Applied Intelligence (2023) 53:16263–16274
https://doi.org/10.1007/s10489-022-04285-7

Stock index prediction based on multi-time scale learning with multi-graph attention networks

Yuxia Liu[1] · Qi Zhang[2] · Tianguang Chu[1]

Accepted: 19 October 2022 / Published online: 3 December 2022

Abstract

We present a stock index prediction model based on a multi-time scale learning (MTSL) and the multi-graph attention network (MGAT) approach. Instead of dealing with individual stock markets, we consider a group of stock markets simultaneously and exploit the effects of their interactions when making predictions. Our model consists of a boosting Hodrick-Prescott (bHP) filter and MGATs, along with MTSL processes. The bHP filter decomposes the stock index series into the slow-varying growth trend and the fast-varying cyclical volatility for training the MGAT to facilitate the learning of multi-time scale data. The MGAT exploits the interplays of stock markets with three different types of graphs: a regionality graph that qualitatively describes the linkages among stock markets within the same region or similar financial systems; a correlativity graph that quantifies the Pearson correlation between stock markets; and a causality graph that measures the convergence cross mapping (CCM) causality between stock markets. Particularly, the last graph is essentially a directed graph, which captures the nonreciprocal relationship between different stock markets. Experimental results on real stock indices reveal the effectiveness and merit of the proposed model in comparison with other models adopted in this paper.

Keywords Stock index prediction · Multi-time scale learning · bHP filter · Multi-graph attention network

1 Introduction

Stock market prediction is a long-standing problem on which numerous studies are available in the literature (see, e.g., [1, 2] and the references therein). A major challenge in predicting the values of stock price concerns the nonlinear, nonstationary, and multi-time scale behavior inherent in the underlying dynamics of stock time series [3, 4].

In the past decade, researchers have widely used artificial neural network (ANN)-based models to predict time series of stock markets [5–8]. Indeed, the nearly standard architecture with simple nonlinear computing units and the universal approximation ability of general nonlinear functions render ANN an appealing mathematical device to handle nonlinear time series arising from various areas [9–11]. Regarding the problem of predicting stock time series, the often used types of ANNs include, for instance, the radial basis function (RBF) neural network model [5], the deep neural network (DNN) model [6], the stochastic time strength neural network (STNN) model [7], and the long-short term memory model (LSTM) [8]. These ANN-based forecasting models have been confirmed as very competent alternatives in nonlinear time series processing.

Apart from nonlinearity, a striking feature of stock time series is its multiple time scale or multiple frequency behavior, which renders the dynamics of the time series very complicated and increases the difficulty of data processing. Intuitively, it would be difficult to train an ANN model satisfactorily by using input data with various time scales or frequencies because of the wide magnitude of distributions. A reasonable idea is to decompose the stock time series according to different time scales and process the subseries respectively. To date, however, only a few works have addressed this issue in research on financial data prediction approaches [7, 8, 12]. In particular, [7] employed empirical mode decomposition (EMD), [8] used variational mode decomposition (VMD), and [12] made use of discrete

✉ Yuxia Liu
1901111599@pku.edu.cn

1 College of Engineering, Peking University, Beijing, 100871, China

2 School of Information Technology & Management, University of International Business & Economics, Beijing, 100029, China

wavelet transform (DWT) decomposition, in addition to the respective ANN models.

Moreover, due to the increasing interconnections of modern economies, interactions and mutual influences between stock markets have become much more important and should be considered in stock market prediction. This has inspired growing interest in employing graph neural networks to improve stock market prediction [13, 14]. In particular, the authors of [13] made use of a GCN with a Spearman rank-order correlation graph to specify the relations between stocks. In [14], a multi-modality graph neural network was introduced to forecast financial time series. Generally, interactions between stock markets can be complicated and diverse, e.g., direct or indirect, qualitative or quantitative. It would be better to employ multiple graphs to exploit more available stock market interactions to perform predictions. This remains to be studied in detail.

In this paper, we intend to propose a new model of stock market prediction by combining multi-time scale learning (MTSL) and multi-graph attention network (MGAT). The basic idea is as follows. We first decompose the original time series into slow- and fast-varying components and learn the resulting subseries with MGAT. The final prediction is then yielded by aggregating the results of the subseries.

To process with the MTSL, we decompose the stock time series into slow- and fast-varying components by using the boosting Hodrick-Prescott (bHP) filter, which is commonly used for analyzing economic time series [15]. The slow-varying part represents the growth trend of the stock index, and the fast-varying part describes its volatility. The basic graph attention network (GAT) module of MGAT consists of two GAT layers and a linear layer with a graph describing certain types of interplays between stock markets. Here, we are concerned with three types of interplay graphs. Specifically, we use a binary graph to qualitatively describe the stock market regional connections within the same region or similar financial systems. To capture the quantitative relations of the stock markets, we also introduce the Pearson correlation graph and the convergence cross mapping (CCM) causality graph. These two graphs are determined solely by the stock time series data. Additionally, the concept of CCM causality was originally proposed for the study of ecosystems [16] and later applied further to other problems, including financial data analysis [17, 18]. In contrast to the regionality graph and the correlation graph, whose adjacent matrices are symmetric, the causality graph is essentially a directed graph with an asymmetric adjacent matrix describing the nonreciprocal relationship of different stock markets. Thus, the constructed MGAT enables the use of more useful information for stock price predictions.

The advantages of our MTSL-MGAT model are as follows: 1) We decompose stock time series according to different time scales and process each subseries separately. This helps to exploit the multiple time scale characteristics of the stock index. 2) We introduce multi-graph approach to find the relationship between different stock markets, which facilitates the integration of the information from interconnected economic data and achieves better prediction performance. We evaluate the performance of our model by using the empirical data from the Shanghai Composite Index (SHCI), the Shenzhen Component Index (SZCI), the Hang Seng Index (HSI), the Standard & Poor's 500 Index (S&P500) and the NASDAQ Composite Index (NASDAQ) and make comparisons with other models. The data are available in Yahoo Finance [19].

The paper is organized as follows. Section II introduces the preliminaries. Section III presents the MTSL-MGAT model and its computational procedures. Section IV shows the experimental results and Section V concludes this paper.

2 The preliminaries

In this section, we introduce the necessary notions and methods that are used in our study. First, we present the bHP filter, which decomposes the stock time series into slow- and fast-varying components for training the MGAT model. Then, we introduce the concept of CCM for establishing the causality graph of the stock markets, as well as the basic GAT module of the MGAT model.

