Stock Movement Prediction Based on Bi-Typed Hybrid-Relational Market Knowledge Graph via Dual Attention Networks

Yu Zhao , Huaming Du , Ying Liu, Shaopeng Wei , Xingyan Chen , Fuzhen Zhuang , *Member, IEEE*, Qing Li , *Member, IEEE*, and Gang Kou , *Member, IEEE*

Abstract—Stock Movement Prediction (SMP) aims at predicting listed companies' stock future price trend, which is a challenging task due to the volatile nature of financial markets. Recent financial studies show that the momentum spillover effect plays a significant role in stock fluctuation. However, previous studies typically only learn the simple connection information among related companies, which inevitably fail to model complex relations of listed companies in real financial market. To address this issue, we first construct a more comprehensive Market Knowledge Graph (MKG) which contains bi-typed entities including listed companies and their associated executives, and hybrid-relations including the explicit relations and implicit relations. Afterward, we propose DᴀɴSᴍᴘ a novel Dual Attention Networks to learn the momentum spillover signals based upon the constructed MKG for stock prediction. The empirical experiments on our constructed datasets against nine SOTA baselines demonstrate that the proposed DᴀɴSᴍᴘ is capable of improving stock prediction with the constructed MKG.

Index Terms—Bi-typed hybrid-relational market knowledge graph, dual attention network, stock movement prediction

◆

1 INTRODUCTION

Sᴛᴏᴄᴋ Movement Prediction (SMP) is a hot topic in *Fintech* area since investors continuously attempt to predict the stock future trend of listed companies for seeking maximized profit in the volatile financial market [1], [2], [3], [4]. The task has spurred the interest of researchers over the years to develop better predictive models [4]. In particular, the application of machine learning approaches yields a promising performance for SMP task [5], [6]. Previous studies in both finance and AI research fields predicting a stock movement rely on

- *Yu Zhao, Ying Liu, Xingyan Chen, and Qing Li are with the Financial Intelligence and Financial Engineering Key Laboratory of Sichuan Province, Institute of Digital Economy and Interdisciplinary Science Innovation, Southwestern University of Finance and Economics, Chengdu 610074, China. E-mail: {zhaoyu, xychen, liq_t}@swufe.edu.cn, liuying2200@gmail.com.*
- *Huaming Du, Shaopeng Wei, and Gang Kou are with the School of Business Administration, Southwestern University of Finance and Economics, Chengdu 610074, China. E-mail: {dhmfcc, 2191201z5003}@smail.swufe.edu.cn, kougang@swufe.edu.cn.*
- *Fuzhen Zhuang is with the Institute of Artificial Intelligence, Beihang University, Beijing 100190, China, and also with Zhongguancun Laboratory, Beijing 100190, China. E-mail: zhuangfuzhen@buaa.edu.cn.*

time-series analysis techniques using its own historical prices (e.g., *opening price, closing price, volume*, etc) [7], [8]. According to the Efficient Market Hypothesis (ETH) that implies financial market is informationally efficient [9], therefore, besides these stock trading factors, other researchers mine more indicative features from its outside-market data such as web media [1], including news information [10], [11], [12] and social media [13], [14], [15], while ignoring the stock fluctuation diffusion influence from its related companies, which is also known as momentum spillover effect [16] in finance.

Recent studies attempt to model stock momentum spillover via Graph Neural Networks (GNN) [17]. However, most of them only consider the simple explicit relations among related companies [5], [6], [18], [19], which inevitably fail to model the complex connections of listed companies in real financial market, such as the implicit relation, and the associated executives-based meta-relations [20], [21].

To address this issue, we construct a more comprehensive Market Knowledge Graph (MKG), which consists of a considerable amount of triples in the form of (*head entity*, *relation*, *tail entity*), indicating that there exists a relation between the two entities. Different from previous graphs in other SOTA works [3], [5], [6], [19], [22], [23], the newly constructed MKG develops two essential characteristics: *(1) Bi-typed*, i.e., containing the significant associated executive entities aside from the ordinary company entities; *(2) Hybrid-relational*, i.e., providing an additional implicit relation among listed companies aside from their typical explicit relations. Fig. 1 shows a toy example of MKG (See Section 3.1 for more details).

Afterward, to learn the stock[1] momentum spillover signals on such bi-typed hybrid-relational MKG for stock

---

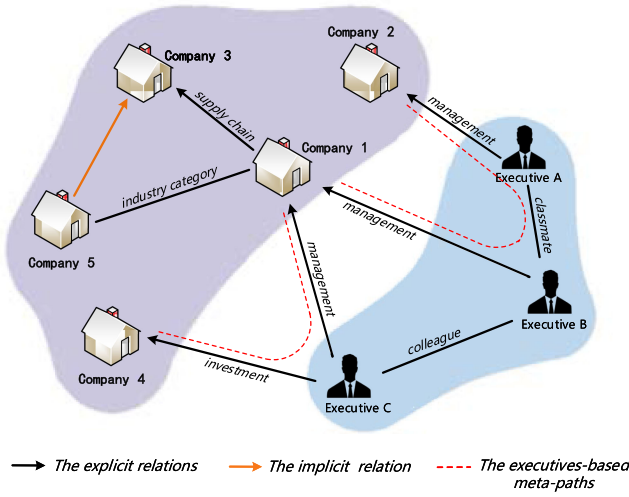1. The term "listed company" and "stock" are used interchangeably.

Fig. 1. Example of a bi-typed (i.e., *listed companies*, and *executives*) hybrid-relational (i.e., *explicit relations*, and *implicit relation*) market knowledge graph (MKG). The relational information in MKG is essential for stock prediction but has not been well utilized in previous works.

movement prediction, we pertinently propose a novel **D**ual **A**ttention **N**etworks (DANSMP), as shown in Fig. 2 II. Specifically, the proposed model DANSMP is equipped with dual attention modules that are able to learn the inter-class interaction among listed companies and associated executives, and their own complex intra-class interaction alternately. Different from previous methods that can only model homogeneous stock graph [3], [22] or heterogeneity of stock explicit relations [23], [24], [25], our method is able to learn bi-typed heterogeneous entities and hybrid-relations in newly constructed market graph of stock for its spillover effects. The comprehensive comparison between the existing state-of-the-art (SOTA) methods with our newly proposed DANSMP model in terms of used market signals and main ideas is shown in Table 1, demonstrating the distinguished advantage of our work.

We collect public data and construct two new datasets (named *CSI100E*[2] and *CSI300E*) based on Chinese Stock Index to evaluate the proposed method, since no existing benchmark datasets can satisfy our need. Aside from the typical stock historical prices and media news, our newly published benchmark datasets also provide rich market knowledge graph as mentioned above. The empirical experimental results on CSI100E and CSI300E against nine SOTA methods demonstrate the better performance of our model DANSMP with MKG. The ablation studies reaffirm that the performance gain mainly comes from the use of the associated executives, and additional implicit relation among companies in MKG via the proposed DANSMP.

The contributions of this paper are threefold:

- To model stock momentum spillover via the complex relations among companies in real market, we first construct a novel market knowledge graph. To the best of our knowledge, this study is the first attempt to explore such bi-typed hybrid-relational knowledge graph of stock via heterogeneous GNNs for its spillover effects.

- We then propose DANSMP, a novel Dual Attention Networks to learn the stock momentum spillover features based on the newly constructed bi-typed hybrid-relational MKG for stock prediction, which is also a non-trivial and challenging task.
- We propose two new benchmark datasets (CSI100E and CSI300E) to evaluate our method, which are also expected to promote Fintech research field further. The empirical experiments on our constructed datasets demonstrate our method can successfully improve stock prediction with bi-typed hybrid-relational MKG via the proposed DANSMP.[3]

2 RELATED WORK

In this section, we evaluate the existing relevant research on stock prediction. Stock movement prediction (SMP) has received a great deal of attention from both investors and researchers since it helps investors to make good investment decisions [26], [27], [28]. In general, traditional SMP methods mainly can be categorized into two classes: technical analysis and fundamental analysis, according to the different types of the available stock own information they mainly used. Another major aspect for yielding better stock prediction is to utilize the stock connection information [3], [6], [18], [22]. We review them in the following.

