及模型在训练数据上表现良好,但在新数据上表现不佳。在时间序列预测中,由于数据随时间的变化而变化,基于机器学习的模型可能无法很好地适应所有类型的数据。

近几年来,深度学习技术不断发展,在语音和图像识别方面取得的成果,远远超过先前相关技术。研究表明,深度学习方法能够更好地预测股票未来走势,有更高的泛化能力,如文献[7-11]分别使用单一结构的深度学习模型,对股票价格进行预测。然而,单一结构网络模型预测精度不高,因此,学者们对单一网络结构进行了改进。Kim等^[12]利用CNN(Convolutional Neural Network)提取股票蜡烛图特征、LSTM提取股票时间特征,CNN-LSTM模型在标准普尔500指数上的实验均优于单个模型。Lee等^[13]提出一种结合门递归单元(Gate Recurrent Unit, GRU)和注意力机制(Attention Mechanism, AM)的股票预测模型(GRU-Attention),将股票基础数据作为模型输入,预测未来交易日的涨跌概率。Lu等^[14]使用一维卷积神经网络提取输入数据的特征,将提

取的特征输入基于注意力机制的双向长短期记忆网络,最后通过全连接层输出预测值,实验结果表明 CNN-BiLSTM-AM 模型具有较好的泛化能力.但上述预测模型功能单一,大多只针对特定的股票或单一股指,同时,预测模型存在时效性问题.因此,股票预测模型的网络结构在精度和深度方面还有很大的优化空间.

有鉴于此,本文提出双向选股策略选取股票实验对象,以长短期记忆网络(Long Short-Term Memory, LSTM)为基础,提出融合自注意力机制(Self-Attention, SA)和时间卷积网络(Temporal Convolution Network, TCN)的双向长短期记忆神经网络(BiLSTM-SA-TCN)股票预测模型,通过在多个股票数据集的实验,验证了本文模型的可行性和有效性.实验结果表明,本文模型具有良好的泛化能力,并解决了股票预测模型存在的时效性问题.

1 算法介绍

1.1 BiLSTM

LSTM 是循环神经网络(Recurrent Neural Network, RNN)的一类变种结构,采取递归的网络架构,结合适当的基于梯度的学习算法,解决了 RNN 模型由于输入序列过长而产生的梯度消失问题^[15]. LSTM 层由一组循环连接的存储单元组成,每个存储单元包含一个或多个细胞结构,每个细胞结构通过三个“门”(即输入门、输出门、遗忘门)保留或遗忘数据信息.

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

$$i_t = \sigma(W_i * h_{t-1} + W_i * x_t + b_i), \quad (2)$$

$$\tilde{C} = \tanh(W_c * h_{t-1} + W_c * x_t + b_c), \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (4)$$

$$o_t = \sigma(W_o * h_{t-1} + W_o * x_t + b_o), \quad (5)$$

$$h_t = o_t * \tanh(C_t). \quad (6)$$

首先,遗忘门根据上个时刻的 h_{t-1} 和当前的输入数据 x_t 得到 f_t ,来决定数据的保留与舍弃;其次,输入门通过 σ 层产生信息的更新值 i_t ,并通过 \tanh 层生成新的候选向量 \tilde{C}_t ;然后,遗忘门将舍弃的数据 f_t 与上一时刻的细胞状态 C_{t-1} 相乘,并与细胞状态更新的数据相结合,得到新的细胞状态 C_t ;最后,输出门通过 σ 层,得到初始输出 o_t ,并把 C_t 输入到 \tanh 层,得到输出部分 h_t .

1.2 自注意力机制

注意力机制起源于人脑,人类视觉可以快速聚焦在注意力权重高的关键区域.注意力机制可以计算输入序列中每个元素的重要性,并根据重要性为不同的特征分配不同的权重^[16].注意力机制的计算图 1 所示,主要为如下 3 个部分:

1) 计算每个查询(Query)和关键字(Key)之间的相似度,以获得相应的权重值(Value); 2) 使用 Softmax 函数对权重进行归一化,获得权重系数; 3) 将权重系数和相应值加权求和,以获得最终的注意力值.如式(7)所示:

