

基于Transformer模型的多尺度股票预测

李文琦, 高乐*

五邑大学智能制造学部, 广东 江门

收稿日期: 2022年5月23日; 录用日期: 2022年6月15日; 发布日期: 2022年6月27日

摘要

随着经济全球化所带来的股票交易市场扩大, 股票交易数据日渐激增, 如何使用更有效的方法在众多股票数据中选取优质股成为股民愈发关注的问题。而采用人工智能技术来分析和处理交易数据可以有效地辅助股民进行合理的股票选择。近些年, 不同学者尝试使用人工智能技术的深度学习方法来解决股票预测问题, 虽然取得了不错的效果, 但是缺少对不同预测周期的股票价格预测能力的稳定性。本文为解决以上存在的问题, 采用深度学习技术中的变换神经网络(transformer)作为网络基本架构对股票价格的开盘价进行预测, 将不同时间尺度的股票价格输入到网络模型中, 能够处理更全局的时间序列特征信息, 然后通过CNN技术将不同尺度的特征进行融合, 使得模型能够满足不同周期数据的预测变化。本研究选取乐普医疗(300003)、亿纬锂能(300014)、浦发银行(600000)和东风汽车(600006)四组股票作为实验数据。实验结果表明, 在5个基本交易指标(开盘价、最高价、最低价、收盘价、成交量)的影响因素下, 本研究的多尺度股票价格预测模型具有更好的性能及泛化能力。

关键词

深度学习, 股票预测, 多尺度, Transformer

Multi-Scale Stock Prediction Based on Transformer Model

Wenqi Li, Le Gao*

Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen Guangdong

Received: May 23rd, 2022; accepted: Jun. 15th, 2022; published: Jun. 27th, 2022

Abstract

With the expansion of the stock trading market brought about by economic globalization, the

*通讯作者。

stock trading data is increasing day by day. How to use a more effective method to select high-quality stocks from a large number of stock data has become a more and more concerning issue for investors. The use of artificial intelligence technology to analyze and process transaction data can effectively assist investors in making reasonable stock selections. In recent years, different scholars have tried to use the deep learning method of artificial intelligence technology to solve the problem of stock prediction. Although they have achieved good results, they lack the stability of stock price forecasting ability in different forecasting periods. In order to solve the above problems, this paper uses the transform neural network (transformer) in the deep learning technology as the basic network structure to predict the opening price of the stock price, and input the stock price of different time scales into the network model, which can deal with more global. Then, the features of different scales are fused by CNN technology, so that the model can meet the prediction changes of different periodic data. This study selects four groups of stocks of Lepu Medical (300003), Yiwei Lithium Energy (300014), Shanghai Pudong Development Bank (600000) and Dongfeng Motor (600006) as experimental data. The experimental results show that under the influence of five basic trading indicators (opening price, high price, low price, closing price, and trading volume), the multi-scale stock price prediction model of this study has better performance and generalization ability.

Keywords

Deep Learning, Stock Prediction, Multi-Scale, Transformer

Copyright © 2022 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

在股票市场中, 尽早地了解股票涨跌模式有助于股民进行明智地风险投资。对于投资者而言, 股票满足了不同的投资意愿和投资需求, 扩大了投资选择的范围, 拓宽了投资渠道, 在一定程度上满足了投资者获得相应收益的可能性, 也在一定程度上增强了资金的灵活性和流动性[1]。股票市场的客观风险在给投资者带来收益的同时, 也可能造成经济损失, 也可能对参股公司的经营状况产生负面影响, 甚至给国民经济建设带来副作用。这些问题在所难免, 因此股票走势预测成为各方高度关注的问题[2]。股票走势预测的研究也成为了金融大数据的一个应用研究方向, 许多学者采用深度神经网络的方法来预测股票走势, 成为当前学术领域的热门研究问题之一。近年来, 随着计算机技术的飞速发展, 神经网络应用在金融信息领域[3] [4] [5] [6]成为当今重点研究对象。股票价格在金融领域是一个高度波动的时间序列数据, 股票价格受利率、汇率、通货膨胀、货币政策、投资者情绪等多种因素的影响。对股票价格与这些因素之间的关系进行建模并预测股票价格趋势对于研究人员来说是一项具有挑战性的任务。

目前, 越来越多的研究人员开始转向使用机器学习技术来分析股票数据并创建性能良好的方法模型来预测未来的股票走势[7] [8] [9]。股市预测模型是通过从历史价格数据中学习来预测未来价格而建立的[10]。但是, 由于从股价中挖掘出的基础市场信息太少, 大部分方法通常假设时间序列是从线性过程中生成的, 无法从固定长度时间序列特征中提取有用信息, 因此在非线性股票价格预测中表现不佳。虽然深度学习具有很好的非线性映射能力, 但是股票交易数据间具有较强的时间关联和数据波动的短期连续性。为了打破时间序列差距, CHEN 等人[11]提出了一种基于单个时间点和多个时间点的双重数据特征提取方法, 将短期市场特征与长期时间特征相结合, 以提高预测的准确性。但由于依赖于时间周期匹配阈值,