2.1 Technical Analysis

Technical analysis takes time-series historical market data of a stock, such as trading price and volume, as features to make prediction [25], [29]. The basic idea behind this type of approach is to discover the hidden trading patterns that can be used for SMP. Most recent methods of this type predict stock movement trend using deep learning models [7], [24], [30]. To further capture the long-term dependency in time series, the Recurrent Neural Networks (RNN) especially Long Short-Term Memory networks (LSTM) have been usually leveraged for prediction [31]. [30] presented a deep learning framework for stock forecasting using stacked auto-encoders and LSTM. [24] studied the usage of LSTM networks to predict future trends of stock prices based on the price history, alongside with technical analysis indicators. [7] proposed an end-to-end hybrid neural networks that leverage convolutional neural networks (CNNs) and LSTM to learn local and global contextual features respectively for predicting the trend of time series. [32] proposed a state frequency memory recurrent network to capture the multi-frequency trading patterns for stock price prediction. [8] proposed to employ adversarial training and add perturbations to simulate the stochasticity of price variable, and train the model to work well under small yet intentional perturbations. [33] decomposed financial times series into inherent mode functions of multiple levels and used the attention-based long short-term memory to prediction. [34] proposed an attention-based long short-term memory model to predict stock price movement and made
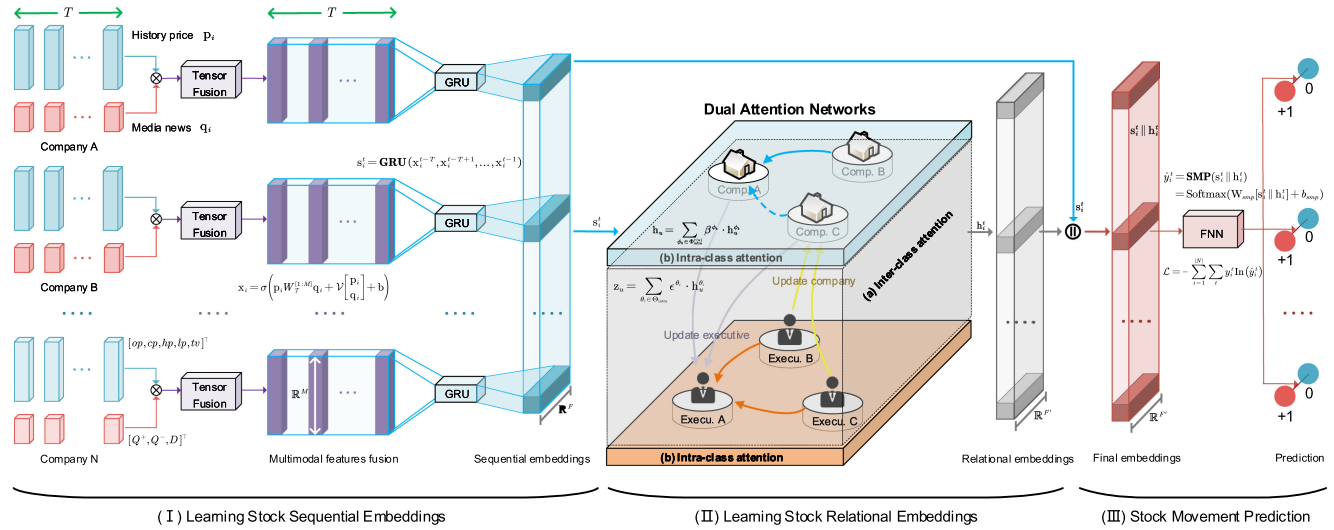
Fig. 2. The overall framework of the proposed method. *(I) Learning Stock Sequential Embeddings* based on Tensor Fusion and GRU. Tensor Fusion is the Neural Tensor Network (NTN) to learn the fusion of the technical indicators vector $\mathbf{p}_i$ and media news sentiment features $\mathbf{q}_i$. The GRU is designed to learn the sequential embedding $\mathbf{s}_i^t$. *(II) Learning Stock Relational Embeddings* by a dual mechanism to model the mutual affects and inner interaction among the bi-typed entities (i.e., companies and executives) alternately, including: **(a)** inter-class attention, and **(b)** intra-class attention. The former aims to deal with the interaction between listed companies and their associated executives and the latter aims to learn the interaction among the same type of entities. *(III) Stock Movement Prediction* via Feed-forward Neural Network (FNN) with the learned firm embeddings.

trading strategies based on the stock price data and some technical indicators. [35] proposed a hybrid model that consists of stochastic recurrent networks, the sequence-to-sequence architecture, the attention mechanism, and convolutional neural network to predict trends of financial markets. Despite their achieved progress, however, technical analysis faces an issue that it is incapable of unveiling the rules that govern the fluctuation of the market beyond stock price data.

2.2 Fundamental Analysis

On the contrary, fundamental analysis takes advantage of information from outside market price data, such as economic and financial environment, and other qualitative and quantitative factors [2], [36]. Many methods are proposed to explore the relation between stock market and web media, e.g., news information, and social media opinion [1], [37]. For instance, [10] mined text information from Wall Street Journal for SMP. [37] presented a deep learning method for stock

TABLE 1
Comparison Between Several SOTA Methods and the Proposed Model in Terms of Used Market Signals and Main Ideas

Literature	Main ideas	Market	Metrics	Method	GNN Types	Stock Market Signals				
						Historical Data	Media News	Market Knowledge Graph		
								Explicit Relation	Implicit Relation	Executives Relation
EB-CNN [45]	• Neural tensor network for learning event embedding • Deep CNN to model the combined influence	S&P500	DA, MCC, Profit	Open IE, CNN	-	✗	✓	✗	✗	✗
SFM [32]	• Extending LSTM by decomposing the hidden memory states • Modeling the latent trading patterns with multiple frequencies	NASDAQ	Average square error	LSTM, DFT	-	✓	✗	✗	✗	✗
Chen's [22]	• Single investment relation	CSI300	DA	LSTM, GCN	Homogeneous GNNs	✓	✗	✓	✗	✗
TGC [5]	• Single historical data • Dynamically adjust predefined firm relations	NASDAQ, NYSE	MSE, MRR, IRR	LSTM, Temporal Graph Convolution	Homogeneous GNNs	✓	✗	✓	✗	✗
HATS [42]	• Hierarchical aggregate different types of firm relational data	S&P500	SR, F1, DA, Return	GAT	Homogeneous GNNs	✓	✗	✓	✗	✗
ALBERT+eventHAN [46]	• ALBERT enhanced event representations • Event-enhanced hierarchical attention network	S&P500, DOW, NASDAQ	DA, Annualized return	Open IE, ALBERT, HAN	-	✗	✓	✗	✗	✗
MAN-SF [44]	• Multi-modal market information • Hierarchical graph attention method	S&P500	F1, MCC	GAT, GRU	Homogeneous GNNs	✓	✓	✓	✗	✗
STHAN-SR [23]	• A neural hypergraph architecture for stock selection • Temporal hawkes attention mechanism	NASDAQ, NYSE, TSE	SR, IRR, NDCG5	Hypergraph Convolution	Homogeneous GNNs	✓	✗	✓	✗	✗
AD-GAT [3]	• Multi-modal market information • Attribute-driven graph attention network	S&P500	DA, AUC	GAT	Homogeneous GNNs	✓	✓	✗	✓	✗
DANSMP (ours)	• Multi-modal market information • Bi-typed hybrid-relation data • Dual attention network	CSI100E, CSI300E	DA, AUC, SR, IRR	Neural Tensor Network, GRU, Dual Attention Network	Heterogeneous GNNs	✓	✓	✓	✓	✓

TABLE 2
Description of the Information Data

Information data	Features
Entities	companies, executives.
Relations	explicit relations (industry category, supply chain, business partnership, investment), implicit relation.
Historical price	opening price (op), closing price (cp), highest price (hp), lowest price (lp), trade volume (tv).
Media news	positive media sentiment of a stock $Q(i)^+$, negative media sentiment of a stock $Q(i)^-$, media sentiment divergence of a stock $D(i)$.

prediction using numerical and textual information. [38] proposed a deep learning method for stock market prediction from financial news articles. [36] put forward a novel deep generative model jointly exploiting text and price signals for this task. [11] presented a hierarchical complementary attention network for predicting stock price movements with news. [12] proposed a multimodal event-driven LSTM model for stock prediction using online news. Some researchers mined market media news via analyzing its sentiment, and used it for SMP [39], [40]. For instance, [13] analyzed twitter mood to predict the stock market. [14] exploited topics based on twitter sentiment for stock prediction. [15] incorporated the sentiments of the specific topics of the company into the stock prediction model using social media. [41] investigated the sentiment of annual disclosures of companies in stock markets to forecast volatility. [40] proposed a multimodal method that takes CEO's vocal features, such as emotions and voice tones, into consideration. In this paper, we extract the signals from stock historical prices and media news sentiments as the sequential embeddings of stocks.

2.3 Stock Relations Modeling

Recent SMP studies take stock relations into consideration [3], [18], [19], [22]. For instance, [22] proposed to incorporate corporation relationship via graph convolutional neural networks for stock price prediction. [5] captured the stock relations in a time-sensitive manner for stock prediction. [42] proposed a hierarchical attention network for stock prediction using relational data. [43] presented a multi-task recurrent neural network (RNN) with high-order Markov random fields (MRFs) to predict stock price movement direction using stock's historical records together with its correlated stocks. [19] proposed a LSTM Relational Graph Convolutional Network (LSTM-RGCN) model, which models the connection among stocks with their correlation matrix. [6] encoded multiple relationships among stocks into graphs based on financial domain knowledge and utilized GCN to extract the cross effect based on these pre-defined graphs for stock prediction. [18] proposed a spatio-temporal hypergraph convolution network for stock movement forecasting. [3] proposed to model the momentum spillover effect for stock prediction via attribute-driven graph attention networks. Despite the substantial efforts of these SOTA methods, surprisingly, most of them only focus on modeling the momentum spillover via the explicit relations among stocks, while ignoring their complex relations in real market.

