

Stock Trend Prediction based on Wide & Deep Asymmetrical Bidirectional Legendre Memory Units

Yong Wang, Yisheng Li^{*}, Zhiyu Xu

School of Artificial Intelligence, Chongqing University of Technology, Chongqing 401135, China

Keywords: Stock trend prediction; Legendre memory unit; Asymmetrical bidirectional recurrent neural network; Wide & Deep model

Citation: Wang Y., Li Y.S., Xu Z.Y.: Stock Trend Prediction based on Wide & Deep Asymmetrical Bidirectional Legendre Memory Units. *Data Intelligence* 6(4), 1014-1031 (2024). doi: <https://doi.org/10.3724/2096-7004.di.2024.0017>

Submitted: June 06, 2023; Revised: August 24, 2023; Accepted: January 09, 2024

ABSTRACT

Deep learning technology has been widely applied in the finance industry, particularly in the study of stock price prediction. This paper focuses on the prediction accuracy and performance of long-term features and proposes a Wide & Deep Asymmetrical Bidirectional Legendre Memory Units that captures long-term dependencies in time series through the immediate backpropagation of bidirectional recurrent modules and Legendre polynomial memory units. The proposed model achieves superior stock trend prediction capabilities by combining the memory and generalization capabilities of the Wide & Deep model. Experimental results on the daily trading data set of the constituents of the CSI 300 index demonstrate that the proposed model outperforms several baseline models in medium and long-term trend prediction.

1. INTRODUCTION

The stock market plays a crucial role in the economic development of a country. The fluctuation of stock prices is influenced by various factors, including the development of the real economy, exchange rates, market trends, technological advancements, and investor sentiment. Predicting the price trends of financial time series data in the stock market and exploring their patterns can help investors formulate more rational investment strategies, promote the stable operation of the capital market, and drive the development and theoretical innovation of time series prediction techniques in finance and other fields. Due to the nonlinearity and instability of stock price trends, predicting them has always been a challenge

^{*} Corresponding author: Yisheng Li (E-mail: yamisora@2020.cqut.edu.cn; ORCID: 0000-0002-3585-7667).

for many researchers. However, a considerable amount of research has shown that stock prices do not follow the theory of random walk and have a certain degree of predictability [1, 2, 3].

With the continuous development and widespread use of deep learning technology, its applications in various fields are becoming increasingly widespread. As a typical data-driven field, the financial sector has gradually started using deep learning technology for modeling and prediction, with particular attention being paid to stock market prediction. The stock market is highly volatile, involves complex and diverse factors, and has a considerable amount of randomness and uncertainty. Therefore, accurately predicting changes in the stock market has always been an important topic of discussion in both the financial industry and academia.

Traditional stock prediction methods are primarily based on statistical models and machine learning methods. While these methods provide some predictive ability, they often fall short when dealing with nonlinear and complex market conditions. However, the emergence of deep learning techniques has provided new ideas and tools for stock prediction. Deep learning can automatically learn features from data and establish models with powerful expressive ability, which can better address the nonlinearity and complexity of the stock market.

In recent years, many scholars and financial companies have begun to study the application of deep learning in the stock market, and have achieved certain results. As stock data is a typical time-series data, extracting key temporal features from it is the focus of research. The commonly used method for time-series data is mainly the Recurrent Neural Network (RNN) [4], which is naturally applied to stock price prediction [5]. RNN inputs the input features of each time step and the hidden state features of the previous time step into a single network module to capture the long-term dependency relationship of the input features. However, the standard RNN has problems such as gradient vanishing or gradient explosion, which limits its performance on long sequence data. In order to overcome this problem, Long Short-Term Memory (LSTM) [6] and Gated Recurrent Unit (GRU) [7] have been proposed. LSTM and GRU both adopt gate mechanisms, which allow the network to selectively forget or retain historical information, thus better handling long sequence data. The gate mechanism in LSTM includes input gate, output gate, and forget gate, which respectively control the weights of the current input, current output, and historical information, making the network flexible in handling historical information. GRU adopts simpler gate mechanisms than LSTM, with only reset gate and update gate, reducing the complexity of the network, but the performance is relatively weaker. Based on this memory module idea, the Legendre Memory Unit (LMU) [8], which can handle longer sequences and has higher performance, has been proposed. LMU models and predicts time-series data by using Legendre polynomials as memory units. LMU combines nonlinear filters and time delays to better capture the long-term dependency relationships of time-series data while maintaining efficiency. This paper applies it to stock trend prediction for the first time, proving its predictive effect on long-term sequence data. Based on Rozenberg et al.'s research ideas [9]. In the development of the field of stock prediction, deep learning models, such as Hierarchical Template-Transformer Model [10] and hybrid models such as MCANet [11], use some of the concepts in them to evaluate stock market predictions. They provide a method that hopes to have a positive impact

on future stock development. Helps formulate more intelligent and personalized investment strategies, thereby providing more support for future stock market development. Among them, MCANet has demonstrated powerful feature learning and generalization capabilities in the Poly (A) signal prediction task through adaptive learning of spatio-temporal context dependencies. Different from this, Wide & Deep Asymmetrical Bidirectional Legendre Memory Units focuses on the financial field and achieves excellent stock price trend prediction by capturing long-term dependencies in time series and combining memory and generalization capabilities. Although the tasks are different, MCANet's powerful feature learning and generalization capabilities may provide an effective model design idea for stock price trend prediction, especially when dealing with financial data with spatiotemporal dependence. In order to predict the stock price change more accurate, this paper proposes a new framework based on Wide Deep Asymmetrical Bidirectional Legendre Memory Units. Our contributions are summarized as follows:

