Check for updates

Stock price forecast based on improved Transformer

Hongwei Zhang

Shanghai Maritime University, Shanghai China

Email: hwzhang_smu@163.com

Abstract. Nowadays, the stock trading market is expanding day by day due to economic globalization, and the stock trading data is increasing day by day. How to use more effective methods to select high quality stocks from numerous stock data has become an increasingly concerned issue for shareholders. Time series prediction has broad application prospects, which attracts more and more researchers to study it deeply. Moving average is a kind of technical index used to observe the trend of stock price change. It is one of the most widely used technical indexes and is often used in the task of predicting the trend of time series. But the original moving average indicator is obtained by calculating equal or preset weights assigned to the data of time series. Ignoring nuances of importance at different points in time; In addition, the same weight is used for various data of time series, ignoring the discrepancies in the intrinsic properties of different time series. The use of deep learning and machine learning methods to analyze and deal with stock trading data can effectively help shareholders to choose reasonable stocks for trading. Nowadays, many researchers have attempted to use deep learning methods of artificial-intelligence technology to deal with the question of stock price and trend prediction, though they have achieved good results, but they lack the stability of stock price prediction ability for different forecasting cycles. This paper adopts Transformer in deep learning technology as the basic network architecture to solve the above problems, to predict the closing price of stock prices, input stock prices at different time scales into the model, which can process more global time series feature information, and then use GRU technology to advance the features at different scales Row fusion enables the model to meet the predicted changes of data in different periods, this model is named MS-Transformer. In this study, the data of China Securities 50 stocks were selected as the experimental data. The results prove that under the effect factors of 5 basic trading indicators (opening price, trading volume, highest price, lowest price and closing price), MS-Transformer model can be proved to have good performance in stock price prediction.

Keywords: stock, transformer, predicted, deep learning methods

A. Haldorai et al. (eds.), *Proceedings of the 2024 3rd International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID 2024)*, Atlantis Highlights in Intelligent Systems 11,
https://doi.org/10.2991/978-94-6463-490-7_20

1 INTRODUCTION

In the stock exchange market, early understanding of the patterns of stock ups and downs helps investors make wise risk investments. Stock trading satisfies the investment needs and investment intentions of different investors, increases the scope of investment choices, expands investment channels, to a certain extent meets the possibility of investors to obtain corresponding returns, and also enhances the liquidity and flexibility of funds to a certain extent[1]. While the stock market brings benefits to investors, due to the existence of objective risks, it may cause economic losses to investors and have a negative influence on the operating conditions of the participating companies. These problems are difficult to avoid, so stock trend prediction becomes a highly concerned task for operators [2]. The method of deep neural network is used to predict the stock trend, and many scholars participate in the stock trend prediction, which has become one of the popular research tasks in the current academic field. In recent years, a variety of computer technologies have developed rapidly, the application of deep neural networks in the field of financial information [3][4] has become a key research object.

Based on the traditional time series modeling methods, which mainly rely on statistical theory, the linear model between stock characteristics and stock prices is established, including the early moving average method, exponential average method and the Auto-regressive model (AR), moving average model (MA) and differential Auto-regressive Integrated Moving Average model (ARIMA). This method is very strict to the rules of data distribution and integrity, and believes that there is only a linear relationship between the factors that cause stock price changes. However, it has been proved that there is a complex nonlinear relationship between stock price changes and its influencing factors, which also leads to the unsatisfactory accuracy of the final prediction results of the traditional time series modeling method. Based on machine learning methods, researchers have used random forest algorithm, XGBoost algorithm and support vector machine algorithm to predict stocks and achieved better results than ARIMA model, because these classical machine learning algorithms can simulate the complex nonlinear relationship between stock characteristics and prices.