Table 1 summarizes the key advantages of our model, comparing with a variety of previous state-of-the-art (SOTA) studies in terms of the used market signals, their methods and GNN types. (1) Different from previous studies, our method takes advantage of all three types of stock market signals, including stock historical data, media news, and market knowledge graph. In particular, we construct a more comprehensive heterogeneous market graph that contains explicit relations, implicit relations and executive relations. (2) Different from most existing models that can only model homogeneous stock graph [3], [22], or heterogeneity of stock explicit relations [5], [6], [23], [44], which fall down in modeling heterogeneity of entities in real market, we propose a novel dual attention networks that is able to model bi-typed heterogeneous entities and hybrid-relations in newly constructed market graph of stock for its spillover effects. (3) To the best of our knowledge, this work is the first attempt to study stock movement prediction via heterogeneous GNNs.

3 MARKET SIGNALS

In this section, we introduce the significant signals of stocks in real financial market. We first give the details of the newly proposed bi-typed hybrid-relational market knowledge graph. Afterward, we introduce the historical price data and media news of stocks. Most of previous works focus on partial financial market information, which makes their modeling insufficient. In this work, we take advantage of all three types of market data, fusing numerical, textual, and relational data together. The features of each data used in this study are summarized in Table 2.

3.1 Bi-Typed Hybrid-Relational Market Knowledge Graph (MKG)

3.1.1 Bi-Typed Entities

Most existing SMP methods solely learn from company relationships in market [6], [19], [42]. In fact, in most stock markets there are also significant associated executives for listed companies, with rich associative information about these companies [20], [21]. Hence, our constructed MKG contains not only company entities but also executive entities. The executive entities can act as the intermediary among companies to build the meta-relations involved in company entities (e.g., *Company-Executive-Company (CEC)*, *Company-Executive-Executive-Company (CEEC)*). For example, in Fig. 1 we show the associated executives of the companies sampled from Chinese Stock Index. The stock spillover signals can pass from neighboring companies to a target company through the meta-relations established by their connected executives, such as Company 4-Executive C-Company 1; Company 2-Executive A $\overset{classmate}{\longleftrightarrow}$ Executive B-Company 1. In sum, the newly constructed MKG contains bi-typed entities, i.e., *listed companies* and their *associated executives*.

3.1.2 Hybrid-Relations

Most existing methods only take the explicit relations among companies, such as *industry category, supply chain, business partnership* and *investment*, into consideration [22]. However, the limited explicit company relationships are always insufficient for market knowledge graph construction due to the complex nature of financial market. To solve the MKG incompleteness issue, here we propose an attribute-driven method to conduct MKG completion by inferring missing implicit correlative relation among stocks, which employs stocks attributions.

Specifically, the attribute-driven implicit unobserved relation is calculated based on the features from both its historical prices and news information filtered by a normalized threshold. A single-layer feed-forward neural network is adopted to calculate the attention value α_{ij}^t between company i and j for inferring their implicit relation.

$$\alpha_{ij}^t = \text{LeakyRelu}\Big(\mathbf{u}^\top [\mathbf{s}_i^t \parallel \mathbf{s}_j^t]\Big), \tag{1}$$

where $\mathbf{s}_i^t$ and $\mathbf{s}_j^t$ are fused market signals of i and j, which are calculated by the Equation (6) (Section 4.1). $\parallel$ denotes the concatenate operation. $\mathbf{u}$ denotes the learnable matrix and LeakyReLU is a nonlinearity activation function. Borrowing gate mechanism in [3], we set an implicit relation between i and j if $\alpha_{ij}^t > \eta$. η denotes a pre-defined threshold. In short, the constructed MKG contains hybrid-relations, i.e., *explicit relations* and *implicit relation*.

3.2 Historical Price and Media News

3.2.1 Technical Indicators

Transactional data is the main manifestation of firms' intrinsic value and investors' expectations. We collect the daily stock price and volume data, including *opening price (op), closing price (cp), highest price (hp), lowest price (lp), and trade volume (tv)*. In order to better compare and observe the fluctuation of stock price, the stock price is transferred to the return ratio, and the trade volume is transferred to the turnover ratio before being fed into our model. The return ratio is an index reflecting the level of stock return, the higher the return ration is; the better the profitability of the stock is. The turnover is the total value of stocks traded over a period of time; the higher the share turnover could indicate that the share haves good liquidity. $\mathbf{p}_i \in \mathbb{R}^5$ indicates the technical indicators of company i, as follows:

$$\mathbf{p}_i = [op(i),\ cp(i),\ hp(i),\ lp(i),\ tv(i)]^\top. \tag{2}$$

3.2.2 Sentiment Signals

Modern behavioral finance theory [1] believes that investors are irrational, tending to be influenced by the opinions expressed in the media. Media sentiment reflects investors' expectations concerning the future of a company or the whole stock market, resulting in the fluctuations of stock price. To capture media sentiment signals, we extract the following characteristics: *positive media sentiment, negative media sentiment* and *media sentiment divergence* [12]. They are denoted respectively as follows:

$$Q(i)^+ = \frac{N(i)^+}{N(i)^+ + N(i)^-},$$

$$Q(i)^- = \frac{N(i)^-}{N(i)^+ + N(i)^-},$$

$$D(i) = \frac{N(i)^+ - N(i)^-}{N(i)^+ + N(i)^-}, \tag{3}$$

where $N(i)^+$ and $N(i)^-$ are the sum of the frequency of each positive and negative sentiment word found in the financial news articles of company i, respectively. $D(i)$ denotes the sentiment divergence. Since many negative sentiment words in the general sentiment dictionary no longer express negative emotional meanings in the financial field, we resort to a finance-oriented sentiment dictionary created in previous study [27]. $\mathbf{q}_i \in \mathbb{R}^3$ indicates the news sentiment signals of company i.

$$\mathbf{q}_i = [Q(i)^+,\ Q(i)^-,\ D(i)]^\top. \tag{4}$$

Note that we do not have everyday news for all companies since the randomness of the occurrence of media news. In order to make the technical indicators aligned with the media sentiment signals and keep pace with the real situation, the sentiment feature $\mathbf{q}_i$ of the firm i is assigned to zero on the day when there are no any media news about it.

Note that, to link the financial news with a company, we first match the title of the financial news with the company name, abbreviation and ticker symbol, and then match the content of the financial news with the company in the same way. In this way, the news can be well linked to the corresponding company through this two-stage matching method. To handle multiple company news on the same day, we first extract sentiment indicators from the company's news on the same day, and then we average the sentiment indicators from the news on the same day and finally obtain the company's market signals on that day.

4 METHODOLOGY

In this section, we introduce the details of our proposed method. Fig. 2 gives an overview of the proposed framework. (I) First, the stock sequential embeddings are learned with historical price and media news via multi-modal feature fusion and sequential learning. (II) Second, a Dual Attention Networks is proposed to learn the stock relational embeddings based upon the constructed MKG. (III) Last, the combinations of sequential embeddings and relational embeddings are utilized to make stock prediction.

4.1 Learning Stock Sequential Embeddings

The stocks are influenced by multi-modal time-series market signals. Considering the strong temporal dynamics of stock markets, the historical state of the stock is useful for predicting its future trend. Due to the fact that the influence of market signals on the stock price would last for some time, we should consider the market signals in the past couple of days when predicting stock trend $\hat{y}_i^t$ of company i at date t. We first capture the multimodal interactions of technical indicators and sentiment signals. We then feed the fused features into a one-layer GRU and take the last hidden

state as the sequential embedding of stock i which preserves the time dependency, as shown in Fig. 2 I.

4.1.1 Multimodal Features Fusion

To learn the fusion of the technical indicators vector $\mathbf{p}_i$ and media news sentiment features $\mathbf{q}_i$, we adopt a Neural Tensor Network which replaces a linear neural network layer with a M-dimensional bilinear tensor layer that directly relate the two features across multiple dimensions. The fused daily market signals[4] of stock i, $\mathbf{x}_i \in \mathbb{R}^M$, are calculated by tensor-based formulation

$$\mathbf{x}_i = \sigma\left(\mathbf{p}_i W_{\mathcal{T}}^{[1:M]} \mathbf{q}_i + \mathcal{V}\begin{bmatrix}\mathbf{p}_i \\ \mathbf{q}_i\end{bmatrix} + \mathbf{b}\right), \tag{5}$$

σ is an activation function, $W_{\mathcal{T}}^{[1:M]} \in \mathbb{R}^{5 \times 3 \times M}$ is a trainable tensor, $\mathcal{V} \in \mathbb{R}^{8 \times M}$ is the learned parameters matrix and $\mathbf{b} \in \mathbb{R}^M$ is the bias vector. Three parameters are shared by all stocks.

4.1.2 Sequential Learning

We feed the fused daily market signals in the past T days into the GRU to learn its sequential embedding $\mathbf{s}_i^t$, as follows:

$$\mathbf{s}_i^t = \mathbf{GRU}\left(\mathbf{x}_i^{t-T}, \mathbf{x}_i^{t-T+1}, \dots, \mathbf{x}_i^{t-1}\right), \tag{6}$$

where $\mathbf{s}_i^t \in \mathbb{R}^F$ denotes the last hidden state of GRU. F is the hidden size of GRU.