- Building a new asymmetric bidirectional Legendre memory unit (ABiLMU): This paper introduces the basic structure of Legendre memory unit and constructs a new module in an asymmetric bidirectional manner by imitating the structure of U-RNN. This innovative structure provides the model with better memory and expression capabilities, and help improving prediction accuracy.
- Constructing a framework with the asymmetric bidirectional Legendre memory unit (ABiLMU) and the wide-deep concept: We use the ABiLMU as the depth part to effectively extract the temporal features of daily data. At the same time, a multi-layer perceptron (MLP) is used as the width part to capture the trend information in the technical factors with feature selection.
- This paper construct an innovative stock trend prediction method, which performs well on real data of the Shanghai and Shenzhen Stock Exchange 300 constituent stocks and brings higher predication accuracy.

2. RELATED WORKS

Studies in the stock market have shown that stock price trend prediction is closely related to the characteristics of stock time series data [1]. Therefore, early stock prediction methods were mainly based on the study of historical stock price sequences. Financial scholars have introduced various evaluation indicators for predicting stock market price fluctuations. Direct indicators such as opening price and closing price have gradually developed into indirect indicators such as momentum that effectively display certain characteristics. For example, Jegadeesh et al. [12] proposed that the price of a stock has a trend of continuing its original direction of movement, and momentum factors derived from trading volume and turnover rate are used to predict stock price trends. Fama et al. [13] constructed a factor pricing model using derived indicator factors such as total market value and book-to-market ratio to explain the cross-sectional changes in expected stock returns.

Some scholars have also started with statistical methods to study changes in stock price trends. The basic idea is to use historical data to fit stock prices and then use these models to predict future prices. Common statistical models include time series decomposition [14], ARMA [15], ARIMA [16], ARCH [17],

GARCH [18], exponential smoothing [19], etc. These methods usually require the manual selection of parameters, so their prediction accuracy is greatly limited.

With the development of machine learning technology, more and more scholars are using various machine learning methods for stock prediction. Machine learning methods usually do not require manual feature engineering of data in advance. They can automatically learn features from data and use these features for prediction. Common machine learning methods include decision trees [20], support vector machines [21], random forests [22], XG-Boost [23], etc.

With the continuous development and popularization of deep learning technology, research on stock prediction based on deep learning has made significant progress. Compared with traditional machine learning methods, deep learning has achieved better results in stock prediction due to its ability to automatically learn features based on large data sets. RNN has a memory function for time series data due to its unique directed recurrent connections between hidden layers, and has been widely used in natural language processing and time series data analysis [4]. However, RNN models are gradient-based algorithms, and may encounter problems such as gradient vanishing or exploding when processing long time series data. Therefore, LSTM, GRU and other derivative variants are more commonly used.

As the deep learning technology continues to develop and become more widespread, research on stock prediction based on deep learning has also made significant progress. Compared to traditional machine learning methods, deep learning has achieved better results in stock prediction due to its ability to automatically learn features based on large datasets. Recurrent neural networks (RNNs), with their unique directed cyclic connections between hidden layers, have memory capabilities for time series data and have been widely used in natural language processing and time series data analysis. However, RNN models are gradient-based algorithms and may face problems of gradient vanishing or explosion when processing long time series data. Therefore, LSTM and GRU, which are derivative variants, are more commonly used. Chen et al. [24] used LSTM to predict prices based on the time series data generated from the history of the stock market. Zhang et al. [25] proposed a method based on State Frequency Memory to decompose the hidden state of LSTM storage units into multiple frequency components in order to discover multi-frequency trading patterns. Hu [26] used PCA and LASSO methods to perform dimensionality reduction and selection of five major technical analysis indicators, and then used LSTM models to improve the accuracy of predictions. Li [27] used the Whale Optimization Algorithm (WOA) to optimize the important parameters of the LSTM network, reducing the impact of human factors and improving the accuracy of model predictions. Xie et al. [28] combined empirical mode decomposition (EMD) and singular spectrum analysis (SSA) with the LSTM-SVR model to achieve higher prediction accuracy. In contrast to existing methods, the Wide & Deep Asymmetrical Bidirectional Legendre Memory Units (WD-ABiLMU) captures longterm dependencies in time series through an instant backpropagation bidirectional cyclic module and a memory unit based on Legendre polynomials, achieving superior accuracy in stock trend prediction.

