Pattern Recognition 121 (2022) 108119

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Relation-aware dynamic attributed graph attention network for stocks recommendation

Shibo Feng[a], Chen Xu[b], Yu Zuo[c], Guo Chen[d], Fan Lin[a,*], Jianbing XiaHou[a,e]

[a] School of Informatics, Xiamen University, Xiamen, China
[b] China Energy Engineering Group Shaanxi Electric Power Design Institute Co., LTD, Xian, China
[c] Department of Economics, New York University, New York, USA
[d] State Grid Shaanxi Information and Telecommunication Company, LTD, Xian, China
[e] Quanzhou Normal University, Quanzhou, China

ARTICLE INFO

Article history:
Received 30 July 2020
Revised 20 April 2021
Accepted 14 June 2021
Available online 6 July 2021

Keywords:
Financial market
Attributed graph attention network
Correlation coefficient
Chinese stock recommendation

ABSTRACT

The inherent properties of the graph structure of the financial market and the correlation attributes that actually exist in the system inspire us to introduce the concept of the graph to solve the problem of prediction and recommendation in the financial sector. In this paper, we are adhering to the idea of recommending high return ratio stocks and put forward an attributed graph attention network model based on the correlation information, with encoded timing characteristics derived from time series module and global information originating from the stacked graph neural network(GNN) based models, which we called Relation-aware Dynamic Attributed Graph Attention Network (RA-AGAT). On this basis, we have verified the practicality and applicability of the application of graph models in finance. Our innovative structure first captures the local correlation topology information and then introduce a stacked graph neural network structure to recommend Top-N return ratio of stock items. Experiments on the real China A-share market demonstrate that the RA-AGAT architecture is capable of surpassing the previously applicable methods in the prediction and recommendation of stock return ratio.

1. Introduction

With the continuous development of financial markets, a strategy for determining the stock selection strategy with the lowest risk and the highest return ratio has become a focus of attention. Due to the high volatility and time-varying characteristics of the stock market, investors and researchers have historically introduced two methods in the process of forecasting stock trends, along with basic statistical analysis [1] and technological analysis [2]. The former focuses on the market composite index, a standard that uses index income to analyze prediction results in the stock market. The last pays more attention to understanding how investing in technological innovation can have commercial benefits. Researchers use *R&D* indicators to quantitatively analyze the trend of stock prices and qualitatively compare the performance of companies with different technical capabilities in the financial market by analyzing the strength of a company's technical patent potential and the momentum change of the stocks historical data. However, these methods of posterior analysis through external conditions reveals its instability and hysteresis today.

As we all know, traditional statistical methods depend on three aspects to analyze financial market, such as the correlation relationship [3], the trend of the stock itself and characteristic relationship of the influencing factors. However, the support of statistical analysis lies in the selection of artificial feature information and empirical selection with greater uncertainty, which greatly reduces the possibility of accurately predicting the future trends.

The complex spatial concepts in financial markets have made the concept of time series studied by most of the previous work [4,5]. Traditional machine learning methods such as Autoregressive Model(AM) [6], Kalman Filters [7], and Threshold Models(TM) [8] all treat the past trend of the stock as a random process and apply linear methods to make the historical stock index fit it. However, stock market is affected by complex factors, the stochastic process assumed in the traditional method may not be compatible with the reality, meaning that these models have only theoretical practicability. Gradually, time series models based on recurrent neural networks(RNN) [9] attracts most researcher's pursuit. How-

* Corresponding author.
E-mail address: iamafan@xmu.edu.cn (F. Lin).

https://doi.org/10.1016/j.patcog.2021.108119

ever, the spatial dimension behind the concept of time series in the financial market largely affects the accuracy of the time series models.

Naturally, relevant financial networks of different standards plus some types of supplements in form of financial events and company correlations [10] have opened up the exploration of spatial concepts in financial markets. The vast majority of researchers study the Chinese stock market through daily and intra-day scale data [11–13], market indices [14,15], and the degree of influence of the foreign exchange market [16]. These studies indicate that there are complex dependencies between economic activities, price changes, market indexes, and impact factors in reality market. The dependence between these external market factors is essentially based on the characteristic factors of the stock itself. Therefore, the linear and non-linear correlations of factor characteristics have become the focus point.

Aiming at the characteristics of non-Euclidean data in the financial market, most researchers introduce graph-embedded representation methods (Representation Learning on Graph [17] such as DeepWalk [18], LINE [19], and node2vec [20]) to learn the distributed representation between corresponding companies that corresponds to the changing trend of the stocks behind. However, these methods cannot synchronize the aggregation of the relationship and characteristic factors in the stock network. The emergence of the graph neural network(GNN) allows the complex relationships and stock feature factors in the financial market to be expressed simultaneously.

In this article, we decompose the financial market through the two dimensions of time and space. Based on the characteristics of the former, we embed the attention mechanism into the time series model to distinguish the time characteristics of different periods at multiple scales. Observed the latter has the complex correlation characteristics and topological structure attributes, we first establish a complex network of financial markets based on non-linear correlation relationships, on this basis, we introduce a model based on graph neural network to transmit the topological structure of complex financial networks and also integrates the factor characteristics between stocks simultaneously. Generally, Our RA-AGAT introduces the correlation feature as the topological structure and the temporal feature as the content structure to perform the fusion representation learning of timing and correlation simultaneously which enables time series models and correlation models that exist in the financial market to be fused and learned in an end-to-end form.

The remainder of this paper is organized as follows. We will briefly introduce some related work that exists in the financial market in Section 2. We introduce the basic structural knowledge used in our article and the detailed process of our proposed method in Section 3. It is followed by an experimental evaluation in Section 4, which describes our experiment parameters and evaluation indicators, and explores the performance of our RA-AGAT model in stock recommendation task. Finally, we summarize the paper and offer our own outlook for future work in the financial field in Section 5. The three key contributions of the paper are summarized as:

(1) Replacing the task of predicting stock prices and trends in the traditional financial field with a new way of recommending the return ratio of stocks.
(2) Introducing the graph convolutional network as a guide to the graph attention network for information aggregation, whose attention mechanism is expanded from node features to topology information to make the stock correlation integrated into the message passing of the stock features.
(3) Applying the factor strategy mechanism into the complex stocks network to select the important factor components.

2. Related work

2.1. Time series model in financial markets

From the beginning to the end, the time series model is a powerful model group in the financial field. The traditional time series model relies on econometric theory, e.g., The autoregressive model(AR) [6], autoregressive moving average model(ARMA) [21], autoregressive integrated moving average model(ARIMA) [22] and their extensions. One of the thorny issues in their solution is relying on the linear equations and introducing a series of predefined linear mathematical models to fit univariate time series in stocks historical data which ignores the high complexity and relevance characteristics of financial markets. Due to the fact that non-linear correlation actually exists in financial markets, some typical non-linear models such as autoregressive conditional heteroskedasticity model (ARCH) [23], generalized autoregressive conditional heteroskedasticity model(GARCH) [24], recursive neural network(RNN) [25], convolutional neural network(CNN)-based models [5,26,27], Long Short-Term Memory(LSTM) [28] and its variations. In view of the long-period and high-frequency time-varying characteristics of stocks, models based on long and short-period time series have received more attention. Thomas Fischer [28] introduced the LSTM model to the stock prediction in the financial field for the first time, and explained its applicability in stock prediction based on the long-period sequence memory and distinguishing characteristics of the LSTM model itself. Li et al. [29] embedded the Naive Bayes model into the LSTM network to construct a hybrid neural network model for predicting stock trends. The Naive Bayes model serves as a framework for extracting external investor information and market factors to increase the effectiveness of the LSTM network. Althelaya et al. [30] introduced bidirectional and stacked LSTM network structures and strengthened the model's capabilities by analyzing bidirectional data change trends and stacking multilayer LSTM networks.

However, it is undoubtedly insufficient to propose the concept of time dimension alone to describe the financial market. The spatial correlation and transitivity in the complex financial network should be taken into consideration.

2.2. Relevance analysis of financial markets

Many previous research works have proposed numerous analytical methods for the description of the spatial concept of financial markets. One of the most eye-catching being is correlation-based network portrayal. Jiang and Zhou [31], Li et al. [32], Song et al. [33], Zhang et al. [34] Existing literature research work provides a variety of criteria to construct a filtered financial network to study its internal structure. The correlation based financial complex networks tend to be divided into linear and non-linear according to the way they measure correlation. Wang et al. [35] constructed the minimum spanning tree of the stock market according to the pearson coefficient and determined the stock trend not only through the structure and distribution information of the minimum spanning tree but also the evolution of the world stock markets. Guo [36] calculated the mutual information(MI) of returns between different stocks, constructed the minimum spanning tree of the mutual information of returns to analyzed the topological structure characteristics such as degree, power-law distribution, and complexity, only to conclude that the structure of the mutual information network is more suitable for real financial markets. Wang et al. [37] concluded that the DCCA coefficient not only has the nonlinear property and scale resolution, but the DCCA coefficient at different scalars could help investors make effective risk management and optimal portfolio selection.

