Information Fusion 104 (2024) 102165

Contents lists available at ScienceDirect

Information Fusion

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

Full length article

Quantitative stock portfolio optimization by multi-task learning risk and return

Yu Ma [a,1], Rui Mao [b,1], Qika Lin [c], Peng Wu [d], Erik Cambria [b,*]

[a] School of Economics and Management, Nanjing University of Science & Technology, 200 Xiaolingwei Road, Nanjing, 210094, Jiangsu, China
[b] School of Computer Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore
[c] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, Shaanxi, China
[d] School of Intelligent Manufacturing, Nanjing University of Science & Technology, 200 Xiaolingwei Road, Nanjing, 210094, Jiangsu, China

ARTICLE INFO

Keywords:
Stock portfolio optimization
Quantitative stock selection
Multi-task learning
Stock relation graph
Graph-based learning

ABSTRACT

Selecting profitable stocks for investments is a challenging task. Recent research has made significant progress on stock ranking prediction to select top-ranked stocks for portfolio optimization. However, the stocks are only ranked by predicted stock return, ignoring the stock price volatility risk — a critical aspect for stock selection and investments. Moreover, they preliminarily attempted to capture the effects of related stocks from a singular relation, disregarding the rich information regarding multiple spillover effects from related stocks and the distinctions in effects among various relations. Thus, we propose a risk and return multi-task learning model with a heterogeneous graph attention network (HGA-MT) to predict stock ranking for portfolio optimization. First, to aggregate the multiple spillover effects of related stocks, we introduce graph convolutional networks to fuse the effects of related stocks in each relation and design an attention network to allocate varying weights to different types of relationships. Second, we use a multi-task learning paradigm to learn stock return and volatility risks jointly. The stock ranking results are calculated by simultaneously considering the risk and return. Thus, Top-K ranked stocks are recommended in the portfolio for the next trading day to achieve higher and more stable profits. Extensive experiments prove that HGA-MT outperforms previous state-of-the-art methods in stock ranking and backtesting trading evaluation tasks.

1. Introduction

Investing in stock markets to achieve profits is an extremely attractive topic for traders and investors. Currently, machine learning-based methods have shown great success in assisting financial investments [1–3]. Developing a portfolio is an important way for financial investments because compared with investing in a single stock, carefully selecting multiple stocks as a portfolio can spread the investment risks of individual stocks [4,5]. However, there exist multiple factors that affect stock prices [6]. The stock markets are known for their intricate and volatile nature [7]. Therefore, it is a challenging task to incorporate a wide range of factors as well as consider return and volatility risk to select profitable stocks, thereby establishing the portfolio for investors. Recently, there have been several deep learning-based portfolio optimization studies that aimed to maximize stock investment profits [5,8–13]. They have formulated stock selection as a stock ranking prediction task. The objective is to directly obtain a stock list ranked by the anticipated return ratio for the next trading day.

The top-ranked stocks are expected to earn higher investment revenues, which are selected in a portfolio. However, previous works only ranked and recommended stocks based on higher stock return ratios while ignoring stock volatility risk. Risk is a crucial aspect that should always be considered in financial investment [4]. Hence, risk-oriented criteria should be introduced as much as possible in stock ranking for stock selection [9]. Among other factors, stock price volatility serves as a reliable measure for assessing trading risk [14]. It is prudent to steer clear of stocks exhibiting high price volatility, despite the potential for a higher return ratio on the subsequent trading day, to avoid losses in the long run. Therefore, in order to achieve higher and more stable profits, it is crucial to take into account both risk and return to rank stocks for portfolio optimization. On the other hand, stock prices are affected by multiple information sources due to the complicated stock market environment [15,16]. Extensive stock prediction research has been conducted using quantitative indicators, including fundamental and technical indicators [17,18].

* Corresponding author.
E-mail addresses: mayu@njust.edu.cn (Y. Ma), rui.mao@ntu.edu.sg (R. Mao), linqika@stu.xjtu.edu.cn (Q. Lin), wupeng@njust.edu.cn (P. Wu), cambria@ntu.edu.sg (E. Cambria).
[1] These authors contributed equally.

https://doi.org/10.1016/j.inffus.2023.102165
Received 8 May 2023; Received in revised form 27 November 2023; Accepted 28 November 2023
Available online 30 November 2023

Moreover, incorporating financial news also benefits stock price prediction [6,19–21], because investment is likely to be affected by news sentiment [22,23]. Finally, financial research has discovered crisis spillover effects that the adverse effects of a crisis not only affect a target company but also spread to its related enterprises [24]. This is because various relationships interconnect listed companies in stock markets, such as the supply chain [25], the shareholding chain [26], and the industry competition relationships [27].

Recently, several studies have attempted to model stock relation information and capture the effects of related stocks by using the graph-based method [8,10,26,28]. Each stock is depicted as a node in a graph, and the edge between two nodes is defined as a specific stock relation. However, these studies only incorporated a single type of stock relations, such as shareholding [26] or industry [8]. Furthermore, prior studies have failed to give ample consideration to differentiating the effects of related stock information on target stocks based on distinct relations. Given different company relationships, the news and stock trading impacts of related companies on a target company may be different [5,29]. For example, the negative news about industry competitors and the related stocks in shareholding chains may have different impacts on the target stocks [30–32]. Therefore, it is essential to aggregate multiple spillover effects and assign different importance weights to different relations.

In light of the above limits, we propose a novel multi-task learning model with a heterogeneous graph attention network (HGA-MT) that leverages risk and return as well as stock relationships to improve ranking-based portfolio optimization. The heterogeneous graph attention (HGA) module is introduced to fuse multi-source information, encompassing not only the information concerning a target stock but also information about related stocks that share multiple relationships with the target stock. Multi-task learning is used to simultaneously acquire knowledge regarding the risk and return associated with a target stock.