4.2 Learning Stock Relational Embeddings via Dual Attention Networks

In real market, the stock fluctuation is partially affected by its related stocks which is known as momentum spillover effect in finance [16]. In this section, based upon our newly constructed bi-typed hybrid-relational MKG, we propose a Dual Attention Networks to learn the relational embeddings of stocks that represent their received spillover signals. Specifically, we employ a *dual mechanism* to model the mutual affection and inner influence among the *bi-typed* entities (i.e., companies and executives) alternately, including inter-class interaction and intra-class interaction, as shown in Fig. 2 II.

4.2.1 Inter-Class Attention Networks

The inter-class attention aims to deal with the interaction between listed companies and their associated executives, as shown in Fig. 2 II-a. Since they are different types of entities, their features usually lie in different space. Hence, we first project their embeddings into a common spaces. Specifically, for a company entity $u \in \mathcal{E}_1$ with type $\tau(u)$ and an executive entity $v \in \mathcal{E}_2$ with type $\tau(v)$, we design two type-specific matrices $\mathbf{W}^{\tau(\cdot)}$ to map their features $\mathbf{h}_u, \mathbf{h}_v$ into a common space.

$$\mathbf{h}'_u = \mathbf{W}^{\tau(u)}\mathbf{h}_u,$$
$$\mathbf{h}'_v = \mathbf{W}^{\tau(v)}\mathbf{h}_v, \tag{7}$$

where $\mathbf{h}'_u \in \mathbb{R}^{F'}$ and $\mathbf{h}_u \in \mathbb{R}^F$ denote the original and transformed features of the entity u, respectively. $\mathcal{E}_1$ and $\mathcal{E}_2$ are

the sets of listed companies and the executives, respectively. Here, the original vectors of company entities ($\mathbf{h}_u, \mathbf{h}_v$) are initialized by learned sequential embeddings ($\mathbf{s}_u, \mathbf{s}_v$) learned in Section 4.1, which can bring rich semantic information in downstream learning. The initial features of executives are then simply an average of the features of the companies which they work for.

We assume that the target company entity u connects with other executives via a relation $\theta_i \in \Theta_{\text{inter}}$ which denotes the set of inter-class relations, so the neighboring executives of a company u with relation θ_i can be defined as $\mathcal{N}_{\text{inter}}^{\theta_i}(u)$. For entity u, different types of inter-class relations contribute different semantics to its embeddings, and so do different entities with the same relation. Hence, we employ attention mechanism here in entity-level and relation-level to hierarchically aggregate signals from other types of neighbors to target entity u.

We first design entity-level attention to learn the importance of entities within a same relation. Then, to learn importance $e_{uv}^{\theta_i}$ which means how important an executive v for a company u under a specific relation θ_i, we perform self-attention [47] on the entities

$$\begin{aligned} e_{uv}^{\theta_i} &= att_{node}(\mathbf{h}_u, '\mathbf{h}'_v; \theta_i) \\ &= \text{LeakyRelu}(\mathbf{a}_{\theta_i}^\top \cdot [\mathbf{h}'_u \| \mathbf{h}'_v]), \end{aligned} \tag{8}$$

where $\mathbf{h}'_u$ and $\mathbf{h}'_v$ are the transformed representations of the node u and v. $\mathbf{a}_{\theta_i} \in \mathbb{R}^{2F'}$ is a trainable weight vector. $\|$ denotes the concatenate operation. LeakyReLU is a nonlinearity activation function. To make $e_{uv}^{\theta_i}$ comparable over different entities, we normalize it using the softmax function.

$$\gamma_{uv}^{\theta_i} = \text{softmax}_v(e_{uv}^{\theta_i}) = \frac{\exp(e_{uv}^{\theta_i})}{\sum_{\bar{v} \in \mathcal{N}_{\text{inter}}^{\theta_i}(u)} \exp(e_{u\bar{v}}^{\theta_i})}, \tag{9}$$

where $\gamma_{uv}^{\theta_i}$ denotes the attention value of entity v with relation θ_i to entity u. $\mathcal{N}_{\text{inter}}^{\theta_i}(u)$ denotes the specific relation-based neighbors with the different type. We apply entity-level attention to fuse inter-class neighbors with a specific relation θ_i

$$\mathbf{h}_u^{\theta_i} = \sigma\left(\sum_{v \in \mathcal{N}_{\text{inter}}^{\theta_i}(u)} \gamma_{uv}^{\theta_i} \cdot \mathbf{h}'_v\right), \tag{10}$$

σ is a nonlinear activation, and $\mathbf{h}'_v$ is the projected feature of v.

Once we learned all relation embeddings $\{\mathbf{h}_u^{\theta_i}\}$, we utilize relation-level attention to fuse them together to obtain the inter-class relational embedding z_u for entity u. We first calculate the importance of each relation w^{θ_i} as follows:

$$w^{\theta_i} = \frac{1}{|\mathcal{E}_1|}\sum_{u \in \mathcal{E}_1} \mathbf{q}^{\tau(u)} \cdot \mathbf{h}_u^{\theta_i} + \frac{1}{|\mathcal{E}_2|}\sum_{v \in \mathcal{E}_2} \mathbf{q}^{\tau(v)} \cdot \mathbf{h}_v^{\theta_i}, \tag{11}$$

$$\epsilon^{\theta_i} = \frac{\exp(w^{\theta_i})}{\sum_{\theta_j \in \Theta_{\text{inter}}} \exp(w^{\theta_j})}, \tag{12}$$

where $\mathbf{q}^{\tau(\cdot)} \in \mathbb{R}^{F' \times 1}$ is learnable parameter. We fuse all relation embeddings to obtain the inter-class relational embedding $\mathbf{z}_u \in \mathbb{R}^{F'}$ of entity u.

---

4. Here, the superscript t is omitted for simplicity.

$$\mathbf{z}_u = \sum_{\theta_i \in \Theta_{\text{intra}}} \epsilon^{\theta_i} \cdot \mathbf{h}_u^{\theta_i}. \tag{13}$$

In inter-class attention, the aggregation of different entities' embedding are seamlessly integrated, and they are mingled and interactively affected each other, as shown in Fig. 2 II-a.

4.2.2 Intra-Class Attention Networks

The intra-class attention aims to learn the interaction among the same type of entities, as shown in Fig. 2 II-b. Specifically, given a relation $\phi_k \in \Phi_{\text{intra}}^{\tau(u)}$ that starts from entity u, we can get the intra-class relation based neighbors $\mathcal{N}_{\text{intra}}^{\phi_k}(u)$. $\Phi_{\text{intra}}^{\tau(u)}$ indicates the set of all intra-class relations of u. For instance, as shown in Fig. 1, Company 5 is a neighbor of Company 3 based on an implicit relation, and Company 4 is a neighbor of Company 1 based on meta-relation *CEC*. Each intra-class relation represents one semantic interaction, and we apply relation-specific attention to encode this characteristic. We first calculate the attention value of entity $\tilde{u}$ with relation ϕ_k to entity u as follows:

$$\alpha_{u\tilde{u}}^{\phi_k} = \frac{\exp(\text{LeakyRelu}(\mathbf{a}_{\phi_k}^\top \cdot [\mathbf{W}\mathbf{z}_u \| \mathbf{W}\mathbf{z}_{\tilde{u}}]))}{\sum_{u' \in \mathcal{N}_{\text{intra}}^{\phi_k}(u)} \exp(\text{LeakyRelu}(\mathbf{a}_{\phi_k}^\top \cdot [\mathbf{W}\mathbf{z}_u \| \mathbf{W}\mathbf{z}_{u'}]))}, \tag{14}$$

where $\mathbf{z}_u$ and $\mathbf{z}_{\tilde{u}}$ are output representations of the inter-class attention, respectively. $\mathbf{W} \in \mathbb{R}^{F' \times F'}$ is a trainable weight matrix which is shared to every node of the same type. $\mathbf{a}_{\phi_k} \in \mathbb{R}^{2F'}$ is the node-level attention weight vector for relation ϕ_k. $\mathcal{N}_{\text{intra}}^{\phi_k}(u)$ denotes the intra-class neighbors of u under relation ϕ_k. The embedding $\mathbf{h}_u^{\phi_k}$ of entity u for the given relation ϕ_k is calculated

$$\mathbf{h}_u^{\phi_k} = \sigma \left(\sum_{\tilde{u} \in \mathcal{N}_{\text{intra}}^{\phi_k}(u)} \alpha_{u\tilde{u}}^{\phi_k} \cdot \mathbf{W}\mathbf{z}_{\tilde{u}} \right), \tag{15}$$

where σ is a non-linear activation. In total, we can get $|\Phi_{\text{intra}}^{\tau(u)}|$ embeddings for entity u. Then, we conduct relation-level attentions to fuse them into the relational embedding $\mathbf{h}_u \in \mathbb{R}^{F'}$

$$\mathbf{h}_u = \sum_{\phi_k \in \Phi_{\text{intra}}^{\tau(u)}} \beta^{\phi_k} \cdot \mathbf{h}_u^{\phi_k}, \tag{16}$$

where $\Phi_{\text{intra}}^{\tau(u)}$ denotes the set of all intra-class relationships of entity u. β^{ϕ_k} denotes the importance of intra-class relation ϕ_k.

$$\mathbf{g}^{\phi_k} = \frac{1}{|\mathcal{E}_1|} \sum_{u \in \mathcal{E}_1} \mathbf{q}^{\tau(u)} \cdot \mathbf{h}_u^{\phi_k},$$

$$\beta^{\phi_k} = \frac{\exp(\mathbf{g}^{\phi_k})}{\sum_{\phi_l \in \Phi_{\text{intra}}^{\tau(u)}} \exp(\mathbf{g}^{\phi_l})}. \tag{17}$$

where $\mathbf{q}^{\tau(u)} \in \mathbb{R}^{F'}$ is a learnable parameter.