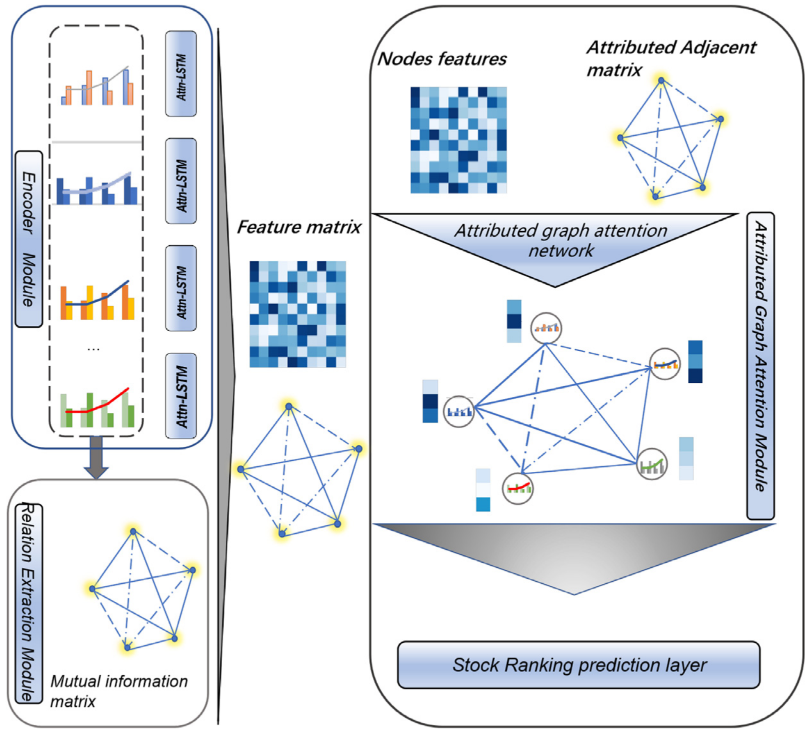

Fig. 1. The upper-left diagram represents the time encoder module, where "$Attn-LSTM$" represents the LSTM module based on the attention mechanism. The lower -left and the right module represent the correlation based financial network and the recommendation module based on the graph neural network, respectively.

Due to the differences in the spatial structure embodied in different markets, we will compare the measurement standards of linearity and nonlinearity in the following experiments. The biggest difference between our work and the previously established correlation-based financial network is that we not only pay attention to the attributes of the model itself, but also introduce the model as the basis for the integration and transmission of stock characteristics.

2.3. GNN-based model in financial markets

The non-European characteristics of the financial market determine that the structure of the model established in the minimum spanning tree in the previous part cannot fully reflect the diversity of spatial concepts. Traditional machine learning methods in graph filed have difficulty in aggregating network structures such as heterogeneous relationships, complex networks and edge information. Most of the traditional graph analysis algorithms involve artificial feature processing and statistical analysis of graph structures. Matsunaga et al. [38] introduced the GNN into the stock market, utilizing financial events and enterprise-related knowledge graphs to establish a relationship graph, and proposed a dynamic graph concept to predict stock trends. Chen et al. [39] constructed an enterprise relationship graph based on the facts of corporate financial investment released by financial media and completed the task of predicting the stock market through a graph CNN. Feng et al. [40] first turned the task of stock prediction into a novel ranking system, collecting the relevant data of events, relationships, and policy forms between companies as prior knowledge to embedded in a GNN-based model (Fig. 1).

However, the establishment of the above models largely rely on the cost-intensive relationship acquisition strategy. We go in the opposite direction and combine the topology and feature representation of the model through a stacked graph neural network architecture to complete the recommendation task relying only on stock features.

Table 1
Notations and descriptions of terms and symbols.

$p(.)$	Probability distribution
v	The learnable parameter
$p(x_i, y_j)$	Joint probability distribution of random variable
$P(.\|.)$	Conditional probability distribution
$[.;.]$	Concatenation of two variables
$b_{(i,f,o,c)}$	Bias of each unit
α_t	Attention coefficient at time t in attention based LSTM
$H = [h_1, h_2, \ldots, h_N] \in \mathcal{R}^{N*F'}$	Embedding representation in attention based LSTM at time T, F' is the output size
$\widetilde{D}, \widetilde{A}$	$\widetilde{D}$ is the degree of $\widetilde{A}$, ($\widetilde{A} = A + I_N$), A is the adjacency matrix
$Z \in \mathcal{R}^{N*F'}, \Theta \in \mathcal{R}^{F*F'}$	Z is the embedding state in GCN network, F' is the output dimension
$W \in \mathcal{R}^{N*1}, \Theta_{att} \in \mathcal{R}^{F*1}$	W is the attention coefficient, Θ_{att} is the learnable parameter
$\vec{a}^T \in \mathcal{R}^{2*F'}, \psi \in \mathcal{R}^{F*F'}$	$\vec{a}^T$ is a weighted vector, ψ parameter in graph attention mechanism
σ_{GCN}, σ	σ_{GCN} is the tanh function,σ is sigmoid or softmax method

3. Methodology

In this section, we will elaborate on how we model the time and space concepts in the financial market, and focus on proposing the correlation basis and stacked graph neural network structure in the space concept. Specifically, we combine the ideas of dynamically analyzing the historical trends of stocks, capturing relevant information of stock returns. On this basis, we established an attributed graph attention network model to recommend stocks of high return ratio. All of the notations and corresponding explanations related in this chapter are listed in Table 1.

3.1. Technical indicators

The characteristics of stock factors have been shown by most of the previous research literature to be the basis of financial mar-

Table 2
with $N = [5, 10, 15]$, $n^* = [5, 10]$, Cl= Closing Price, Op = Opening Price, Hg = High Price, Lw = Low Price, $n = [14]$ is the time range of observation, *period_sinceHighestHigh* and *period_sinceLowestLow* represent the number of days after the highest price and the lowest price in the observation time period, respectively.

Moving Average (MA)	$MA_{t,N} = \frac{1}{N}\sum_{i=t-n+1}^{t} Cl_i$
Exponential Moving Average (EMA)	$EMA = (Cl(t) - EMA(t-1)) * mult + EMA(t-1)$ $\Delta =$ timeperiod EMA $mult = \frac{2}{\Delta+1}$
Momentum_n^*(MOM_ n^*)	$MOM_{-n^*} = Cl_t - Cl_{t-n}$
Relative Strength Index (RSI_n^*)	$RSI\_n^* = 100 - \left(\frac{100}{1+RS}\right)$ $RS = \frac{\text{Average of n days'up close}}{\text{Average of n days'down close}}$
MACD	$MACD = 12EMA - 16EMA$
Average Price (AP)	$AP = Op + Hg + Lw + Cl/4$
Typical Price (TP)	$TP = Hg + Lw + Cl/3$
Weighted Close Price (WCP)	$WCP = Hg + Lw + Cl * 2/4$
AROON	$AROON_{UP} = n-$period_sinceHighestHigh$/n$ $AROON_{down} = n-$period_sinceLowestLow$/n$
On Balance Volume (OBV)	$IFX_t > X_{t-1} \rightarrow OBV_t = OBV_{t-1} + V_t$ $IFX_t < X_{t-1} \rightarrow OBV_t = OBV_{t-1} - V_t$ $IFX_t = X_{t-1} \rightarrow OBV_t = OBV_{t-1}$

ket forecasts. We introduce several technical indicators that are highly correlated to stock returns and utilize stock price changes to capture and evaluate stock trends. Table 2 describes the technical indicators introduced in this paper and explains the correlation between these indicators and stock returns in the following statement.

In view of our task of recommending high-yield stocks, we solve the uncertainty of characteristic factors in the financial market by adding different types of technical indicators, e.g., the price factor(AP, TP, WCP, $MACD$) which is a strong correlation factor that cannot be ignored for the return ratio recommendation task, momentum factor(MOM_n^*, RSI_n^*) as a key factor that describes market fluctuations and trends to reflects the changes and trends of return ratio, trading volume(OBV) is the variable displayed by the stock return result, focused by pattern extraction module.

3.2. Correlation analysis in the stock market

3.2.1. Normalized mutual information based on logarithmic rate of return

Mutual Information originates from Shannons entropy theory, which is a broad measure of correlation defined in information theory. The entropy of any discrete random variable X is defined as

$$H(X) = -\sum_i p(x_i)\log_2 p(x_i) \tag{1}$$

The joint entropy of two random variable X, Y is defined as

$$H(X,Y) = -\sum_i\sum_j p(x_i, y_j)\log_2 p(x_i, y_j) \tag{2}$$

Mutual Information is captured by combining the above two formulas

$$\begin{aligned} I(X,Y) &= H(X) + H(Y) - H(X,Y) \\ I(X,Y) &= H(X) - H(X \mid Y) \end{aligned} \tag{3}$$

where $H(X \mid Y)$ represents the conditional entropy of X under the conditional Y, and is defined as

$$H(X \mid Y) = -\sum_i\sum_i p(x_i, y_j)\log_2 p(x_i \mid y_j) \tag{4}$$

In general, Mutual Information I(X, Y) describes the degree of information sharing between random variable X, Y. When I(X, Y) = 0, it indicates that X and Y are independent. Normalized Mutual Information(NMI) is defined as

$$\mathrm{NMI}(X,Y) = \frac{2I(X,Y)}{H(X) + H(Y)} \tag{5}$$

In our prediction task, we create a recommendation list based on the stock return ratio and introduce the concept of mutual information into stock return. This clearly expresses that we introduce the mutual information between stock returns as the modeling basis of our predictive model.