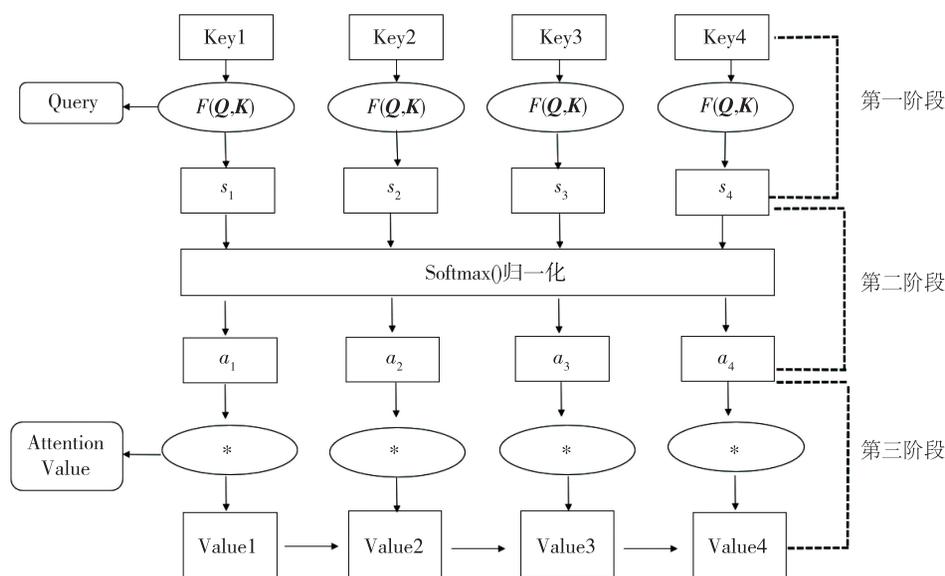


图 1 注意力机制

Fig. 1 Attention mechanism

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}. \quad (7)$$

而自注意力机制是注意力机制的一种特例,它的 Query、Key、Value 3 个矩阵为同一个输入,自注意力机制能把输入序列的不同信息联系起来,更擅长捕获重要数据。

1.3 时间卷积网络

TCN 是 Lea 等^[17]在卷积神经网络的基础上提出的一种可以用于处理时间序列数据的网络结构。TCN 的网络结构是由堆叠的一维全连接卷积层组成,每个卷积层具有独特的因果卷积结构,如图 2 所示,它可以接受任意长度的输入序列,并输出相同长度,在 T 时刻的输出只与 T 之前的元素进行卷积,这避免了未来信息的“泄露”。因果卷积计算公式如下:

$$F(s) = \sum_{i=0}^{k-1} f(i)x_{s-di}, \quad (8)$$

其中, x 为输入, f 为过滤器, d 为扩张系数, k 为卷积核大小, $s - di$ 决定了只对过去的输入数据做卷积操作。

1.4 BiLSTM-SA-TCN 预测模型

为了解决股票预测模型预测功能单一化和时效性问题,本文在 LSTM 神经网络基础上进行改进,提

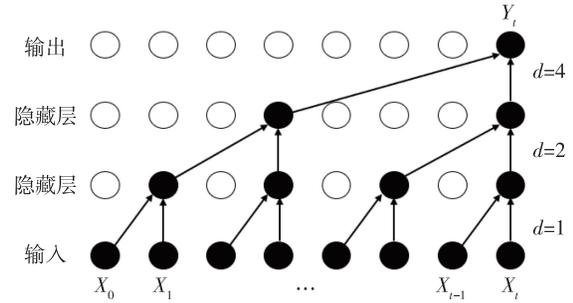


图 2 时间卷积网络

Fig. 2 Temporal convolution network

出一种融合自注意力机制和时间卷积网络的 BiLSTM 股票预测模型。通过引入自注意力机制可以从众多股票特征信息中选择出当前任务中更重要的信息,即能够有效解决循环神经网络的时效性问题;通过引入时间卷积网络,使得模型可以并行处理时间序列信息,提高模型预测准确度和训练效率。预测模型结构如图 3 所示。

本文使用 BiLSTM 作为输入序列数据的特征提取和预处理的网络层,输入序列数据为 3D 张量,即样本 (Samples)、时间步长 (Time_steps)、特征 (Features)。BiLSTM 不仅能将长输入序列转换成强化集合特征,而且,通过使用两个 LSTM 网络,一个从前