3. WIDE & DEEP ASYMMETRICAL BIDIRECTIONAL LEGENDRE MEMORY UNITS

3.1 Stock Technical Indicator

Stock technical factors are a type of factor in stock analysis. They are quantitative indicators calculated based on market data such as stock prices and trading volume, used to analyze trends and movements in the stock market. The study and application of stock technical indicators were proposed by the school of stock technical analysis. Stock technical indicators can be divided into two categories: price-based and volume-based. Price-based technical indicators mainly include moving averages, relative strength index (RSI), stochastic oscillator (KDJ), moving average convergence divergence (MACD), etc. These indicators mainly reflect information on stock price trends, volatility, overbought/oversold conditions, and other related indicators. Volume-based technical indicators mainly include average volume, volume standard deviation, volume oscillation, and others, which reflect information on buying/selling pressure, liquidity, and other related indicators in the stock market. In this paper, 13 daily data and 55 technical indicators were used as input data for the prediction model. The daily data includes 13 indicators: opening price, closing price, highest price, lowest price, yesterday's closing price, trading volume, turnover, total share capital, circulating share capital, free float share capital, total market value, and circulating market value. The technical indicators include 55 commonly used technical indicators with a total of 93 indicators, which are listed in Appendix A Table 4.

3.2 Wide & Deep Model

Wide & Deep was first proposed in the field of recommendation systems [29], which essentially involves training both a linear model and a deep neural network to combine the advantages of memorization and generalization. The model's "memorization ability" refers to its ability to learn and utilize the "co-occurrence frequency" of items or features in historical data, while its "generalization ability" refers to its ability to propagate feature correlations and discover the correlation between rare or previously unseen features and the final label. The wide model shown in Figure 1(a) uses memorization ability to handle stock technology factors that have already been widely used. This means that the wide model can rely on existing experience and knowledge for prediction and analysis. The deep model shown in Figure 1(c) uses feature extraction capabilities to process daily data with implicit trend information. In other words, the deep model can predicts and analyzes hidden patterns and trends in the data. The Wide & Deep model shown in Figure 1(b) combines the Wide model and the Deep model to handle prediction and analysis tasks in the stock market. The comprehensive use of Wide and Deep models can fully utilize existing stock technical factors and implicit trend information, thereby improving the accuracy and comprehensive ability of prediction and analysis. The combination of this model helps to handle complex situations in the stock market, providing more comprehensive predictions and analysis results.

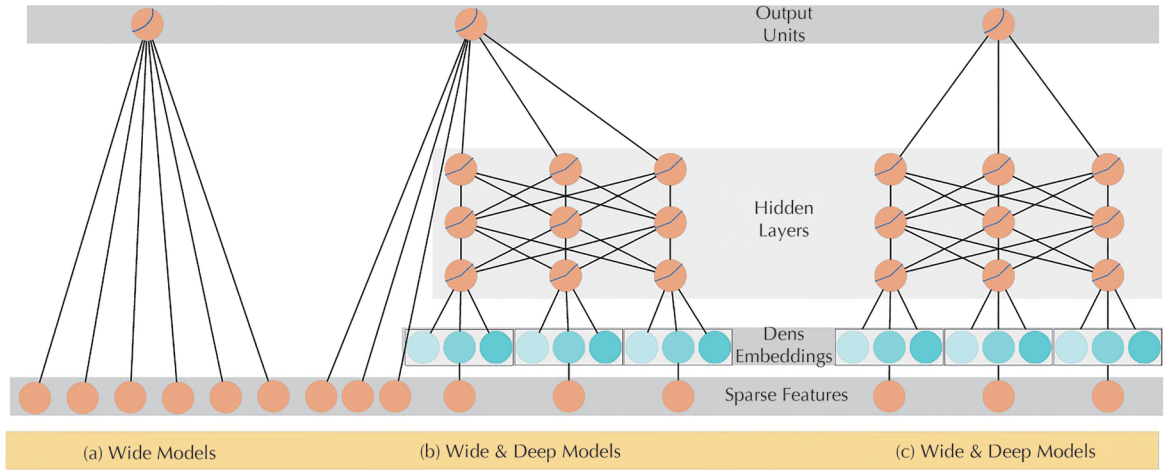


Figure 1. The left structure shows the wide model, the right structure shows the deep model, and the middle structure shows the wide and deep model with the combination of the two model.

3.3 Legendre Memory Units

The Legendre Memory Unit (LMU) is a type of recurrent neural network model that uses Legendre polynomials as memory units to model and predict time series data [8]. By combining nonlinear filters with time delays, the LMU can capture long-term dependencies in time series data and is more efficient than traditional recurrent neural networks such as Long Short-Term Memory (LSTM). The main idea behind the LMU is to map input data through a Legendre polynomial basis function and then pass it to the next layer of the network using matrix multiplication and nonlinear activation functions. This approach allows for the input time series data to be processed in segments and represents long-term temporal relationships as polynomial coefficients.