We denote $C_{i,t}$ and $R_{i,t}$ as the closing price of stock i in tth trading day and log-return of stock i in tth trading day, respectively. $R_{i,t}$ is defined as $R_{i,t} = \ln\frac{C_{i,t}}{C_{i,t-1}}$, $(t = 2, \ldots, T, i = 1, \ldots, N)$ The above results can be divided into a defined interval which is described as $\left[\min R_{i,t}, \max R_{i,t}\right]$ and then divided equally into k sub-intervals. We replace "possibility" $p_{i,n}$ with the frequency at which the log-return of stock i falls into the sub-interval n

$$p_{i,n} \approx \frac{f_{i,n}}{T}(i = 1, \ldots, N, n = 1, 2, \ldots, k) \tag{6}$$

The entropy of return of stock i is defined by

$$H(X_i) = -\sum_{n=1}^{k} p_{i,n}\log_2 p_{i,n} \tag{7}$$

For joint entropy $H(X_i, X_j)$ in the calculation of Mutual Information calculation, we divide the square of log-return $\left[\min R_{i,t}, \max R_{i,t}\right] \times \left[\min R_{j,t}, \max R_{j,t}\right]$ into $k \times k$ shape. The joint probability distribution of stock i and stock j can be replaced by the frequency of joint log-returns falling into (m, n)

$$p_{i,j,m,n} \approx \frac{f_{i,j,m,n}}{T}(i, j = 1, \ldots, N, m, n = 1, \ldots, k) \tag{8}$$

$$H(X_i, X_j) = -\sum_{m=1}^{k}\sum_{n=1}^{k} p_{i,j,m,n}\log_2 p_{i,j,m,n} \tag{9}$$

In summary, the Mutual Information between stocks i and j can be expressed as

$$I(X_i, Y_i) = H(X_i) + H(Y_i) - H(X_i, Y_i) \tag{10}$$

Normalized Mutual Information of stock i and j is defined by

$$\mathrm{NMI}\,(X_i, Y_i) = \frac{2I(X_i, Y_i)}{H(X_i) + H(Y_i)} \tag{11}$$

3.2.2. Detrended cross-correlation analysis (DCCA) coefficient

Because of the highly nonlinear, non-stationary and long-range correlations within the stock market, the cross-correlation between stock markets has become the focus of our research. Consider the returns from two stocks i, j in the time range T as $r_i(t), r_j(t)$ respectively, where $t = 1, 2, \ldots, T$. We will describe in detail how to calculate the cross-correlation between stocks through the DCCA coefficient [37]

Step 1. Determine the envelope and generate a new form of time series relationship

$$R_i(t) = \sum_{k=1}^{t}(r_i(k) - \langle r_i\rangle), R_j(t) = \sum_{k=1}^{t}\left(r_j(k) - \langle r_j\rangle\right), t = 1, 2, \ldots, T \tag{12}$$

Step 2. Split $R_i(t)$ and $R_j(t)$ into $T_s = \mathrm{int}(T/s)$ non-overlapping segments and each has the same length s. Because the time length T of the stocks does not necessarily have a divisible relationship with the time scale s, in most cases, there will be a residue at the end of the envelope. To not ignore the recording data of this part, repeat the above process at the end of the remaining part. Therefore, we obtain $2T_s$ segments. In this paper, we choose $30 \le s \le T/10$, where s is set to 90.

Step 3. Calculate the least squares fit of the data for each segment in the $2T_s$ range to get obtain the local trend. Calculate the variance of the detrending time series of each $2T_s$ segment by averaging all data points i in the νth part.

$$F^2(s,\nu) = \frac{1}{s}\sum_{k=1}^{s}\left(R_i^{(\nu-1)s+k}(k) - \tilde{R}_i^{\nu}(k)\right)\left(R_j^{(\nu-1)s+k}(k) - \tilde{R}_j^{\nu}(k)\right) \quad (13)$$

where v represents each segment $= [1, 2, \ldots, T_s]$ and

$$F^2(s,\nu) = \frac{1}{s}\sum_{k=1}^{T}\left(R_i^{T-(\nu-T_s)s+k}(k) - \tilde{R}_i^{\nu}(k)\right)\left(R_j^{T-(\nu-T_s)s+k}(k) - \tilde{R}_j^{\nu}(k)\right) \quad (14)$$

where each segment v, v $= [T_s+1, T_s+2, \ldots, 2T_s]$, $\tilde{R}_i^{\nu}(k)$ and $\tilde{R}_j^{\nu}(k)$ both represent the fitting polynomial.

Step 4. Average all of the segment elements to capture the detrending covariance function.

$$F^2_{\mathrm{DCCA}}(s) = \frac{1}{2T_s}\sum_{\nu=1}^{2T_s} F^2(s,\nu) \quad (15)$$

When R_i is identical to R_j, $F^2_{\mathrm{DCCA}}(s)$ evolves into a detrending variance function $F_{\mathrm{DFA}}(s)$

$$F_{\mathrm{DFA}}(s) = \left\{\frac{1}{2T_s}\sum_{\nu=1}^{2T_s} F^2(s,\nu)\right\}^{1/2} \quad (16)$$

Step 5. The DCCA coefficient between the returns from stock i, j is calculated by the ratio of $F^2_{\mathrm{DCCA}}(s)$ and $F_{\mathrm{DFA}}(s)$ divided.

$$\rho_{ij}(s) = \frac{F^2_{\mathrm{DCCA}}(s)}{F_{\mathrm{DFA}\{r_i(t)\}}(s)F_{\mathrm{DFA}\{r_j(t)\}}(s)}, \quad (17)$$

where the range of $\rho_{ij}(s)$ is [−1,1], If the returns of the two stocks are well-fitted in the time series, or they are in a state of opposition to each other, then the value of $\rho_{ij}(s)$ is the interval endpoint 1 or −1. If the $\rho_{ij}(s)$ is equal to 0, this means that there are no cross-correlations between the returns from the two stocks. We clearly see that $\rho_{ij}(s)$ can quantify the dynamic correlation characteristics between different stocks, which is suitable for the characteristics of the long-term correlation of financial markets.

3.3. Attention-based long short term memory (Attn LSTM)

The reality of financial data has long-term and short-term information relevance and different forms of expression in different time spans. A single correlation index model cannot describe the changing characteristics of the stock in the time dimension, so we introduce the Long Short Term Memory(LSTM) model based on the attention mechanism to selectively superimpose feature information at different moments.

Actually, the financial data itself has the characteristics of large time span, high time variability and high volatility. The naive LSTM model has the problem that the length of the input data is too long which makes the model cannot learn an accurate vector representation at a point in time, resulting in lower accuracy of downstream tasks. The introduction of the attention mechanism allows the model to determine what needs to be learned through the size of the attention score in data with a large time span, and avoid learning too much redundant information and causing undesirable results.

Given stock data at T times $x_1, x_2, x_3, \ldots, x_T$, the attention mechanism calculation process at time T is

$$\begin{aligned} h_T &= \delta\left(W[h_{T-1}; x_t] + b_{(i,f,o,c)}\right) \\ h'_t &= \sum_{t=T-s}^{T} \alpha_t h_t \\ \alpha_t &= \frac{\exp\left(h^T t\nu\right)}{\sum_{k=T-s}^{T}\exp\left(h^T k\nu\right)} \end{aligned} \quad (18)$$

Since the attention-based model can update the corresponding variables according to the correlation. In the encode module of stock time series concept, there exists strong correlation characteristics between stock features that makes the attention mechanism better map the characteristics to the low space (Fig. 2).

3.4. Attributed graph attention network (AGAT)

We introduce variants of GNN structures to deal with financial stock networks. The notations appearing in the section are listed in Table 1. To fully consider the complex relationships and stock characteristics in the stock market, we introduce a stacked GNN-based model to analyze the internal connections behind the finance. We divide the aggregation process of the attributed graph attention network model into two steps. Firstly, aggregating the topology and feature information of different stocks through a graph convolutional network, followed by exploiting this score as a guide for the self-attention mechanism in the following graph attention network to integrates topology and content information. Fig. 3 describes the detailed structure.