4.3 SMP With Stock Final Embeddings

Finally, the stock final embeddings by combining learned sequential embeddings and relational embeddings are utilized to make stock prediction by a dense layer feed-forward neural network (FNN) and a softmax function, as shown in Fig. 2 III.

$$\hat{y}_i^t = \mathbf{SMP}(\mathbf{s}_i^t \| \mathbf{h}_i^t)$$
$$= \text{Softmax}(\mathbf{W}_{smp}[\mathbf{s}_i^t \| \mathbf{h}_i^t] + b_{smp}), \tag{18}$$

where $\mathbf{W}_{smp}$ is a trainable weight matrix, and b_{smp} is the bias vector. We leverage the Adam algorithm [48] for optimization by minimizing the cross entropy loss function $\mathcal{L}$.

$$\mathcal{L} = -\sum_{i=1}^{|N|} \sum_t y_i^t ln(\hat{y}_i^t), \tag{19}$$

where y_i^t and $\hat{y}_i^t$ represent the ground truth and predict stock trend of stock i at t day, respectively. $|N|$ is the total number of stocks.

5 EXPERIMENTS

In this section, we present our experiments, mainly focusing on the following research questions:

• RQ1: Can our model achieve better performance than the state-of-the-art stock prediction methods?

• RQ2: Can our model achieve a higher investment return and lower risk in the investment simulation on real-world datasets?

• RQ3: How is the effectiveness of different components in our model?

• RQ4: Are all firm relations equally important for SMP? How do different parameters influence our model's performance?

Next, we first present experimental settings and then answer these research questions by analyzing the experimental results.

5.1 Experimental Settings

5.1.1 Data Collection

Since no existing benchmark datasets can satisfy our need to evaluate the effectiveness of our method, we collect public available data about the stocks from the famous China Securities Index (CSI) and construct two new datasets. We name them *CSI100E* and *CSI300E* with different number of listed companies, respectively. 185 stocks in CSI300E index without missing transaction data and having at least 60 related news articles during the selected period are kept. Similarly, 73 stocks in CSI100E index are kept. First, we get historical price of stocks[5] from November 21, 2017 to December 31, 2019 which include 516 transaction days. Second, we collect web news published in the same period from four financial mainstream sites,[6] including *Sina*, *Hexun*, *Sohu* and *Eastmoney*. Last, we collect four types of company relations[7] and the connections of executives[8] for CSI300E and CSI100E. The basic statistics of the datasets are summarized in Table 3. The usage details of the multimodal market signals are described in Section 3.2.

TABLE 3
Statistics of Datasets

	CSI100E	CSI300E
#Companies(Nodes)	73	185
#Executives(Nodes)	163	275
#Investment(Edges)	7	44
#Industry category(Edges)	272	1043
#Supply chain(Edges)	27	37
#Business partnership(Edges)	98	328
#Implicit relation(Edges)	dynamic	dynamic
#meta-relation CEC	18	42
#meta-relation CEEC	134	252
#Classmate(Edges)	338	592
#Colleague(Edges)	953	2224
#Management(Edges)	166	275
#Investment(Edges)	1	8
#Train Period	21/11/2017- 05/08/2019	21/11/2017- 05/08/2019
#Valid Period	06/08/2019- 22/10/2019	06/08/2019- 22/10/2019
#Test Period	23/10/2019- 31/12/2019	23/10/2019- 31/12/2019

5.1.2 Evaluation Protocols

SMP is usually treated as a binary classification problem. If the closing price of a stock i is higher than its opening price at day t, the stock movement trend is defined as "upward" ($y_i^t = 1$), otherwise as "downward" ($y_i^t = 0$). According to statistics, there are 46.7% "upward" stocks and 53.3% "downward" ones in CSI100E, and 47.8% "upward" and 52.2% "downward" stocks in CSI300E. Hence, the datasets are roughly balanced.

Some indicators [6], [49] are selected to demonstrate the effectiveness of the proposed method, i.e., Directional Accuracy (DA), Precision, (AUC), Recall, F1-score. We use the *Directional Accuracy (DA)* and *AUC* (the area under the precision-recall curve) to evaluate classification performance in our experiments, which are widely adopted in previous works [3], [12]. Similar to [18], [23], to evaluate DANSMP's applicability to real-world trading, we assess its profitability on CSI100E and CSI300E using metrics: cumulative investment return rate (*IRR*) and *Sharpe Ratio* [50]. Similar to previous method [3], [12], we use the market signals of the past T trading days (also called lookback window size) to predict stock movement on t^{th} day. The DA, IRR and SR are defined as follows:

$$DA = \frac{n}{N},$$

$$IRR^t = \sum_{i \in S^{t-1}} \frac{p_i^t - p_i^{t-1}}{p_i^{t-1}},$$

$$SR_a = \frac{E[R_a - R_f]}{std[R_a - R_f]}, \qquad (20)$$

where n is the number of predictions, which witness the same direction of stock movements for the predicted trend and the actual stock trend and N is the total number of predictions. S^{t-1} denotes the set of stocks on day $t-1$, and p_i^t is the price of stock i at day t. R_a denotes an asset return and R_f is the risk-free rate. In this study, the risk-free rate is set

as the one-year deposit interest rate of the People's Bank of China in 2019, i.e., $R_f = 1.5\%$.

Note that, to ensure the robustness of the evaluation, we repeat the testing procedure 10 times with different initialization for all the experimental results and the average performance is reported as the final model result.

5.1.3 Baselines

To demonstrate the effectiveness of our proposed model DANSMP, we compare the results with the following baselines.

- LSTM [51]: a typical RNN model that has promising performance on time-series data. In the evaluation, two-layer LSTM networks are implemented.
- GRU [52]: a simpler RNN that achieves similar performance with LSTM. In the comparison, two-layer GRU networks are implemented.
- GCN [53]: It performs graph convolutions to linearly aggregate the attributes of the neighbor nodes. In this study, two-layer GCN network was implemented.
- GAT [17]: It introduces attention mechanism which assigns different importance to the neighbors adaptively. Two-layer GAT networks are implemented.
- RGCN [54]: It designs specialized mapping matrices for each relations. Two-layer RGCN network was implemented.
- HGT [55]: It uses transformer architecture to capture features of different nodes based on type-specific transformation matrices.
- MAN-SF [44]: It fuses chaotic temporal signals from financial data, social media and stock relations in a hierarchical fashion to predict future stock movement.
- STHAN-SR [23]: It uses hypergraph and temporal Hawkes attention mechanism to rank stocks with only historical price data and explicit firm relations. We only need to slightly modify the objective function of MAN-SF to predict future stock movement.
- AD-GAT [3]: a SOTA method to use an attribute-driven GAT to capture attribute-sensitive momentum spillover of stocks, which can modeing market information space with feature interaction to further improve stock movement prediction.

These baselines cover different model characters. Specifically, the sequential-based LSTM [51] and GRU [52] can capture the time dependency of stock data, and the fused market signals were used as the input to the LSTM and GRU model. The homogeneous GNNS-based GCN [53], GAT [17], RGCN [54], HGT [55], MAN-SF [44], STHAN-SR [23] and AD-GAT [3] can capture the influence of related stocks based on the fused market signals and simple firm relations. Note that, for fair comparison, we do not select the methods that are incapable of dealing with all fused multi-modal market signals (i.e., historical price, media news and stock relations) as baselines.

5.1.4 Parameter Settings

All trainable trainable parameters vectors and matrices are initialized using the Glorot initialization [56]. The parameters of the comparison methods were set as follows. For all baselines, the number of the network layer is set to 2, and the hidden dimension is 64. For the LSTM, GRU, GCN, GAT, and STHAN-SR, the learning rate is 8e-4. For the

TABLE 4
The Hyper-Parameter Settings on Two Datasets

Parameter	CSI100E	CSI300E
Lookback window size T	20	20
The slice size of NTN M	10	10
Attention layer hidden size F'	39	22
GRU hidden size F	78	44
Learing rate	0.0008	0.00085
Implicit relation threshold η	0.0054	0.0052
Maximum number of epochs	400	400

RGCN, the learning rate is 5e-4. For the HGT, the learning rate is 12e-4. For the MAN-SF, the learning rate is 10e-4 and the attention heads are 2. For the AD-GAT, the attention heads are 3 and the learning rate is 6e-4.