导致模型方法在不同阈值下具有较差的预测稳定性。本文结合历史股票交易数据的五个基本交易指标,设计并实现了一个股票走势预测系统,为投资者推荐股票和股票走势预测服务,以降低或规避投资风险,从而为投资者带来相对稳定的经济回报。以 transformer 模型为网络主体架构,同时输入不同周期长度的时间序列特征,经过多个分支的编码器处理,最后通过 CNN 将输出特征进行融合,得到具有多个尺度的特征信息。

2. 相关工作

股票预测通过分析历史数据对未来股市参数进行预测与分析。涵盖了包括工商业、经济学、环境科学和金融等行业。根据数据的时间序列,预测问题可归类为短期预测(预测几秒,几分钟,几天,几周或几个月)、中期预测(预测 1 至 2 年)和长期预测(预测超过 2 年) [12]。即定义所选数据变量的时间顺序及周期做为观察序列,使算法模型具有对未来股票价格的预测能力,不同周期数据的预测问题具有不同序列长度的预测能力。

机器学习算法的主要问题是它们的性能在很大程度上取决于给定数据的表示[13]。股票市场生成的时间序列数据最好被描述为随机游走,这使得股票数据的特征工程变得更加困难。因此,使用机器学习算法预测股票价格要困难得多。由于深度学习模型不需要单独执行特征工程,因此这些模型被广泛用于从大量过去数据中预测股价或股价趋势的方法[14]。Huang 等人[15]提出了一种名为双向门控循环单元(BGRU)的神经网络架构,用于预测标准普尔 500 指数上市的股票,并将 BGRU 与 LSTM 和 GRU 网络的性能进行了比较。为了评估金融新闻随着时间的推移对股票价格的影响,研究人员在许多时间间隔(即 1 天、2 天、5 天、7 天和 10 天)上检查了预测方法。实验表明,在前 24 小时内获得了最高的准确度,使用 BGRU 实现的准确度高于 LSTM 和 GRU 实现的准确度。Dang Lien Minh 等人[16]提出了双流门控循环单元(TGRU)和训练于股票新闻和情绪字典的嵌入模型 Stock2Vec 预测日内交易的短期股票走势,应用三个技术指标作为股票价格趋势预测的附加功能集,分析新闻是否立即影响或是短期(2 天、1 周)影响股价。类似地, Lu 等人[17]选取具有不同特征的神经网络模型,从短期股价波动数据中提取有效的特征组合并且利用注意力机制进一步探索了预测 K 线模式。Chen 等人[11]提出了一种基于编码器-解码器框架的股票价格趋势预测模型(TPM),采用分段线性回归方法和卷积神经网络提取金融时间序列在不同时间跨度的长期时间特征和短期市场特征。在编码器和解码器中都引入了注意力机制,以便自适应地选择和合并所有时间点特征的最相关维度,能够自适应地预测股价走势及其持续时间。Yang 等人[18]提出了一种基于金融时间序列历史信息预测价格走势方向的深度学习框架。结合了用于特征提取的卷积神经网络(CNN)和用于预测的长短期记忆(LSTM)网络,为 CNN 构建了一个改进的三维输入张量(时间序列信息、技术指标信息以及股票指数)。使用多变量时间序列进行预测的目的是在给定先前和当前的几个单变量时间序列数据的情况下预测未来值,由于在快速波动的金融时间序列数据中很难测量噪声与信息信号混合的程度,因此设计一个好的预测模型并不是一项简单的任务。Park 等人[19]提出了一种基于深度神经网络和趋势过滤的新预测框架,将嘈杂的时间序列数据转换为分段线性方式。

3. 多尺度的 Transformer 网络架构

Transformer 是由 Vaswani 等人[20]提出来用于神经网络机器翻译任务的序列到序列的模型,目的是提取感兴趣区域在数据全局的重要程度。由于金融市场的不确定性,在不同时间周期下预测股票等金融证券的价格变化是一个重要且具有挑战性的任务。

3.1. Transformer 模型结构

Transformer 网络结构细节解析(图 1):