2.1 The bHP filter

We first recall the HP filter [20], which decomposes l observations of a variable y_t into the following form:

$$y_t = g_t + c_t, \quad t = 1, 2, \ldots, l, \tag{1}$$

where g_t represents the growth trend and c_t represents the cyclical volatility, which correspond to the low-frequency and high-frequency components of y_t, respectively. The trend g_t is determined by the following minimization problem:

$$\min_{g_t} \left\{ \sum_{t=1}^{l} (y_t - g_t)^2 + \eta \sum_{t=3}^{l} (g_t - 2g_{t-1} + g_{t-2})^2 \right\}, \tag{2}$$

where $\eta \geq 0$ is a tuning parameter. Furthermore, let $\boldsymbol{y} = [y_1, y_2, \cdots, y_l]^\top$, $\boldsymbol{g} = [g_1, g_2, \cdots, g_l]^\top$, and

$$F = \begin{bmatrix} 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}_{(l-2)\times l},$$

one can rewrite the problem (2) in a compact form as follows:

$$\min_{\boldsymbol{g}} \left\{ \|\boldsymbol{y} - \boldsymbol{g}\|^2 + \eta \|F\boldsymbol{g}\|^2 \right\}, \tag{3}$$

where $\|\cdot\|$ represents the Euclidean norm of a vector. The following solution to the optimization problem (3) is given:

$$\boldsymbol{g} = S^{-1}\boldsymbol{y},$$

where $S = I_l + \eta F^\top F$, and I_l is the $l \times l$ identity matrix.

The bHP filter successively applies the HP filter to the residual $\boldsymbol{c}$ until it satisfies the Bayesian-type information criterion [15]. After p operations of the filtering, the growth trend and cyclical volatility components are given:

$$\begin{cases} \boldsymbol{g}^{(p)} = \boldsymbol{y} - \boldsymbol{c}^{(p)}, \\ \boldsymbol{c}^{(p)} = (I_l - S^{-1})^p \boldsymbol{y}. \end{cases} \tag{4}$$

This provides the slow- and fast-varying variable decomposition of the primary time series and is processed separately with the MGAT model.

2.2 Convergent Cross Mapping (CCM)

The notion of CCM was proposed in [16] to identify causality in complex ecosystems based on the state space reconstruction approach. Later, this approach was further applied to investigate causal relationships in economic systems [17, 18], meteorological data [21], and energy consumption [22]. Here, we use the CCM approach to establish the causality graph for the stock markets under consideration. The basic idea of the CCM algorithm is as follows [16]. Consider two time series of length l, $\boldsymbol{u} = [u_1, u_2, \ldots, u_l]$ and $\boldsymbol{z} = [z_1, z_2, \ldots, z_l]$. To calculate the cross mapping of $\boldsymbol{u}$ to $\boldsymbol{z}$, one begins by forming the lagged-coordinate vectors $\underline{\boldsymbol{u}}_t = [u_t, u_{t-\tau}, u_{t-2\tau}, \ldots, u_{t-(m-1)\tau}]$ for $t = 1 + (m-1)\tau$ to $t = l$, where τ is the time lag and m is the embedded dimension [16]. This set of vectors is referred to as the "shadow manifold" $\boldsymbol{M_u}$. To generate a cross-mapped estimate of z by $\boldsymbol{M_u}$, denoted as $\hat{z}|\boldsymbol{M_u}$, we first locate the contemporaneous lagged-coordinate vector on $\boldsymbol{M_u}$ and find its $m+1$ nearest neighbors. Next, the time indices are denoted (from closest to farthest) as the $m+1$ nearest neighbors by $t_1, \ldots, t_{m+1}$. These time indices corresponding to the nearest neighbors on $\boldsymbol{M_u}$ are used to identify a set of neighbors in z, denoted as $z_t = \{z_{t_1}, \ldots, z_{t_{m+1}}\}$. Finally, the estimate of z from a locally weighted mean of the $m+1$ z_{t_i} values is denoted as follows:

$$\hat{z}|\boldsymbol{M_u} = \sum_{i=1}^{m+1} r_i z_{t_i},$$

where r_i is a weight based on the distance between $\underline{\boldsymbol{u}}_t$ and its ith nearest neighbor on $\boldsymbol{M_u}$, determined by:

$$r_i = \frac{r_i'}{\sum_{j=1}^{m+1} r_j'},$$

where

$$r_i' = \exp\left(-\frac{d(\underline{\boldsymbol{u}}_t, \underline{\boldsymbol{u}}_{t_i})}{d(\underline{\boldsymbol{u}}_t, \underline{\boldsymbol{u}}_{t_1})}\right),$$

and $d(\underline{\boldsymbol{u}}_t, \underline{\boldsymbol{u}}_s)$ is the Euclidean distance between two vectors. The CCM causality coefficient of $\boldsymbol{u}$ to z is denoted as:

$$\zeta = \frac{\mathrm{cov}(z, \hat{z}|\boldsymbol{M_u})}{\sqrt{\mathrm{var}(z)}\sqrt{\mathrm{var}(\hat{z}|\boldsymbol{M_u})}}, \tag{5}$$

where $\mathrm{cov}(\cdot)$ is the covariance function and $\mathrm{var}(\cdot)$ is the variance function.

As suggested in [16], if $\boldsymbol{u}$ is dynamically coupled to z, the nearest neighbors of $\boldsymbol{M_u}$ should identify the corresponding nearest neighbors' time indices on z. With the increase in the time series length l, the attractor manifold on $\boldsymbol{M_u}$ fills in and the distances among the $m+1$ nearest neighbors shrinks. As a result, $\hat{z}|\boldsymbol{M_u}$ converges to z. In this sense, convergence of the nearest neighbors is used to test whether there is a correspondence from $\boldsymbol{u}$ to z.

The cross mapping from z to $\boldsymbol{u}$ and the CCM causality coefficient of z to $\boldsymbol{u}$ are defined analogously.