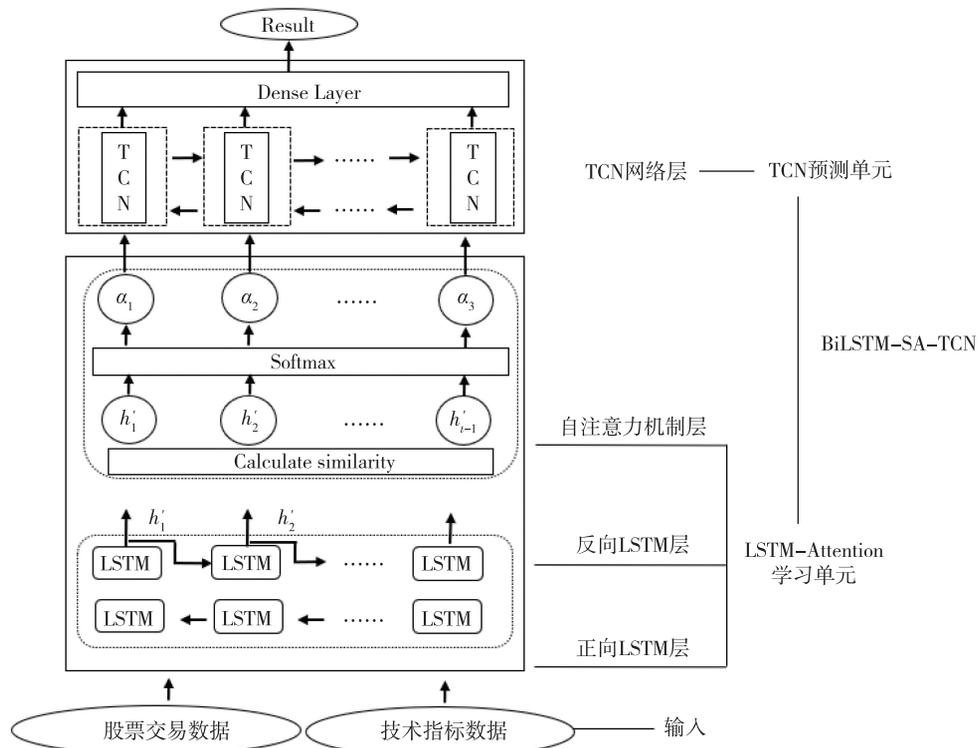


图 3 股票预测模型结构

Fig. 3 Structure of the proposed stock price prediction model

向后处理序列,另一个从后向前处理序列,可以捕获序列中的双向信息.因为前面的股票数据可能会对后面的产生影响,可以有效学习股票数据中长期依赖关系,挖掘输入序列的特征规律,从而提高序列建模的准确性.

在股票预测分析的过程中,股票序列数据的时间范围通常较大,LSTM 在处理较长的时序序列时会出现梯度衰减过快的问题,这会影响 LSTM 网络学习股票数据中的重要特征信息,导致模型出现重要特征信息丢失的情况,从而导致模型的预测准确率下降.针对上述情况,本文改进了网络结构,引入自注意力机制,使得模型可以优先考虑隐藏层中与股票数据的重要特征信息,忽略其他无关信息,通过计算多源特征融合股票序列中每个元素的重要性,并根据重要性为不同的特征分配不同的权重,使得模型可以关注到时间序列其他位置的内容,并将这些内容作为帮助指引,提高模型的预测准确率.

随着网络层数的增加,以及迭代次数增加,神经网络在训练过程中梯度更新的幅度过大,导致权重变化过快,从而导致模型出现网络退化效应,即模型在处理新数据时表现不佳,因为它没有很好地捕捉训练数据之间的潜在关系.为此,本文引入 TCN 网络层处理时间序列,通过 TCN 中的卷积层来捕获序列的依赖关系,揭示数据内部的隐藏关系.同时,残差连接可以减少网络深度和参数数量,提高模型的泛化能力和训练效率.最后,将 TCN 网络层计算结果通过 Dense 层输出,输出结果为次交易日的股票收盘价预测值.

2 实验结果及分析

2.1 股票数据集构建

通过 Tushare(<https://tushare.pro/>)提供的接口获取股票历史交易数据,其中,历史交易数据主要包括开盘价(open)、收盘价(close)、最低价(low)、最高价(high)、成交额(amount)、成交量(volume)、涨跌幅(change_percent)和涨跌额(change),并通过每日收盘价计算得到5日均线 and 布林线,两个技术指标分别表示股价的周、月趋势.以往的研究^[18-21]仅选取沪深指数或沪深股市中的几只股票进行实验,所选股票的价格趋势较为稳定,缺乏波动性,且实验覆盖面较差,导致模型实验结果不具说服力.因此,本文提出双向的选股策略选取实验对象,使得实验覆盖面更广.横向,通过公司市值大小进行选择,分为

大盘股和中小盘股;纵向,按照股票分类进行选择,包括金融、房地产、煤炭、钢铁、有色金属、石油化工、汽车等7大总类.选取的股票如表1所示.

表1 实验所选股票

Table 1 Selected stocks for experiment

| 所属类别 | 大盘股 | 中小盘股 |
|------|----------------|----------------|
| 金融 | 招商银行(SH600036) | 国金证券(SH600109) |
| 房地产 | 保利发展(SH600048) | 西藏城投(SH600773) |
| 煤炭 | 中国神华(SH601088) | 电投能源(SZ002128) |
| 钢铁 | 中信特钢(SZ000708) | 方大炭素(SH600516) |
| 有色金属 | 天齐锂业(SZ002466) | 云南铜业(SZ000878) |
| 石油化工 | 中国石油(SH601857) | 岳阳兴长(SZ300164) |
| 汽车 | 比亚迪(SZ002594) | 东风汽车(SH600006) |

本实验选取每只股票最近2300个交易日的股票数据进行实验,时间范围2012年8月—2022年8月.