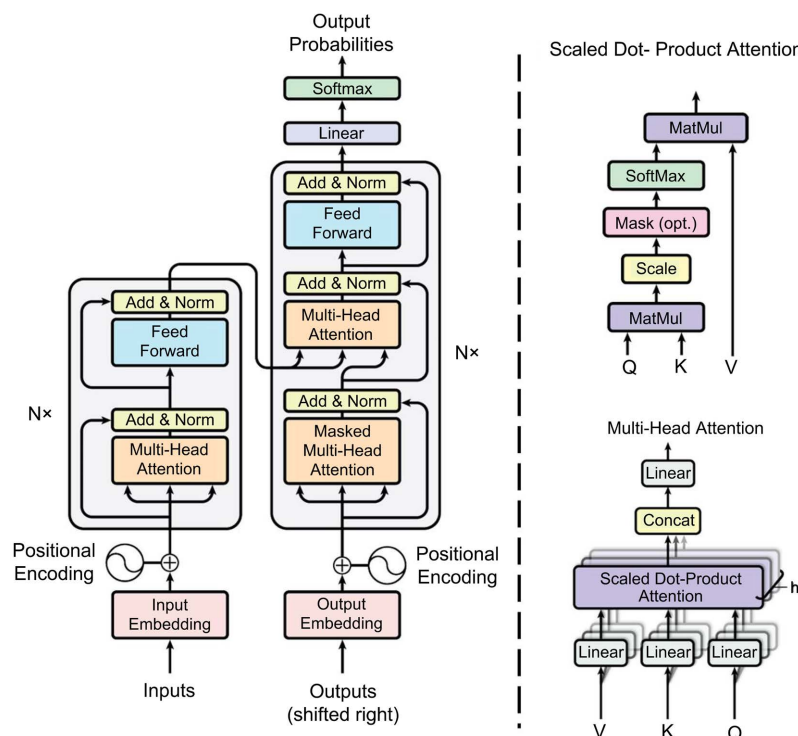


Figure 1. Transformer network standard architecture [20]

图 1. Transformer 网络标准结构[20]

1) 自注意力机制

注意力机制的公式定义如式 1 所示:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

其中 $\mathbf{Q} \in \mathbb{R}^{n \times d_k}$, $\mathbf{K} \in \mathbb{R}^{m \times d_k}$, $\mathbf{V} \in \mathbb{R}^{m \times d_v}$, 这就是传统的缩放点乘注意力(Scaled Dot-Product Attention), 把这个注意力理解为一个神经网络层, 将 $n \times d_k$ 的序列 \mathbf{Q} 编码成了一个新的 $n \times d_v$ 的序列。因为对于较大的 d_k , 内积会数量级地放大, 太大的话 softmax 可能会被推到梯度消失区, softmax 后就非 0 即 1 (那就是 hardmax), 所以 $q \cdot k = \sum_{i=1}^{d_k} q_i k_i$ 按照比例因子 $\sqrt{d_k}$ 缩放。而自注意力机制即 $\text{Attention}(\mathbf{X}, \mathbf{X}, \mathbf{X})$, 通过在序列自身做 Attention, 寻找序列自身内部的联系。

2) 残差和正则化(Add&Norm)

残差(Add)和正则化(Norm)的计算公式如下:

$$\text{Add\&Norm} = \text{LayerNorm}(\mathbf{X} + \text{MultiHeadAttention}(\mathbf{X})) \quad (2)$$

其中 Add 指 $\mathbf{X} + \text{MultiHeadAttention}(\mathbf{X})$, 是一种残差连接, 通常用于解决多层网络训练的问题, 可以让网络只关注当前差异的部分。Norm 指网络层正则化(Layer Normalization), 通常用于 RNN 结构, Layer Normalization 会将每一层神经元的输入都转成均值方差都是一样的, 这样可以加快收敛。

3) 位置编码(Position Embedding)

在解码时序信息时, LSTM 模型通过时间步的概念以输入到输出流一次一个的形式编码的。而 Transformer 选择把时序编码为正弦波。这些信号作为额外的信息加入到输入和输出中以表达时序信息。

这种编码使模型能够感知到当前正在处理的是输入(或输出)序列的哪个部分。位置编码可以学习或者使用固定参数。作者进行了测试(PPL, BLEU), 显示两种方式表现相似。文中作者选择使用固定的位置编码参数:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (3)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (4)$$

其中 pos 是位置, i 是维度。

3.2. 多尺度的 Transformer 模型结构

多尺度的 transformer 由三个 transformer encoder-decoder 分支组成, 如图 2 所示, 将输入的时间序列股票数据通过 Word2Vec 映射为模型输入特征, 并将输入特征分别从 20、40 和 60 时间序列周期长度的三种尺度提取数据特征。然后对特征序列进行卷积运算, 并对横纵坐标的位置编码进行加权处理, 增强了模型保持特征序列顺序的能力; 编码器将输入的时间序列股票数据 (x_1, x_2, \dots, x_n) 映射为模型输入特征 (z_1, z_2, \dots, z_n) , 然后, 将顶部编码器的输出转换为一组注意向量 K (Key 向量)和 V (Value 向量)。每个解码器在编码 - 解码注意力层中使用这些注意向量, 有助于解码器将注意力集中在输入序列中的适当位置。