2.3 Graph Attention Network (GAT)

GAT introduces the attention mechanism into graph neural networks by calculating the attention coefficient and weighted summation, which enhances feature learning capabilities [23]. Let $\mathcal{G} = (v, \epsilon, \Theta)$ be a directed weighted graph, where v represents a finite set of n vertices, ϵ denotes a set of edges, and $\Theta = [\Theta_{ij}] \in \mathbb{R}^{n\times n}$ is an adjacency matrix with Θ_{ij} denoting the weight of the edge between node i and node j. Let $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ and $\boldsymbol{h} = \{h_1, h_2, \ldots, h_n\}$ be the input and output of GAT, respectively, where $x_i \in \mathbb{R}^{D_{in}}$ is the input feature vector, and $h_i \in \mathbb{R}^{D_{out}}$ is the output feature vector for $i = 1, 2, \ldots, n$. The attention coefficient between node i and its neighbor node j is given as:

$$e_{ij} = a^\top([Wx_i \| Wx_j])\tilde{\Theta}_{ij}, \quad j \in \mathcal{N}_i,$$

where $W \in \mathbb{R}^{D_{out}\times D_{in}}$ is a shared linear transformation used to learn the attention coefficient between nodes i and j, $[\cdot\|\cdot]$ represents the splicing of the transformed features of nodes

i and j, $a(\cdot)$ represents the mapping of the spliced high-dimensional features to a real value, $\tilde{\Theta} = \Theta + I_n = [\tilde{\Theta}_{ij}]$ is the adjacency matrix of the graph $\mathcal{G}$ with added self-connections, I_n is the $n \times n$ identity matrix, and $\mathcal{N}_i$ is some neighborhood of node i in the graph. The normalized e_{ij} is as follows:

$$\gamma_{ij} = \frac{\exp(\mathrm{LeakyReLU}(e_{ij}))}{\sum_{j' \in \mathcal{N}_i} \exp(\mathrm{LeakyReLU}(e_{ij'}))},$$

where $\mathrm{LeakyReLU}(\cdot)$ represents the activation function. Then, the features are weighted and summed as follows:

$$h_i = \sum_{j \in \mathcal{N}_i} \gamma_{ij} W x_j, \tag{6}$$

which represents the new feature of GAT's output for node i.

For a more robust process of self-attention learning, we further introduce the multi-head attention mechanism by modifying the (6) as follows:

$$h_i(K) = \frac{1}{K} \sum_{k=1}^{K} \sum_{j \in \mathcal{N}_i} \gamma_{ij}^k W^k x_j, \tag{7}$$

where K represents the number of attention heads, γ_{ij}^k is the normalized attention coefficient for the kth-head, and W^k is the corresponding input linear transformation matrix. For more details see, e.g., [23].

3 The MTSL-MGAT prediction model

In this section, we specify the structure of the MGAT and the related multi-time scale prediction procedure.

3.1 The basic GAT module

The basic GAT module for MGAT consists of three layers, with two GAT layers and a linear layer, as shown in Fig. 1. The GAT layer aggregates neighboring node features through a graph attention mechanism, characterizing certain kinds of interactions between different stock markets. The linear layer predicts the current moment data by a linear combination of historical data and consider the time dependence of the stock market itself. Let the dimensions of the signals in the input, hidden, and output layers be D_{in}, D_h, and D_{out}, respectively. For the input signal $\boldsymbol{x}$, the output of the first GAT layer is $\boldsymbol{h}^{(1)} = \mathrm{GAT}(\Theta, \boldsymbol{x}, W^{(1)})$, and the layer is activated by ReLU as $\boldsymbol{o}^{(1)} = \mathrm{ReLU}\left(\boldsymbol{h}^{(1)}\right)$, where $\mathrm{GAT}(\cdot)$ represents a graph attention network using a multi-head attention mechanism, Θ is the weight coefficient matrix in the graph $\mathcal{G}$, and $W^{(1)} \in \mathbb{R}^{D_h \times D_{in}}$ are trainable parameters of the first GAT layer. The output of the second GAT layer is given by $\boldsymbol{h}^{(2)} = \mathrm{GAT}(\Theta, \boldsymbol{o}^{(1)}, W^{(2)})$, where

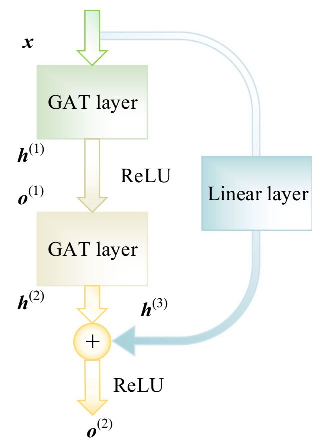


Fig. 1 The structure of the GAT module

$W^{(2)} \in \mathbb{R}^{D_{out} \times D_h}$ are trainable parameters of the second GAT layer. The output of $\boldsymbol{x}$ through the linear layer is $\boldsymbol{h}^{(3)} = W^{(3)}\boldsymbol{x} + b$, where $W^{(3)} \in \mathbb{R}^{D_{out} \times D_{in}}$ and $b \in \mathbb{R}^{D_{out}}$ are the learnable weights and biases of the linear layer. The final output is $\boldsymbol{o}^{(2)} = \mathrm{ReLU}\left(\boldsymbol{h}^{(2)} + \boldsymbol{h}^{(3)}\right)$.

3.2 The architecture of the MGAT

The proposed MGAT model for stock market prediction is illustrated in Fig. 2. Here, we consider five stock markets, namely, SHCI, SZCI, HSI, S&P500, and NASDAQ, and assign them as vertices v_1, v_2, v_3, v_4, and v_5 of a graph $\mathcal{G}$. To better exploit the complex interplay of these stock markets to fulfill the prediction task, we introduce the following three types of graphs to describe the primary interconnections among them.

The regionality graph $\mathcal{G}_1 = (v, \epsilon, \Theta_1)$: We use this type of graph to qualitatively describe the regional economic relationship between stock markets in the same economic region or similar financial systems. As a rule, these markets tend to have strong mutual influences, such as the case of SHCI and SZCI in China. Of the five concerned stock markets, HSI in Hong Kong is something special. It is geographically part of the Chinese economic circle and has a financial system similar to that in the West. Therefore, the HSI should have a connection with all other markets. Thus, we adopt the following regional connectivity matrix

$$\Theta_1 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

The correlativity graph $\mathcal{G}_2 = (v, \epsilon, \Theta_2)$: We further quantify the relations among the stock markets by