3.4.1. Self-attention graph convolutional network

The fundamental challenge in graph attention network is incorporating the self-attention mechanism into topology information. Inspired by Lee et al. [41], who proposed a new type of attention calculation method that applied to graph neural network based models. Exploiting the principle of aggregating topological structure and adjacent features in the graph convolutional neural network, we deformed it in a new fashion.

$$Z = \widetilde{D}^{-1/2} A \widetilde{D}^{-1/2} X \Theta \quad (19)$$

where Z describes the convolved signal matrix that aggregates the topology information and feature signals. The self-attention mechanism, often referred to as the intra-attention mechanism, uses the node features as the criteria in the GNN. Suppose we embed the state of the stocks historical data into the GCN model, named $H = [h_1, h_2, \ldots, h_N]$. We then introduce the self-attention mechanism to the convolution operation

$$W = \sigma_{GCN}\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}} H \Theta_{att}\right) \quad (20)$$

A generic illustration of the proposed GCN structure for attention calculation is shown in Fig. 4. Note that we changed the dimensional information of the learnable parameter Θ_{att} to obtain the attention score between different nodes.

3.4.2. Attributed graph attention network (AGAT)

For the graph attention network model, the attention mechanism performs a weighted summation of information based on the characteristics of the nodes in the graph structure. However, the graph attention network ignores an important element: the edge information in the stock graph structure, which describes the correlation between the different stocks. To overcome this problem, we use the result of the convolution operation as a guide for

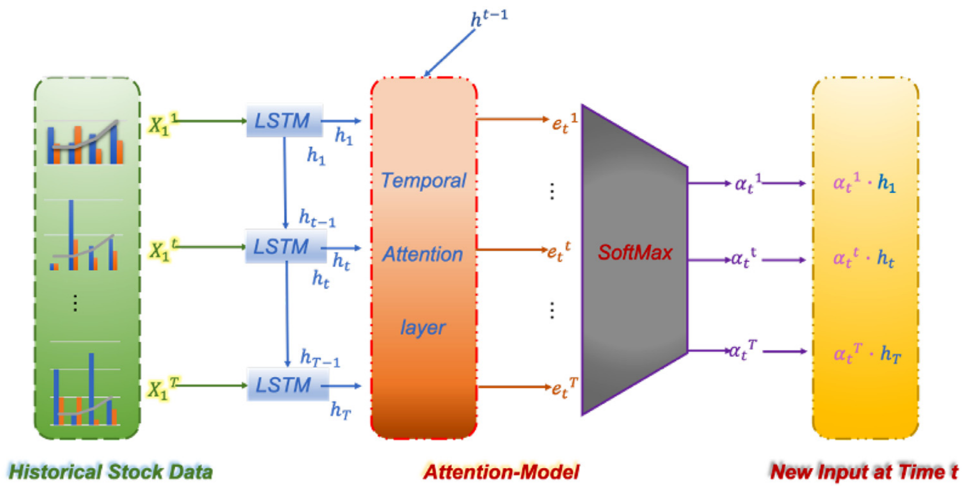

Fig. 2. Temporal attention LSTM network structure. For X_i in timestamp t, Temporal Attention the temporal attention layer calculates the attention score based on the hidden state h_{t-1} of X_i, and the updated state of h_t' is obtained by attention coefficient $a_t * h_t$.

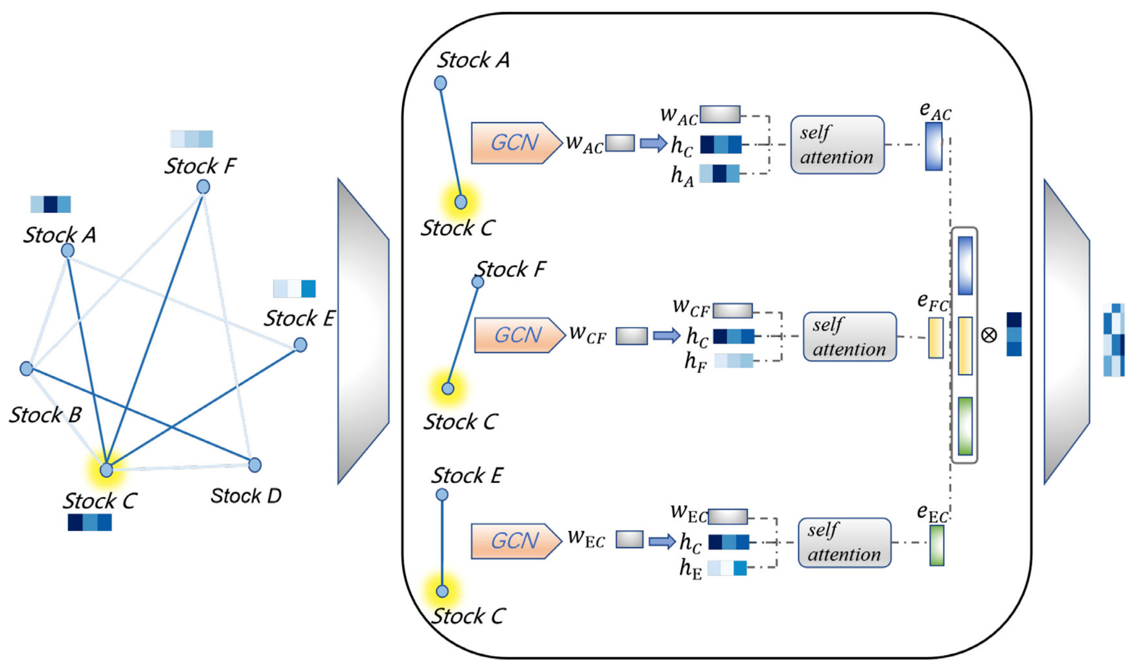

Fig. 3. Illustration of the attributed based graph attention network model, the "self-attention" is represented as the attention mechanism in the graph attention network.

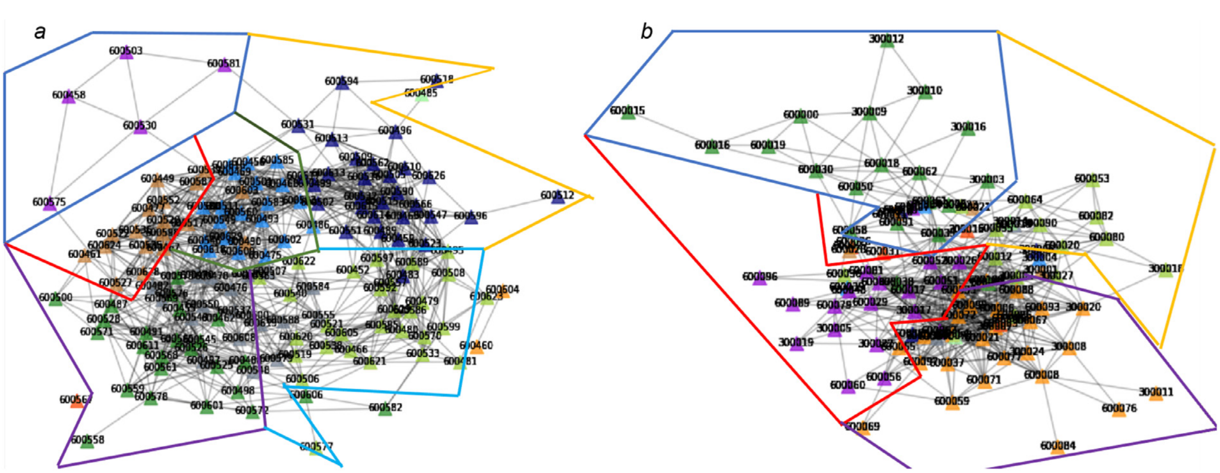

Fig. 4. a, b represent the complex network structure based on Normalized Mutual Information, DCCA coefficient respectively.

our graph attention network to selectively aggregate neighbor features. Specifically, the attention coefficient between stocks i, j in our AGAT model is defined as

$$e_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{a}^T \psi\left[W_i \vec{h}_i \| W_j \vec{h}_j\right]\right)\right)}{\sum_{k \in N_i} \exp\left(\text{LeakyReLU}\left(\vec{a}^T \psi\left[W_i \vec{h}_i \| W_k \vec{h}_k\right]\right)\right)} \tag{21}$$

Based on the attention coefficient, we update the features of the global nodes in the graph structure. In view of we exploit the graph convolutional network and graph attention network structures as a hybrid model to update the node parameters, the final representation contains the weighted summation of different node features and the internal connection of the topology relationship. In addition, the multi-head attention strategy is widely proposed in methods based on the attention network. Considering the multi-head strategy includes the correlation of features in different dimensions is suitable for the characteristics of multi-dimensional correlation in complex stock networks, we try to exploit it to improve the accuracy of our results.

$$\begin{aligned} h_i' &= \sigma\left(\sum_{j \in N_i} e_{ij} \psi h_j\right) \\ h_i' &=_{m=1}^{M} \| \sigma\left(\sum_{j \in N_i} e_{ij}^m \psi^m h_j\right) \end{aligned} \tag{22}$$

3.5. Loss function in the stock prediction layer

Because we converted the prediction task to a ranking problem, we propose an objective function containing point-wise regression loss, ranking loss, and loss of correlation for the optimization in our structure.

$$\begin{aligned} \text{Loss}(\hat{p} - p) = \|\hat{p} - p\|^2 + \alpha \sum_{i=0}^{N} \sum_{j=0}^{N} \max\left(0, -\left(\hat{p}_i - p_i\right)\left(\hat{p}_j - p_j\right)\right) \\ + \beta \|\Theta_{att}\|^2 \end{aligned} \tag{23}$$

where $\hat{p} = \left[\hat{p}_1^{t+1}, \hat{p}_2^{t+1}, \ldots, \hat{p}_N^{t+1}\right]$ and p describes the predicted ranking score and the target ranking result respectively. We additional proposed the calculation of the loss of the correlation score between different stock pairs to point out the importance of the correlation relationship in our module.

