Pattern Recognition 121 (2022) 108218

Contents lists available at ScienceDirect

Pattern Recognition

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

Financial time series forecasting with multi-modality graph neural network

Dawei Cheng[a,*], Fangzhou Yang[b], Sheng Xiang[c], Jin Liu[b]

[a] Department of Computer Science and Technology, Tongji University, Shanghai, China
[b] AI Lab, Emoney Inc, Shanghai, China
[c] Center for Artificial Intelligence, University of Technology Sydney, Sydney, Australia

ARTICLE INFO

Article history:
Received 15 December 2020
Revised 22 July 2021
Accepted 31 July 2021
Available online 2 August 2021

Keywords:
Graph neural network
Graph attention
Deep learning
Quantitative investment

ABSTRACT

Financial time series analysis plays a central role in hedging market risks and optimizing investment decisions. This is a challenging task as the problems are always accompanied by multi-modality streams and lead-lag effects. For example, the price movements of stock are reflections of complicated market states in different diffusion speeds, including historical price series, media news, associated events, etc. Furthermore, the financial industry requires forecasting models to be interpretable and compliant. Therefore, in this paper, we propose a multi-modality graph neural network (MAGNN) to learn from these multimodal inputs for financial time series prediction. The heterogeneous graph network is constructed by the sources as nodes and relations in our financial knowledge graph as edges. To ensure the model interpretability, we leverage a two-phase attention mechanism for joint optimization, allowing end-users to investigate the importance of inner-modality and inter-modality sources. Extensive experiments on real-world datasets demonstrate the superior performance of MAGNN in financial market prediction. Our method provides investors with a profitable as well as interpretable option and enables them to make informed investment decisions.

1. Introduction

The financial market capitalization of US domestic listed companies reaches 30 trillion dollars in 2019, account over 1.5 times the Gross Domestic Product (GDP) in the United States [1]. In this massive yet volatile market, forecasting the price movement of equities is very important for both financial institutions and investors. According to the efficient market hypothesis (EMH) [2], ideally, the stock's prices reflect all available information in an efficient market, which includes historical prices, news, events, etc. However, in a real-world situation, different equities responding to different events are non-intuitive and non-synchronized. Thus, it is challenging to model this intricate phenomenon, named the lead-lag effect [3], in a time series forecasting framework.

The financial industry has researched price prediction models since the beginning of the twentieth century [4] and has perfected these technologies ever since, investing millions of dollars in this process. Traditional quantitative methods rely on historical time-series price data for stock price movement prediction [5,6]. These models aim to reduce the stochasticity and capture consistent patterns by extracting meaningful technical indicators [7] and/or latent features [8]. Lately, with the development of social media and natural language processing technologies, unstructured news has been leveraged to improve the prediction model capability [9]. But these technologies do not capture internal relations among equities, which limits their potentials for the forecasting model. For example, the term-level feature of an event "Qualcomm files lawsuit against Apple" cannot differentiate the appellor "Qualcomm" and appellee "Apple", so it is difficult to infer the corresponding price movements of the related equities, Qualcomm and Apple Inc.

Recently, researchers [10] tend to improve the representation of market information by extracting structural event tuples and indicators (i.e., sentiment indicator) [11] from media news. The main idea is to learn distributed representations that similar events or similar sentiment news could have similar features. These features are then linked to listed companies and integrated with historical time series for price prediction [12]. But two similar events may be quiet unrelated, such as "Steve Jobs quits Apple" and "David Peter leaves Starbucks". To overcome this, studies [13,14] employ external information from knowledge graphs (KG) in the feature learning process [15]. Then, the above two events can have different representations according to the semantic differences in KG, because Steve Jobs is the founder of Apple while David Peter is more like to be a customer in Starbucks.

* Corresponding author.
E-mail address: dcheng@tongji.edu.cn (D. Cheng).

https://doi.org/10.1016/j.patcog.2021.108218
0031-3203/

However, the stock's price movements in the financial market not only rely on individual events of itself but also related to the connections of other equities [16]. These multi-modality inputs, including numerical time series, unstructured texts and relational graphs, contribute differently as a synergy effect on the price movement. For instance, an event "Qualcomm suits against Apple" will also influence other players (i.e., competitors, upstream and downstream firms) of the smartphone market in different diffusion speeds, such as Samsung, Foxconn, and Google, etc. Effectively forecasting the prices of related equities from the lead-lag effects is challenging, due to the incompleteness of financial domain knowledge and intricate sequential patterns.

Therefore, in this paper, we propose a multi-modality graph neural network model for forecasting the price movements by incorporating sources of lead-lag relationships, including historical prices, media events, and corresponding knowledge from KG. In particular, we first extract relations of linked entities from raw news by and then store them in our financial knowledge graphs (FinKG). Then, we propose a heterogeneous graph attention network to learn the unified representation of target time series, in which multi-modality sources are defined as source nodes and the predicted equity as target node. We leverage a two-phase attention mechanism (inner-modality and inter-modality attention) to infer the internal sequential patterns and inter-source lead-lag relations. Inner-modality attention mechanism is designed to automatically learn different contributions of graph-structured sources to the target node within each modality inputs. While inter-modality attention is proposed to learn weights among different modalities dynamically for a decent price movement prediction of target nodes, as different modality contributes differently in different time period. Afterwards, the learned informative features are fed into prediction layer for price movement forecasting. Extensive experiments on real market data show the effectiveness of our method and interpretability of the proposed two-phase attention mechanism.

In a nutshell, the main contribution of this paper includes:

- We formalize the problem of lead-lag effects in financial time series forecasting and identify their unique challenges arising from real financial industry applications.
- We propose a novel multi-modality graph neural network (MAGNN) to learn the lead-lag effects for financial time series forecasting, which preserves informative market information as inputs, including historical prices, raw news text and relations in KG. To our best knowledge, this is the first study to explore the lead-lag effects by embedding informative sources in a unified graph neural framework for price movements prediction.
- In order to follow highly regulated processes in the financial industry, we design and implement a two-phase attention mechanism to infer the interpretability from both the inner-modality and inter-modality sources. We also validate the effectiveness of designed attention technologies in learning the internal sequential patterns and inter-source lead-lag relations through empirical studies.
- Extensive experimental results on 3714 stocks demonstrate the superior performance of our proposed method. Furthermore, Our model has been deployed in a major financial service provider of China and we validate its performance of profitability and interpretability in real-world scenarios. The source codes will be released in near future.

2. Preliminaries

In this section, we introduce the background of lead-lag effect and the construction process of heterogeneous graphs.

2.1. Lead-lag effect

In an informational efficient market, price movements of stocks can be deemed as the reaction of financial events or news [17]. However, when a new event hit the stock market, prices of some stocks response faster than others. This phenomenon of correlated yet asynchronous price movement is referred to as lead-lag effect [3]. For example, in Fig. 1, when a new event ("Qualcomm suites against Apple") hit the market, it will not only bring price fluctuation of "Qualcomm" and "Apple", but will also influence upstream and downstream companies, such as Samsung (supplier and major competitor of Apple in smart phone market) and Foxconn (manufacturer of Apple). But their price movements are asynchronous because the event diffusion speed is different over different entities. Therefore, it is a challenging task to learn from this lead-lag relationship in financial market.