As we all know, the traditional neural network is affected by its topology, the generalization ability is weak, and it is toilless to fall into local optimal in the iterative process. With the rapid growth of the era of large data, deep learning technology, with its more powerful learning ability and the ability of feature extraction has made breakthrough progress, mainly includes the convolutional neural network (CNN), Cyclic Neural Network (RNN), and both Short-term and Long-Term Memory Networks(LSTM), etc. Deep learning has the feature of extracting features from a large amount of original data without relying on prior probability, so it has great potential for the research of stock data. RNN [5] is often used as the most effective method for price time series. However, with the increase of sequence length, RNN model is prone to the problem of gradient disappearance. LSTM[6] is a deep learning network designed based on RNN model, and this model solves the problem of memory storage and forgetting of RNN model by adding gate mechanism, and has the forecasting ability of long price series. With the continuous improvement of the prediction result of the social requirement model, various of deep learning methods have also been constructed and added to the network model. The function of the convolutional layer in the CNN model [7] is a method to capture the internal relationship of the input data. However, according to the characteristics of the convolutional kernel, the convolutional layer performs the same convolution for all the input data, and does not identify the correlation difference between different input data. Attention mechanisms[8] are constructed to identify differences between input data, selectively connect relevant information, and better extract characteristic information between data. Probabilistic prediction model [9] is a prediction model built on the basis of probabilistic knowledge. This model has the advantages of modeling uncertainty, analyzing the relationship between variables, realizing causal reasoning and randomly generating sample data. DeepAR[10] model is a method of adding probabilistic prediction based on LSTM model. Achieve higher prediction accuracy and smaller prediction error.

Due to the lack of basic marketing information explored from the stock prices, most methods usually assume that the time series is produced from a linear process, and cannot extract useful information from the fixed length time series features, so they do not perform well in nonlinear stock price prediction. Although deep learning has good nonlinear mapping ability, stock trading data has strong temporal correlation and short-term continuity of data fluctuations. In order to break the time series gap, Chen et [11] came up with an extraction method based on two data features of multiple time points and single time points, which combined short-term time features with long-term market features to enhance the accuracy of stock prediction. However, due to the dependence on time period matching threshold, the model method has poor predictive stability under different thresholds. This paper designs a stock prediction system based on five basic trading indicators of historical stock trading data, which provides investors with stock prediction services to avoid or reduce investment risks, so as to bring more stable returns to investors. This paper use transformer model as the main network architecture, time series features of different cycle lengths are input at the same time, which are processed by encoders of multiple branches, and finally the output features are fused through GRU to obtain feature information with multiple scales.

2 RELATED WORK

Early stock prediction methods mainly analyzed stock prices based on statistics and economic theories, and selected relatively single stock characteristic data for stock prediction. Auto-regressive model (AR), moving average auto-regressive model (ARMA) and differential integrated moving average auto-regressive model (ARIMA) were commonly used. The auto-regressive model based on statistics has certain advantages in computational efficiency due to its single feature, but the model is too simple and the input dimension is low, which limits its modeling ability for nonlinear non-stationary financial time series and anti-interference ability to abnormal data. With the development of computer technology, researchers have gradually begun to apply machine learning technology to the problem of stock prediction. The massive stock data accumulated in the financial market over the years also provides sufficient data basis for machine learning. Stock forecasting forecasts and analyzes future stock market parameters by analyzing historical data. It covers industries such as business, economics, environmental science and finance. Based on the time series of the data, stock forecasting tasks can be divided into short-term forecasting (forecasting seconds, minutes, days, etc.), medium-term prediction(forecasting 1 to 2 years), and long-term prediction(forecasting more than 2 years)[12]. That is, the time order and period of the selected data variables are defined as the observation series, so that the algorithm model has the ability to forecast the future stock price, and the prediction problem of the data of different periods has the ability to predict the sequence length.