In our DANSMP, we set the lookback window size T among [10, 15, 20,..., 40]. We search the learning rate from [0.00005, 0.0001, 0.00015,..., 0.002]. The slice size of NTN M and attention layer hidden size F' are determined in [5,10,15,..., 50] and [10, 11, 12,..., 50], respectively. The GRU hidden size F is set from [20, 22, 24,..., 100]. In our model, all hyperparameters were optimized with the validation set, and Table 4 shows the hyper-parameter settings of our method. The proposed DANSMP is implemented with PyTorch[9] and PyTorch Geometric,[10] and each training process costs 1.5 hrs averagely using a GTX 1080 GPU. To prevent overfitting, we use early stopping based on AUC (the area under the precision-recall curve) over the validation set.

5.2 Experimental Results and Analysis (RQ1)

Table 5 reports the average and standard deviation of all metrics on two datasets against nine state-of-the-art (SOTA) baselines, from which we observe that our proposed method outperforms all baselines for stock movement prediction in terms of all metrics on CSI100E and CSI300E. It confirms the capability of our method in modeling the comprehensive market signal representations via dual attention networks.

Analysis. (1) The LSTM and GRU, which only consider historical prices and media news, perform largely worse than our method. The results indicate that the relation datas contribute to stock movement prediction and the proposed method can take full advantage of the relational information in MKG to improve performance. (2) The graph-based methods, such as GCN and GAT, are homogeneous GNNs which are incapable of modeling heterogeneous market graph. Although being able to model multi-relational graph, RGCN can not sufficiently encode bi-typed heterogeneous graph become of the fact that it ignores the heterogeneity of node attributes and calculates the importance of neighbors within the same relation based on predefined constants. HGT focuses on handling web-scale heterogeneous graphs via graph sampling strategy, which thus is prone to overfitting when dealing with relative sparse MKG. HGT can not learns multi-level representation by sufficiently utilize interactions between two types of nodes. We believe that is the reason they perform worse than our model DANSMP which is pertinently designed to model bi-typed hybrid-relational

TABLE 5
Stock Prediction Results of Different Models

Methods	CSI100E		CSI300E	
	Accuracy	AUC	Accuracy	AUC
LSTM [51]	51.14±0.44	51.33±0.52	51.78±0.27	52.24±0.44
GRU [52]	51.66±0.31	51.46±0.16	51.11±0.47	52.30±0.49
GCN [53]	51.58±0.92	52.18±0.69	51.68±0.43	51.81±0.31
GAT [17]	52.17±2.11	52.78±1.81	51.40±0.24	52.24±0.79
RGCN [54]	52.33±0.89	52.69±1.09	51.79±0.09	52.59±0.37
HGT [55]	53.01±0.29	52.51±0.03	51.70±0.27	52.19±0.21
MAN-SF [44]	52.86±0.54	52.23±0.07	51.91±0.35	52.48±0.53
STHAN-SR [23]	52.78±0.85	53.05±0.46	<u>52.89±0.42</u>	53.48±0.40
AD-GAT [3]	<u>54.56±2.41</u>	<u>55.46±1.28</u>	52.63±1.30	<u>54.29±1.49</u>
DANSMP(ours)	**57.75±1.50**	**60.78±1.38**	**55.79±1.04**	**59.36±1.47**

MKG. (3) The proposed DANSMP consistently outperforms three other SMP competitors, including AD-GAT, STHAN-SR and MAN-SF. Specifically, it exceeds the second place by approximately 3.19% and 5.32% in terms of Accuracy and AUC in CSI100E, and 3.16% and 5.07% in CSI300E. The results clearly demonstrate the effectiveness of DANSMP and the explicit relation and executives relation are meaningful for stock movement prediction.

5.3 Investing Simulation (RQ2)

To test whether a model can make a profit, we set up a back-testing via simulating the stock investment in CSI100E and CSI300E over the test period, during which the CSI100 and CSI300 index increased by 4.20% and 5.14% (from 4144.05 to 4317.93, and 3896.31 to 4096.58), respectively. During the same period, the A-share market index increased by 3.31% (from 3093.68 to 3195.98). Specifically, the top-15 stocks with the highest predicted ranking score in each model are bought and held for one day. We choose RMB 10,000 as the investment budget, and take into account a transaction cost of 0.03% when calculating the investment return rate, which is in accordance with the stock market practice in China. The cumulative profit will be invested into the next trading day. From Table 6 and Fig. 3, we can find that DANSMP achieves stable and continuous positive returns throughout the back-testing. DANSMP can more accurately predict the future trend of the stock market, sell stocks that are going down and buy stocks with rising signal, thus gaining more cumulative return. Particularly, the advantage of DANSMP over all baselines mainly lies in its superior performance when the stock market is in a bear stage. The proposed DANSMP achieves significantly higher returns than all baselines with the cumulative rate of 9.09% and 17.06% in CSI100E and CSI300E. In addition, DANSMP has a more desirable risk-adjusted return with Sharpe Ratio of 3.019 and 4.195 in CSI100E and CSI300E, respectively, which demonstrate the superiority of the proposed method in terms of the trade-off between the return and the risk.

5.4 Ablation Study (RQ3)

To examine the usefulness of each component in DANSMP, we conduct ablation studies on CSI100E and CSI300E. We design six variants: (1) DANSMP*w/o executives*, which deletes the executive entities. MKG is degraded into a simple uni-

TABLE 6
Profitability of All Methods in Back-Testing

Methods	CSI100E		CSI300E	
	IRR	SR	IRR	SR
LSTM [51]	-3.45±1.02%	-1.597±0.46	-0.81±2.62%	-0.123±1.27
GRU [52]	-2.62±0.40%	-1.151±0.16	-3.04±0.76%	-1.079±0.24
GCN [53]	2.82±2.22%	2.423±1.52	1.99±1.56%	0.958±0.87
GAT [17]	1.60±1.12%	0.808±0.70	2.69±3.78%	1.810±2.26
RGCN [54]	5.79±1.81%	2.819±1.80	-0.43±0.50%	-0.087±2.31
HGT [55]	4.43±0.71%	2.324±1.11	2.68±2.80%	1.412±1.69
MAN-SF [44]	-2.60±0.39%	-1.723 ±0.24	1.28±0.64%	0.564±0.28
STHAN-SR [23]	0.44±1.32%	0.047 ±1.07	2.37±0.55%	1.497±1.14
AD-GAT [3]	1.47±2.41%	0.742±1.27	14.63±4.37%	2.257±2.58
DANSMP(ours)	9.09±1.09%	3.019±0.55	17.06±2.21%	4.195±0.91

TABLE 7
The Ablation Study Over DANSMP

Variants	CSI100E		CSI300E	
	Accuracy	AUC	Accuracy	AUC
DANSMP	**57.75**	**60.78**	**55.79**	**59.36**
w/o executives	52.71	52.62	53.80	55.88
w/o implicit rel.	53.52	54.38	52.13	53.37
w/o explicit rel.	55.12	57.05	54.10	55.49
w/o technical ind.	52.55	52.79	51.66	53.57
w/o sentiment sig.	54.21	55.60	53.92	55.83
w/o dual	56.12	58.60	55.43	57.85

type knowledge graph. (2) DANSMP*w/o implicit relation*, which removes the implicit relation when model the stock momentum spillover effect. (3) DANSMP*w/o explicit relation*, which deletes the explicit relations and only uses the implicit relation to predict stock movement. (4) DANSMP*w/o technical indicators*, which deletes the technical indicators and only uses the sentiment signals to predict stock movement. (5) DANSMP*w/o sentiment signals*, which deletes the sentiment signals and only uses the technical indicators to predict stock movement. (6) DANSMP*w/o dual*, which replaces the dual module by conventional attention mechanism and does not distinguish the node intra-class and inter-class relation.

From Table 7, we observe that removing any component of DANSMP would lead to worse results. The effects of the four components vary in different datasets, but all of them contribute to improving the prediction performance. Specifically, removing executives relations and implicit relations leads to

the most performance drop, compared to the other two, which means a company can influence the share price of other companies through interactions between executives. Technical indicators and sentiment signals are helpful for stock prediction, and technical indicators seem to be more valuable. In contrast, using the conventional attention mechanism produces the least performance drop. Compared with conventional attention mechanism, the dual attention module enables DANSMP to adaptively select more important nodes and relations. This finding further proves that the proposed DANSMP fully leverages bi-typed hybrid-relational information in MKG via dual mechanism for better stock prediction.