将股票交易数据和相关技术指标(5日均线、布林线)作为股票数据集.但由于股票数据集的数据存在较大差异,为了消除量纲,需要对数据作归一化处理^[22],即将数据按比例缩放到一个较小的特定区间内.数据的归一化可以方便数据的处理,以及保证模型训练时收敛加快,有助于提高模型的性能.

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (9)$$

其中, X_{\min} 和 X_{\max} 分别为数据集中的最小值和最大值.

2.2 神经网络模型训练策略

本文实验在 Windows 10 操作系统下搭建,使用 GPU(Graphics Processing Unit)版本的 Tensorflow 框架,实验环境配置如表2所示.

表2 实验环境配置

Table 2 Experimental environment configuration

| 实验环境 | 环境配置 |
|--------|-----------------|
| 操作系统 | Windows 10 |
| 内存 | 16 GB |
| GPU | Geforce RTX2060 |
| 编程语言 | Python3.8 |
| 深度学习框架 | Tensorflow |

在本实验中,滑动窗口周期大小为10,即通过前10天的股票数据预测后一天的股票收盘价.在训练配置参数中,首先需要更新神经网络中的权重和偏置,使用随机梯度下降法使损失函数最小化,其次需

要控制训练过程的迭代次数以便逐步提高模型的准确性,epoch 设置为 100.为了有效地利用计算资源和加速模型的训练,batch size 设置为 5,学习率设置为 0.001,其大小会影响模型的训练速度和稳定性以及最终的训练效果.具体模型参数设置如表 3 所示.

表 3 模型参数设置

Table 3 Model parameter setting

| 模型参数 | 设置值 |
|-----------------|---------|
| BiLSTM 层神经元 | 64 |
| BiLSTM 层激活函数 | RELU |
| Attention 层激活函数 | Softmax |
| TCN 层神经元 | 32 |
| TCN 层激活函数 | RELU |

评价指标包括均方根误差 (RMSE)、平均绝对误差 (MAE) 和决定系数 (R^2):

$$e_{RMSE} = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (10)$$

$$e_{MAE} = \frac{1}{n} \sum_{i=1}^n | \hat{y}_i - y_i |, \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (12)$$

其中, \hat{y}_i 为预测值, y_i 为真实值.

2.3 实验及结果分析

首先使用本文模型对不同时期的股票数据进行实验,其次使用现有的股票预测模型与不同的基准模型进行对比.

2.3.1 股票预测

使用本文模型对 14 只股票进行实验,并通过不同时期(2012-08—2021-08、2012-08—2022-02、2012-08—2022-07、2012-08—2022-08)的股票数据,来验证本文模型的有效性,同时检验本文模型是否存在时效性的问题,实验结果如图 4 所示.由图 4 可以发现,BiLSTM-SA-TCN 模型在不同时期的 RMSE 的平均值分别为 0.087、0.085、0.083 和 0.083, R^2 的平均值分别为 0.938、0.939、0.937 和 0.94, MAE 的平均值分别为 0.035、0.037、0.032 和 0.032,表明 BiLSTM-SA-TCN 模型处理新数据时表现良好,预测结果较为平稳,能够捕捉训练数据之间的潜在关系,具有较高的准确性和泛化能力,且不存在时效性问题.