The machine learning algorithm of the main problem is that their prediction results heavily depend on the characteristics of the given data[13]. Time series data generated by stock markets are characterized by volatility, which leads feature engineering of stock data more difficult. As a result, it is much more difficult to predict stock prices employing machine learning algorithms. Because deep learning algorithms do not perform separate feature engineering, deep learning models are widely used as a method to predict stock trends and stock prices from large amounts of historical data [14]. Huang et al[15] proposed a deep neural network model called Bidirectional Gated Recurrent Unit (BGRU) for forecasting stocks of the S&P 500 index, and contrasted the performance of BGRU with LSTM and GRU networks. In order to evaluate the effect of the passage of time on stock prices, forecasting methods were examined over many time intervals (1, 5, 7, and 10 days). Experiments have shown that the topmost accuracy is achieved in the previous 24 hours, and The accuracy using BGRU is better than the accuracy of GRU and LSTM. Dang et al.[16] put forward a two-current gated cycle unit (TGRU) and an embedded model, Stock2Vec, trained on stock-related news and sentiment dictionary to forecast short-term stock trends in Intraday trading, and applied 3 technical indicators as an additional function set for stock price trend forecasting. Analyze whether this news affects the stock price at once or in the short term (2 days,7 days). Similarly, Lu et al. [17] selected neural network models with different features to acquire effective feature group from short term stock price data and further explored the predictive K-line model by using attention mechanism. Wang et al[10] put forward a stock price trend prediction model (TPM) based on the encoder-decoder framework. The convolutional neural network and linear regression method were used to acquire the long-term time characteristics and short-term market characteristics of stock time series in different time spans. An attention mechanism has been introduced in both the decoder and the encoder to adaptive choose and merge the optimal-related dimensions of features at all points in time, enabling adaptive prediction of stock price movements and their duration. Yang et al[18] proposed a deep learning model for predicting price trend direction based on financial time series historical stock information. By combining the convolutional neural network for feature acquisition and the Long Short Term memory network for forecast, an improved three-dimensional input tensor (time series information, technical indicator information, and stock index) is constructed for CNN. The purpose of using multivariate time series for forecasting is to predict future values given several the current and the previous univariate time-series data. Since it is hard to calculate the degree of noise mixing with information signals in rapidly fluctuating stock time-series data, designing a better forecasting model is not a simple task. Park et al[19] proposed a fresh prediction model based on trend filtering and deep neural networks to convert noisy time-series data into a piecewise linear approach.

Recently, a complete Encoder-Decoder transformer architecture used univariate time-series prediction: Li et al[20] show superior performance compared with the classical statistical method ARIMA. The recent matrix decomposition method TRMF, RNN based auto-regression model (AR) and RNN based state space model (Deep state) 4 common prediction data sets, Wu et al.[21]used transformers to predict influenza epidemics compared with ARIMA, which also showed performance benefits. Bryan et al[22] used a transformer model for multi-level univariate prediction to support the interpretation of time dynamics.

3 TIME SERIES ANALYSIS AND MODEL DESCRIPTION

Time series problems are distinguished into regression and classification by using historical series values as input data. Given the sliding window features of the training sequence $X=(X_1,X_2,...,X_T)$hand $X_t=(X_{1t},X_{2t},...,X_{Lt}),X_t \in X$ sequence, define the interval length of the time step length L and give the historical value $y=(y1,y2,...,y_{T-1})$. To forecast the future trend and features of the sequence, the nonlinear mapping function is usually used to learn the historical sequence feature X and the corresponding target value y to predict the future value y_T, which corresponds to the model formula: $y_T=f(X,y)$.

4 TRANSFORMER NETWORK ARCHITECTURE

Transformer is a sequence-to-sequence model put forward by Vaswani et al. [19] for neural network machine translation tasks, with the purpose of extracting the importance of the region of interest in the global data. The Transformer model is a sequence modeling model based on a self-attention mechanism, consisting of coding blocks and decoding blocks, each of which is stacked with multiple independent coding layers (Encoder or Decoder). Each encoding layer contains a Multi-Head Attention layer, a fully connected layer (FFN), and a regularization layer (Add&Norm), while each decoding layer contains two multi-head attention layers. The Transformer model uses Positional Encoding strategy to obtain relative position information in the input sequence, and uses multi-head attention mechanism to focus on different details in different subspaces.