4. Experiment and comparisons

In this section, we investigate the advantages and limitations of our RA-AGAT model. In order to significantly improve the practicability of our model in the real financial field, we compared our model with the methods applied in the financial field among the recent years and applied it to the Shanghai Stock Exchange (SSE).

4.1. Relevance analysis in stock market

To demonstrate the effectiveness of our relevance analysis strategy in an actual stock market, we conducted experiments on the Chinese stock market that involved 738 stocks, 2223 trading days, and five basic features (i.e., opening price, closing price, highest price, lowest price, and trading volume). For relevance analysis, three alternative algorithms were selected as baseline methods, and we briefly describe them below (Table 3). **Data preprocessing** In view of the high complexity of the data in the Chinese stock market, we have eliminated specific anomalies and missing values through several specific methods.(1) Remove stocks with a listing time of less than six months, delete the PT or ST stocks on the stock picking day and remove stocks that were suspended from trading on the day of stock selection.(2) Based on the days of the Shanghai and Shenzhen 300 index data, if individual stocks lack data for a certain day, the closing price of the most recent day of that day is used as the opening price, highest price, lowest price, and closing price of that day, and the trading volume is set to 0. Use the day of the index data as the standard, if the index data does not exist on a certain day, the data of the individual stocks of the day will be removed.(3) We performed a five day sliding slice of individual stocks and index. Specifically, the opening price, highest price, lowest price, closing price, and trading volume of each stock are used as a slice matrix, and the sliding interval is a day. According to the data preprocessing, we selected 738 stocks in the SSE market. The detailed descriptions in our dataset are shown in Table 4.

Baseline Methods We compared our correlation analysis method with three state-of-the-art methods for obtaining the correlation matrix: Pearson Coefficient [42]; Partial Correlation Coefficient; and DCCA coefficient [37]

Measurement To quantify the accuracy of different methods for dividing the stock network, we conducted a qualitative analysis of the experimental results based on five measurement indicators. Modularity (Q) [43] is used to measure whether the division of a community is a better result. A relatively good result has a higher similarity of nodes within the community and a lower similarity of nodes outside the community. Clustering coefficient (C) is used to reflect the degree of aggregation of nodes within graph. The higher aggregation degree means the higher of the integration of the nodes in the cluster that indicates the integration of feature change trends between stocks. Transitivity (T) quantitatively analyzes the connectivity characteristics between nodes in a complex network. Due to the large number of nonlinear associations in the financial market, it is suitable for it to better reflects the ability of different methods to retain relevant information. Performance (P) [43] is a "correctness" measurement indicator to calculate the number of correctly divided node pairs, i.e., there are edge connections between nodes belonging to the same cluster, and there is no direct connection between nodes of different divisions. Coverage (V) is the ratio of the number of intra-community edges to the total number of edges. It is an effective structure where the clusters are disconnected from each other. The calculation details of all evaluation indicators are shown in Table 5.

Discussion To verify the applicability of different correlation extraction methods to the Chinese SSE, we randomly selected 100 stocks in the dataset, which can be seen in Figs. 4 and 5. We used the Planar Maximally Filtered Graph (PMFG), a tool for filtering information in complex systems, to extract skeletons from the correlation matrix in order to express the hierarchical structure of the stock market. The results of the five measurement indicators are presented in Table 6. We observe that the performance of the complex network based on mutual information exceeds that of the other methods. At the same time, the DCCA coefficient expressed as a non-linear characteristic is similar to the Pearson coefficient with a linear characteristic. The reason for our thinking is that the time period of the stocks we selected is 10 years(2009–2019), it will be difficult for DCCA to determine the Scope of observation and find general rules. According to the above phenomenon, we also used the stocks selected in the above experiment to analyze the correlation between the Pearson coefficient and the DCCA coefficient and the experimental results are shown in Table 8. We found that the above two have certain similarities in the properties of Mean, Maximum, Minimum, standard deviation, Skewness and Kurtosis, which expresses that the DCCA coefficient may not always reflect the characteristics of the nonlinear relationship. The partial correlation coefficient, which considers the linear relationship between two variables, is often affected by other variables, and the simple correlation coefficient cannot fully reflect the lin-

Table 3
Notations and detailed descriptions in our experiment section.

L_i^t	Normalized closing price of stock i at day t
rr_i	The return ratio of stock i $rr_i = \frac{L_i^{t+1} - L_i^t}{L_i^t}$
L_{target}	Target ranking list based on real return ratio
N	Number of stocks
k	Length of ranking list
$rank_i$	Position of stock i in the ranking list
Ω_k	Items in the prediction list
M	Total number of edges in the graph G
N	Number of nodes in graph G
P_{ij}	Expected number of edges between vertices i and j in the null model.
$\delta(C_i, C_j)$	Yield one if vertices i and j are in the same community($C_i = C_j$); otherwise, zero
$\lambda_G(v)$	Number of triangles in graph G
$\tau_G(v)$	Number of triples in graph G
P_i	Partition i in graph G
e^i_{intra}	Number of intra edges in partition i
e	Total number of edges in graph G

Table 4
Statistics of the historical data.

Market	Stocks	Training Days	Validation Days	Testing Days
SSE	738	1714 days 2009-01-01 2017-02-18	236 days 2017-02-19 2019-08-14	273 days 2018-08-15 2019-07-12

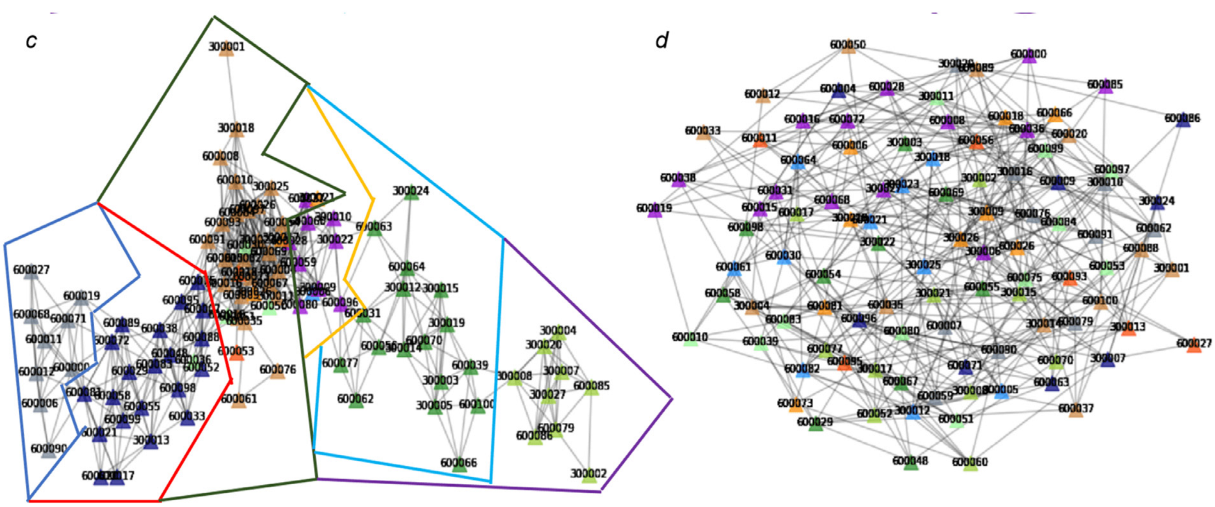

Fig. 5. c, d represent the complex network structure based on the Pearson Coefficient, and Partial Correlation Coefficient respectively.

Table 5
Notations of the measurements in a complex stock market.

Modularity (Q)	$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(C_i, C_j)$
Clustering coefficient (C)	$\bar{C} = \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda_G(v)}{\tau_G(v)}$
Transitivity (T)	$T = 3 * \frac{\lambda_G(v)}{\tau_G(v)}$
Performance (P)	$P(\mathcal{P}) = \frac{\lvert\{(i,j)\in E, C_i=C_j\}\rvert + \lvert\{(i,j)\notin E, C_i\neq C_j\}\rvert}{\frac{n(n-1)}{2}}$ $P = \frac{1}{n}\sum_{i=1}^{n}P(\mathcal{P}_i)$
Coverage (V)	$V = \frac{1}{n}\sum_{i=1}^{n}\frac{e^i\text{intra}}{e}$

ear relationship between the two. However, when calculating the correlation of the log returns of any two stocks, if the log returns of the remaining stocks are used, that may violate the rule that the log returns of the stock market cannot be used to calculate the effect of the log returns on any two stocks. We argue that the reason for the strong results of the mutual information method is that it is the long time span of our dataset that enables the stock data to retain economic homogeneity.