2.3.2 预测模型对比实验

为了进一步验证 BiLSTM-SA-TCN 模型的有效

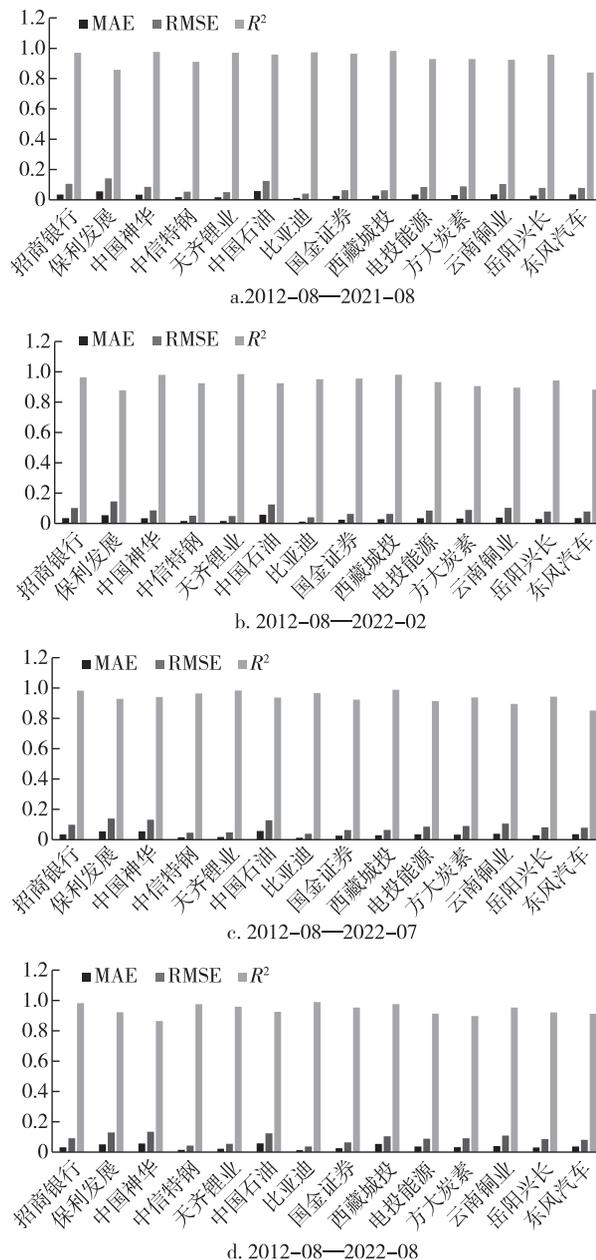


图 4 BiLSTM-SA-TCN 模型的预测性能

Fig. 4 Prediction performance of the proposed BiLSTM-SA-TCN model

性和先进性,需要与其他基准模型进行对比实验.基准模型选择如下:CNN-LSTM^[12],利用长短期记忆模型(LSTM)建立一个预测未来股票市场价值的模型;GRU-Attention^[13],结合门控循环神经网络和注意力机制的预测模型,对股票未来的涨跌进行预测;CNN-BiLSTM-AM^[14],CNN 用于提取输入数据的特征,BiLSTM 用来提取的特征数据来预测股票第 2 天的收盘价,利用 AM 捕捉特征状态对过去不同时刻股票收盘价的影响.实验结果如表 4—10 所示.

表 4 不同预测模型在招商银行和国金证券的评价指标对比

Table 4 Prediction performance comparison of different models for stock prices of China Merchants Bank Co., Ltd. and Sinolink Securities Co., Ltd.

| 模型 | 招商银行 | | | | 国金证券 | | | |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|
| | RMSE | MAE | R^2 | t/s | RMSE | MAE | R^2 | t/s |
| LSTM | 0.104 | 0.037 | 0.968 | 191 | 0.062 | 0.025 | 0.952 | 181 |
| CNN-LSTM | 0.162 | 0.092 | 0.742 | 159 | 0.014 | 0.009 | 0.951 | 167 |
| GRU-Attention | 0.104 | 0.035 | 0.723 | 172 | 0.062 | 0.023 | 0.961 | 176 |
| CNN-BiLSTM-AM | 0.157 | 0.089 | 0.768 | 187 | 0.062 | 0.025 | 0.981 | 187 |
| 本文 | 0.106 | 0.035 | 0.972 | 164 | 0.063 | 0.025 | 0.966 | 169 |

表 5 不同预测模型在保利发展和西藏城投的评价指标对比

Table 5 Prediction performance comparison of different models for stock prices of Poly Developments and Holdings Group Co., Ltd. and Tibet Urban Development and Investment Co., Ltd.

| 模型 | 保利发展 | | | | 西藏城投 | | | |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|
| | RMSE | MAE | R^2 | t/s | RMSE | MAE | R^2 | t/s |
| LSTM | 0.143 | 0.058 | 0.748 | 193 | 0.064 | 0.028 | 0.979 | 197 |
| CNN-LSTM | 0.143 | 0.058 | 0.822 | 166 | 0.065 | 0.028 | 0.975 | 178 |
| GRU-Attention | 0.143 | 0.056 | 0.858 | 202 | 0.063 | 0.027 | 0.975 | 206 |
| CNN-BiLSTM-AM | 0.144 | 0.057 | 0.816 | 191 | 0.063 | 0.028 | 0.962 | 179 |
| 本文 | 0.142 | 0.056 | 0.859 | 171 | 0.063 | 0.027 | 0.984 | 181 |

表 6 不同预测模型在中国神华和电投能源的评价指标对比

Table 6 Prediction performance comparison of different models for stock prices of China Shenhua Energy Company Limited and Inner Mongolia Dian Tou Energy Corporation Limited

| 模型 | 中国神华 | | | | 电投能源 | | | |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|
| | RMSE | MAE | R^2 | t/s | RMSE | MAE | R^2 | t/s |
| LSTM | 0.085 | 0.034 | 0.974 | 198 | 0.084 | 0.035 | 0.909 | 201 |
| CNN-LSTM | 0.088 | 0.038 | 0.965 | 176 | 0.085 | 0.036 | 0.831 | 179 |
| GRU-Attention | 0.085 | 0.033 | 0.972 | 201 | 0.084 | 0.034 | 0.905 | 204 |
| CNN-BiLSTM-AM | 0.087 | 0.034 | 0.943 | 186 | 0.086 | 0.036 | 0.827 | 191 |
| 本文 | 0.086 | 0.034 | 0.977 | 178 | 0.085 | 0.035 | 0.935 | 185 |