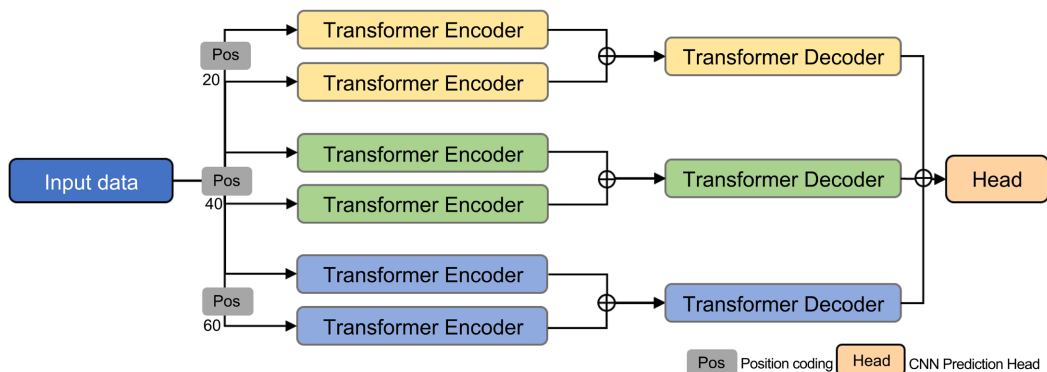


Figure 2. Multi-scale transformer model architecture

图 2. 多尺度的 transformer 网络结构

在编码过程中, 将 $t-1$ 时刻的状态 h_{t-1} 和 t 时刻的数据 x_t 输入到 t 时刻的编码单元中, 从而得到 t 时刻的状态 h_t , 经过 T 个时间片以后, 可以得到长度等于隐节点数量的特征向量 c 。在解码过程中, 将特征向量 c 和上一时间片预测的输出 y_{t-1} 输入到解码单元中, 得到该时刻的输出 y_t , 如此经过 T 个时间段后, 得到最终的输出结果。将每个尺度分支的模型输出进行逐一元素相加, 然后分别通过 1×3 和 1×5 的卷积层对融合数据进行处理, 并通过 RELU 激活函数消除负值的影响。最后通过 MLP 层将特征从嵌入维度转化为输出标签的特征维度, 如公式 5 所示:

$$f^{out} = \text{Proj}(f^{emb} \in \mathbb{R}^d) \quad (5)$$

其中 Proj 表示由 CNN 和 MLP 组成的神经元传播方式, 特征维度 emb 使网络模型定义的超参, 本文使用 512 的特征嵌入维度。

4. 实验分析

4.1. 实验数据

本文获取的原始数据来自 baostock 证券数据平台。选取了四只不同行业的股票数据, 分别是乐普医疗(300003), 亿纬锂能(300014), 浦发银行(600000)和东风汽车(600006)。选取 2016 年 1 月 8 日到 2022 年 5 月 6 日共 1542 个交易日的数据, 前 1342 个数据作为训练值, 后 200 个数据作为测试值来预测开盘价。考虑对股票价格影响比较重要的一些指标, 选取了 5 个基本交易指标, 分别是开盘价、最高价、最低价、收盘价、成交量, 选取的不同行业股票数据开盘价如图 3~图 6 所示。