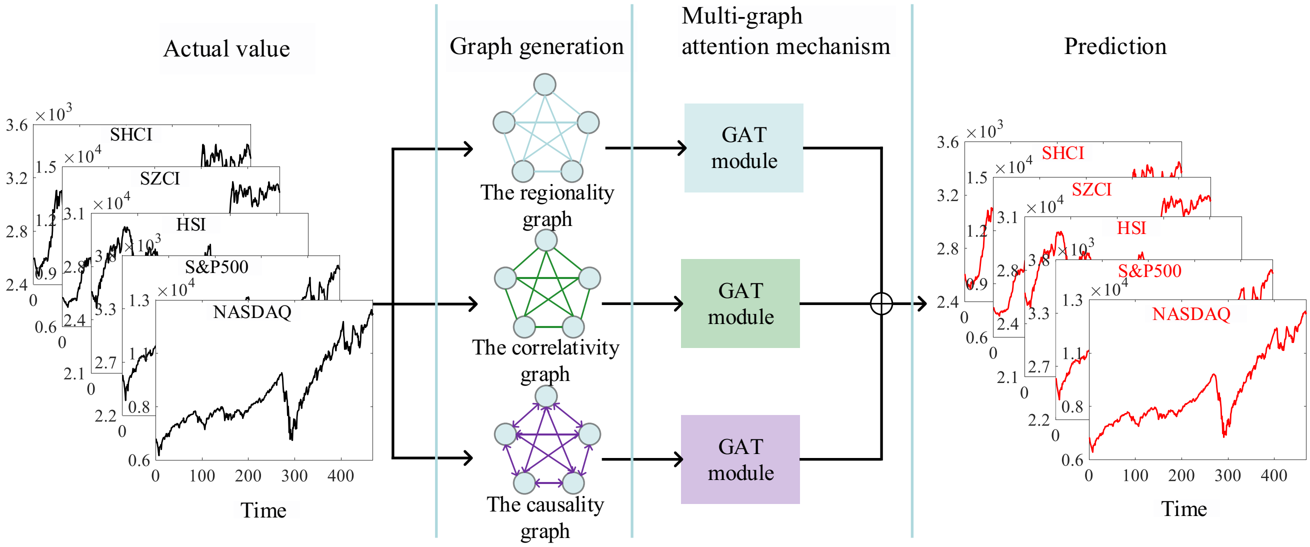


Fig. 2 The architecture of the MGAT

calculating the Pearson correlation coefficients between the stock markets, which are denoted as follows:

$$\Theta_{2,ij} = \frac{\mathrm{cov}(x_i, x_j)}{\sqrt{\mathrm{var}(x_i) \times \mathrm{var}(x_j)}},$$

where x_i and x_j denote signals (the time series data) for vertices v_i and v_j, $\mathrm{cov}(\cdot)$ is the covariance function, and $\mathrm{var}(\cdot)$ is the variances function [24]. Using the data available in Yahoo Finance [19], we can calculate the weight coefficient matrix as follows:

$$\Theta_2 = \begin{bmatrix} 0 & 0.9305 & 0.6996 & 0.3626 & 0.3375 \\ 0.9305 & 0 & 0.6664 & 0.2716 & 0.2874 \\ 0.6996 & 0.6664 & 0 & 0.7399 & 0.7014 \\ 0.3626 & 0.2716 & 0.7399 & 0 & 0.9853 \\ 0.3375 & 0.2874 & 0.7014 & 0.9853 & 0 \end{bmatrix}.$$

The causality graph $\mathcal{G}_3 = (v, \epsilon, \Theta_3)$: We also use the CCM method mentioned in Section 2.2 to calculate the causality between the stock markets, and obtain the weight coefficient matrix:

$$\Theta_3 = \begin{bmatrix} 0 & 0.9366 & 0.7456 & 0.4910 & 0.5069 \\ 0.9307 & 0 & 0.6825 & 0.3626 & 0.3794 \\ 0.6504 & 0.6795 & 0 & 0.7371 & 0.7001 \\ 0.6104 & 0.4904 & 0.7815 & 0 & 0.9930 \\ 0.6341 & 0.6160 & 0.8503 & 0.9918 & 0 \end{bmatrix},$$

whose entries represent the CCM causality coefficients for a pair of vertices in the graph, which are defined by (5).

At this point, we would like to make some comments as follows. It can be observed that the first two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are undirected graphs, with symmetric connectivity matrices Θ_1 and Θ_2 describing reciprocal interactions between two stock markets. Furthermore, it can be seen from Θ_2 that the Pearson value between SHCI and SZCI is greater than 0.9, reflecting a very strong coupling. The case of the S&P500 and NASDAQ is similar. However, the Pearson value between SHCI/SZCI and S&P500/NASDAQ is less than 0.4, indicating a relatively weak coupling relation between the stock markets in different economic regions. In the case of HIS, it is clear by Θ_2 that the Pearson values between HIS and the other four stock indices all range from 0.6 to 0.7, indicating a strong coupling between HIS and each of the others, as well. These results are consistent with our choice of the regional connectivity matrix Θ_1.

In contrast to graphs $\mathcal{G}_1$ and $\mathcal{G}_2$, the causality graph $\mathcal{G}_3$ is a directed graph, with an asymmetric connectivity matrix Θ_3 capturing the disparity in mutual influences of two stock markets. The entry values of Θ_3 indicate that, in general, stock markets in the same regions have strong dynamic dependence on each other, and S&P500/NASDAQ has a relatively greater impact on the variation in SHCI/SZCI.

In the sequel, we integrate all the above defined graphs into a multi-graph attention mechanism and use a fully connected layer to process the features obtained from the multi-graph attention layer to predict the stock markets. This enables us to take advantage of different kinds of interaction effects between stock markets to achieve better predictions.

3.3 The complete computational procedures

Overall, the MTSL-MGAT model for stock market prediction can be completed by the following steps (also see

Fig. 3) after the data collection and preprocessing for the training stage.

Step 1: Generate the three types of connectivity graphs $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$ in Section 3.2, where we take the parameter values $m = 2$ and $\tau = 1$ for the causality graph $\mathcal{G}_3$.

Step 2: Decompose the pre-processed stock index time series into growth trend and economic volatility components, i.e., the slow- and fast-varying parts, by using the bHP filtering method described in Section 2.1, with the parameter $\eta = 1 \times 10^5$ and $p = 50$ following [25, 26].

Step 3: Establish the MGAT model composed of basic GAT modules as shown in Figs. 1 and 2 by integrating the three graphs in *Step 1* into the multi-graph attention mechanism for processing the slow- and fast-varying components of the time series.

Step 4: Create the training set, the validation set, and the test set; Set the parameter values $K = 1$, $D_{in} = 5 \times 6$, $D_h = 10$, and $D_{out} = 1$ in the MGAT model; initialize the GAT layers with trainable parameters $W^{(1)}$ and $W^{(2)}$, and the linear layer with learnable parameters $W^{(3)}$ and b; choose ReLU as the activation function, MSELoss as the loss function, and Adam as the optimizer; train the MGAT model.