表 7 不同预测模型在中信特钢和方大炭素的评价指标对比

Table 7 Prediction performance comparison of different models for stock prices of Citic Pacific Special Steel Group Co., Ltd. and Fangda Carbon New Material Co., Ltd.

| 模型 | 中信特钢 | | | | 方大炭素 | | | |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|
| | RMSE | MAE | R^2 | t/s | RMSE | MAE | R^2 | t/s |
| LSTM | 0.055 | 0.019 | 0.764 | 199 | 0.089 | 0.033 | 0.916 | 208 |
| CNN-LSTM | 0.057 | 0.019 | 0.896 | 167 | 0.090 | 0.038 | 0.902 | 176 |
| GRU-Attention | 0.055 | 0.018 | 0.874 | 197 | 0.089 | 0.032 | 0.919 | 201 |
| CNN-BiLSTM-AM | 0.054 | 0.018 | 0.902 | 191 | 0.089 | 0.033 | 0.912 | 205 |
| 本文 | 0.054 | 0.018 | 0.912 | 172 | 0.085 | 0.032 | 0.935 | 181 |

从表 4—10 可以发现, LSTM 模型的评价指标大多低于 CNN-BiLSTM-AM、GRU-Attention 和 BiLSTM-SA-TCN 模型,且训练效率较低.虽然 LSTM 能够保留或遗忘序列特征信息,但单一的 LSTM 模型的网络

结构存在缺陷,导致 LSTM 模型只能大致拟合股票走势,预测值与真实值存在较大误差,模型的拟合度也较低.对于 CNN-LSTM 模型而言,通过 CNN 提取股票数据特征,再通过 LSTM 网络进行学习,能较好

表 8 不同预测模型在天齐锂业和云南铜业的评价指标对比

Table 8 Prediction performance comparison of different models for stock prices of Tianqi Lithium Corporation and Yunnan Copper Co., Ltd.

| 模型 | 天齐锂业 | | | | 云南铜业 | | | |
|---------------|-------|-------|-------|-----|-------|-------|-------|-----|
| | RMSE | MAE | R^2 | t/s | RMSE | MAE | R^2 | t/s |
| LSTM | 0.051 | 0.017 | 0.953 | 201 | 0.104 | 0.039 | 0.906 | 210 |
| CNN-LSTM | 0.065 | 0.024 | 0.898 | 180 | 0.104 | 0.040 | 0.897 | 191 |
| GRU-Attention | 0.050 | 0.016 | 0.865 | 197 | 0.101 | 0.037 | 0.939 | 199 |
| CNN-BiLSTM-AM | 0.050 | 0.017 | 0.458 | 188 | 0.105 | 0.038 | 0.924 | 204 |
| 本文 | 0.050 | 0.016 | 0.972 | 185 | 0.105 | 0.037 | 0.925 | 205 |

表 9 不同预测模型在中国石油和岳阳兴长的评价指标对比

Table 9 Prediction performance comparison of different models for stock prices of Petro China Company Limited and Yueyang Xingchang Petro-chemical Co., Ltd.

| 模型 | 中国石油 | | | | 岳阳兴长 | | | |
|---------------|-------|-------|-------|-----|-------|-------|-------|-----|
| | RMSE | MAE | R^2 | t/s | RMSE | MAE | R^2 | t/s |
| LSTM | 0.125 | 0.058 | 0.924 | 205 | 0.165 | 0.062 | 0.839 | 210 |
| CNN-LSTM | 0.122 | 0.058 | 0.928 | 189 | 0.165 | 0.064 | 0.835 | 195 |
| GRU-Attention | 0.122 | 0.057 | 0.949 | 206 | 0.164 | 0.061 | 0.901 | 201 |
| CNN-BiLSTM-AM | 0.121 | 0.058 | 0.953 | 197 | 0.164 | 0.062 | 0.896 | 204 |
| 本文 | 0.121 | 0.057 | 0.965 | 199 | 0.079 | 0.028 | 0.959 | 200 |

表 10 不同预测模型在比亚迪和东风汽车的评价指标对比

Table 10 Prediction performance comparison of different models for stock prices of BYD Company Limited and Dongfeng Automobile Co., Ltd.