The main component of LMU is a storage unit that orthogonalizes the input signal $u(t) \in \mathbb{R}$ over a sliding window of length $\theta \in \mathbb{R}_{>0}$ in continuous time. The unit is derived from a linear transfer function $F(s) = e^{-\theta s}$ with continuous-time delays, which is best-approximated by d coupled ordinary differential equations (ODEs): $\theta \dot{\mathbf{m}}(t) = \mathbf{A}\mathbf{m}(t) + \mathbf{B}u(t)$

$\mathbf{m}(t) \in \mathbb{R}^d$ is a d -dimensional state vector and the ideal state space matrix $(\mathbf{A}, \mathbf{B})$ is approximated using the Padé [30] estimate [31]:

$$\mathbf{A} = [a]_{ij} \in \mathbb{R}^{d \times d}, \quad a_{ij} = (2i+1) \begin{cases} -1 & i < j \\ (-1)^{i-j+1} & i \geq j \end{cases} \quad (1)$$

$$\mathbf{B} = [b]_i \in \mathbb{R}^{d \times 1}, \quad b_i = (2i+1)(-1)^i, \quad i, j \in [0, d-1]. \quad (2)$$

The main feature of this dynamical system is the sliding window in which m at the previous moment represents u through the highest Legendre [32] polynomial of order $d-1$:

$$u(t-\theta') \approx \sum_{i=0}^{d-1} \mathcal{P}_i\left(\frac{\theta'}{\theta}\right) m_i(t), \quad 0 \leq \theta' \leq \theta, \quad \mathcal{P}_i(r) = (-1)^i \sum_{j=0}^i \binom{i}{j} \binom{i+j}{j} (-r)^j \quad (3)$$

$\mathcal{P}_i(r)$ is the i th shifted Legendre polynomial [33], which gives the unique and optimal decomposition where the function of m corresponds to the computation of a window of spanning length θ projected onto the d -orthogonal basis function.

Embedding the above equation into the recurrent neural network structure: at time point $t \in \mathbb{N}$, for $\mathbf{m}_t \in \mathbb{R}^d$, given some input $u_t \in \mathbb{R}$, we have $\mathbf{m}_t = \bar{\mathbf{A}}\mathbf{m}_{t-1} + \bar{\mathbf{B}}u_t$ where $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ is the discretization matrix provided by the ODE solver at some time step Δt with respect to the window length θ . For example, assuming that Δt is sufficiently small by the Eulerian method: $\bar{\mathbf{A}} = (\Delta t / \theta)\mathbf{A} + \mathbf{I}$, $\bar{\mathbf{B}} = (\Delta t / \theta)\mathbf{B}$

When $d=1$, the LMU is similar to an LSTM without a gating mechanism in that its storage capacity relative to the frequency content increases as d . The memory of the LMU represents the entire sliding window of the input history as a linear combination of these proportionally invariant polynomials. Increasing the number of dimensions supports the storage of higher frequency inputs relative to the time scale.

As shown in Figure 2, the each single LMU layer maintains its own hidden state vector and memory vector. With the input vector $\mathbf{X}_t$, LMU use previous hidden state $\mathbf{h}_{t-1}$ and previous memory state $\mathbf{m}_{t-1}$ to generates the new memory state $\mathbf{m}_t$ with a linear function. Similar to NRU [32], the hidden state vector is a combination of input, previous state, and current memory vector with a nonlinear function:

$$\mathbf{h}_t = f(\mathbf{W}_x \mathbf{x}_t + \mathbf{W}_h \mathbf{h}_{t-1} + \mathbf{W}_m \mathbf{m}_t) \quad (4)$$

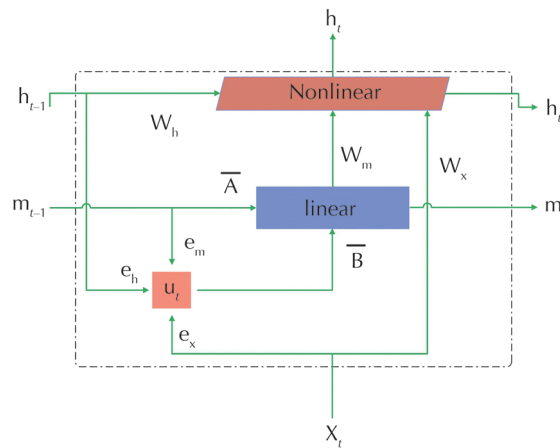


Figure 2. Legendre Memory Unit single layer structure after time step expansion.

where f is a nonlinear function, $\mathbf{W}_x, \mathbf{W}_h, \mathbf{W}_m$ is a learnable parameter matrix.

Although both the hidden state $\mathbf{h}_t$ and the memory state $\mathbf{m}_t$ are obtained using the input, previous state and current memory vector, they are calculated differently. The input signal u_t written to the memory $\mathbf{m}_t$ is:

$$U_t = \mathbf{e}_x^T \mathbf{x}_t + \mathbf{e}_h^T \mathbf{h}_{t-1} + \mathbf{e}_m^T \mathbf{m}_{t-1} \quad (5)$$

where $\mathbf{e}_x, \mathbf{e}_h, \mathbf{e}_m$ are the learnable encoding vectors. The kernel ($\mathbf{W}$) learns to compute the nonlinear functions in the memory module, while the encoder ($\mathbf{e}$) learns to translate the relevant information into the memory module.

3.4 Asymmetrical Bi-RNN

The basic principle of Bidirectional Recurrent Neural Networks (BiRNN) [34] is to process input sequences in both forward and backward directions on the basis of Recurrent Neural Networks (RNN) in order to leverage information from both directions and obtain global contextual information (Figure 3). Specifically, BiRNN consists of two RNNs, one processing the forward sequence and the other processing the backward sequence. The two directions of RNN each output a hidden state, which are concatenated as the final output at each time step. As a result, BiRNN is able to fully utilize information from both directions of the sequence, providing global contextual information for each time step.