Step 5: Aggregate the results from *Step 4* of the slow- and fast-varying components to yield the final prediction.

4 Experiments

We apply the MTSL-MGAT model to the analysis and prediction of the five stock indices SHCI, SZCI, HSI, S&P500, and NASDAQ and evaluate the performance of the proposed approach in comparison with existing models. The data used in our experiments are available in [19]. We first collect opening values, closing values, highest values, lowest values, and trade volumes of the stock indices for each day from February 18, 2004 to December 11, 2020. Then, we exclude holiday trading data, and obtain the stock index datasets of 3969 trading days, including the training set of 3100 trading days from February 18, 2004 to April 11, 2017, the validation set of 400 trading days from April 12, 2017 to December 13, 2018, and the test set of 469 trading days from December 14, 2018 to December 11, 2020. In our experiments, we normalize the data in the range [0, 1] by:

$$\boldsymbol{x}' = \frac{\boldsymbol{x} - \min(\boldsymbol{x})}{\max(\boldsymbol{x}) - \min(\boldsymbol{x})}$$

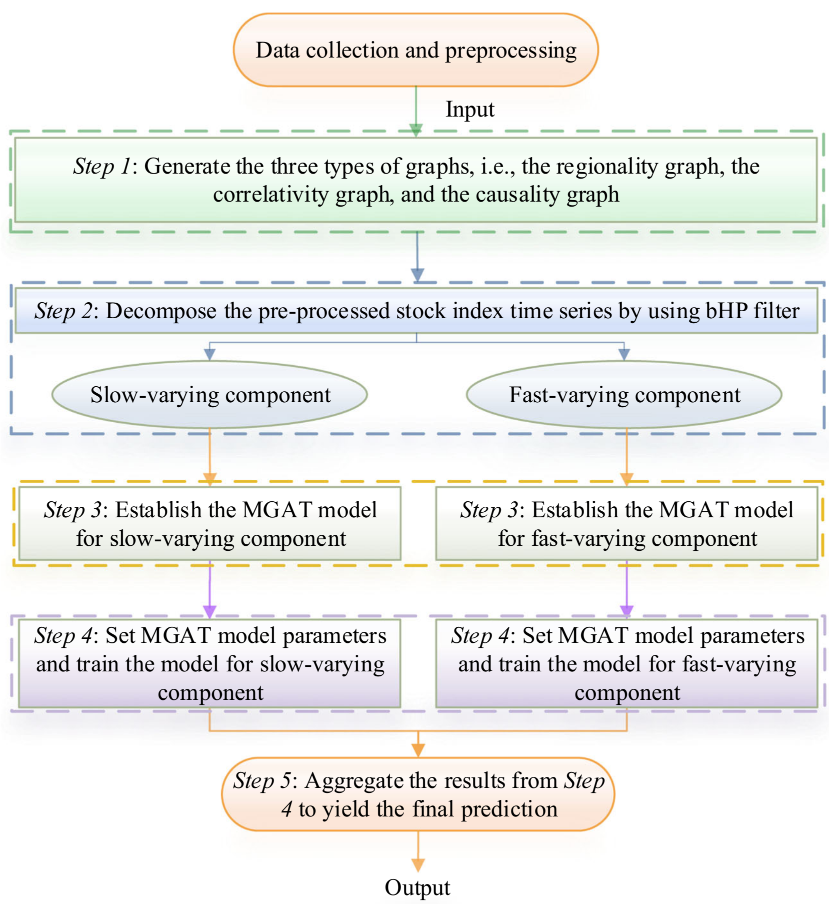
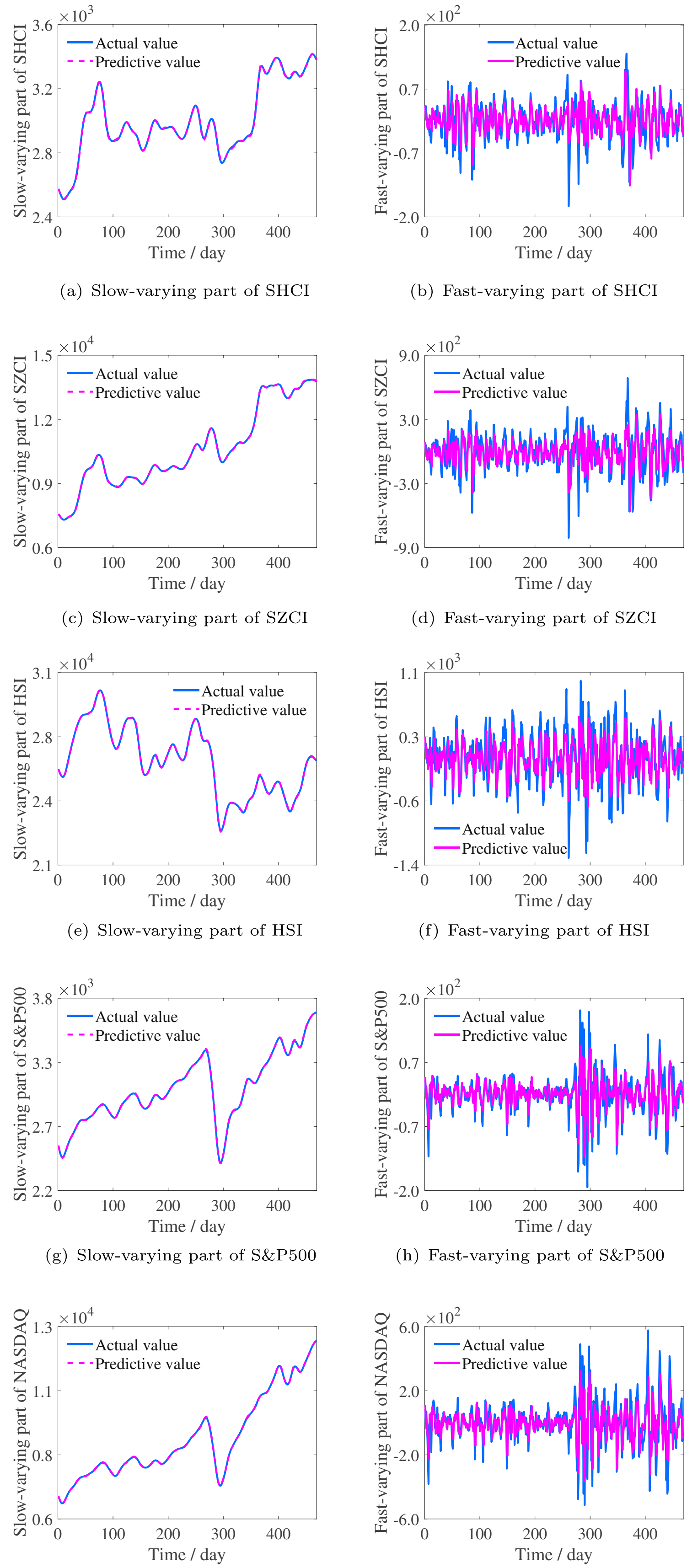