| 模型 | 比亚迪 | | | | 东风汽车 | | | |
|---------------|-------|-------|-------|-----|-------|-------|-------|-----|
| | RMSE | MAE | R^2 | t/s | RMSE | MAE | R^2 | t/s |
| LSTM | 0.041 | 0.013 | 0.831 | 198 | 0.079 | 0.037 | 0.707 | 187 |
| CNN-LSTM | 0.126 | 0.084 | 0.166 | 181 | 0.081 | 0.039 | 0.537 | 170 |
| GRU-Attention | 0.047 | 0.013 | 0.454 | 186 | 0.079 | 0.036 | 0.704 | 190 |
| CNN-BiLSTM-AM | 0.042 | 0.012 | 0.951 | 194 | 0.079 | 0.037 | 0.791 | 181 |
| 本文 | 0.041 | 0.012 | 0.975 | 187 | 0.079 | 0.036 | 0.841 | 173 |

地提高股票预测模型预测精度,且模型训练效率较高,但在部分数据集出现了预测失准的情况,如比亚迪和东风汽车.如何提高预测模型的精度,这需要神经网络模型重点关注与股票价格相关的特征信息,利用注意力机制可以很好地解决这一问题.

从实验中能够发现 CNN-BiLSTM-AM 和 GRU-Attention 模型的各项评价指标在 64%的数据集上明显优于 LSTM 和 CNN-LSTM 模型,表明结合了注意力机制的神经网络模型在一定程度综合了局部重要信息特征,提高了模型的泛化能力,能够让模型的预测结果更加准确,同时提高模型的训练效率.但股票价格处于骤涨或骤跌的时段,CNN-BiLSTM-AM 和 GRU-Attention 模型拟合度较低,出现较大的预测误差,表明这两个模型无法学习到股票价格骤涨或骤

跌的特征信息.鉴于以上的基准预测模型的实验结果,它们可能不适用于发展中经济体的股票市场.

对于 BiLSTM-SA-TCN 模型,实验结果表明在 78%的股票数据集中 R^2 评价指标最优,在 71%的数据集中 MAE 评价指标最优, RMSE 的平均指标比 LSTM、CNN-LSTM、GRU-Attention 和 CNN-BiLSTM-AM 模型分别提高 14%、25%、14.2% 和 7.8%,在模型训练方面,BiLSTM-SA-TCN 模型展现了卓越的特性,其收敛速度较快,加速了模型训练过程,从而更迅速地解决了问题并提高了预测性能.表明 BiLSTM-SA-TCN 模型比其他基准模型拥有更准确的表达能力,符合股票价格变动的基本趋势.同时,在股票价格骤涨或骤跌的时段下,BiLSTM-SA-TCN 模型的拟合效果更优,表明 BiLSTM-SA-TCN 模型可以重点关

注并学习时序序列的股票重要特征信息,使得神经网络模型能够在股票价格发生剧烈波动的情况下,预测结果更加精准。

3 结论

在 LSTM 神经网络的基础上,提出了融合自注意力机制和时间卷积网络的 BiLSTM 神经网络股票预测模型.在股票预测模型中加入自注意力机制能够有效学习股票数据相关性,同时,加入 TCN 可以更容易处理长时间序列信息,提高模型预测准确率.通过在不同股票数据集上与多个模型实验对比,结果表明本文模型具有更小的误差、更高的准确率、更好的拟合度和较高的训练效率,解决了股票预测模型预测功能单一和存在时效性的问题。

下一步将从两个方面进一步研究:一是增加次新股或上市几年的股票数据,对现有的预测模型进行修正和优化,以提高其对这些股票的预测效果;二是重点考虑政策、突发新闻事件、经济形势和公司财报等因素对于股票价格的影响,建立一个更为全面的预测模型,以更好地应对市场变化。