4.1 Transformer Model structure

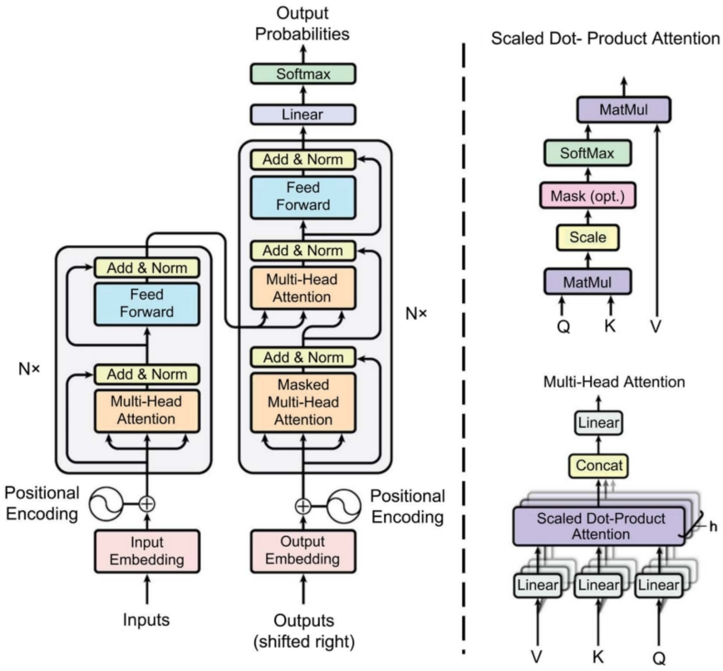

Fig. 1[19]. Detail analysis of Transformer network structure

(1) Self-attention mechanism

The formula definition of attention mechanism is shown in equation (1):

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

Where Q, K, V respectively represent the query, key, value, this is the traditional Scaled Dot Product Attention (Scaled dot Product attention), this attention is understood as a neural network layer, the $n*d_k$ sequence Q encoded into a new $n*d_v$ sequence. Because for larger DKS, the inner product will be amplified by orders of magnitude, too large Softmax may be pushed to the gradient disappearance zone, Softmax is either 0 or 1(that is hardmax), so qk scales according to the scale factor. The self-attention mechanism, Attention(X,X,X), looks for connections within the sequence itself by paying Attention to it **in Fig1**.

(2) Residuals and regularization (Add&Norm)

The formula for calculating residual (Add) and regularization (Norm) is as follows:

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

Where add refers to $\mathbf{X} + \mathrm{MultiHeadAttention}(\mathbf{X})$, which is a residual connection usually used to solve the problem of multi-layer network training, so that the network can only focus on the part that is currently different. Norm network refers to layer regularization, which is commonly used in RNN structures. Layer Normalization converts the input of each layer of neurons to mean and variance, which speeds up convergence.

(3) Position Embedding

When decoding timing information, the LSTM model is encoded in the form of input to output streams one at a time through the concept of time steps. Transformer chooses to encode the timing as sine waves. These signals are added to the input and output as additional information to represent timing information. This encoding enables the model to sense which part of the input (or output) sequence is currently being processed. Location coding can be learned or used with fixed parameters. The authors conducted tests (PPL,BLEU) that showed similar performance in both ways. This article chooses to use a fixed location encoding parameter:

$$\mathrm{PE}_{(\mathrm{pos},2\mathrm{i})} = \sin\left[\frac{\mathrm{pos}}{10000^{\frac{2\mathrm{i}}{\mathrm{d}_{\mathrm{model}}}}}\right] \tag{3}$$

$$\mathrm{PE}_{(\mathrm{pos},2\mathrm{i}+1)} = \cos\left[\frac{\mathrm{pos}}{10000^{\frac{2\mathrm{i}}{\mathrm{d}_{\mathrm{model}}}}}\right] \tag{4}$$

4.2 Multi-scale Transformer model structure

Multi-scale transformer is composed of multiple encoder-decoder branches of transformer. As shown in the **Fig. 2** below, the input time series stock data is mapped into model input features through Word2Vec. The input features are extracted from three scales of 20, 40 and 60 time series cycle length. Then the feature sequence is convolved and the horizontal and vertical coordinates are weighted to enhance the ability of the model to maintain the sequence of feature sequences. The encoder maps the input time series stock count ($x1,x2,x3...x_n$) to the model input feature ($z1,z2,z3...z_n$), and then converts the output of the top encoder into a set of attention vectors K(key vector) and V(value vector). Each decoder uses these attention vectors in the coding-decoding attention layer, helping the decoder focus its attention at the appropriate place in the input sequence.