2.2. Heterogeneous graph construction

In MAGNN, multi-modality heterogeneous graph extends the conventional heterogeneous graph [18] with multi-modality sources. Graph nodes are divided into six types (source, news, events, market, bridge and target nodes) with three modality inputs (numeral time series, media texts, and relations). We give the definition as follows:

Definition 1Heterogeneous graph. . A heterogeneous graph is denoted as

$$\mathcal{G} = (\mathcal{V}_\mathcal{T}, \mathcal{V}_\mathcal{S}, \mathcal{E}),$$

where $\mathcal{V}_\mathcal{T}$ represents the set of target nodes, $\mathcal{V}_\mathcal{S}$ denotes the set of source nodes and $\mathcal{E}$ is the set of links connecting between nodes.

Definition 2Source nodes. $\mathcal{V}_\mathcal{S}$ are associated with different modality by a mapping function $\Psi : \mathcal{V}_\mathcal{S} \rightarrow \Phi$, where Φ denotes the set of modalities, including numeral market data, media texts, and relations.

Definition 3Target nodes. $\mathcal{V}_\mathcal{T}$ are our predicted equities in the graph, which is designed to receive and aggregate messages from other nodes via directed links.

Definition 4Bridge node. denotes the connected nodes between multi-modality sources and target nodes. They are extracted from the domain knowledge graph FinKG.

Definition 5Attributed nodes. include news, event and market nodes, which only connect to their subject companies.

Multi-modality inputs are seemed as nodes in a heterogeneous graph, in which they can pass messages to other nodes via links. A company might be one of source, target or bridge node, while attributed (news, event, market) nodes only connect to its subject company. For example, the market node (M) of Apple only connect target node Apple, as shown in Fig. 1.

Definition 6. Edges ($\mathcal{E}$) are a set of links connecting between nodes, which include directed and undirected edges. The relationship among companies (source, target or bridge nodes) are directed, which arrow from the subject to the object. The connection between company and its attributed nodes are undirected.

Fig. 1 shows a running example of heterogeneous graph and multi-modality inputs. When an event (or news) "Qualcomm suits against Apple" hit the market, we extract the relation between its subject (Qualcomm) and object (Apple), and establish an edge (suit against) directed from the subject (Qualcomm) to the object (Apple). Then, if we want to forecast Apple's price movements

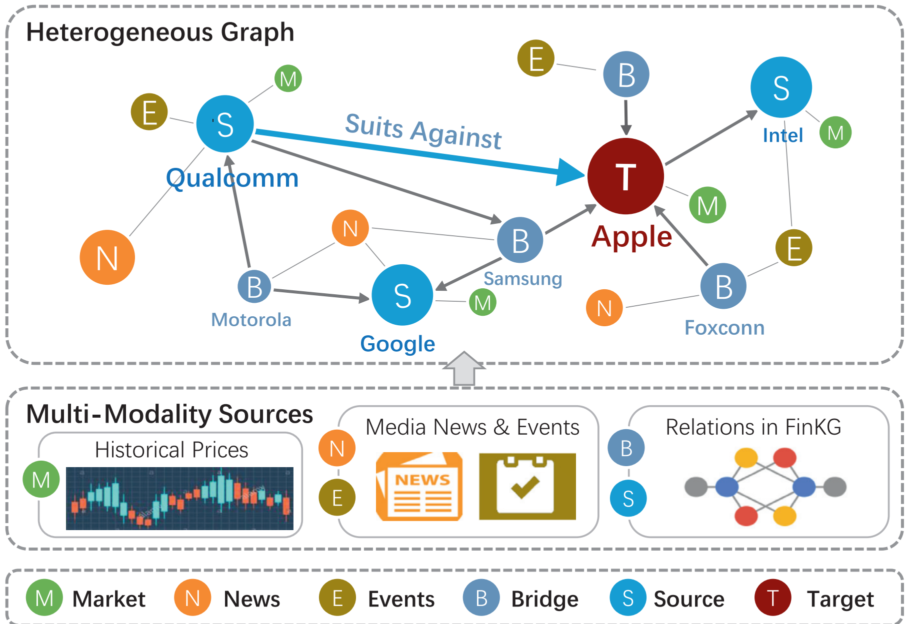

Fig. 1. An illustration of multi-modality inputs and heterogeneous graph.

in the following day, we set it as target node and extract multi-modality inputs accordingly, which includes event (news) semantic embeddings, linked source node (Qualcomm) and its historical prices, target node's (Apple) market data, and the relations (including edges and bridge nodes, such as Samsung, Motorola, Foxconn, etc.) of source and target nodes. For example, in smart phone market, Samsung is competitor of Apple and a downstream customer of Google. Thus, it is a bridge node of Google and Apple in this scenario. As shown in Fig. 1, normally, each event (news) is accomplished with a source node and a target node. We construct heterogeneous graphs by multi-modality inputs of the linked nodes and corresponding relations in FinKG. The detailed methods of relation and graph construction are presented in Section 3.1. These informative inputs are then fed into MAGNN for joint and interpretable learning.

3. Methodology

In this section, we first introduce the general framework and multi-modality inputs of our proposed approach, and then present inner-modality graph attention and inter-modality source attention, respectively. Lastly, we introduce the target forecasting network and model optimization.

3.1. Model framework and inputs

Fig. 2 shows the general framework of the proposed multi-modality graph neural network for financial time series forecasting. We construct the heterogeneous graph first by the events, news, relations in KG and the market data, as shown Fig. 2a. Then, multi-modality inputs are fed into inner-modality graph attention layer (InnGAT) in parallel, in which each modality input is learned by InnGAT independently over the heterogeneous graph. The inter-modality source attention (IntSAT) takes the output of InnGAT and learn high-order representations from all modalities. Finally, the learned features are fed into a feed-forward and classification network for target forecasting.

In the implementation, we employ a pretrained BERT[1] [19] as our news embedding model, and finetune BERT model from our large-scale financial news corpus. For event tuple extraction, we leverage the widely-used OpenIE [20] and utilize the embedding of structured tuples learned by tensor neural network [21] as event feature. In the FinKG construction, we employ OpenNRE[2] to extract relations from massive news text and store them in our knowledge graph FinKG. If the entity of an event (or news) is a listed company, we mark them as the source node. The rest entities are denoted as bridge nodes in the knowledge graph. When a set of events hit the FinKG, we extract the adjacent nodes and corresponding relations of the mentioned entities as base graph. Then, we mark the predicted stocks as target nodes. Afterwards, the news, events and market data are linked to each entity and finally form the heterogeneous graph, as shown Fig. 2a. We update the heterogeneous graph by every trading day.