5.5 Analysis of Firm Relation (RQ4)

To investigate the impact of different types of relations for stock prediction, we show the learned attention scores of our model in Fig. 4. The attention score is learned parameter for different firm relations. The main findings are as follows: (1) We can observe Fig. 4 that the learned attention score of implicit relation gains more weight than other relations, and the implicit relation which contains a lot of valuable information proved to be helpful for stock movement prediction. (2) The industry category, supply chain and investment get almost the same attention scores, and those can improve the performance of model. (3) Compared with other relations, the business partnership has the lowest score. Although the number of business partnership relation is greater than that of investment and supply chain, the relatively dense business partnership relation may carry some noise, which adds irrelevant information to the representations of target nodes. The results further demonstrate the necessity of considering the implicit relation in modeling the momentum spillover effect. In addition, our model can adaptively weight important company relations to obtain
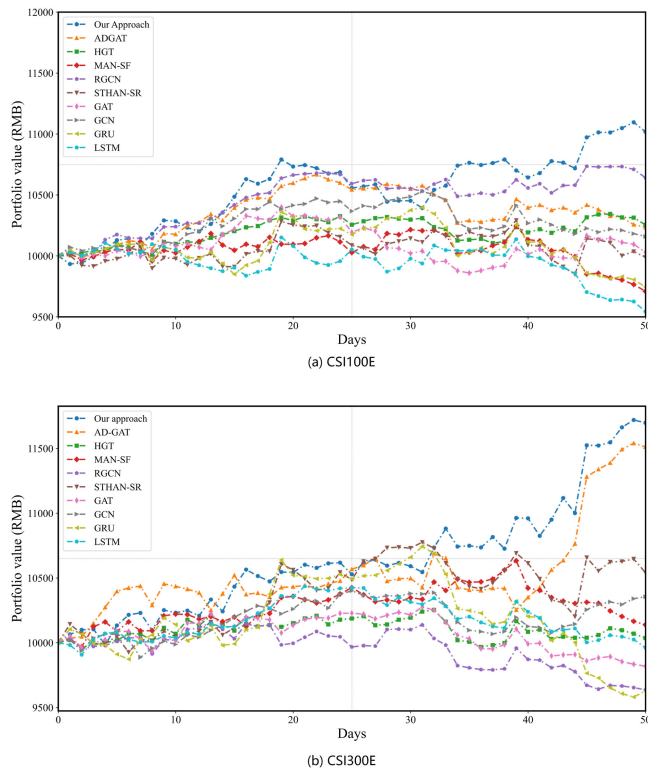

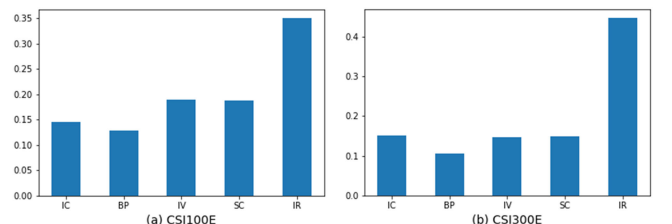Fig. 3. Profitability analysis on CSI100E and CSI300E.

Fig. 4. The presentation of the learned attention scores of DANSMP on CSI100E and CSI300E. Here, IC denotes the industry category; BP stands for the business partnership; IV denotes the investment; SC is the supply chain; IR denotes the implicit relation.
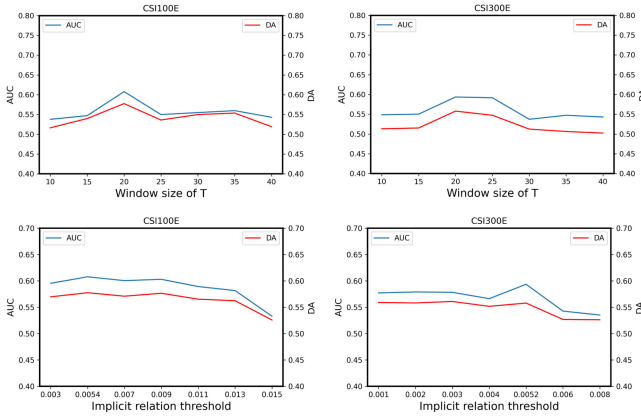
Fig. 5. Sensitivity to parameters T and η.

better representations of target nodes, which can improve the performance of the model for stock prediction.

5.6 Parameter Sensitivity Analysis (RQ4)

We also investigate on the sensitivity analysis of two parameters in DANSMP. We report the results of DANSMP under different parameter settings on CSI100E and CSI300E and experimental results are shown in Fig. 5.

Lookback Window Size T. We analyze the performance variation with different lookback window size T in Fig. 5a. Our model performs best when T is set to about 20 in both datasets. Lower performance indicates that the shorter window size can't capture stock market information, such as public information which requires time to absorb into stock price movement [57]. As the T increases, the larger window size, including some stale information from past days [58], may deteriorate the model performance.

Implicit Relation Threshold η. The results of our model with different implicit relation thresholds are reported in Fig. 5b. The performance of our proposed model grows with the increment of η and achieves the best performance when η is set to 0.0054 in CSI100E. With the increment of η, the performance raises at first and then drops gradually in the dataset CSI300E. The performance first increases because the relatively weak edges which have lower influence on price will be filtered, and then when the η becomes bigger, the performance decreases possibly because some meaningful implicit edges are neglected.

Number of Convolution Layers l. We bulid DANSMP with different number of convolution layers and analyze the impact of the number of convolution layers on the stock prediction. As the l increases, the prediction performance increases and then decreases accordingly. Our model provides the best prediction performance with $l = 2$.

6 CONCLUSION AND FUTURE WORK

In this paper, we focus on stock movement prediction task. To model stock momentum spillover in real financial market, we first construct a novel bi-typed hybrid market knowledge graph. Then, we propose a novel Dual Attention Networks, which are equipped with both inter-class attention module and intra-class attention module, to learn the stock momentum spillover features on the newly constructed MKG. To evaluate our method, we construct two

new datasets CSI100E and CSI300E. The empirical experiments demonstrate our method can successfully improve stock prediction with bi-typed hybrid-relational MKG via the proposed DANSMP. The ablation studies reaffirm that the performance gain mainly comes from the use of the associated executives, and additional implicit relation between companies in MKG. An interesting future work direction is to explore web media about the executives including: (i) the negative facts from news, such as accusation of crime, health issue, etc; (ii) the improper speech on social media, such as Twitter and Weibo. We believe these factual event information of executives can be detected and utilized to feed into graph-based methods for better SMP performance.

REFERENCES

[1] Q. Li, Y. Chen, J. Wang, Y. Chen, and H. Chen, "Web media and stock markets: A survey and future directions from a Big Data perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 2, pp. 381–399, Feb. 2018.

[2] Z. Hu, W. Liu, J. Bian, X. Liu, and T.-Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, 2018, pp. 261–269.

[3] R. Cheng and Q. Li, "Modeling the momentum spillover effect for stock prediction via attribute-driven graph attention networks," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 55–62.

[4] G. Wang, L. Cao, H. Zhao, Q. Liu, and E. Chen, "Coupling macro-sector-micro financial indicators for learning stock representations with less uncertainty," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 4418–4426.

[5] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, and T.-S. Chua, "Temporal relational ranking for stock prediction," *ACM Trans. Inf. Syst.*, vol. 37, no. 2, pp. 1–30, 2019.

[6] J. Ye, J. Zhao, K. Ye, and C. Xu, "Multi-graph convolutional network for relationship-driven stock movement prediction," 2020, *arXiv:2005.04955*.

[7] T. Lin, T. Guo, and K. Aberer, "Hybrid neural networks for learning the trend in time series," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 2273–2279.

[8] F. Feng, H. Chen, X. He, J. Ding, M. Sun, and T.-S. Chua, "Enhancing stock movement prediction with adversarial training," in *Proc. Int. Joint Conf. Artif. Intell.*, 2019, pp. 5843–5849.

[9] B. G. Malkiel and E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[10] F. Ming, F. Wong, Z. Liu, and M. Chiang, "Stock market prediction from WSJ: Text mining via sparse matrix factorization," in *Proc. IEEE Int. Conf. Data Mining*, 2014, pp. 430–439.

[11] Q. Liu, X. Cheng, S. Su, and S. Zhu, "Hierarchical complementary attention network for predicting stock price movements with news," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 1603–1606.

[12] Q. Li, J. Tan, J. Wang, and H. Chen, "A multimodal event-driven LSTM model for stock prediction using online news," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 10, pp. 3323–3337, Oct. 2021.

[13] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011.

[14] J. Si, A. Mukherjee, B. Liu, Q. Li, H. Li, and X. Deng, "Exploiting topic based twitter sentiment for stock prediction," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, 2013, pp. 24–29.

[15] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9603–9611, 2015.

[16] U. Ali and D. Hirshleifer, "Shared analyst coverage: Unifying momentum spillover effects," *J. Financial Econ.*, vol. 136, no. 3, pp. 649–675, 2020.

[17] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Representations*, 2018.

[18] R. Sawhney, S. Agarwal, A. Wadhwa, and R. R. Shah, "Spatiotemporal hypergraph convolution network for stock movement forecasting," in *Proc. IEEE Int. Conf. Data Mining*, 2020, pp. 482–491.

[19] W. Li, R. Bao, K. Harimoto, D. Chen, J. Xu, and Q. Su, "Modeling the stock relation with graph network for overnight stock movement prediction," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, 2020, pp. 4541–4547.

[20] J. Cai, R. A. Walkling, and K. Yang, "The price of street friends: Social networks, informed trading, and shareholder costs," *J. Financial Quantitative Anal.*, vol. 51, no. 3, pp. 801–837, 2016.

[21] W. Jing and X. Zhang, "Online social networks and corporate investment similarity," *J. Corporate Finance*, vol. 68, 2021, Art. no. 101921.

[22] Y. Chen, Z. Wei, and X. Huang, "Incorporating corporation relationship via graph convolutional neural networks for stock price prediction," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 1655–1658.

[23] R. Sawhney, S. Agarwal, A. Wadhwa, T. Derr, and R. R. Shah, "Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 497–504.

[24] D. M. Nelson, A. C. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2017, pp. 1419–1426.