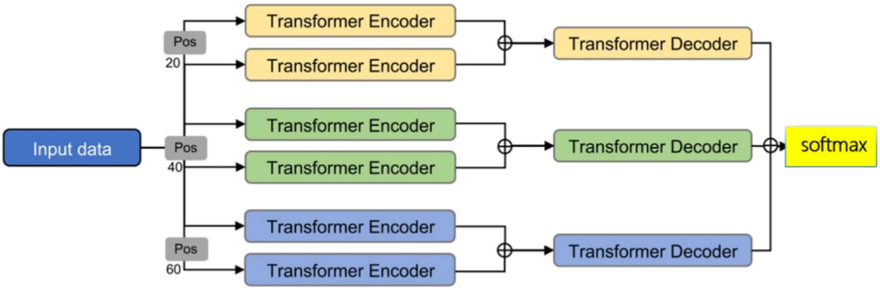

Fig. 2. Multi-scale Transformer model structure(MS-Transformer)

In the coding process, the state h_{t-1} at time t−1 and the data x_t at time t are input into the coding unit at time t, thus obtaining the state ht at time t. After T time slices, the feature vector c whose length is equal to the number of hidden nodes can be obtained. In the decoding process, the characteristic vector c and the output yt−1 predicted by the previous time slice are input into the decoding unit to obtain the output yt of the moment, so that the final output result is obtained after T time periods. The model output of each scale branch is added element by element, and then the fusion data is processed through the convolution layer of 1×3 and 1×5 respectively, and the negative effect is eliminated by the RELU activation function. Finally, the feature is transformed from the embedding dimension to the feature dimension of the output label through the MLP layer.

5 EXPERIMENTAL ANALYSIS

5.1 Experimental data

The original data obtained in this paper comes from the flush securities data platform. Select the stock data of China Securities 50. The data includes minute-level transaction data and daily transaction data. The data of 1542 trading days from January 2016 to December 2022 were selected, the trading data of the first 1342 days were used as the training value, and the data of the last 200 trading days were used as the test value to predict the closing price. Considering some important indicators that affect the stock price, five basic trading indicators are selected, which are the opening price, the highest price, the lowest price, the closing price and the trading volume.

5.2 Data normalization

On this basis, considering that stock data is multi-variable data, there are differences in different dimensions and values, direct application will affect the training and testing of the model. Therefore, the data normalization method is used to map the stock sample data to the interval [0,1] to ensure the same value range between different variables, reduce the preference for data and weight allocation, and ensure the accuracy and stability of the model prediction. The formula is as follows:

$$X' = (X - X\_min) / (X\_max - X\_min) \tag{5}$$

5.3 Hyper-Parameters

The hyper-parameter settings of this experiment are shown in Table 1.

Table 1. The hyper-parameters set for this article are as follows:

Hyper-Parameters	set
batch size	12
learning rating	0.001
loss	MSE
epoch	200
multi-head	6
GPU	GeForce GTX 1060

5.4 Evaluation indicators

To evaluate the prediction effect of different models, this paper uses four different evaluation indexes: mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE). These four evaluation indicators are usually used as criteria for evaluating regression problems. Given the predicted value and true value of the model, the evaluation value is obtained through the following calculation formulas:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^{-})^2} \tag{6}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i^{-}| \tag{7}$$

5.5 Result Analysis

In order to verify the prediction effect of MS-Transformer, five baseline models, namely MS-Transformer and SVR, CNN, LSTM, Transformer and DST2V-Transformer, are used to compare and forecast the China Securities 50 stocks. At the same time, in order to eliminate the contingency of a single test, the test results are obtained by averaging multiple tests. The prediction results of different models are shown in Table 2.

Table 2. Comparison results of different models under MAE and RMSE indexes

Model	MAE	RMSE
SVR	3.826	4.855
CNN	2.829	3.233

LSTM	1.903	1.745
Transformer	1.227	1.634
DST2V-Transformer	1.137	1.601
MS-Transformer	**0.843**	**1.184**