4.2. Recommendations in the stock market

Baseline methods We compared our proposed method RA-AGAT with six methods for stock return ratio ranking. These methods are autoregressive integrated moving average model (*ARIMA*) [44]; state frequency memory (*SFM*); logistics regression (*LR*); attention-based LSTM (*Attn-LSTM*) [45]; attention-based LSTM-

Table 6
Results of the measurements from different relation extraction methods.

ModelMeasurement	Q	C	T	P	V
Partial Correlation Coefficient	0.1681	0.0477	0.0471	0.0152	0.1312
DCCA Coefficient	0.5525	0.6264	0.6410	0.5228	0.6312
Pearson Coefficient	0.5751	0.6638	0.6848	0.5413	0.6420
Mutual Information	**0.6695**	**0.7123**	**0.7010**	**0.6203**	**0.6948**

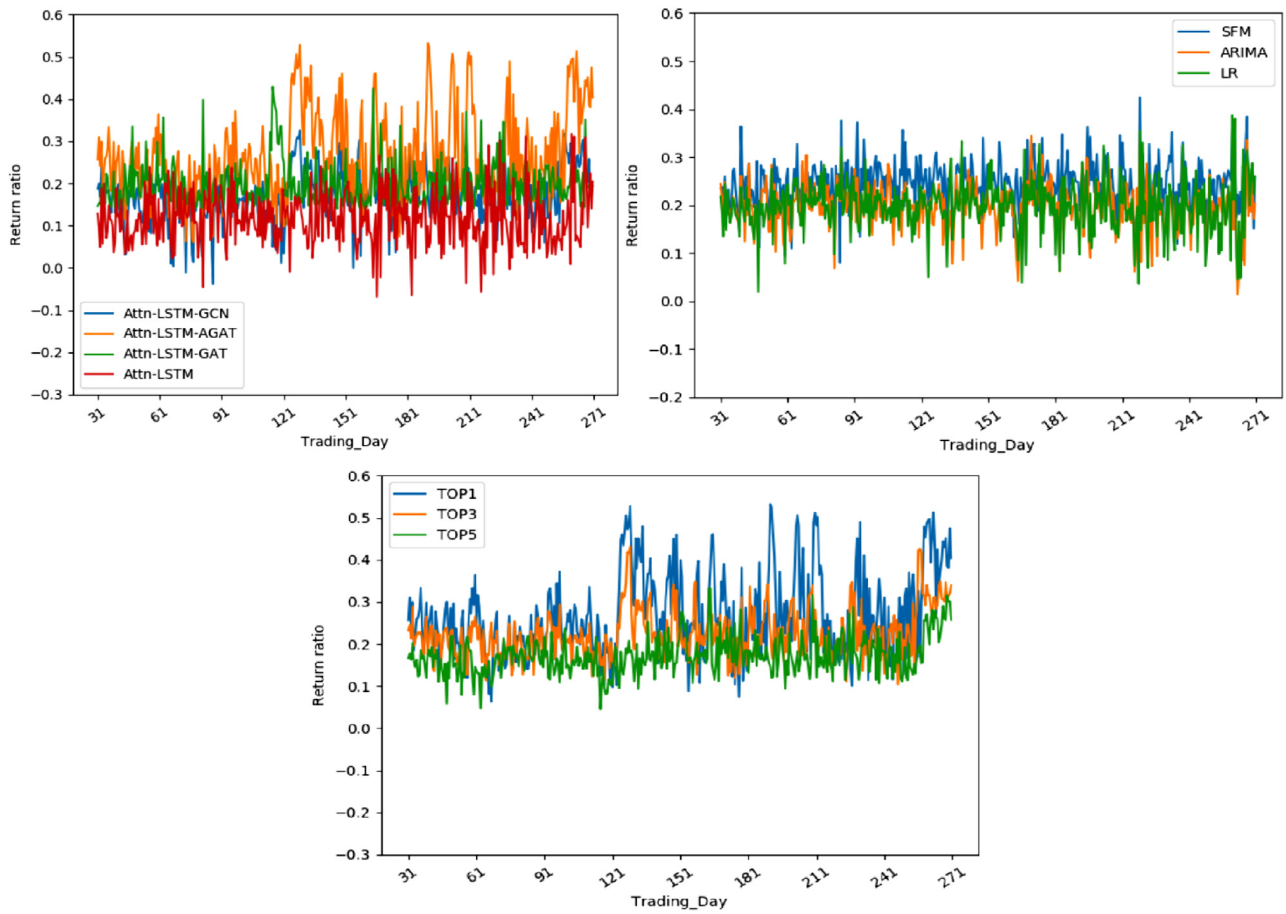

Fig. 6. The highest return ratio curve obtained by different analysis functions in the testing date. The bottom part is the result of our back-testing strategy in the Ra-AGAT model, where the Top N curve represents the weighted average result of the top n stocks with return ratio on each day of the testing date.

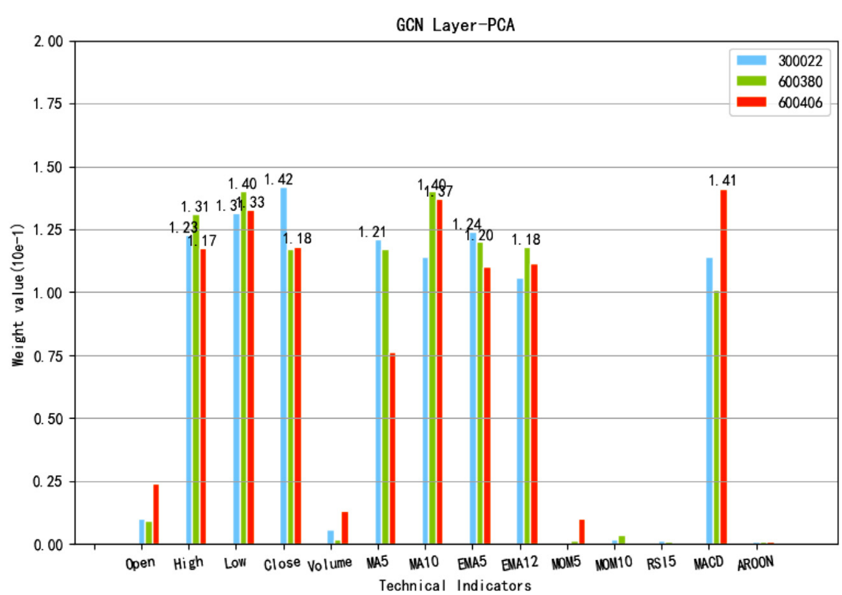

Fig. 7. The weight coefficient of each technical indicator in the first layer of our attribute graph attention network.

graph attention network (*Attn-LSTM-GAT*); and attention-based LSTM graph convolutional network (*Attn-LSTM-GCN*). within the first three methods, each method describes only the change trend of data in a specific time series, which expresses these methods only aggregate the characteristics displayed on the time axis. However, due to the fact that the trend and results of stocks in the financial market are greatly affected by the complexity relationship, only the model based on the analysis of the time dimension appears to have a large deviation. The latter three models, we combine the time series model with the graph neural network based

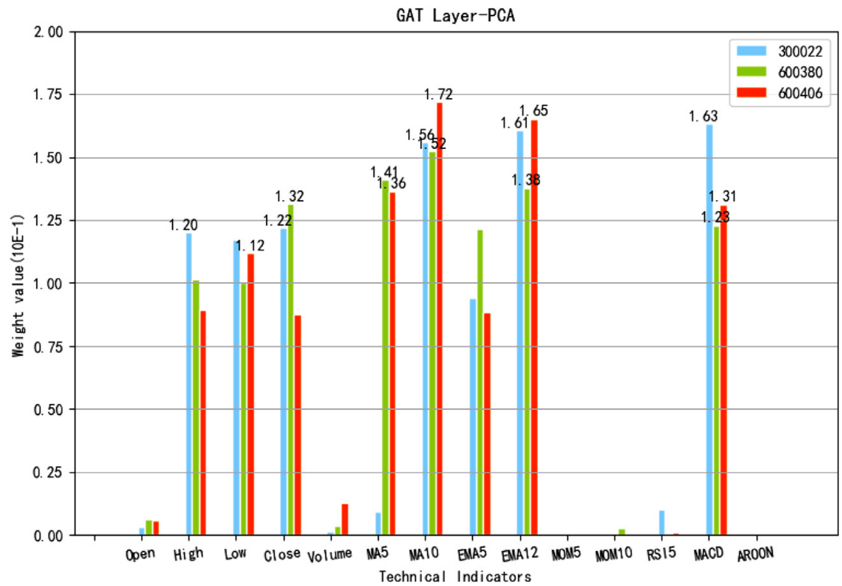

Fig. 8. The weight coefficient of each technical indicator in the last layer of our attribute graph attention network.