However, as a type of RNN, BiRNN still has some illogical aspects when processing time series information, namely its symmetry in the time direction. BiRNN is usually used for natural language processing, where word order is almost entirely determined by grammar rules rather than time series. Nevertheless, in some cases, data may have a preferred direction in time, i.e., the forward direction. Another potential drawback of BiRNN is that its output is merely a simple concatenation of the two directions of the input, meaning that it never reads the input with knowledge of future information. On the other hand, the idea of an Asymmetrical Bidirectional Recurrent Neural Network (U-RNN) [9] is to first propagate backward information and then use future information in the forward propagation process.

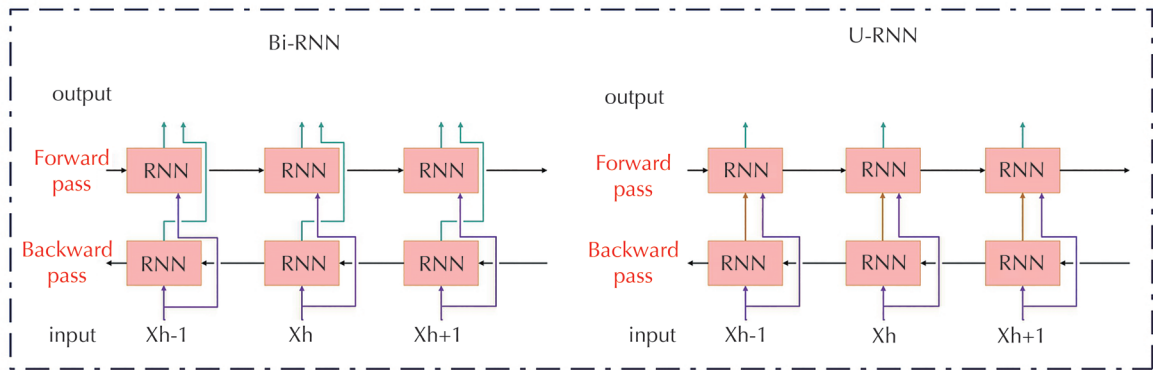


Figure 3. The overall structure of Bi-RNN and U-RNN.

This backward information propagation strategy is more consistent with temporal logic if the data has an obvious forward tendency.

3.5 Wide & Deep Asymmetrical Bidirectional Legendre Memory Units Structure

The LMU module is constructed in an asymmetrical bidirectional structure, similar to URNN, to form the entire ABiLMU temporal feature extraction module, which serves as the deep part for extracting temporal features from daily data. The MLP is used as the wide part to capture the trend information of the technical factors that have been memorized after feature selection, where M technical factors with the lowest intercorrelations are selected using the Spearman correlation coefficient matrix to remove redundant features with high correlation. Finally, the two parts of information are fused to obtain the model's prediction output. The specific network structure is shown in the Figure 4.

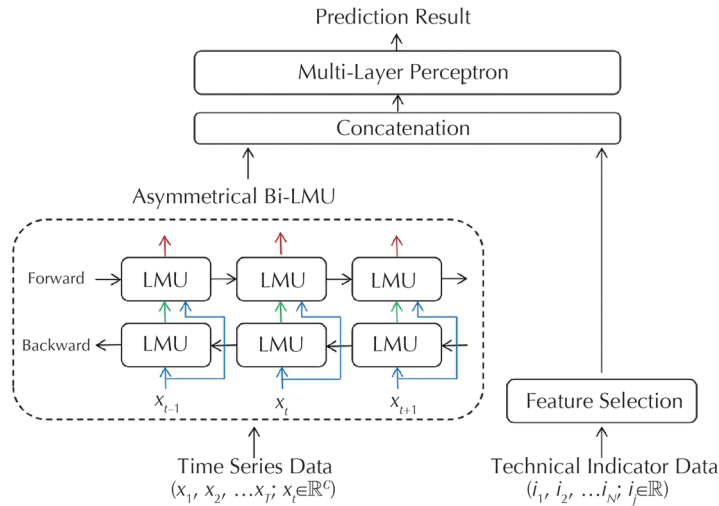


Figure 4. The overall structure of Wide & Deep Asymmetrical Bidirectional Legendre Memory Units.

The input data consists of T days of C -dimensional daily trading data represented by $\mathbf{x}_t \in \mathbb{R}^C$, $1 \leq t \leq T$, which are first processed by the backward LMU module. The hidden state output at each time step is then combined with the corresponding original input at that time step and fed into the forward LMU module. The extracted time series features are then concatenated with the selected technical factors data and passed through a final multi-layer perceptron to obtain the final prediction output.