3.2. Inner-modality graph attention

Given each modality input feature and the constructed heterogeneous graph, inner-modality graph attention is designed to propagate and aggregate information from source nodes to the target node. As shown in Fig. 2b, the inputs of InnGAT include the pre-trained embeddings of the source node $e_i^{S_\phi}$ and the target node e_s^t, where $\phi \in \{n, e, p\}$ denotes the modality type and $i \in \mathcal{N}_s$ indicates the ith neighbors of node S. $\mathcal{N}_s$ is the set of neighbors.

We design two-phase projections for mapping the multi-modality inputs into latent representations, named source projection and target projection. They are parameterized by weight matrix $W_S^\phi \in \mathbb{R}^{d_h \times d_\phi}$ and $W_T^\phi \in \mathbb{R}^{d_h \times d_t}$, respectively. d_ϕ, d_t and d_h denote the dimension of source node embedding, target node embedding and projected hidden features. Then, a shared attention mechanism is introduced to compute node-level attention coefficients, which is parameterized by a weight vector $\vec{a_\phi} \in \mathbb{R}^{2d_h}$. Finally, The inner-modality attention coefficient for source type ϕ

[1] https://github.com/google-research/bert
[2] https://github.com/thunlp/OpenNRE

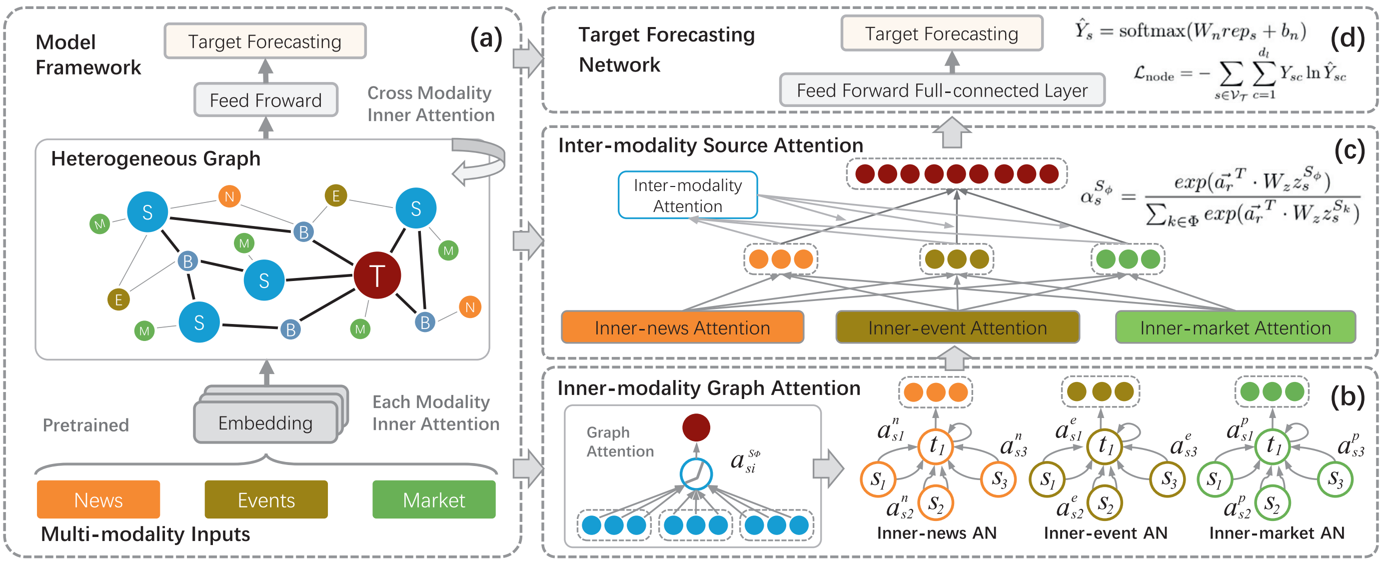

Fig. 2. The general framework of the proposed multi-modality graph neural network. It includes multi-modality inputs, inner-modality graph attention layer, inter-modality source attention layer and the target forecasting network. In the heterogeneous graph, the symbol of S, B, E, N, M, T denotes the source, bridge, events, news, market and target nodes respectively.

between source node i and target node s is formulated as:

$$\alpha_{si}^{S_\phi} = \frac{\exp(\mathrm{LeakyReLU}(\vec{a_\phi}^{\top}[W_T^\phi \cdot e_s^t || W_S^\phi \cdot e_i^{S_\phi}]))}{\sum_{k \in \mathcal{N}_s^{S_\phi}} \exp(\mathrm{LeakyReLU}(\vec{a_\phi}^{T}[W_T^\phi \cdot e_s^t || W_S^\phi \cdot e_k^{S_\phi}]))}, \tag{1}$$

where $\vec{a_\phi}^{\top}$ represents transposition of $\vec{a_\phi}$ and $||$ is the concatenation operation.

Afterward, we compute the output features of the target node S for modality type ϕ as weighted average of the source hidden features with sigmoid function, which is formulated as:

$$z_s^{S_\phi} = \sigma\left(\sum_{i \in \mathcal{N}_s^{S_\phi}} \alpha_{si}^{S_\phi} W_S^\phi e_i^{S_\phi}\right), \tag{2}$$

where W_S^ϕ denotes the learned weights, and σ is the sigmoid function. $\mathcal{N}_s$ indicates the neighbors set of node S. $z_s^{S_\phi}$ denotes the output feature of InnGAT for node S in modality ϕ. In the implementation, we extend the InnGAT with multi-head attention in order to stabilize the learning process.

3.3. Inter-modality source attention

Inter-modality source attention (IntSAT) is proposed to selectively aggregate the information from multi-modality sources for target node representation. As illustrated in Fig. 2c, the inputs of IntSAT for target node include the output features $z_s^{S_\phi}$ of InnGAT from all modalities, where $\phi \in \{n, e, p\}$. In the inter-modality source attention network, a shared linear transformation parameterized by a weight matrix $W_z \in \mathbb{R}^{d_r \times d_z}$ and a multi-source attention mechanism parameterized by a weight vector $\vec{a_r} \in \mathbb{R}^{d_r}$ are employed to compute source attention coefficients, respectively. d_z indicates the dimension of $z_s^{S_\phi}$ and d_r is the dimension of the transformed hidden features. Mathematically, the attention coefficient of modality type ϕ for target node can be formulated by:

$$\alpha_s^{S_\phi} = \frac{\exp(\vec{a_r}^{\top} \cdot W_z z_s^{S_\phi})}{\sum_{k \in \Phi} \exp(\vec{a_r}^{\top} \cdot W_z z_s^{S_k})}, \tag{3}$$

where $\vec{a_r}$ and W_z are the learned weights, and $\alpha_s^{S_\phi}$ denotes the attention coefficient of modality type ϕ.