[25] C. Chen, L. Zhao, J. Bian, C. Xing, and T.-Y. Liu, "Investment behaviors can tell what inside: Exploring stock intrinsic properties for stock trend prediction," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 2376–2384.

[26] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Knowledge-driven event embedding for stock prediction," in *Proc. Int. Conf. Comput. Linguistics*, 2016, pp. 2133–2142.

[27] Q. Li, Y. Chen, L. L. Jiang, P. Li, and H. Chen, "A tensor-based information framework for predicting the stock market," *ACM Trans. Inf. Syst.*, vol. 34, no. 2, pp. 1–30, 2016.

[28] S. Deng, N. Zhang, W. Zhang, J. Chen, J. Z. Pan, and H. Chen, "Knowledge-driven stock trend prediction and explanation via temporal convolutional network," in *Proc. World Wide Web Conf.*, 2019, pp. 678–685.

[29] R. D. Edwards, J. Magee, and W. C. Bassetti, *Technical Analysis of Stock Trends*. Boca Raton, FL, USA: CRC Press, 2018.

[30] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS One*, vol. 12, no. 7, 2017, Art. no. e0180944.

[31] Q. Gao, "Stock market forecasting using recurrent neural network," Ph.D. Dissertation, University of Missouri–Columbia, MO, 2016.

[32] L. Zhang, C. Aggarwal, and G.-J. Qi, "Stock price prediction via discovering multi-frequency trading patterns," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2017, pp. 2141–2149.

[33] L. Chen, Y. Chi, Y. Guan, and J. Fan, "A hybrid attention-based EMD-LSTM model for financial time series prediction," in *Proc. 2nd Int. Conf. Artif. Intell. Big Data*, 2019, pp. 113–118.

[34] L.-C. Cheng, Y.-H. Huang, and M.-E. Wu, "Applied attention-based LSTM neural networks in stock prediction," in *Proc. IEEE Int. Conf. Big Data*, 2018, pp. 4716–4718.

[35] J. Wang, T. Sun, B. Liu, Y. Cao, and H. Zhu, "CLVSA: A convolutional LSTM based variational sequence-to-sequence model with attention for predicting trends of financial markets," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 3705–3711.

[36] Y. Xu and S. B. Cohen, "Stock movement prediction from tweets and historical prices," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 1970–1979.

[37] R. Akita, A. Yoshihara, T. Matsubara, and K. Uehara, "Deep learning for stock prediction using numerical and textual information," in *Proc. IEEE/ACIS 15th Int. Conf. Comput. Inf. Sci.*, 2016, pp. 1–6.

[38] M. R. Vargas, B. S. D. Lima, and A. G. Evsukoff, "Deep learning for stock market prediction from financial news articles," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl.*, 2017, pp. 60–65.

[39] M. Sedinkina, N. Breitkopf, and H. Schutze, "Automatic domain adaptation outperforms manual domain adaptation for predicting financial outcomes," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 346–359.

[40] Y. Qin and Y. Yang, "What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 390–401.

[41] N. Rekabsaz, M. Lupu, A. Baklanov, A. Hanbury, A. Dür, and L. Anderson, "Volatility prediction using financial disclosures sentiments with word embedding-based IR models," in *Proc. Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 1712–1721.

[42] R. Kim, C. H. So, M. Jeong, S. Lee, J. Kim, and J. Kang, "HATS: A hierarchical graph attention network for stock movement prediction," 2019, *arXiv:1908.07999*.

[43] C. Li, D. Song, and D. Tao, "Multi-task recurrent neural networks and higher-order Markov random fields for stock price movement prediction: Multi-task RNN and higer-order MRFs for stock price classification," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 1141–1151.

[44] R. Sawhney, S. Agarwal, A. Wadhwa, and R. R. Shah, "Deep attentive learning for stock movement prediction from social media text and company correlations," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2020, pp. 8415–8426.

[45] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 2327–2333.

[46] X. Wu, "Event-driven learning of systematic behaviours in stock markets," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2020, pp. 2434–2444.

[47] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[49] M. G. Sousa, S. Kenzo, L. de Souza Rodrigues, P. H. Moraes, E. R. Fernandes, and E. T. Matsubara, "BERT for stock market sentiment analysis," in *Proc. IEEE 31st Int. Conf. Tools Artif. Intell.*, 2019, pp. 1597–1601.

[50] W. F. Sharpe, "The sharpe ratio," *J. Portfolio Manage.*, vol. 21, no. 1, pp. 49–58, 1994.

[51] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[52] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*.

[53] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations*, 2017.

[54] M. Schlichtkrull et al., "Modeling relational data with graph convolutional networks," in *Proc. Eur. Semantic Web Conf.*, 2018, pp. 593–607.

[55] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous graph transformer," in *Proc. Web Conf.*, 2020, pp. 2704–2710.

[56] G. Xavier and B. Yoshua, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[57] R. Luss and A. d'Aspremont, "Predicting abnormal returns from news using text classification," *Quantitative Finance*, vol. 15, no. 6, pp. 999–1012, 2015.

[58] D. Bernhardt and J. Miao, "Informed trading when information becomes stale," *J. Finance*, vol. 59, no. 1, pp. 339–390, 2004.

Yu Zhao received the BS degree from Southwest Jiaotong University, in 2006, and the MS and PhD degrees from the Beijing University of Posts and Telecommunications, in 2011 and 2017, respectively. He is currently an associate professor with the Southwestern University of Finance and Economics. His current research interests include machine learning, natural language processing, knowledge graph, Fintech. He has authored more than 30 papers in top journals and conferences including *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Transactions on Mobile Computing*, ACL, ICME, etc.

Huaming Du received the MS degree from the China University of Petroleum, Beijing, China, in 2018. He is currently working toward the PhD degree with the Southwestern University of Finance and Economics, Chengdu, China. His research interests include Fintech, reinforcement learning, and graph representation learning.

Ying Liu received the BS degree from Southwest Jiaotong University, in 2018. She is currently working toward the MS degree with the Southwestern University of Finance and Economics, Chengdu, China. Her current research interests include Fintech and graph learning.

Shaopeng Wei received the BS degree from Huazhong Agricultural University, in 2019, and now is working toward the PhD degree with the Southwestern University of Finance and Economics. His research interests include graph learning and relevant applications in recommendation system and Fintech.
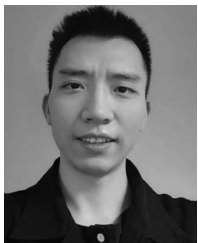
Xingyan Chen received the PhD degree in computer technology from the Beijing University of Posts and Telecommunications (BUPT), in 2021. He is currently a lecturer with the School of Economic Information Engineering, Southwestern University of Finance and Economics. He has published papers in well-archived international journals and proceedings, such as the *IEEE Transactions on Mobile Computing*, *IEEE Transactions on Circuits and Systems for Video Technology*, *IEEE Transactions on Industrial Informatics*, and IEEE INFOCOM, etc. His research interests include multimedia communications, multi-agent RL.

Fuzhen Zhuang (Member, IEEE) received the PhD degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences. He is currently a full professor with the Institute of Artificial Intelligence, Beihang University. His research interests include machine learning and data mining, including transfer learning, multi-task learning, multi-view learning and recommendation systems. He has published more than 100 papers in the prestigious refereed conferences and journals, such as KDD, WWW, SIGIR, ICDE, IJCAI, AAAI, EMNLP, *Nature Communications*, *IEEE Transactions on Knowledge and Data Engineering*, *ACM Transactions on Knowledge Discovery from Data*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Neural Networks and Learning Systems*, *ACM Transactions on Intelligent Systems and Technology*, etc.

Qing Li (Member, IEEE) received the BS and MS degrees from Harbin Engineering University, China, and the PhD degree from the Kumoh National Institute of Technology, in February of 2005, Korea. He is a postdoctoral researcher with Arizona State University and the Information & Communications University of Korea. He is a professor with the Southwestern University of Finance and Economics, China. His research interests include natural language processing, FinTech. He has published more than 70 papers in the prestigious refereed conferences and journals, such as *IEEE Transactions on Knowledge and Data Engineering*, *ACM Transactions on Information Systems*, AAAI, SIGIR, ACL, WWW, etc.

Gang Kou (Member, IEEE) received the BS degree from the Department of Physics, Tsinghua University, China, the master's degree from the Department of Computer Science, University of Nebraska at Omaha, and the PhD degree in information technology from the College of Information Science & Technology, University of Nebraska at Omaha. He is a distinguished professor of Chang Jiang Scholars Program with the Southwestern University of Finance and Economics, managing editor of *International Journal of Information Technology & Decision Making (SCI)* and managing editor-in-chief of *Financial Innovation (SSCI)*. He is also editors for other journals, such as: *Decision Support Systems*, and *European Journal of Operational Research*. Previously, he was a professor of the School of Management and Economics, University of Electronic Science and Technology of China, and a research scientist in Thomson Co., R & D. He has published more than 100 papers in various peer-reviewed journals. His h-index is 57 and his papers have been cited for more than 10000 times. He is listed as the Highly Cited Researcher by Clarivate Analytics (Web of Science).

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.