3.6 Data

The training and testing sets are divided in chronological order, rather than randomly. For the selection of the dataset, we used all constituent stocks of the Shanghai and Shenzhen 300 Index, and used daily

data from 2020 to 2022 for a total of three years. By rolling the time series, we use continuous 60 day data as input to the training set, and use the highest and lowest prices of the next 20 days as corresponding labels for training. Therefore, the final input data are divided into two groups: daily time series data $x_1, x_2, \dots, x_{60} \in \mathbb{R}^{13}$ and technical factor data $i_1, i_2, \dots, i_{93} \in \mathbb{R}$; the labels of the output data are the highest and lowest prices of the last 20 days:

$$p_{\text{highest}} = \max_{t < i \leq t+20} \{\text{high}_i\}, p_{\text{lowest}} = \min_{t < i \leq t+20} \{\text{low}_i\} \quad (6)$$

We divided the training and testing sets using a 9:1 ratio. Because in the stock data, the previous stock prices and data mainly have an impact on subsequent stock price changes, the first 90% samples of the entire dataset in chronological order were used to train the network, and the last 10% samples were used as the test set. In the experiment, we chose the best performing result, which means we chose the model parameter value that performed best in multiple random initializations.

3.7 Evaluation

In this paper, the four most commonly used indicators in regression models are selected as indicators for the evaluation of the model. There are Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

For the predicted value: $y = y_1, y_2, \dots, y_n$ with the true value: $y = \{y_1, y_2, \dots, y_n\}$ there are: (1) Mean Squared Error (MSE) refers to the average of the sum of squares of the differences between the predicted value and the actual value, which is usually used to measure the prediction accuracy of continuous variables.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (7)$$

(2) Root Mean Squared Error (RMSE) refers to the square root of the mean square error, usually used to measure prediction accuracy for continuous variables.

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

(3) Mean Absolute Error (MAE) refers to the average value of the absolute value of the difference between the predicted value and the actual value. It is usually used to measure the prediction accuracy of continuous variables.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (9)$$

(4) Mean Absolute Percentage Error (MAPE) refers to the difference between the predicted value and the actual value. It is usually used to measure the prediction accuracy of discrete variables.

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (10)$$

3.8 Results

The ablation experiments were conducted by selecting the basic models, including the single time-series data without the width part LMU and ABiLMU models, and the Wide & Deep network model, which was combined with the LMU and ABiLMU module structures proposed in this paper. The experiment was conducted on a machine equipped with a single GeForce RTX 3090 GPU. PyTorch was employed as the deep learning framework, and the CUDA version used was 11.7. The prediction results on the test set data of the CSI 300 component stocks are shown in Table 1. Regardless of whether the Wide & Deep module is added or not, the ABiLMU model outperformed the LMU model in the ablation experiments. Moreover, the performance was significantly improved when the ABiLMU model was combined with the WD module to integrate the two types of data.

Table 1. Comparison of ablation experiment results.

Model	MSE	MAE	RMSE	MAPE
LMU (w/o Wide)	0.00778982	0.06166816	0.08825999	0.05932125
ABiLMU (w/o Wide)	0.00762684	0.06166960	0.08733183	0.06015784
Wide & Deep	0.00834562	0.06448297	0.09135436	0.06204423
WD-LMU	0.00655185	0.05760284	0.08094352	0.05717582
WD-ABiLMU	0.00628084	0.05506376	0.07925178	0.05426137

The comparative experiment selected RNN [4], GRU [35], LSTM [24], and the ABiLMU module proposed in this paper to be combined with the basic Wide & Deep network model. The prediction results on the test set data of the CSI 300 constituents are shown in Table 2. In the comparative experiment results, the proposed model in this paper performs better than several commonly used time series neural network models in terms of prediction accuracy. To analyze the stability of the proposed model in detail, this paper chooses to train the model on different datasets and compare its performance on the test set. We use three datasets with the same length but different time periods to train the model, and the data after each time period is used as the test set to observe its performance. The prediction results on the three training sets with a length of two years are shown in Table 3. The experimental results show that the proposed model has a certain stability in terms of performance on different training sets.

Table 2. Comparison of prediction effects of different baseline models.

Model	MSE	MAE	RMSE	MAPE
WD-RNN	0.00759073	0.06196116	0.08712480	0.05935024
WD-GRU	0.00684365	0.05634511	0.08272637	0.05584129
WD-LSTM	0.00665541	0.05779845	0.08158070	0.05600557
WD-ABiLMU (ours)	0.00628084	0.05506376	0.07925178	0.05426137

Table 3. Comparison of prediction results after data training in different time periods.

Training data time period	MSE	MAE	RMSE	MAPE
January 2020-December 2021	0.00633321	0.05571948	0.07958150	0.05478986
April 2020-March 2022	0.00631663	0.05642781	0.07947724	0.05540507
July 2020-June 2022	0.00638765	0.05611271	0.07992281	0.05435620

The goal of our method is to predict the highest and lowest prices in the next 20 days. When use it in practical applications, the available strategy is to buy a stock when the price is close to the predicted lowest price and sell a stock when its' price is close to the predicted highest price. In view of the difficulty of accurately predicting prices, a error tolerance range should be introduced: when the price is higher than 90% of the highest price range or lower than 10% of the lowest price range, the corresponding buy or sell operation is performed.