Finally, we construct the representation of the target node rep_s by the concatenation of the attention-weighted projected features from all three modalities, formulated as:

$$rep_s = [\alpha_s^{S_n} W_z z_s^{S_n} || \alpha_s^{S_e} W_z z_s^{S_e} || \alpha_s^{S_p} W_z z_s^{S_p}], \tag{4}$$

where the $\alpha_s^{S_n}$, $\alpha_s^{S_e}$ and $\alpha_s^{S_p}$ are the attention coefficient of IntSAT. W_z denotes the learned weights and rep_s means the output representation of inter-modality source attention network.

3.4. Target forecasting network and optimization

Given the learned representation of target node from InnGAT and IntSAT, we then employ a shallow neural network for the target price forecasting, as shown in Fig. 2d. In particular, we formulate the forecasting task as a classification problem, which means we divide the trend of price movements into three categories $\{up, neural, down\}$. We will detailed describe the settings in experiment section. The forecasting network consists of two full connect layers and one softmax layer. They are defined as:

$$\hat{Y}_s = \mathrm{softmax}\big(\mathrm{NN}_f(W_n rep_s + b_n)\big) \tag{5}$$

where NN_f denotes a shallow neural network with two-layers of full connection. $W_n \in \mathbb{R}^{d_s \times d_l}$ and $b_n \in \mathbb{R}^{d_l}$ are the weight matrix and bias respectively. d_l is the number of target categories. In this paper, we set the $d_l = 3$.

Finally, we define the loss function of the proposed model by the cross-entropy of the likelihood in output layer as below:

$$\mathcal{L}_{\mathrm{target}} = -\sum_{s \in \mathcal{V}_T} \sum_{c=1}^{d_l} Y_{sc} \ln \hat{Y}_{sc} \tag{6}$$

where Y_{sc} is the ground-truth label of c_{th} movement category for stock s, which is marked as 1 for the "up" price movements, 0 for the "neural" and -1 for the "down" movement, respectively. $\mathcal{V}_T$ denotes the set of target nodes.

Our proposed multi-modality graph neural network can be trained in an end-to-end manner by minimizing the classification cross-entropy loss. Theoretically, we can optimize the model by the standard stochastic gradient descent process. In practice, we employ Adam algorithm [22] as the optimizer of our model. We set the initial learning rate to 0.001, and the batch size to 64 by default.

4. Experiments

In this section, we conduct extensive experiments to validate the effectiveness of our proposed technologies. We introduce the data acquisition and experimental settings first, and then report the result of each experiment in turn.

4.1. Datasets and experimental settings

Data acquisition Generating informative datasets from massive multi-modality sources is challenging in our experiment. To ensure fairness, we collect financial events, news, market prices and the knowledge graph for all 3714 public companies listed in China A-shares market,[3] from Jan 01 2018 to Dec 31 2019. In particular, we crawl the public announcements and leverage event extraction methods to construct events for list companies. There are total 143,884 structured events across 41 categories in our dataset, such as seasonal/annual reports, asset restructuring, increase/decrease of credit ratings, change of the chairman or board members, production accident, etc. For financial news, we crawl information from 87 major websites that cover most important reports in the market. There are 5.13 million news during the time interval. We leverage named entity recognition (and linking) and neural relation extraction technologies to extract entities and relations from raw texts. These linked listed companies are stored as nodes in the knowledge graph, in which each relation is stored as an edge between nodes. Finally, there are total 5.26 million entities and 6.93 million relations in the FinKG.

We gather the stock price data of China A-shares listed companies from the Shanghai and Shenzhen Stock Exchange sources from 2018 to 2020, including 500 trading days. The daily market data includes stock prices (open, close, high, and low) and the trading information (trading volume and turnover rate) of that day for each stock. In the experiment, we remove the trading suspension stock and untradable prices (such as limit-up, limit-down stocks) from the dataset. In the China stock market, investors need to follow a 10% limit up-limit down mechanism strictly.

Experimental settings We forecast the price movement into three categories $\{up, neural, down\}$. For the stock i in day t, the return rate can be computed by $R_{r,i} = p_i^t / p_i^{t-1} - 1$. We set the ground-truth label of the price movements as:

$$f(R_r) = \begin{cases} up & R_r \geq r_{up}, \\ neural & r_{down} < R_r < r_{up}, \\ down & R_r \leq r_{down} \end{cases} \tag{7}$$

where we set $r_{up} = 0.01$ and $r_{down} = -0.01$. In our dataset, there are 226,585 samples in “up” category, 327,851 “neural” and 238,630 samples in “down” category.

In the experiment, we employ the data of the year 2018 as the training set and evaluate the performance in the year 2019. Particularly, we construct features from multi-modality data in the recent 60 trading days and apply the next day’s price movement as the label. Then, we apply a sliding window of each trading day and report the average result of 2019 in the experiment. In the trading strategy settings, we simply buy the forecasted “up”, sell the “down” equities, and keep no action on “neural” stocks. The trading percentage is allocated by the linear weights of predicted probability. Please note that there are many techniques for developing a trading strategy, which is beyond the scope of this paper. We ignore the transaction costs of all compared methods for simplicity and fairness in the experiment.

Compared methods and evaluation metrics We utilize the following widely used approaches as baselines to validate the effectiveness of our proposed method: Stock-LSTM [13], News-ATT [9], Stock-GAT [23], and Event-NTN [21]. All parameters are set based on their default suggestions in the paper. For instance, Stock-LSTM is set as two layers with a hidden size of 100 and 50. Our method has two variations: MAGNN-G and MAGNN-S, which only employ the inner-modality graph attention or the inter-modality source attention alone. MAGNN-all denotes the full version of our proposed techniques.

Table 1
The comparison of forecasting performance.

	Micro-F1	Macro-F1	Weighted-F1
Stock-LSTM	0.4540	0.4233	0.4489
News-ATT	0.4551	0.4551	0.4502
Stock-GAT	0.4656	0.4396	0.4654
Event-NTN	0.4720	0.4478	0.4718
MAGNN-G	0.4815	0.4607	0.4798
MAGNN-S	0.4813	0.4604	0.4793
MAGNN-all	**0.4838****	**0.4627****	**0.4825****

** indicates that the improvements are statistically significant for $p < 0.01$ judged by paired t-test.

For the evaluation, we apply the Micro-F1, Macro-F1 and Weighted-F1 score to measure the performance of forecasting accuracy. For the constructed portfolio, we employ asset accumulate return (A Return), average daily return (D Return) and widely-used Sharpe Ratio [24] as evaluation metrics. A Return is formulated as

$$\mathrm{AR}^t = \frac{1}{|S^{t-1}|} \sum_{i \in S^{t-1}} \frac{p_i^t - p_i^{t-1}}{p_i^{t-1}}, \tag{8}$$

where S^{t-1} denotes the set of stocks in portfolio at time $t-1$. p_i^t is the price for stock i at time t and $|\cdot|$ denotes the number of set items. Sharpe ratio (SR) is the average return earned in excess of the risk-free rate per unit of volatility, which is expressed as: $\mathrm{SR} = (R_p - R_f)/\sigma_p$ where R_p is the return of the portfolio, R_f is the risk-free rate, σ_p is the standard deviation of the portfolio’s excess return. We use 1-year China Government Bond Yield[4] as the risk free rate.