Table 7
A comparison of the performance between six baseline methods and our proposed Ra-AGAT on SSE financial market. In the second column, our method is slightly inferior to Attn-LSTM-GCN. However, our proposed method exceeds all other methods in terms of the remaining evaluation indicators.

Shanghai Stock Exchange (SSE)			
	MSE	*MRR*	*CRoI*
LR	$2.92e-3 \pm 2.21e-3$	$2.70e-2 \pm 1.72e-2$	0.17 ± 0.15
ARIMA	$2.81e-3 \pm 2.10e-3$	$3.75e-2 \pm 2.17e-2$	0.20 ± 0.12
SFM	$2.70e-3 \pm 2.19e-3$	$4.12e-2 \pm 1.34e-2$	0.21 ± 0.14
Attn-LSTM	$3.63e-3 \pm 2.42e-3$	$3.02e-2 \pm 2.15e-2$	0.09 ± 0.13
Attn-LSTM-GCN	$2.61e-3 \pm 2.49e-3$	$4.40e-2 \pm 2.31e-2$	0.24 ± 0.12
Attn-LSTM-GAT	$2.43e-3 \pm 2.37e-3$	$4.01e-2 \pm 1.97e-2$	0.26 ± 0.15
Attn-LSTM-AGAT	$2.62e-3 \pm 2.32e-3$	$4.71e-2 \pm 2.34e-2$	0.36 ± 0.17

Table 8
Descriptive statistics of cross-correlation coefficients. Pearson coefficient P_{ij} and DCCA coefficient $DCCA_{ij}$.

Measurement Correlation Coefficients	P_{ij}	$DCCA_{ij}$
Mean	0.4514	0.4261
Maximum	0.9018	0.8723
Minimum	−0.0329	−0.0275
Standard deviation	0.1251	0.1374
Skewness	0.0727	0.0292
Kurtosis	4.127	3.962

methods to verify the characteristics and applicability of the graph structure in the financial market

Network structure configuration For historical data analysis, we used the following sets of hyperparameters for our Attn-LSTM: the period of sequential stock data within [5, 14] and the number of LSTM units in [16]. For the AGAT model in our embedding layer, the hyperparameters were set as follows. a): the hidden dimension and attention head within [4] and [4], respectively. b); were set within $[1, 0.5]$, which were applied to the loss function for balancing the relationship between points and pairs loss; and the top-k in the final ranking list based on the rate of return within [1, 3, 5]. **Measurement** Based on our translation of the stock prediction problem into a ranking task, we evaluated our model in three aspects: ranking results accuracy, the quality of ranking result and ranking system performance.

- Mean Squared Error(MSE) to describe the accuracy of the ranking model. The smaller the value, the smaller the difference between the prediction and the target. Its calculation formula is

$$MSE = \frac{1}{N}\sum_{1}^{N}\left(L_{\text{target}} - L_{\text{pred}}\right)^2 \tag{24}$$

- Mean Reciprocal Rank(MRR) measures the accuracy of the recommendation algorithm. Because the final result in our experiment is a unique ranking list, we made some changes.

$$MRR = \frac{1}{\Omega_k}\sum_{i=1}^{k}\frac{1}{rank_i} \tag{25}$$

- The Cumulative Return on Investment($CRoI$) is one of the most important indicators in the field of financial investment. It has been referred to in previous literature [40]. $CRoI$ was calculated by adding up the RoI of the stocks selected in our test set over all test days. Because it is the most direct indicator of how to make an optimal investment portfolio selection, it is the most important element in our final recommendation results.

Of the model evaluation indicators, the smallest MSE and the largest other two indicators prove the strong performance of the model. We repeated all the experiments several times to avoid the effects of accidental factors and hyperparameters. The quality of the model and the possibility of application to an actual investment strategy were determined by showing the scores of all the indicators mentioned above.

Discussion The recommendation results obtained with different methods are shown in Table 7, in which the boldfaced values indicate the best result in each column. It is clear that the linear functions that ignore the relationship in the financial market, such as ARIMA, SFM, LR and their variants, performed poorly in the CRoI evaluation indicator. The reason we believe is that the characteristics of the highly nonlinear changes in the stock market make the methods that consider only the numerical changes in the time dimension appear as a large deviation.

Of the GNN-based methods, our proposed RA-AGAT method ranked the highest by improving the performance of the traditional graph attention network by 0.12. The main reason we consider is that the traditional graph attention network(GAT) ignores edge information in the process of aggregating adjacent information. Besides, the graph attention network contains the same parameters in the iteration, whereas most popular networks use different parameters in different layers, which serve as a hierarchical feature extraction function. Compared with the graph convolutional network(GCN) in CRoI indicator, our methods ranked the highest by improving the performance by 0.17. Compared with the previous results based on the graph attention network, it is not difficult to find that the association relationship extracted in the topological relationship of the graph convolution structure is necessary, but the multi-dimensional association relationship will cause the redundancy of the information in the aggregation process which reduces the influence of the characteristic factors of the stock itself. When different graph neural network models become a stacked structure, the upper-layer topology-based network selectively filter and check information, which is equivalent to guiding the lower-layer graph attention network based on feature factors to adaptively fusing important feature factors.

What's more, we experimented the back-testing function on our proposed method, namely Top1, Top3, and Top5 stocks with the highest return ratio for trading. We hope to utilize simulated transactions to judge whether the results predicted based on the level of return ratio can generate greater returns. Specifically, if we chose the Top3 investment strategy, then we allocated the budget evenly to the test date of our experiment to buy the top three stocks and calculated the average rate of return. As can be seen in Fig. 6, the results of the Top1 and Top3 strategies are roughly the same, but the latter has a slightly lower cumulative return rate. The reason seems that when the algorithm itself has accurate prediction capabilities and high robustness, once the relative order of the return ratio prediction is determined, the top ranked trading stocks will show a higher cumulative return. However, the cumulative return curve has a small fluctuation range and lower result in the Top5 strategy which means that the gap of return ratio between different stocks in the Chinese stock market is large, and the volatility of returns is large. Therefore, our future work will introduce the prior knowledge of external events into the graph-based model.

Factor selection strategy The interpretability in the financial market lies in the selection of key factors to conduct strategic analysis of future trends. We screened the top 3 stocks in the test set for factor analysis through principal component analysis(PCA). Figs. 7 and 8 respectively display the weight values corresponding to different factor features in our stacked graph network structure. Among our stacked graph structure, The upper layer is based on the graph convolution network module to selectively transfer and integrate the factor features according to the topology structure, and the lower layer is based on the graph attention network module to adaptively integrate the factor features in the financial network. Therefore, the importance of factor features is intuitively displayed through parameter weight values in our stacked graph structure. From the above two figures, we can clearly observe that the price factor and the moving average signal factor have high weight coefficients between layers, and the top five key signal factors corresponding to the top three return ratio stocks roughly display the same pattern. There are significant differences in the weight parameters of different signal factors in the subsequent GAT layer. The moving average factor of different periods has higher signal strength. From the perspective of time dimension, the longer period factor reflects the stronger the signal strength, which is also in line with the characteristics of the longer period of our data itself. There is an obvious weak signal with the volume factor and momentum factor, which is also in line with the premise that momentum factor and volume factor need to be combined in strategic analysis. We will also conduct more in-depth analysis in future work.

5. Conclusion

In this article, we make up for the shortcomings of the correlation coefficient model that cannot carry out the multi-scale fusion and dissemination of financial data by fusing the single correlation coefficient model into the model based on graph neural network(GNN). We propose a new graph attention network whose attention mechanism is based on the graph convolutional network to be applied into the financial market. Our main goal is to recommend the Top-N return ratio of the stock items and propose the factor analysis strategies to filter out the important factors reflected in the stocks, hoping to achieve better prediction accuracy through the graph-based learning process.

We plan to extend our work in a number of ways. Firstly, we will incorporate the real-world relationships in financial markets, newsletters, investor sentiment and other real-world knowledge into the GNN-based model to make up for the lack of prior knowledge in our model. Secondly, we look for coding tools that are more suitable for the highly time-varying characteristics of the stock market, so as to capture a more accurate change process in the time dimension. Finally, we hope to promote the application of heterogeneous information networks in financial markets on the basis of existing models, which more effectively describe the complex relationships that exist in the stock network, and to propose a general model that can be applied to the world financial markets with different characteristics. Only through comprehensive consideration of the complexity in the financial field can we improve the interpretability of the application of the GNN-based models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Shibo Feng, Chen Xu and Yu Zuo are co-first authors, which contributed equally to this paper. This work is supported by the National Natural Science Foundation of China (Grant No. 61977055 and U2066213), and the research Funds of State Grid Shaanxi Electric Power Company and State Grid Shaanxi Information and Telecommunication Company (No. SGSNXT00GCJS2100120).