4. CONCLUSIONS

This paper proposes an innovative stock trend prediction method based on an asymmetric bidirectional Legendre memory unit (ABiLMU) integrated with a wide and deep neural network. The method efficiently captures time series features by simultaneously utilizing sequential characteristics in daily stock market data and trend information in stock technical factors. The ABiLMU is utilized to capture the time series features and then combined with trend information to achieve better prediction of stock trends. The detailed ablation and comparative experiments on the real dataset of the CSI 300 constituent stocks demonstrate the effectiveness of the ABiLMU-based method for stock trend prediction. In future work, we plan to integrate short-term trends in stock prices with company information to better predict overall trends and provide optimal trading strategies.

AUTHOR CONTRIBUTION STATEMENT

Yong Wang (email: ywang@cqut.edu.cn, ORCID: 0000-0002-7847-3807) conceptualization, methodology.

Yisheng Li (email: yamisora@2020.cqut.edu.cn, ORCID: 0000-0002-3585-7667) methodology, formal analysis, software, investigation, and manuscript writing.

Zhiyu Xu (email: xzyaer@2020.cqut.edu.cn, ORCID: 0009-0004-3346-8045) investigation, validation.

DATA AVAILABILITY STATEMENT

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY STATEMENT

The code used during the current study is available from the corresponding author upon reasonable request.

CONFLICT OF INTEREST STATEMENT

On behalf of all authors, the corresponding author states that there is no conflict of interest.

APPENDIX A STOCK TECHNICAL INDICATORS

All stock technical indicators used in this paper are listed in Table 4.

Table 4. Stock technical indicators.

Indicator name	Indicator value
Moving Average	MA
Moving Deviation	MD
Exponential Moving Average	EMA
Different of Moving Average	DMA, AMA
Directional Movement Index	+DI, -DI, ADX, ADXR
Moving Average Convergence Divergence	MACD
Stochastic Indicator	K, D, J
Relative Strength Index	RSI
Stochastic Relative Strength Index	StochRSI, fastK, fastD
Volume Relative Strength Index	VRSI

Table 4. *Continued.*

Indicator name	Indicator value
Bollinger bands/BOLL	boll_up, boll_mid, boll_low
Bull and Bear Index Bollinger bands	bbiboll, UPR, DWN
Williams %R	w_r
Bias	BIAS
Accumulation Swing Index	SI, ASI
Volatility Volume Ratio	VR
Popularity Willingness Indicator	AR, BR
Detrended price Oscillator	DPO, DPOMA
Triple Exponential Moving Average	TEMA
Triple Exponentially Smoothed Average	TRIX, TRMA
Momentum Index	MTM, MTMMA, MI
On Balance Volume	OBV
Commodity Channel Index	CCI
Price Oscillator	OSC
Deviation Bias Deviation Rate	DBCD, DBCDMA
Rate of Change	ROC, ROCMA, RC, RCMA
Capabilities Rate	CR
Psychological Line Indicator	PSY
Williams Accumulation/Distribution	WAD, WADMA
Money Flows Indicators	MFI
Price Volume Trend Indicator	PVT
William's Variable Accumulation Distribution	WVAD, WVADMA
Counter Daily Potential	CDP, AH, AL, NH, NL
Ease of Movement Value	env_up, env_low
MIKE Base	WR, MR, SR, WS, MS, SS
Volume Moving Average	VMA
Volume Moving Average Convergence Divergence	VMACD

Table 4. *Continued.*

Indicator name	Indicator value
Volume Oscillator	VOSC
Total Amount Weighted Stock Index	TAPI, TAPIMA
Volume Standard Deviation	VSTD
ADTM	ADTM, ADTMMA
Adaptive Dynamic Trading Money Flow Index	MICD
Sri Ramakrishna Mills Coimbatore Ltd	SRMI
ZDZB	zdz_b_a, zdzb_b, zdzb_d
Average True Range	TR, ATR
Mass Index	MASS
Vertical Horizontal Filter	VHF
Chaikin Volatility	CVLT
Accumulation Distribution	ACD
Coppock Curve	COP
Price/Earnings Trailing Twelve Months	peTTM
Price/Book Most Recent Quarter	pbMRQ
Price/Sales Trailing Twelve Months	psTTM
Price/Cash Flow Net Cash Flow TTM	pcfNcfTTM

REFERENCES

- [1] Lo, A.W., MacKinlay, A.C.: Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, 1(1): 41–66 (1988)
- [2] Huang, B.-N.: Do Asian stock market prices follow random walks? Evidence from the variance ratio test. *Applied Financial Economics*, 5(4): 251–256 (1995)
- [3] Lee, C.F., Chen, G.M., Rui, O.M.: Stock returns and volatility on China’s stock markets. *Journal of Financial Research*, 24(4): 523–543 (2001)
- [4] Lipton, Z.C., Berkowitz, J., Elkan, C.: A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019* (2015)
- [5] Saad, E.W., Prokhorov, D.V., Wunsch, D.C.: Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Transactions on Neural Networks*, 9(6): 1456–1470 (1998)