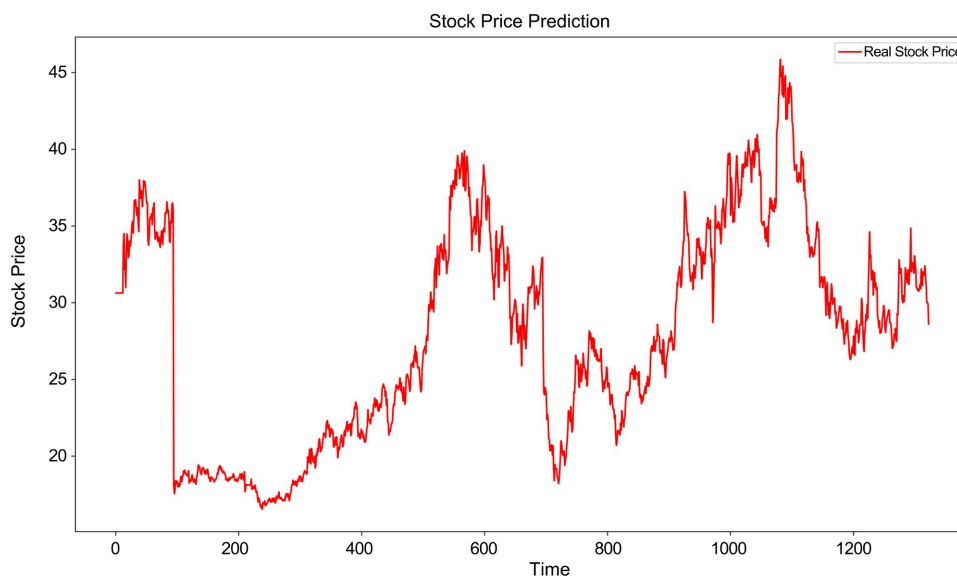


Figure 3. Lepu Medical stock data opening price
图 3. 乐普医疗股票数据开盘价

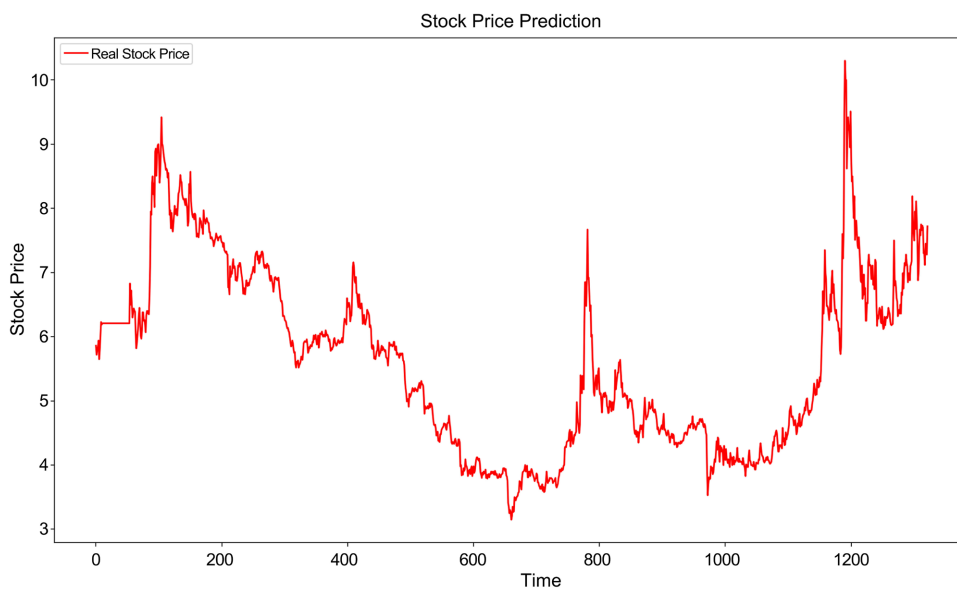


Figure 4. Dongfeng Motor stock data opening price
图 4. 东风汽车股票数据开盘价

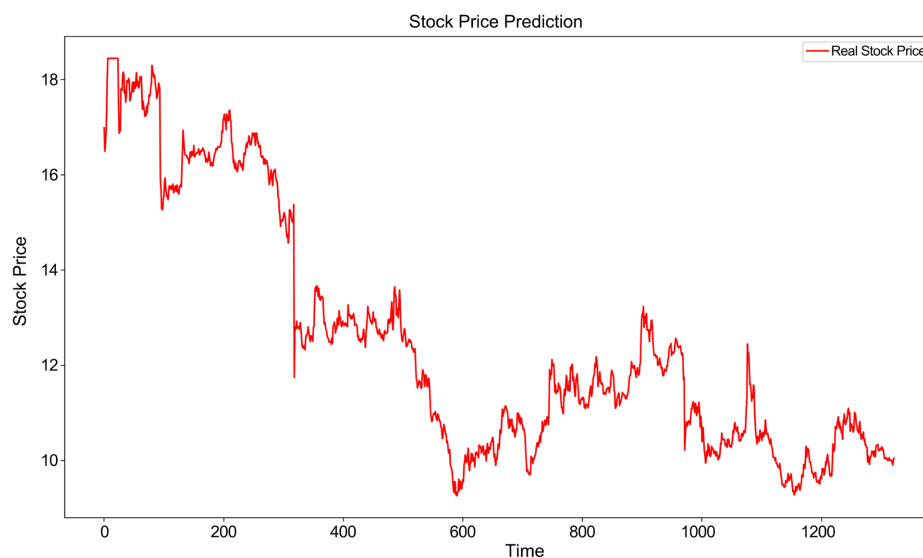


Figure 5. Shanghai Pudong Development Bank stock data opening price

图 5. 浦发银行股票数据开盘价

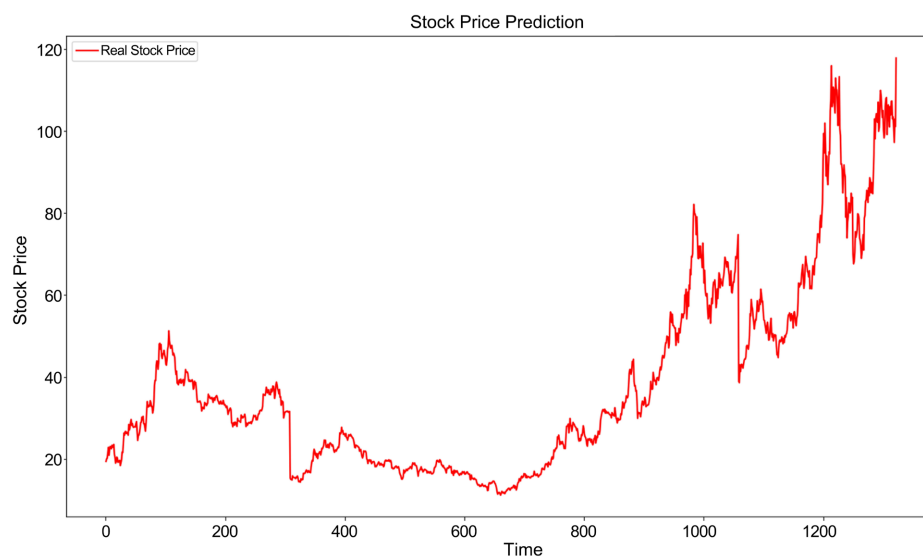


Figure 6. Yiwei Lithium Energy stock data opening price

图 6. 亿纬锂能股票数据开盘价

4.2. 结果分析

在实际问题中,单一数据集的样本特征具有不同的特性,在建立数学模型之前对数据进行规范化处理是十分必要的,使模型训练时避免极值对迭代的性能影响。本文首先将数据进行拟合,对数据进行某种统一处理(将数据映射到某个固定区间),消除极大值和极小值的影响,从而实现数据的标准化和归一化。同时数据的缩放能够避免在模型训练后的过拟合和欠拟合,同时在检验数据建模效果的时候能够更加直观。

在数据处理好之后,输入到 transformer 网络当中,采用前一日的特征指标来预测下一日的股票开盘价格。经设置不同的参数进行实验,发现选用 Adam 作为优化算法,学习率设置为 0.005,迭代次数为 200 次,预测精度最高。同时选用均方根误差(RMSE)、平均绝对误差(MAE)决定系数(R^2 -score)来检验实验精度, R^2 反映因变量的全部变异能通过回归关系被自变量解释的比例。计算公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (6)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

多尺度的 transformer 股票预测网络模型对乐普医疗、东风汽车、浦发银行和亿纬锂能的预测推广 R^2 分数分别是 0.9872、0.9970、0.9897 和 0.9944，与结合 LSTM 和 Attention 模块的 AM-LSTM 网络[14]相比，分别提高了 0.0539、0.0632、0.0607 和 0.0622，同时对不同的时间结点的测试更具有性能稳定性。四个不同行业股票数据开盘价检测效果如图 7~10 所示。