As can be seen from Table 2, SVR, as a classical machine learning model, has the largest MAE and RMSE, which is due to SVR's low fitting degree for time series prediction problems with strong volatility, such as stock prices, and its prediction performance is the worst. Similarly, all evaluation indicators show that the prediction results of CNN are not ideal. This is because CNN does not have the modeling ability of time series and cannot make full use of historical stock price and technical factor data, which inevitably leads to the problem of historical information leakage. In addition, LSTM, as an improved neural network model of CNN, has better predictive performance than CNN. Although the problems of gradient disappearance and explosion during model training can be avoided, the predictive performance of Transformer based on self-attention mechanism is more prominent. Transformer can quickly extract timing information and implement parallel calculation, effectively reducing prediction errors and improving calculation speed. DST2V-Transforme is an improved model based on Transformer. It smooth data by introducing moving average method to identify trend components of time series and introduce time information into data to capture periodic and aperiodic components of time series, thus improving Transformer's long-term forecasting performance. Experimental results show that MAE and RMSE indexes of DST2V-Transformer are improved compared with classic Transformer. In summary, all evaluation indicators of MS-Transformer have the best performance. This is because MS-Transformer has stronger feature extraction and network learning capabilities, so its MAE and RMSE are the smallest, and the MAE and RMSE are 23.2% and 25.7% lower than Transformer respectively. The prediction of stock price is more accurate and the prediction error is smaller, and the prediction effect is significantly better than the other five baseline models.

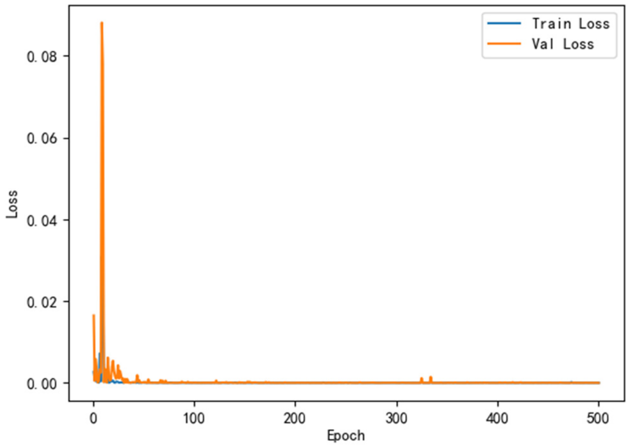

Fig.3. MS-Transformer Loss

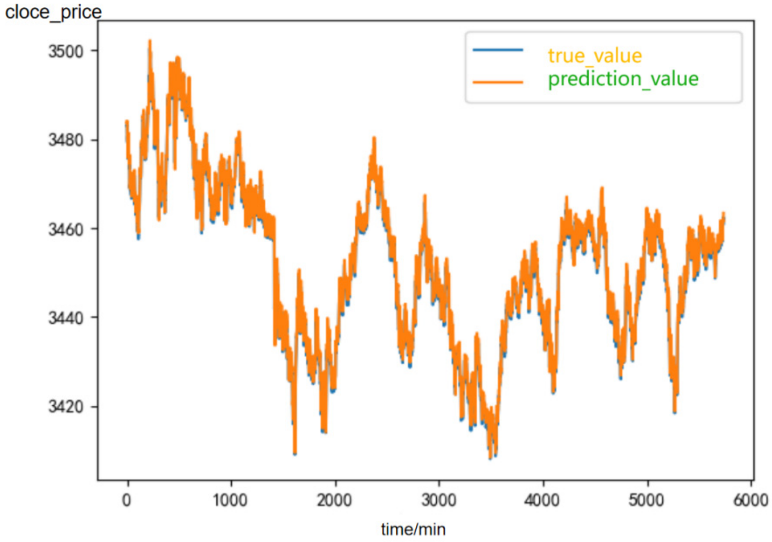

Fig.4. MS-Transformer closing price forecast

As can be seen from **Figure 3**, the training loss tends to 0 after epoch=50, and epoch=100 is set for training and testing. **Figure 4** shows the comparison between the predicted value of MS-Transformer and the real value. And see in figure 5. This is because the predicted value of MS-Transformer based on the attention mechanism is close to the real value, and the forecast of future price is more accurate. In summary, MS-Transformer can accurately capture the inflection point with large trend change amplitude, and the prediction performance is further improved compared with Transformer. The two curves are nearly identical, the error between the predicted value and the true value is minimal, and the regression fitting effect is optimal, which intuitively validates the fitting ability and superiority of MS-Transformer.