4.2. Financial forecasting

In this section, we evaluate the forecasting accuracy of financial time series, which is the main task of this paper. Table 1 reports the Micro-F1, Macro-F1 and Weighted-F1 score of each approach. ** denotes that the improvements are statistically significant for $p < 0.01$ judged by paired t-test.

The first four lines of Table 1 shows the classification result of compared baselines. It is clear that, Stock-LSTM and News-ATT are not satisfactory, demonstrating neither stock nor news alone could achieve optimal performance. Stock-GAT is slightly better than Stock-LSTM, proving the effectiveness of preserving graph structure in a time series forecasting model. In all baseline, Event-NTN is most competitive, which considerably outperforms New-ATT. The process of extracting structured events from raw news shows useful in learning representative embeddings. Line 5 and 6 display the performance of the variations of our proposed method. As we can see, MAGNN-S is similar to MAGNN-G. Both are better than the most competitive baselines. The validity of integrating multi-modality inputs in our task is strongly proved. It is essential to design an innovative model to learn from the above sources, which is the primary motivation of this paper. MAGNN-all outperforms all baselines, demonstrating its superiority in learning from multi-modality inputs for financial forecasting.

[3] https://www.investopedia.com/terms/a/a-shares.asp

[4] http://yield.chinabond.com.cn/

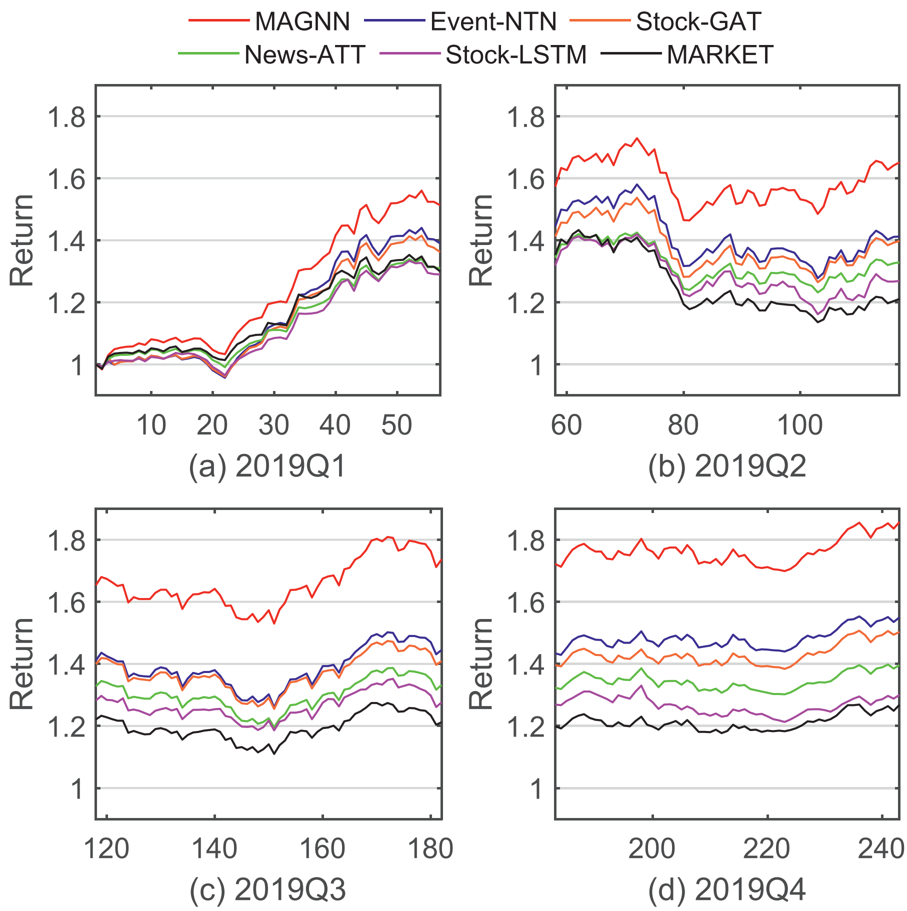

Fig. 3. The accumulated returns gained in the test set (2019) by the proposed method and compared baselines. For better illustration, we divide it into four quarters view.

Table 2
The return of portfolios with different method.

Methods	A Return	D Return	Sharpe ratio
Stock-LSTM	0.3002	0.0012	2.7919
News-ATT	0.3960	0.0015	3.5097
Stock-GAT	0.5035	0.0018	3.4506
Event-NTN	0.5507	0.0019	3.4467
MAGNN-G	0.6521	0.0021	3.5082
MAGNN-S	0.6775	0.0021	3.5561
MAGNN-all	**0.8571****	**0.0027****	**3.7619****

In this experiment, We would like to stress that the market trend prediction is very challenging and a small fraction of improvement can already bring a large amount of revenue in the financial industry. According to the practice from Marcos et al.,[25] even 0.005 improvements in the prediction accuracy is very difficult for new researchers, which could normally lead to over 12% excess profits. Our method improves the best baselines over 1% in Table 1 and consequently leads to near 30% profit improvements in the accumulated returns, as reported in Table 2. Therefore, we can safely claim that our proposed methods significantly outperform state-of-the-art baselines in the forecasting task.

4.3. Performance of the portfolio

In the constructed portfolio performance evaluation, we report the asset return (A Return), averaged daily return (D Return) and Sharpe Ratio first. Then, we present the accumulated return curve across the time interval of the test period. As described above, we buy the predicted "up" stocks and sell the "down" ones. The position is simply set linear to the probability of forecasting model outputs.

Table 2 reports the performance of investment portfolios constructed by our proposed method and other baselines. We can observe that, in all evaluation metrics, our proposed technique outperforms the baseline significantly. Particularly, Stock-LSTM and News-ATT achieve lower performance in "A Return" and "Sharpe Ratio", which indicates the poor profitable returns. By incorporating the knowledge graph and structured events, the return of Stock-GAT and Event-NTN is higher than classic baselines. The same phenomenon is observed in the sharpe ratio metrics. The last three rows display the result of our proposed method and its sub-models. MAGNN constantly performs better than all compared methods in three widely-used evaluation metrics. The effectiveness of our proposed method on constructing profitable portfolios is strongly proved.