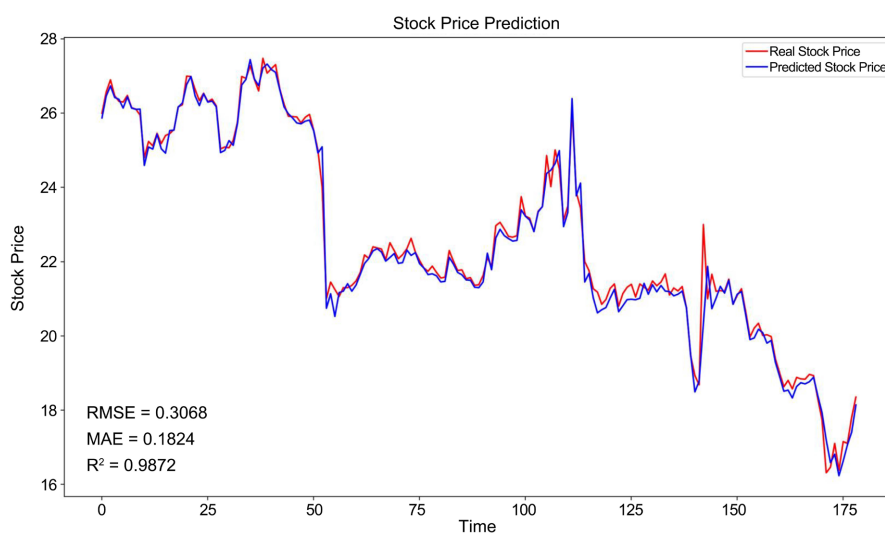


Figure 7. Lepu Medical stock data opening price prediction

图 7. 乐普医疗股票数据开盘价预测

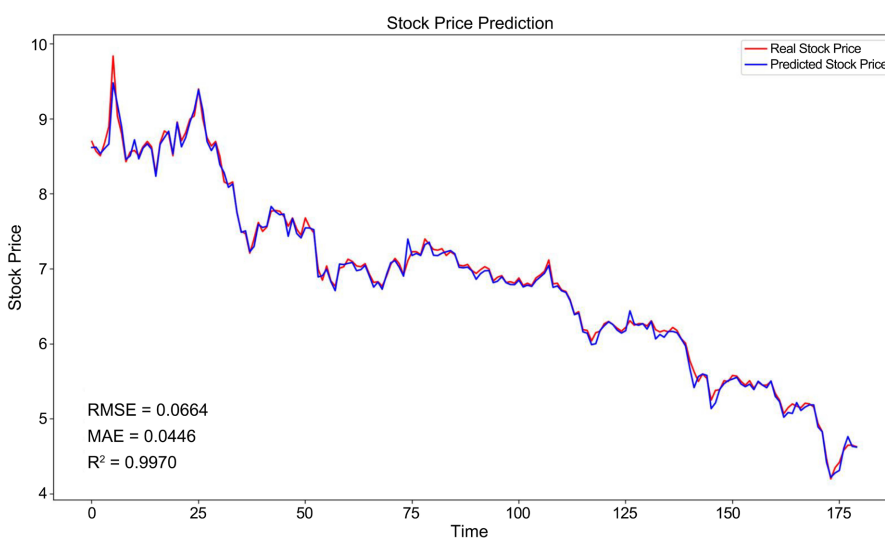


Figure 8. Dongfeng Motor's stock data opening price prediction

图 8. 东风汽车的股票数据开盘价预测

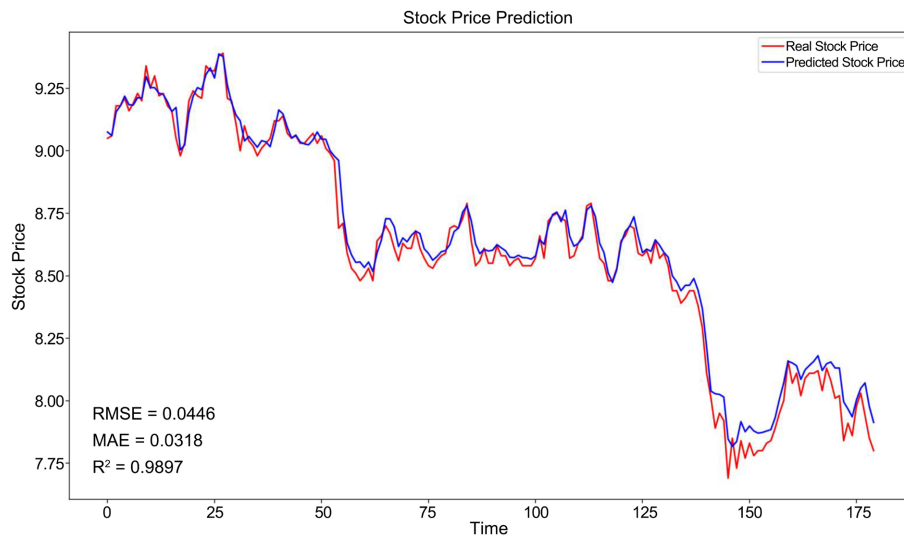


Figure 9. Shanghai Pudong Development Bank's stock data opening price prediction
图 9. 浦发银行的股票数据开盘价预测

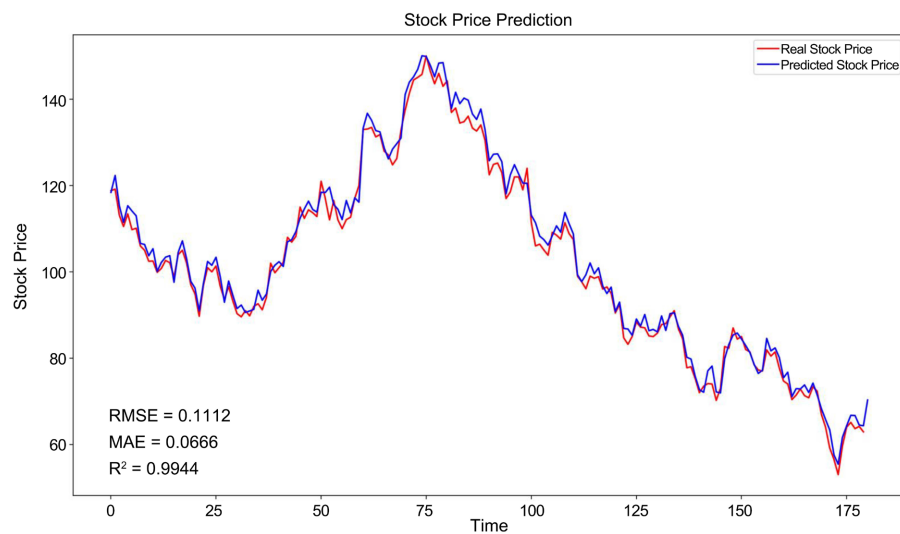


Figure 10. Yiwei Lithium Energy's stock data opening price prediction
图 10. 亿纬锂能的股票数据开盘价预测

5. 结论

传统方法无法提取相关特征来挖掘金融时间序列。为了解决这个问题, 本研究提出了一种基于多尺度 transformer 的金融股票价格预测模型。首先, 在数据预处理阶段, 通过标准化和正则化将时间序列特征和转化为非离散数据, 更有利于提取数据周期的时间序列特征; 其次, 我们从数据中提取不同周期长度的特征信息, 并结合 transformer 模型, 提取出多尺度的时间序列特征信息; 最后将我们的多尺度 transformer 模型训练测试于四支股票中, 实验证明该模型对单支股票具有很好的预测效果和稳定性。