Fig. 3 The procedures of the MTSL-MGAT model

Fig. 4 The predicted and actual values of slow- and fast-varying parts



(a) Slow-varying part of SHCI (b) Fast-varying part of SHCI

(c) Slow-varying part of SZCI (d) Fast-varying part of SZCI

(e) Slow-varying part of HSI (f) Fast-varying part of HSI

(g) Slow-varying part of S&P500 (h) Fast-varying part of S&P500

(i) Slow-varying part of NASDAQ (j) Fast-varying part of NASDAQ

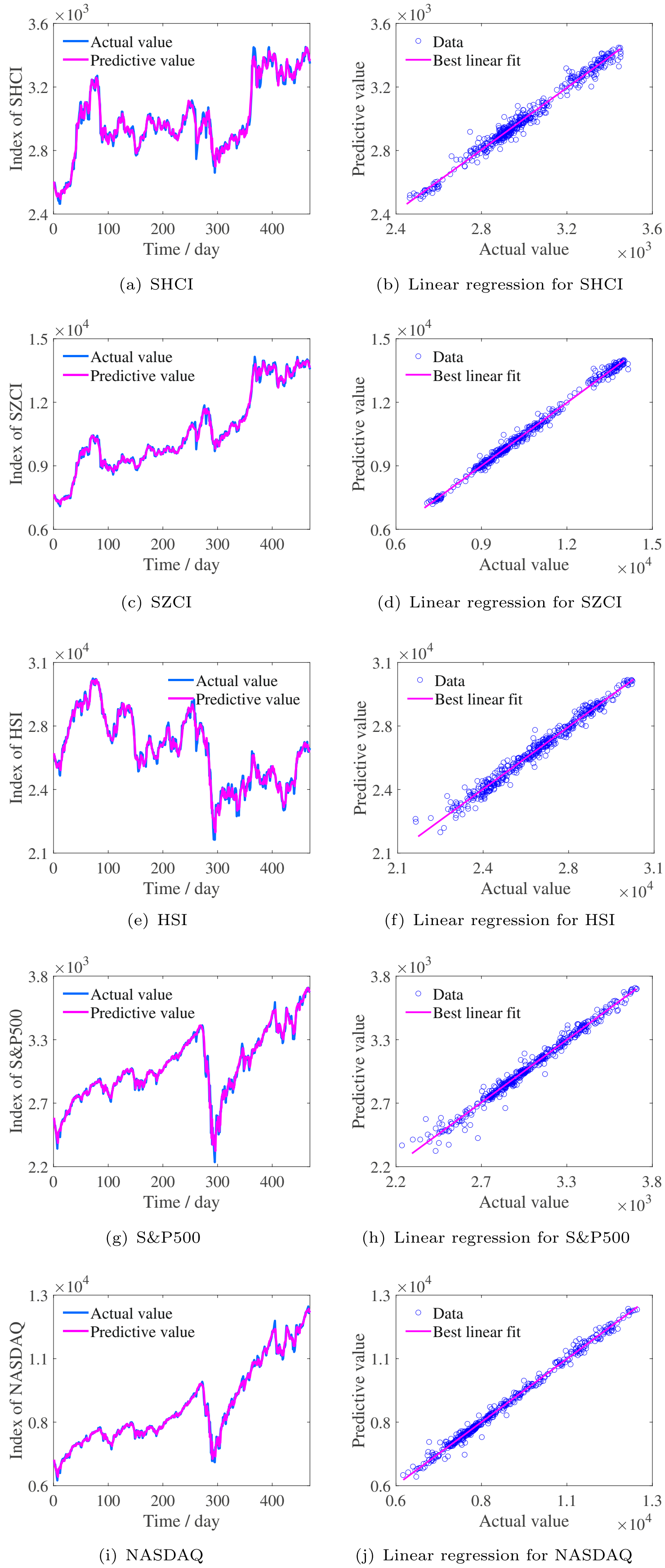


Fig. 5 Comparison of the predicted and actual values

Table 1 Values of correlation coefficients and linear regression parameters for different stock market indices

Parameter	SHCI	SZCI	HSI	S&P500	NASDAQ
ρ	0.9901	0.9970	0.9893	0.9911	0.9967
ϕ	0.9768	0.9929	0.9817	0.9873	0.9942
ω	69.68	62.89	468.80	40.03	50.09

to eliminate data dimensional differences, where $\max(\cdot)$ and $\min(\cdot)$ are the maximum and minimum values in the time series, respectively. Our task is applying the historical data of the previous six days as the input to predict the closing value of the next trading day.

4.1 Prediction results of the MTSL-MGAT model

In this section, we present our experimental results from the MTSL-MGAT model for the cases of SHCI, SZCI, HSI, S&P500, and NASDAQ. Figure 4 shows the predicted and actual values of the slow- and fast-varying components of the time series, respectively, for each case. In general, the stock market indices exhibit multi-time scale characteristics, and their decomposition into slow- and fast-varying components facilitates processing the data more effectively and precisely. In fact, slow-varying components have smooth curves that change gently, and thus, can be predicted more accurately. In contrast, fast-varying components have rough curves that vary violently and cannot be well predicted. Since the proportion of fast-varying components in the original data is far lower than that of slow-varying components, it has less of an impact on the overall predicted values despite the lower prediction accuracy of fast-varying components.

The final stock index predictions are plotted in Fig. 5, which aggregates the predictions of the slow- and fast-varying components for each stock index time series. It can be seen that the predicted results agree with the actual data very well, showing good performance of the trained MTSL-MGAT model.