To further evaluate the return of our proposed method throughout the test time interval, we examined the accumulated return in each trading day and report the compared results in Fig. 3. We can observe that Stock-LSTM is very close to the market CSI 300 index[5] throughout the year 2019. News-ATT are fared better than Stock-LSTM. Since the end of 2019Q1, our method leads the returns and enlarge the gap in 2019Q2 with compared baselines, which is noteworthy. We then conduct empirical studies with financial domain experts on the showcase of return curves. The reason appears to be that when the market goes down (in 2019Q2), our method could forecast the "down" signal in advance, which is learned from the implicitly lead-lag effects in multi-modality sources. As a re-

[5] http://www.csindex.com.cn/en/indices/index-detail/000300

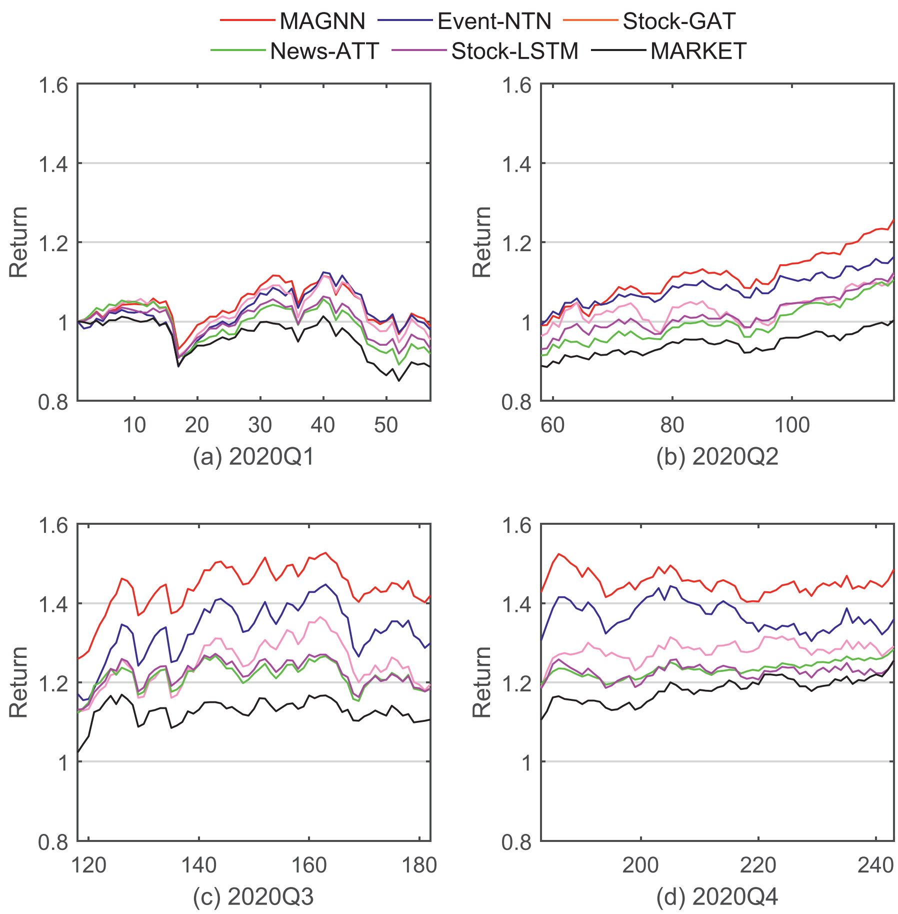

Fig. 4. The accumulated returns gained in the validation window (2020) by the proposed method and compared baselines.

sult, MAGNN performs the best performance ever since the end of 2019Q1 and leads the superiority until the end of 2019Q4, with over 80% of investment profits.

4.4. Performance generalization

In order to observe the performance generalization of our proposed method, we chose a longer duration to evaluate the portfolio's return compared with baselines. In particular, we train our model by the historical data during 2019 and then predict the price movements in 2020. The trading strategy is set as same as previous experiment and the trading percentage is also allocated by the linear weights of predicted probability. Fig. 4 report the performance of accumulated returns by the proposed method and compared baselines. As we can see, the Stock-LSTM is the lowest, which is very close to the market return at the end of 2020. While the News-ATT and Stock-GAT perform better that Stock-LSTM and Market, proving the effectiveness of including news and graph relations in stock price prediction task. In all baselines, Event-NTN is most competitive. Our method achieve the best result in the observation window and steadily leading the performance since the beginning of the second quarter of 2020. The result of model generalization experiment by training data in 2019 is consistent with the result of training the model with 2018, which demonstrates the effectiveness and generalization of our proposed method.

4.5. Interpretability of attention model

As described in the preliminary section, price movements of stocks can be seemed as the reaction of financial events or news, which is also related to its own historical performance. Thus, we need to include multi-modality (news, event, market) sources as the input of our model. However, different sources contribute differently. Inter-modality attention mechanism could automatically learn their weights in price prediction and so that achieve the state-of-the-art performance. Moreover, within each modality, the inner relation of different companies is also very important. For example, an event "Microsoft buys LinkedIn" was reported on Jun 13, 2016. Immediately, the price of LinkedIn raised 46.81% and Microsoft declined 3.2% on that day. Interestingly, the prices of Salesforce, which is the main competitor of LinkedIn, decreased over 6% in the following two weeks. Salesforce is not the direct subject of this event but it is also deep influenced. Inner-modality attention model could learn this graph-structured relations between the input sources and target prediction. Therefore, our method could help to predict stock price movement more accurately and the experiment results strongly demonstrate its superior performance.

Then, in order to explore the interpretability of our proposed method, we visualize the attention weights of both inner-modality graph attention and inter-modality source attention in Fig. 5. We locate each equity according to the predicted return (x-axis) and their situation in the heterogeneous graph or modality source (y-axis) in the heat map. Then, we color it by the averaged attention weights in the forecasting model.

Fig. 5 a displays the learned weights of InnGAT. As we can see, equities with approximately three to six neighborhood nodes generally contribute more important in the model. Besides, there is no noteworthy difference for node structures in different market situations (e.g., across the x-axis). The result proves that the return of a node with about four neighbors in the heterogeneous graph is

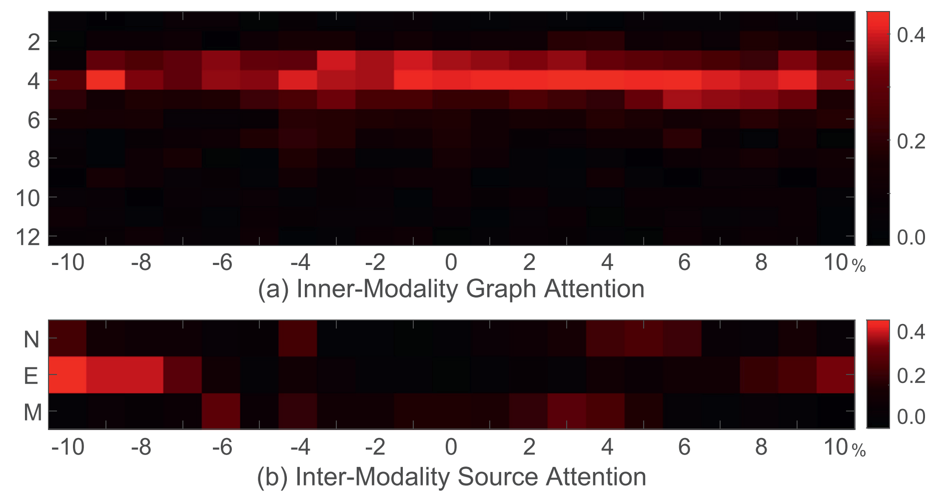

Fig. 5. Visualization of the attentional coefficients. *X*-axis denotes the daily return of equities. *Y*-axis denotes the number of neighbors in subgraph (a) and the modality type in (b), in which N,E,M denotes the news, events and market prices.