基金项目

项目基金由“广东省教育厅项目, 项目编号: GDJX2020009”和“广东省研究生教育创新项目, 项目编号: 2022SQXX040”资助。

参考文献

- [1] Wen, M., Li, P., Zhang, L., *et al.* (2019) Stock Market Trend Prediction Using High-Order Information of Time Series. *IEEE Access*, **7**, 28299-28308. <https://doi.org/10.1109/ACCESS.2019.2901842>
- [2] Zhou, F., Zhou, H.-M., Yang, Z., *et al.* (2019) EMD2FNN: A Strategy Combining Empirical Mode Decomposition and Factorization Machine Based Neural Network for Stock Market Trend Prediction. *Expert Systems with Applications*, **115**, 136-151. <https://doi.org/10.1016/j.eswa.2018.07.065>
- [3] Moghar, A. and Hamiche, M. (2020) Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Computer Science*, **170**, 1168-1173. <https://doi.org/10.1016/j.procs.2020.03.049>
- [4] Zhang, J., Cui, S., Xu, Y., *et al.* (2018) A Novel Data-Driven Stock Price Trend Prediction System. *Expert Systems with Applications*, **97**, 60-69. <https://doi.org/10.1016/j.eswa.2017.12.026>
- [5] Thakkar, A. and Chaudhari, K. (2020) CREST: Cross-Reference to Exchange-Based Stock Trend Prediction Using Long Short-Term Memory. *Procedia Computer Science*, **167**, 616-625. <https://doi.org/10.1016/j.procs.2020.03.328>
- [6] 刘翀, 杜军平. 一种基于深度 LSTM 和注意力机制的金融数据预测方法[J]. 计算机科学, 2020, 47(12): 125-130.
- [7] Sun, W., Zhang, P., Wang, Z., *et al.* (2021) Prediction of Cardiovascular Diseases Based on Machine Learning. *ASP Transactions on Internet of Things*, **1**, 30-35. <https://doi.org/10.5281/OTIOT.2021.100035>
- [8] Nabipour, M., Nayyeri, P., Jabani, H., *et al.* (2020) Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms via Continuous and Binary Data; A Comparative Analysis. *IEEE Access*, **8**, 150199-150212. <https://doi.org/10.1109/ACCESS.2020.3015966>
- [9] Vijh, M., Chandola, D., Tikkiwal, V.A., *et al.* (2020) Stock Closing Price Prediction Using Machine Learning Techniques. *Procedia Computer Science*, **167**, 599-606. <https://doi.org/10.1016/j.procs.2020.03.326>
- [10] Paul, S. and Vishnoi, S. (2020) Real-Time Stock Trend Prediction via Sentiment Analysis of News Article. *Proceedings of the e-Journal-First Pan IIT International Management Conference—2018*. <https://doi.org/10.2139/ssrn.3753015>
- [11] Chen, Y., Lin, W. and Wang, J.Z. (2019) A Dual-Attention-Based Stock Price Trend Prediction Model with Dual Features. *IEEE Access*, **7**, 148047-148058. <https://doi.org/10.1109/ACCESS.2019.2946223>
- [12] Selvin, S., Vinayakumar, R., Gopalakrishnan, E., *et al.* (2017) Stock Price Prediction Using LSTM, RNN and CNN-Sliding Window Model. *Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, 13-16 September 2017, 1643-1647. <https://doi.org/10.1109/ICACCI.2017.8126078>
- [13] Goodfellow, I., Bengio, Y. and Courville, A. (2017) Deep Learning (Adaptive Computation and Machine Learning Series). The MIT Press, Cambridge, MA, 321-359.
- [14] 刘甲, 孙德山. 基于注意力机制和 LSTM 网络的股价预测[J]. 应用数学进展, 2021, 10(12): 4379-4385. <https://doi.org/10.12677/AAM.2021.1012466>
- [15] Huynh, H.D., Dang, L.M. and Duong, D. (2017) A New Model for Stock Price Movements Prediction Using Deep Neural Network. *Proceedings of the 8th International Symposium on Information and Communication Technology, Nha Trang*, 7-8 December 2017, 57-62. <https://doi.org/10.1145/3155133.3155202>
- [16] Minh, D.L., Sadeghi-Niaraki, A., Huy, H.D., *et al.* (2018) Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network. *IEEE Access*, **6**, 55392-55404. <https://doi.org/10.1109/ACCESS.2018.2868970>
- [17] Lu, R. and Lu, M. (2021) Stock Trend Prediction Algorithm Based on Deep Recurrent Neural Network. *Wireless Communications and Mobile Computing*, **2021**, Article ID: 5694975. <https://doi.org/10.1155/2021/5694975>
- [18] Yang, C., Zhai, J. and Tao, G. (2020) Deep Learning for Price Movement Prediction Using Convolutional Neural Network and Long Short-Term Memory. *Mathematical Problems in Engineering*, **2020**, Article ID: 2746845. <https://doi.org/10.1155/2020/2746845>
- [19] Park, Y., Eom, D., Seo, B., *et al.* (2020) Improved Predictive Deep Temporal Neural Networks with Trend Filtering. *Proceedings of the First ACM International Conference on AI in Finance*, New York, 15-16 October 2020, Article No. 8. <https://doi.org/10.1145/3383455.3422565>
- [20] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **30**, 5998-6008.