Next, we evaluate the predictive performance of the proposed model. We first consider the Pearson correlation coefficient ρ between the actual data and the predicted data. The value of ρ measures the strength of a linear relationship between two variables [24]. Particularly, a coefficient of $\rho = 1$ indicates that the two variables have a perfect positive correlation. Table 1 provides the numerical values of the correlation coefficients for SHCI, SZCI, HSI, S&P500 and NASDAQ. It shows that the values of ρ are all around 0.99, signifying a very strong association between the predicted and actual data. We also perform linear regression analysis for the datasets of the MTSL-MGAT model with the regression equation $\boldsymbol{Y} = \phi \boldsymbol{X} + \omega$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the actual data and the predicted data, ϕ represents the slope, and ω represents the intercept of the linear equation. The closer the slope of the linear equation is to 1, the closer the predicted value is to the actual value. We plot the linear regression for SHCI, SZCI, HSI, S&P500, and NASDAQ in Fig. 5(b)(d)(f)(h)(j), which clearly indicate a high fitting degree of the predicted and actual values. Table 1 also provides the regression parameter values for the five stock indices. The slopes of the linear equations are all greater than 0.97, indicating that the deviation between the predicted and actual values is not substantial.

4.2 Comparison with other predicting models

To verify the efficiency of our MTSL-MGAT model, we compare models including the ANN, recurrent neural network (RNN), LSTM, BiGRU [27], GCN (adjacency matrix Θ_1) [28], MGAT, multivariate empirical mode decomposition-long short term memory network (MEMD-LSTM) [29], and multi-time scale learning-graph attention network (MTSL-GAT) with adjacency matrix Θ_2. In particular, the network structure of the ANN is similar to that of our MTSL-MGAT model, which has a two-layer network with an input layer at a size of 30, a hidden layer at a size of 10, and an output layer at a size of 1. Meanwhile, we use ReLU as the activation function, and take MSELoss and Adam as the loss function and the optimizer, respectively. Other compared models have similar configurations. Table 2 presents the the mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and relative root mean square error (rRMSE) values given by these models [7, 30]. The definitions are as follows:

$$\mathrm{MAE} = \frac{1}{l}\sum_{t=1}^{l}|X_t - Y_t|,$$

$$\mathrm{MAPE} = 100 \times \frac{1}{l}\sum_{t=1}^{l}\left|\frac{X_t - Y_t}{X_t}\right|,$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{l}\sum_{t=1}^{l}(X_t - Y_t)^2},$$

$$\mathrm{rRMSE} = \sqrt{\frac{1}{l}\sum_{t=1}^{l}\left(\frac{X_t - Y_t}{X_t}\right)^2},$$

Table 2 The MAE, MAPE, RMSE, and rRMSE values of different models

Index	SHCI	SZCI	HSI	S&P500	NASDAQ
Errors	MAE				
MTSL-MGAT	23.33	104.68	196.06	23.76	80.89
MTSL-GAT	25.04	104.57	210.97	24.74	83.23
MEMD-LSTM	23.59	107.41	203.93	27.11	91.28
MGAT	30.24	132.37	240.86	31.34	107.68
GCN	29.53	142.21	255.42	32.58	117.46
BiGRU	31.33	146.09	305.37	33.82	110.38
LSTM	35.17	155.33	263.76	34.08	121.93
RNN	33.70	149.89	266.11	40.19	131.03
ANN	35.55	163.85	338.11	44.09	141.68
Errors	MAPE				
MTSL-MGAT	0.7790	0.9807	0.7580	0.8105	0.9212
MTSL-GAT	0.8337	0.9786	0.8209	0.8431	0.9458
MEMD-LSTM	0.7835	1.0127	0.7860	0.9126	1.0060
MGAT	1.0080	1.2531	0.9312	1.0706	1.2293
GCN	0.9857	1.3420	0.9861	1.1123	1.3502
BiGRU	1.0403	1.3747	1.1658	1.1462	1.2513
LSTM	1.1704	1.4601	1.0201	1.1600	1.3722
RNN	1.1147	1.4139	1.0252	1.3208	1.4308
ANN	1.1788	1.5320	1.2888	1.4762	1.5915
Errors	RMSE				
MTSL-MGAT	31.87	144.02	261.76	38.40	121.50
MTSL-GAT	34.90	149.33	284.73	40.40	128.62
MEMD-LSTM	32.31	144.59	269.53	39.71	133.21
MGAT	42.01	186.52	333.96	49.35	160.91
GCN	41.23	197.49	349.42	49.26	169.03
BiGRU	44.34	203.97	393.01	49.14	157.47
LSTM	46.79	210.75	354.23	50.07	170.73
RNN	46.43	203.10	362.78	57.33	187.32
ANN	48.67	220.26	432.75	56.52	181.93
Errors	rRMSE				
MTSL-MGAT	0.0107	0.0133	0.0103	0.0138	0.0144
MTSL-GAT	0.0116	0.0137	0.0113	0.0145	0.0153
MEMD-LSTM	0.0107	0.0135	0.0106	0.0140	0.0149
MGAT	0.0140	0.0175	0.0132	0.0178	0.0191
GCN	0.0138	0.0183	0.0137	0.0177	0.0201
BiGRU	0.0147	0.0190	0.0151	0.0174	0.0182
LSTM	0.0156	0.0196	0.0140	0.0179	0.0196
RNN	0.0153	0.0190	0.0142	0.0192	0.0203
ANN	0.0161	0.0204	0.0166	0.0194	0.0204

where X_t and Y_t are the actual value and the predicted value at time t for $t = 1, 2, \ldots, l$ and l is the actual data $\boldsymbol{X}$ length.

From Table 2, the MAE, MAPE, RMSE, and rRMSE values of the MTSL-MGAT model for the SHCI dataset are 23.33, 0.7790, 31.87, and 0.0107, respectively, which are smaller than those of the compared models, indicating that the predicted values of the proposed MTSL-MGAT model are closer to the actual values. Similarly, it can be observed that the MTSL-MGAT model always outperforms the other models on the SZCI, HSI, S&P500, and NASDAQ indices.

5 Conclusion

In this work, we developed an MTSL-MGAT model for the prediction of stock market time series. The merit of the present approach consists of two aspects. One is concerned with the decomposition of stock index series into fast- and slow-varying components by bHP filtering and processing the two parts separately on their own time scales. This facilitates the training of MGAT with multi-time scale data and helps to improve the predictive precision. The other one addresses the effects of generic connections among stock markets with multi-graph descriptions to take advantage of comprehensive information for making predictions. Experiments show that the MTSL-MGAT model achieves lower MAE, MAPE, RMSE, and rRMSE values in the SHCI, SZCI, HSI, S&P500, and NASDAQ test datasets than the other models adopted in this paper, indicating that the MTSL-MGAT model has a higher prediction precision.