Fig. 6. The interface of our proposed MAGNN in web-based portfolio management system, which is deployed in a major financial service provider of China. We translate the key information by handwritten orange and underlined words.

more likely to be influenced by adjacent nodes, which means the lead-lag effects are more prominent in this situation.

A more interpretable output is observed in Fig. 5b. By visualizing each modality's attention weights, we find that all sources (N, E, and M) contribute importantly to the forecasting model, which strongly proves the declaration of this paper. In addition, news performs similarly in different market situations, with a slightly prominent small positive return. The same phenomenon is observed in market modality. On the contrary, events' contribution is significantly higher in large positive and negative returns than the small ones. The reason might be that the sharp change of equity is mostly driven by events, instead of routine news or price data. The result demonstrates the effects of multi-modality input and the proposed attentional model.

4.6. System implementation and deployment

We then deploy our method in real-world scenarios and evaluate its performance in the market tracking experiment. Fig. 6 shows the interface of the portfolio management system of our proposed method. In the homepage navigation view (Fig. 6a), we can investigate the forecasted "up" and "down" stocks, the constructed portfolio, and the leading events. The left navigator provides the effects of each modality sources, including news, events and market prices, and the lead-lag status of stocks on selected events. Fig. 6b reports the forecasting view on a typical equity Spring Airlines (Code: 601021) Ltd., which is China's first and North Asia's largest low fare airline. The upper part shows the ground-truth stock's price, and the lower part displays the predicted price movements ratio since Jan 01, 2020.

As we can see, our method successfully forecasted four significant fluctuations in advance of Spring Airlines. At last, we report the portfolio's performance details in Fig. 6c. The result shows that our method significantly improves the returns with over 60% of excess profit. In addition, by learning the lead-lag effects from multi-modality sources, our method could avoid large losses in the market, which decreases the maximum drawdown from −16.08% to −12.48%.

In the implementation, we employ distributed Scrapy as the web crawler, Redis as the in-memory database. The proposed model is written in Tensorflow on Python and requires two hours for training on two pieces of Telsa P100 GPU. The integrated portfolio management system is implemented by Spring Cloud micro-services and written in Java.

5. Related works

In this section, we introduce some works that are related to our research, including financial time series analysis and multi-modality graph neural network.

Financial time series analysis In recent decades, numerous works have proposed to forecast the financial time series [26]. Conventional approaches includes autoregressive model, moving average method, factor analysis [16], indicator optimization [27], etc. Afterwards, the machine learning techniques have been employed for stock price prediction, such as support vector machine [28], boosting trees [29], and neural network, especially for the RNN and LSTM [30]. Recently, some researchers demonstrated the effectiveness of leveraging unstructured text news and events to learn representative embeddings for stock price prediction [31,32]. Advanced transfer learning [33] and unsupervised learning [34] techniques are also introduced to learn the meaningful embeddings for time series analysis. However, existing approaches fail to learn the internal relations of stock movements on the lead-lag effects of events (or news). The majority of them only employ a single modality source for forecasting, which may dismiss much useful information.

Multi-modality and graph attention network Graph neural network (GNN) has shown its superior performance for representing graph-structured data [35,36]. The graph attention model improves the node representation by adjusting meaningful weights in the aggregation process with adjacent nodes [37], indicating the importance of corresponding nodes [38] and the attributed relations [39]. GNN with attention mechanism has shown its effectiveness in a wide range of fields, including finance [40], healthcare [41], computer vision [42], e-commerce [43,44] etc. Recently, some works explore to apply GNN to learning from multi-modality inputs, such as disease diagnosis by multi-modality images [45]. However, there are few studies on financial forecasting by multi-modality graph neural network.

6. Conclusion

In this paper, we propose a novel multi-modality graph neural network for financial time series forecasting. Our method addresses the key problem of price prediction in the financial industry, interpretable learning the lead-lag effects with informative source, by inner-modality graph attention and inter-modality source attention mechanism. We thoroughly evaluate the proposed method's effectiveness by comparing it with the state-of-the-art baselines on the massive historical datasets. In addition, we deploy the model in real-world applications, and the result proves that our work could avoid significant financial investment losses.

In conclusion, this is the first work to study the financial time series forecasting problem by advanced GNN techniques with informative sources, which may innovate more studies on both the computer science and finance communities. On the one hand, we extend the graph attention model to multi-modality scenarios; on the other hand, we improve financial forecasting with learning on alternative data.

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.

Acknowledgments

The work is supported by the National Key R&D Program of China (2018YFB2100801), the National Science Foundation of China (6210070838), the Shanghai "Scientific and Technological Innovation Action Plan" High and New Technology Projects (19511101300).