References

[1] L.E. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Ann. Math. Stat. 41 (1) (1970) 164–171.
[2] P. Thomas, A relationship between technology indicators and stock market performance, Scientometrics 51 (1) (2001) 319–333.
[3] D. Chen, X. Guo, J. Wang, J. Liu, Z. Zhang, E.R. Hancock, Thermodynamic motif analysis for directed stock market networks, Pattern Recognit. 114 (2021) 107872.
[4] L. Bai, L. Cui, Z. Zhang, L. Xu, Y. Wang, E.R. Hancock, Entropic dynamic time warping kernels for co-evolving financial time series analysis, IEEE Trans. Neural Netw. Learn. Syst. (2020).
[5] Z. Zhang, D. Chen, J. Wang, L. Bai, E.R. Hancock, Quantum-based subgraph convolutional neural networks, Pattern Recognit. 88 (2019) 38–49.
[6] J.S. Tyssedal, D. Tjøstheim, An autoregressive model with suddenly changing parameters and an application to stock market prices, J. R. Stat. Soc. Ser. C (Applied Statistics) 37 (3) (1988) 353–369.
[7] X. Yan, Z. Guosheng, Application of Kalman filter in the prediction of stock price, 5th International Symposium on Knowledge Acquisition and Modeling (KAM 2015), Atlantis Press, 2015.
[8] T. Hirabayashi, H. Takayasu, H. Miura, K. Hamada, The behavior of a threshold model of market price in stock exchange, Fractals 1 (01) (1993) 29–40.
[9] D.M. Nelson, A.C. Pereira, R.A. de Oliveira, Stock market's price movement prediction with LSTM neural networks, in: 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, 2017, pp. 1419–1426.
[10] D. Goldfarb, G. Iyengar, Robust portfolio selection problems, Math. Oper. Res. 28 (1) (2003) 1–38.

[11] P. Cizeau, M. Potters, J.-P. Bouchaud, Correlation structure of extreme stock returns (2001).
[12] B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series, Phys. Rev. Lett. 100 (8) (2008) 084102.
[13] M. Tumminello, T. Di Matteo, T. Aste, R.N. Mantegna, Correlation based networks of equity returns sampled at different time horizons, Eur. Phys. J. B 55 (2) (2007) 209–217.
[14] M. Eryiğit, R. Eryiğit, Network structure of cross-correlations among the world market indices, Physica A 388 (17) (2009) 3551–3562.
[15] D.-M. Song, M. Tumminello, W.-X. Zhou, R.N. Mantegna, Evolution of worldwide stock markets, correlation structure, and correlation-based graphs, Phys. Rev. E 84 (2) (2011) 026108.
[16] K. Jebran, A. Iqbal., Dynamics of volatility spillover between stock market and foreign exchange market: evidence from Asian countries, Financ. Innov. 2 (1) (2016) 1–20.
[17] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint (2013).
[18] B. Perozzi, R. Al-Rfou, S. Skiena, DeepWalk: online learning of social representations, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 701–710.
[19] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, Q. Mei, LINE: large-scale information network embedding, in: Proceedings of the 24th International Conference on World Wide Web, 2015, pp. 1067–1077.
[20] A. Grover, J. Leskovec, node2vec: scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855–864.
[21] G.E. Box, D.A. Pierce, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, J. Am. Stat. Assoc. 65 (332) (1970) 1509–1526.
[22] B.K. Nelson, Time series analysis using autoregressive integrated moving average (ARIMA) models, Acad. Emerg. Med. 5 (7) (1998) 739–744.
[23] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, Econometrica (1982) 987–1007.
[24] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econom. 31 (3) (1986) 307–327.
[25] A. Samarawickrama, T. Fernando, A recurrent neural network approach in predicting daily stock prices an application to the sri lankan stock market, in: 2017 IEEE International Conference on Industrial and Information Systems (ICIIS), IEEE, 2017, pp. 1–6.
[26] L. Sayavong, Z. Wu, S. Chalita, Research on stock price prediction method based on convolutional neural network, in: 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), IEEE, 2019, pp. 173–176.
[27] Z. Zhang, D. Chen, Z. Wang, H. Li, L. Bai, E.R. Hancock, Depth-based subgraph convolutional auto-encoder for network representation learning, Pattern Recognit. 90 (2019) 363–376.
[28] C.K. Thomas Fischer, Deep learning with long short-term memory networks for financial market predictions, Eur. J. Oper. Res. 270 (2) (2018) 654–669.
[29] J. Li, H. Bu, J. Wu, Sentiment-aware stock market prediction: a deep learning method, in: 2017 International Conference on Service Systems and Service Management, IEEE, 2017, pp. 1–6.
[30] K.A. Althelaya, E.-S.M. El-Alfy, S. Mohammed, Evaluation of bidirectional LSTM for short-and long-term stock market prediction, in: 2018 9th International Conference on Information and Communication Systems (ICICS), IEEE, 2018, pp. 151–156.
[31] Z.-Q. Jiang, W.-X. Zhou, Complex stock trading network among investors, Physica A 389 (21) (2010) 4929–4941.
[32] M.-X. Li, Z.-Q. Jiang, W.-J. Xie, X. Xiong, W. Zhang, W.-X. Zhou, Unveiling correlations between financial variables and topological metrics of trading networks: evidence from a stock and its warrant, Physica A 419 (2015) 575–584.
[33] D.-M. Song, Z.-Q. Jiang, W.-X. Zhou, Statistical properties of world investment networks, Physica A 388 (12) (2009) 2450–2460.
[34] Z. Zhang, D. Chen, L. Bai, J. Wang, E.R. Hancock, Graph motif entropy for understanding time-evolving networks, IEEE Trans. Neural Netw. Learn. Syst. (2020).
[35] G.-J. Wang, C. Xie, H.E. Stanley, Correlation structure and evolution of world stock markets: evidence from Pearson and partial correlation-based networks, Comput. Econ. 51 (3) (2018) 607–635.
[36] H. Guo, On the out-of-sample predictability of stock market returns, J. Bus. 79 (2) (2006) 645–670.
[37] G.-J. Wang, C. Xie, S. Chen, J.-J. Yang, M.-Y. Yang, Random matrix theory analysis of cross-correlations in the us stock market: evidence from Pearsons correlation coefficient and detrended cross-correlation coefficient, Physica A 392 (17) (2013) 3715–3730.
[38] D. Matsunaga, T. Suzumura, T. Takahashi, Exploring graph neural networks for stock market predictions with rolling window analysis, in: arXiv preprint arXiv:1909.10660, 2019.
[39] Y. Chen, Z. Wei, X. Huang, Incorporating corporation relationship via graph convolutional neural networks for stock price prediction, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 1655–1658.
[40] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, T.-S. Chua, Temporal relational ranking for stock prediction, ACM Trans. Inf. Syst. (TOIS) 37 (2) (2019) 1–30.
[41] J. Lee, I. Lee, J. Kang, Self-attention graph pooling, in: arXiv preprint arXiv: 1904.08082, 2019.
[42] J. Shen, B. Zheng, Cross-correlation in financial dynamics, EPL (Europhysics Letters) 86 (4) (2009) 48005.
[43] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (3–5) (2010) 75–174.
[44] B.U. Devi, D. Sundar, P. Alli, An effective time series analysis for stock trend prediction using ARIMA model for nifty Midcap-50, Int. J. Data Min. Knowl. Manage. Process 3 (1) (2013) 65.
[45] S. Chen, L. Ge, Exploring the attention mechanism in LSTM-based hong kong stock price movement prediction, Quant. Finance 19 (9) (2019) 1507–1515.

Shibo Feng is now a master student at the software school of Xiamen University, China. His research interests include data mining, machine learning, graph embedding and complex network analysis. His mentor is Fan Lin.

Chen Xu received her B. Sc. degree (2011) in Economics and Finance from The University of York (UK) and a Master degree (2013) of Global Business from Coventry University (UK). She currently works as a business manager in China Energy Engineering Group Shaanxi Electric Power Design Institute Co., LTD, from 2013. Her research interests include financial analysis.

Yu Zuo is currently an undergraduate student in the Department of Economics, New York University, USA, and majors in Economics joint mathematics. Her research interests include financial analysis.

Fan Lin received his M.S and Ph.D. degree from Xiamen University in 2003 and 2013. He is an Associate Professor in the School of Informatics Xiamen University, as Deputyhead for Smart Home Engineering Center of Fujian Province. His research interests cover Internet of Things, Explainable Machine Learning, Recommendation System Reinforcement Learning.

Jianbing Xiahou received his M.S and Ph.D. degrees from Xiamen University in 1999 and 2008, respectively. He is an Associate Professor in the Software School of Xiamen University. He obtained his Master Degrees of Electronic Engineering at the Xiamen University in 1999 and Doctor Degree of Finance in 2008, respectively. He has been working in fields related to IT technology and project since 1999. His research interests cover Big Data, Multi-media, Cloud Computing, Machine Learning, Software Engineering and financial informationizition.