- [6] Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Computation*, 9(8): 1735–1780 (1997)
- [7] Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014)
- [8] Voelker, A., Kajic, I., Eliasmith, C.: Legendre memory units: Continuous-time representation in recurrent neural networks. In: *Advances in Neural Information Processing Systems*, Vol 32 (2019)
- [9] Rozenberg, R., Gesnouin, J., Moutarde, F.: Asymmetrical bi-rnn for pedestrian trajectory encoding. *arXiv preprint arXiv:2106.04419* (2021)
- [10] Yuan, L., Wang, J., Yu, L.-C., Zhang, X.: Hierarchical template transformer for fine-grained sentiment controllable generation. *Information Processing & Management* 59(5), 103048 (2022)
- [11] Guo, Y., Zhou, D., Li, P., et al.: Context-Aware Poly (A) Signal Prediction Model via Deep Spatial-Temporal Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* (2022)
- [12] Jegadeesh, N., Titman, S.: Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1): 65–91 (1993)
- [13] Fama, E.F., French, K.R.: Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1): 3–56 (1993)
- [14] Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: STL: A seasonal-trend decomposition. *Journal of Official Statistics*, 6(1): 3–73 (1990)
- [15] Cao, C.Q., Tsay, R.S.: Nonlinear time-series analysis of stock volatilities. *Journal of Applied Econometrics*, 7(S1): S165–S185 (1992)
- [16] Wang, J.-H., Leu, J.-Y.: Stock market trend prediction using ARIMA-based neural networks. In: *Proceedings of International Conference on Neural Networks (ICNN'96)*, Vol 4, pp. 2160–2165 (1996)
- [17] French, K.R., Schwert, G.W., Stambaugh, R.F.: Expected stock returns and volatility. *Journal of Financial Economics*, 19(1): 3–29 (1987)
- [18] Day, T.E., Lewis, C.M.: Stock market volatility and the information content of stock index options. *Journal of Econometrics*, 52(1-2): 267–287 (1992)
- [19] Montgomery, D.C., Johnson, L.A., Gardiner, J.S.: *Forecasting and Time Series Analysis*. McGraw-Hill Companies (1990)
- [20] Wang, H., Jiang, Y., Wang, H.: Stock return prediction based on Bagging-decision tree. In: *2009 IEEE International Conference on Grey Systems and Intelligent Services (GSIS 2009)*, pp. 1575–1580 (2009)
- [21] Huang, W., Nakamori, Y., Wang, S.-Y.: Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32(10): 2513–2522 (2005)
- [22] Kumar, M., Thenmozhi, M.: Forecasting stock index movement: A comparison of support vector machines and random forest. In: *Indian Institute of Capital Markets 9th Capital Markets Conference Paper* (2006)
- [23] Dey, S., Kumar, Y., Saha, S., Basak, S.: Forecasting to Classification: Predicting the direction of stock market price using Xtreme Gradient Boosting. In: *PESIT South Campus*, pp. 1–10 (2016)
- [24] Chen, K., Zhou, Y., Dai, F.: A LSTM-based method for stock returns prediction: A case study of China stock market. In: *2015 IEEE International Conference on Big Data (Big Data)*, pp. 2823–2824 (2015)
- [25] Zhang, L., Aggarwal, C., Qi, G.: Stock price prediction via discovering multi-frequency trading patterns. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2141–2149 (2017)
- [26] Hu, Y.: Stock Forecast Based on Optimized LSTM Model (in Chinese). *Computer Science*, 48(6A): 151–157 (2021)
- [27] Li, J.: Stock price forecast model based on whale algorithm optimizing LSTM (in Chinese). *Intelligent Computer and Applications*, 13(2): 35–40 (2023)

- [28] Xie, Y., Wang, W.: Stock forecasting model based on EMD and SSA (in Chinese). *Computer Engineering and Applications*, 1–10 (2023)
- [29] Cheng, H.-T., Koc, L., Harmsen, J., et al.: Wide & deep learning for recommender systems. In: *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pp. 7–10 (2016)
- [30] Padé, H.: Sur la représentation approchée d’une fonction par des fractions rationnelles. In: *Annales scientifiques de l’Ecole normale supérieure*, Vol 9, pp. 3–93 (1892)
- [31] Voelker, A.R.: *Dynamical systems in spiking neuromorphic hardware*. University of Waterloo (2019)
- [32] Laplace, P.S.: *Mémoires de mathématique et de physique présentés à l’Académie Royale des Sciences par divers Savans* (1776)
- [33] Rodrigues, O.: *De l’attraction des sphéroïdes, Correspondence sur l’École Impériale Polytechnique*. PhD Thesis, Thesis for the Faculty of Science of the University of Paris (1816)
- [34] Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11): 2673–2681 (1997)
- [35] Zhao, J., Wang, Y.: GRU-based Stock Ups and Downs Forecast (in Chinese). *Journal of Huzhou University*, 42(8) (2020)

AUTHOR BIOGRAPHY

Yong Wang is a professor at School of Artificial Intelligence, Chongqing University of Technology, Chongqing, China. His research interests include Artificial Intelligence; multimedia technology.

Yisheng Li is a student at School of Artificial Intelligence, Chongqing University of Technology, Chongqing, China. His research interests include Financial Data Analysis, Artificial Intelligence, Big data.

Zhiyu Xu is a student at School of Artificial Intelligence, Chongqing University of Technology, Chongqing, China. Her research interests include Financial Data Analysis, Artificial Intelligence.