References

[1] Worldbank, Market capitalization of listed domestic companies, 2020, Accessed 16-March-2021, https://data.worldbank.org/indicator/CM.MKT.LCAP.CD.
[2] E.F. Fama, The behavior of stock-market prices, J. Bus. 38.1 (1965) 34–105 APA.
[3] M.L. O'Connor, The cross-sectional relationship between trading costs and lead/lag effects in stock & option markets, Financ. Rev. 34 (4) (1999) 95–117.
[4] A. Cowles 3rd, Can stock market forecasters forecast? Econometrica 1 (3) (1933) 309–324.
[5] S.J. Taylor, Modelling Financial Time Series[M], World Scientific, 2008.
[6] C.Q. Cao, R.S. Tsay, Nonlinear time-series analysis of stock volatilities[J], J. Appl. Econom. 7 (S1) (1992) S165–S185.
[7] R.D. Edwards, J. Magee, W.H. Bassetti, Technical Analysis of Stock Trends, 2012.
[8] D. Cheng, Y. Tu, Z. Niu, L. Zhang, Learning temporal relationships between financial signals, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 2641–2645.
[9] Z. Hu, W. Liu, J. Bian, X. Liu, T.-Y. Liu, Listening to chaotic whispers: a deep learning framework for news-oriented stock trend prediction, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 261–269.
[10] D. Tashiro, H. Matsushima, K. Izumi, H. Sakaji, Encoding of high-frequency order information and prediction of short-term stock price by deep learning, Quant. Finance 19 (9) (2019) 1499–1506.
[11] S. Bharathi, A. Geetha, Sentiment analysis for effective stock market prediction, Int. J. Intell. Eng. Syst. 10 (3) (2017) 146–154.
[12] Y. Xu, S.B. Cohen, Stock movement prediction from tweets and historical prices, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 1970–1979.
[13] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, Eur. J. Oper. Res. 270 (2) (2018) 654–669.
[14] S. Deng, N. Zhang, W. Zhang, J. Chen, J.Z. Pan, H. Chen, Knowledge-driven stock trend prediction and explanation via temporal convolutional network, in: Companion Proceedings of The 2019 World Wide Web Conference, 2019, pp. 678–685.
[15] D. Cheng, F. Yang, X. Wang, Y. Zhang, L. Zhang, Knowledge graph-based event embedding framework for financial quantitative investments, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 2221–2230.
[16] G. Ganeshapillai, J. Guttag, A. Lo, Learning connections in financial time series, in: International Conference on Machine Learning, 2013, pp. 109–117.
[17] W.S. Chan, Stock price reaction to news and no-news: drift and reversal after headlines, J. Financ. Econ. 70 (2) (2003) 223–260.
[18] R. Hussein, D. Yang, P. Cudré-Mauroux, Are meta-paths necessary? Revisiting heterogeneous graph embeddings, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 437–446.
[19] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: pre-training of deep bidirectional transformers for language understanding, 2018arXiv:1810.04805
[20] S. Saha, et al., Open information extraction from conjunctive sentences, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 2288–2299.
[21] X. Ding, K. Liao, T. Liu, Z. Li, J. Duan, Event representation learning enhanced with external commonsense knowledge, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 4896–4905.
[22] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980(2014).
[23] P. Velickovic, A. Casanova, P. Lio, G. Cucurull, A. Romero, Y. Bengio, Graph attention networks, in: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018.
[24] O. Ledoit, M. Wolf, Robust performance hypothesis testing with the sharpe ratio, J. Empir. Finance 15 (5) (2008) 850–859.
[25] M.L. De Prado, Advances in Financial Machine Learning, John Wiley & Sons, 2018.
[26] T.G. Andersen, R.A. Davis, J.-P. Kreiß, T.V. Mikosch, Handbook of Financial Time Series, Springer Science & Business Media, 2009.
[27] Z. Li, D. Yang, L. Zhao, J. Bian, T. Qin, T.-Y. Liu, Individualized indicator for all: stock-wise technical indicator optimization with stock embedding, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 894–902.
[28] L.-J. Cao, F.E.H. Tay, Support vector machine with adaptive parameters in financial time series forecasting, IEEE Trans. Neural Netw. 14 (6) (2003) 1506–1518.
[29] A. Agapitos, A. Brabazon, M. O'Neill, Regularised gradient boosting for financial time-series modelling, Comput. Manag. Sci. 14 (3) (2017) 367–391.
[30] A. Sagheer, M. Kotb, Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems, Sci. Rep. 9 (1) (2019) 1–16.
[31] D. Zhou, L. Zheng, Y. Zhu, J. Li, J. He, Domain adaptive multi-modality neural attention network for financial forecasting, in: Proceedings of The Web Conference 2020, 2020, pp. 2230–2240.

[32] O.B. Sezer, M.U. Gudelek, A.M. Ozbayoglu, Financial time series forecasting with deep learning: a systematic literature review: 2005–2019, Appl. Soft Comput. 90 (2020) 106181.
[33] R. Ye, Q. Dai, Implementing transfer learning across different datasets for time series forecasting, Pattern Recognit. 109 (2021) 107617.
[34] H. Wang, Q. Zhang, J. Wu, S. Pan, Y. Chen, Time series feature learning with labeled and unlabeled data, Pattern Recognit. 89 (2019) 55–66.
[35] N. Dehmamy, A.-L. Barabási, R. Yu, Understanding the representation power of graph neural networks in learning graph topology, in: Advances in Neural Information Processing Systems, 2019, pp. 15413–15423.
[36] F. Manessi, A. Rozza, M. Manzo, Dynamic graph convolutional networks, Pattern Recognit. 97 (2020) 107000.
[37] B. Knyazev, G.W. Taylor, M. Amer, Understanding attention and generalization in graph neural networks, in: Advances in Neural Information Processing Systems, 2019, pp. 4202–4212.
[38] N. Park, A. Kan, X.L. Dong, T. Zhao, C. Faloutsos, Estimating node importance in knowledge graphs using graph neural networks, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 596–606.
[39] S. Feng, C. Xu, Y. Zuo, G. Chen, F. Lin, J. XiaHou, Relation-aware dynamic attributed graph attention network for stocks recommendation[J], Pattern Recognit. (2021) 108–119, doi:10.1016/j.patcog.2021.108119.
[40] D. Cheng, Z. Niu, Y. Zhang, Contagious chain risk rating for networked-guarantee loans, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 2715–2723.
[41] E. Choi, M.T. Bahadori, L. Song, W.F. Stewart, J. Sun, Gram: graph-based attention model for healthcare representation learning, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 787–795.
[42] J. Wu, S.-h. Zhong, Y. Liu, Dynamic graph convolutional network for multi--video summarization, Pattern Recognit. 107 (2020) 107382.
[43] D. Cheng, S. Xiang, C. Shang, Y. Zhang, F. Yang, L. Zhang, Spatio-temporal attention-based neural network for credit card fraud detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, 34, 2020, pp. 362–369.
[44] D. Cheng, X. Wang, Y. Zhang, et al., Graph neural network for fraud detection via spatial-temporal attention[J], IEEE Trans. Knowl. Data Eng. (2020) https://ieeexplore.ieee.org/abstract/document/9204584/.
[45] X. Fang, Z. Liu, M. Xu, Ensemble of deep convolutional neural networks based multi-modality images for Alzheimer's disease diagnosis, IET Image Proc. 14 (2) (2019) 318–326.

Dawei Cheng is an assistant professor with the department of computer science and technology, Tongji University, Shanghai, China. Before that, Dawei was a post-doctoral associate at MoE key lab of artificial intelligence, department of computer science and engineering, Shanghai Jiao Tong University. He received the Ph.D. Degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2018. His research interests include graph learning, data mining and machine learning.

Fangzhou Yang is a senior research scientist with the laboratory of artificial intelligence (Seek-Data team), Emoney Inc., Shanghai, China. Fangzhou received his master degree from Technical University of Berlin and Shanghai Jiao Tong University. Before that, he received his B.Sc. degree in computer science from Shanghai Jiao Tong University. His research interests include pattern recognition and data mining in finance, portfolio management and quantitative investment.

Sheng Xiang is a Ph.D. candidate in the Center for Artificial Intelligence, major in Computer Science, University of Technology Sydney, Australia. He received his B.Sc. degrees in Bioinformatics Engineering from Shanghai Jiao Tong University. His research interests include big graph processing and analysis, query processing on data stream, uncertain data and graphs.

Jin Liu is a senior research scientist with the laboratory of artificial intelligence (Seek-Data team), Emoney Inc., Shanghai, China. He received his BSc and Master degree in computer science and engineering from Wuhan University, Hubei, China. His research interests include financial data mining, factor analysis and machine learning.