This work provides an interdisciplinary study from economics and computer science. From an economics perspective, our MTSL-MGAT model effectively improves the prediction accuracy of the stock index by enploiting the multi-scale characteristics from various interconnected stock markets. From a computer science perspective, the proposed model verifies the feasibility and efficiency of applying machine learning approaches to deal with nonlinear and unsteady data by incoporating a multi-time scale learning method with a deep learning mechanism based on a multi-graph attention network. In the future, our MTSL-MGAT model may be extended to forecast other types of economic data, such as GDP, commodity prices, and exchange rates, etc.

Acknowledgements Project supported by the National Natural Science Foundation of China (Grant Nos. 61673027 and 62106047), Beijing Natural Science Foundation (No. 19L2037)) and the Industry-University-Research Innovation Fund for Chinese Universities (No. 2020HYB04001).

References

1. Xu Y, Yang C, Peng S, Nojima Y (2020) A hybrid two-stage financial stock forecasting algorithm based on clustering and ensemble learning. Appl Intell 50(11):3852–3867
2. Bazrkar MJ, Hosseini S (2022) Predict stock prices using supervised learning algorithms and particle swarm optimization algorithm. Comput Econ:1–22
3. Zhang J, Li L, Chen W (2021) Predicting stock price using two-stage machine learning techniques. Comput Econ 57(4):1237–1261
4. Wang X, Li X, Li S (2022) A novel stock indices hybrid forecasting system based on features extraction and multi-objective optimizer. Appl Intell:1–24
5. Jia L, Li W, Qiao J (2022) An online adjusting RBF neural network for nonlinear system modeling. Appl Intell:1–14
6. Singh R, Srivastava S (2017) Stock prediction using deep learning. Multimed Tools Appl 76(18):18569–18584
7. Wang J, Wang J (2017) Forecasting stochastic neural network based on financial empirical mode decomposition. Neural Netw 90:8–20. https://doi.org/10.1016/j.neunet.2017.03.004
8. Niu H, Xu K, Wang W (2020) A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network. Appl Intell 50(12):4296–4309
9. Sitte R, Sitte J (2002) Neural networks approach to the random walk dilemma of financial time series. Appl Intell 16(3):163–171
10. Karevan Z, Suykens JA (2020) Transductive LSTM for time-series prediction: an application to weather forecasting. Neural Netw 125:1–9
11. Koo E, Kim G (2022) A new neural network approach for predicting the volatility of stock market. Comput Econ:1–15
12. Khandelwal I, Adhikari R, Verma G (2015) Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. Procedia Comput Sci 48:173–179. https://doi.org/10.1016/j.procs.2015.04.167
13. Chen W, Jiang M, Zhang WG, Chen Z (2021) A novel graph convolutional feature based convolutional neural network for stock trend prediction. Inf Sci 556:67–94
14. Cheng D, Yang F, Xiang S, Liu J (2022) Financial time series forecasting with multi-modality graph neural network. Pattern Recogn 121:108218
15. Phillips PCB, Shi Z (2021) Boosting: why you can use the HP filter. Int Econ Rev 62(2):521–570
16. Sugihara G, May R, Ye H, Hsieh C-H, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. Science 338(6106):496–500
17. Stolbov M, Shchepeleva M (2020) Systemic risk, economic policy uncertainty and firm bankruptcies: Evidence from multivariate causal inference. Res Int Business Finance 52:101172. https://doi.org/10.1016/j.ribaf.2019.101172
18. Wu T, Gao X, An S, Liu S (2021) Time-varying pattern causality inference in global stock markets. Int Rev Financial Anal 77:101806. https://doi.org/10.1016/j.irfa.2021.101806
19. (2020). Yahoo Finance: stock data. [EB/OL]. https://hk.finance.yahoo.com/ Accessed 30 Dec 2020
20. Hodrick RJ, Prescott EC (1997) Postwar U.S. business cycles: an empirical investigation. J Money Credit Bank 29(1):1–16. https://doi.org/10.2307/2953682
21. Wang Y, Yang J, Chen Y, De Maeyer P, Li Z, Duan W (2018) Detecting the causal effect of soil moisture on precipitation using convergent cross mapping. Sci Reports 8(1):1–8
22. Dost F, Maier E (2018) E-commerce effects on energy consumption: a multi-year ecosystem-level assessment. J Ind Ecol 22(4):799–812
23. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2018) Graph attention networks. In: International conference on learning representations (ICLR), pp 1–12
24. Liu Q, Ghosh S, Li J, Wong L, Ramamohanarao K (2018) Discovering pan-correlation patterns from time course data sets by efficient mining algorithms. Computing 100(4):421–437
25. Weron R, Zator M (2015) A note on using the Hodrick-Prescott filter in electricity markets. Energy Econ 48:1–6. https://doi.org/10.1016/j.eneco.2014.11.014
26. Das A (2016) Cyclical behavior analysis of indian market using HP filter and spectral techniques. IUP J Appl Finance 22(2):62–78

27. Weerakody PB, Wong KW, Wang G, Ela W (2021) A review of irregular time series data handling with gated recurrent neural networks. Neurocomputing 441:161–178
28. Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. In: International conference on learning representations (ICLR)
29. Deng C, Huang Y, Hasan N, Bao Y (2022) Multi-step-ahead stock price index forecasting using long short-term memory model with multivariate empirical mode decomposition. Inf Sci:297–321
30. Banerjee T, Sinha S, Choudhury P (2022) Long term and short term forecasting of horticultural produce based on the LSTM network model. Appl Intell 52(8):9117–9147

Yuxia Liu is currently pursuing the Ph.D. degree at the Peking University. Her research interests include nonlinear dynamics and control, nonlinear time series prediction, machine learning, etc.

Qi Zhang received the Ph.D. degree in dynamics and control from Peking University, Beijing, China, in 2017. She was a visiting Ph.D. student in Yale Institute of Network Science, Yale University from Sep.2014 to Sep. 2015. She is currently an assistant professor in the School of Information Technology & Management at University of International Business & Economics. Her research interests include machine learning, data mining, social network, etc.

Tianguang Chu received the Ph.D. degree from Tsinghua University, Beijing, China, in 1993. He was a Visiting Research Fellow with The University of Melbourne, in 2001. He is currently a Professor with the College of Engineering, Peking University, Beijing. His research interests include nonlinear dynamics and control, multiagent systems, evolutionary dynamics, and learning systems.