6 CONCLUSION AND PROSPECT

Traditional methods cannot extract relevant features to mine financial time series. To solve this problem, this paper proposes a financial stock price prediction model based on multi-scale transformer. First, in the data preprocessing stage, the time series features are converted into non-discrete data by normalization and regularization, which is more conducive to extracting the time series features of the data cycle. Secondly, feature information of different cycle lengths is extracted from the data and combined with transformer model, multi-scale time series feature information is extracted. Finally, our multi-scale transformer model is trained and tested in four stocks, and the experiment proves that the model has good prediction effect and stability for a single stock. Since stock prices are difficult to predict due to political, economic, social, psychological and other factors, the stock price prediction model combining financial news, stock opinion, investor sentiment and other sentiment analysis theories will be further studied in the future, and the prediction performance of Transformer will be optimized to provide more accurate and practical reference for investors.

REFERENCE

1. Wen, Li, Zhang, et al 2019 Stock Market Trend Prediction Using High-Order Information of Time Series.IEEAccess,7,28299-28308.https://doi.org/10.1109/ACCESS.2019.2901842
2. Zhou, Zhou, H.-M., Yang, 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. Zhang, Cui, S., Xu, 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
4. Thakkar, 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
5. Liu Chong, Du Junping 2020 A financial data forecasting method based on deep LSTM and Attention mechanism [J]. Computer science, 2020, 47(12): 125-130.
6. Sun, W., Zhang, P., Wang, et al 2021 Prediction of Cardiovascular Diseases Based on Machine Learning. ASP Transactions on Internet of Things, 1, 30-35. https://doi.org/10.52810/TIOT.2021.100035
7. Nabipour, Nayyeri, Jabani, 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
8. Vijh, M., Chandola, D., Tikkiwal, et al 2020 Stock Closing Price Prediction Using Machine Learning Tech niques. Procedia ComputerScience, 167, 599-606. https://doi.org/10.1016/j.procs.2020.03.326
9. Paul, S. and Vishnoi 2020 Real-Time Stock Trend Prediction via Sentiment Analysis of News Article. Proceed ings of the e-Journal-First Pan IIT International Management Conference—2018. https://doi.org/10.2139/ssrn.3753015
10. Chen, Y., Lin, W. and Wang, J.Z 2019 A Dual-Attention-Based Stock Price Trend Prediction Model with Dual Fea tures. IEEE Access, 7, 148047-148058. https://doi.org/10.1109/ACCESS.2019.2946223
11. 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,Commu nications and Informatics (ICACCI), Udupi, 13-16 September 2017,1643-1647. https://doi.org/10.1109/ICACCI.2017.8126078
12. Goodfellow, I., Bengio, Y. and Courville 2017 Deep Learning (Adaptive Computation and Machine Learning Se ries). The MIT Press, Cambridge, MA, 321-359.
13. Liu Jia, Sun Deshan. Stock Price prediction based on Attention Mechanism and LSTM Network[J]. Advances in Applied Mathematics, 2021, 10(12): 4379-4385. https://doi.org/10.12677/AAM.2021.1012466

14. Huynh, H.D., Dang, L.M. and Duong 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
15. Minh, D.L., Sadeghi-Niaraki, A., Huy, et al 2018 Deep Learning Approach for Short-Term Stock Trends Pre diction Based on Two-Stream Gated Recurrent Unit Network. IEEE Access, 6,55392-55404. https://doi.org/10.1109/ACCESS.2018.2868970
16. Lu, R. and Lu 2021 Stock Trend Prediction Algorithm Based onDeep Recurrent Neural Network. Wireless Communications and Mobile Computing, 2021, Article ID: 5694975.https://doi.org/10.1155/2021/5694975
17. Yang, C., Zhai, J. and Tao 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
18. 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
19. Vaswani, A., Shazeer, N., Parmar, N., et al 2017 Attention Is All You Need. Advances in Neural Information Processing Systems, 30, 5998-6008
20. Li. X, Zhou. Z, and Chen. L, "Residual attention-based LSTM for video captioning," World Wide Web. vol. 22, pp. 621-636, January 2018.
21. Neo Wu, Bradley Green, Ben Xue, Shawn O'Banion,Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case.Submitted on 23 Jan 2020.https://arxiv.org/pdf/2001.08317v1.pdf
22. Bryan Lim, Sercan,O Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4):1748–1764, 2021.