参考文献

References

- [1] Algeriani N M S, Majstorovic V D, Kirin S, et al. Risk model for integrated management system [J]. Tehni č ki Vjesnik, 2019, 26(6) : 1833-1840
- [2] Liu Y B, Hu Z Y, Li H B, et al. Does preemption lead to more leveled resource usage in projects? A computational study based on mixed-integer linear programming [J]. Economic Computation and Economic Cybernetics Studies and Research, 2019, 53(4) : 243-258
- [3] Zhou F, Zhang Q, Sornette D, et al. Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices [J]. Applied Soft Computing, 2019, 84: 105747
- [4] Kurani A, Doshi P, Vakharia A, et al. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting [J]. Annals of Data Science, 2023, 10(1) : 183-208
- [5] Sadorsky P. A random forests approach to predicting clean energy stock prices [J]. Journal of Risk and Financial Management, 2021, 14(2) : 48
- [6] 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(自然科学版), 2023, 15(6) : 631-642
HAN Jinlei, XIONG Pingping, SUN Jihong. Stock price time series prediction based on LSTM and grey model [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2023, 15(6) : 631-642
- [7] Naik N, Mohan B R. Intraday stock prediction based on deep neural network [J]. National Academy Science Letters, 2020, 43(3) : 241-246
- [8] Selvin S, Vinayakumar R, Gopalakrishnan E A, et al. Stock price prediction using LSTM, RNN and CNN-sliding window model [C] // 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). September 13 - 16, 2017, Udupi, India. IEEE, 2017: 1643-1647
- [9] Chen K, Zhou Y, Dai F Y. A LSTM-based method for stock returns prediction: a case study of China stock market [C] // 2015 IEEE International Conference on Big Data. October 29 - November 1, 2015, Santa Clara, CA, USA. IEEE, 2015: 2823-2824
- [10] Nelson D M Q, Pereira A C M, de Oliveira R A. Stock market's price movement prediction with LSTM neural networks [C] // 2017 International Joint Conference on Neural Networks (IJCNN). May 14 - 19, 2017, Anchorage, AK, USA. IEEE, 2017: 1419-1426
- [11] Elliot A, Hsu C H. Time series prediction: predicting stock price [J]. arXiv e-print, 2017, arXiv: 1710. 05751
- [12] Kim T, Kim H Y. Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data [J]. PLoS One, 2019, 14(2) : e0212320
- [13] Lee M C. Research on the feasibility of applying GRU and attention mechanism combined with technical indicators in stock trading strategies [J]. Applied Sciences, 2022, 12(3) : 1007
- [14] Lu W J, Li J Z, Wang J Y, et al. A CNN-BiLSTM-AM method for stock price prediction [J]. Neural Computing and Applications, 2021, 33(10) : 4741-4753
- [15] Smagulova K, James A P. A survey on LSTM memristive neural network architectures and applications [J]. The European Physical Journal Special Topics, 2019, 228(10) : 2313-2324
- [16] Niu Z Y, Zhong G Q, Yu H. A review on the attention mechanism of deep learning [J]. Neurocomputing, 2021, 452: 48-62
- [17] Lea C S, Vidal R, Reiter A, et al. Temporal convolutional networks: a unified approach to action segmentation [C] // European Conference on Computer Vision. October 10-16, 2016, Amsterdam, Netherlands. 2016: 47-54
- [18] 邓佳丽,赵凤群,王小侠. MTICA-AEO-SVR 股票价格预测模型 [J]. 计算机工程与应用, 2022, 58(8) : 257-263
DENG Jiali, ZHAO Fengqun, WANG Xiaoxia. MTICA-AEO-SVR model for stock price forecasting [J]. Computer Engineering and Applications, 2022, 58(8) : 257-263
- [19] 李晓杰,崔超然,宋广乐,等.基于时序超图卷积神经网络的股票趋势预测方法 [J]. 计算机应用, 2022, 42(3) : 797-803
LI Xiaojie, CUI Chaoran, SONG Guangle, et al. Stock trend prediction method based on temporal hypergraph convolutional neural network [J]. Journal of Computer Applications, 2022, 42(3) : 797-803
- [20] 张婷婷,唐振鹏,吴俊传.基于优化 KELM 模型的股票指数预测方法 [J]. 统计与决策, 2021, 37(13) : 148-150.

- ZHANG Tingting, TANG Zhenpeng, WU Junchuan. Forecasting method of stock index based on optimized KELM model[J]. Statistics & Decision, 2021, 37(13): 148-150
- [21] Gao Y, Wang R, Zhou E M. Stock prediction based on optimized LSTM and GRU models[J]. Scientific Programming, 2021: 1-8
- [22] Yadav A, Jha C K, Sharan A. Optimizing LSTM for time series prediction in Indian stock market [J]. Procedia Computer Science, 2020, 167: 2091-2100

Application of BiLSTM-SA-TCN time series model in stock price prediction

YANG Zhiyong^{1,2} YE Yuxi² ZHOU Yu³

1 Big Data and Internet of Things School, Chongqing Vocational Institute of Engineering, Chongqing 402260, China

2 College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China

3 College of Finance and Tourism, Chongqing Vocational Institute of Engineering, Chongqing 402260, China

Abstract To address the poor timeliness and simple prediction functions of stock forecasting models, we propose a model abbreviated as BiLSTM-SA-TCN, which combines Bi-directional Long Short-Term Memory (Bi-LSTM) neural network, Self-Attention (SA) and Temporal Convolution Network (TCN). The learning unit and prediction unit in the proposed model can effectively learn important stock data, capture long-term dependency information, and output the predicted next day close price. The experimental results indicate that the BiLSTM-SA-TCN model has more stable prediction results on multiple data sets and has higher model generalization ability. Furthermore, in comparative experiment, the BiLSTM-SA-TCN model achieves the lowest root mean square error, the lowest mean absolute error, and the best fitting degree of R^2 on the majority of datasets.

Key words stock price forecast; long short-term memory (LSTM); attention mechanism; temporal convolution network (TCN)