First, we concatenate 39 quantitative indicators and financial news features to represent the market information of each stock. The concatenated features are delivered to the following Bidirectional Long Short-Term Memory (BiLSTM) [33] encoders to capture the financial temporal hidden states. Second, the heterogeneous graph, which presents the three stock relations, is reconstructed into three relation-specific subgraphs. We employ Graph Convolutional Networks (GCN) [34] to fuse the effects of related stocks in each relation and assign different importance to different relation types by an attention mechanism. Third, the fused information is fed to the multi-task learning module to learn and predict the stock return and volatility risk on the next trading day. The investment-oriented stock ranking results are given by a score function, incorporating the predicted risk and return. Finally, K stocks with the highest scores are selected to invest on the next trading day. We test the effectiveness of the proposed HGA-MT model in Chinese stock markets and execute the experiments with the stock set of CSI100 index constituent stocks.

For stock ranking performance, the results indicate that our model outperforms previous state-of-the-art methods, i.e., MAC [15] with 0.0118 (11.09%) gains in Precision and 0.0221 (8.29%) gains in Mean Reciprocal Rank (MRR) on average. For the financial evaluation, HGA-MT exceeds MAC by 0.0510 (14.94%) in cumulative Investment Return Ratio (IRR) and 0.2055 (11.24%) in Sharpe (SP) ratio on average, achieving higher profits in the backtesting trading. Additionally, compared to MAC, HGA-MT yields a lower Maximum Drawdown (MDD, 8.81% vs. 7.54% on average), indicating that HGA-MT achieves lower investment risk. Our ablation study shows that learning both risk and return can yield more accurate return-based stock ranking predictions. Meantime, ranking stocks by considering both risk and return factors deliver more robust portfolio profits. Finally, learning and distinguishing multiple stock relationships also contribute to higher ranking accuracy and investment profits.

The contributions of this work are threefold:

- We propose a multi-task learning model that jointly learns the stock return and volatility risk for stock ranking prediction, which can achieve higher and more stable profits. To the best of our knowledge, we are the first to incorporate the concept of stock price volatility risk through multi-task learning into the stock ranking prediction model for portfolio optimization.
- We introduce an HGA module based on GCN and attention networks to fuse the multiple effects of related stocks to target stocks. The HGA module can disentangle various stock relationships and effectively integrate the related stock information by assigning varying degrees of importance.
- Extensive experimental results indicate that our proposed model outperforms the state-of-the-art baselines in terms of stock ranking and financial evaluation, demonstrating our model's effectiveness.

The rest of this paper is organized as follows: Section 2 presents the related works; Section 3 describes our proposed HGA-MT method; Section 4 presents our experimental settings; Section 5 reports the evaluation results, ablation, and hyperparameter analysis; Finally, Section 6 concludes this work and suggests future work directions.

2. Related work

2.1. Ranking-based stock portfolio optimization

A crucial task for financial investments is constructing the portfolio by selecting profitable stocks. Extensive research has been conducted for stock selection and portfolio optimization, forming different research directions with different emphases, based on different perspectives and forms [35–39]. Among these approaches is the conceptualization of the portfolio optimization task as a stock ranking task, aimed at ranking stocks with the potential for higher returns, and then incorporating high-ranked stocks into the portfolio [40,41]. The advantage of such a task formulation is that: (1) The stock ranking approach provides a straightforward and transparent mechanism for portfolio optimization. Investors can easily understand the rationale behind including or excluding specific stocks based on their ranking. (2) By explicitly emphasizing the ranking of stocks according to their potential for higher returns, this approach directs attention to one of the key goals of portfolio optimization: maximizing returns. (3) The success of deep learning in financial forecasting can be attributed to its proficiency in complex pattern recognition and holistic feature integration. Harnessing these strengths in a regression-like framework is optimally achieved by framing the portfolio optimization task as a ranking task. Consequently, an intuitive approach is to rank stocks, based on time-series prediction.

As the pioneer, Feng et al. [9] tailored the deep learning models for stock ranking prediction to select top-ranked stocks for portfolio optimization. Sawhney et al. [10] proposed a neural hypergraph framework for stock selection, resulting in a list of stocks prioritized based on their return ratios. The top-ranked stocks with higher expected returns were selected for investment. Ma et al. [5] proposed an attribute-oriented fuzzy hypergraph model for stock recommendations, which also focused on stock ranking prediction to recommend top-ranked stocks. Feng et al. [12] designed a model to recommend Top-N return ratio stocks, which combined a time series module for encoding timing characteristics with attributed GCN for capturing correlation topology information.

These ranking-based models have achieved promising results for stock selection and portfolio optimization [5,8,10,42]. However, these methods obtained the stock ranking results only based on higher expected returns without considering stock price volatility risks. As a crucial aspect, risk should be considered in financial investments [4]. The classical study on stock ranking prediction indicated that risk-oriented criteria should be introduced in stock ranking [9].

Therefore, it is essential to consider both the risk and return concurrently to achieve more reasonable stock ranking outcomes that can optimize the investment portfolio. Stock price volatility can serve as a reliable measure to assess trading risk [14]. This notion has motivated us to construct a multi-task learning model that can simultaneously acquire knowledge about both the stock return and risk.

2.2. Multi-source information for stock price prediction

Stock prices are affected by multi-source information in complex stock trading environments [15,43]. Extensive research indicated that fundamental and technical indicators reflect the profitability of listed companies and stock price patterns, which help to predict stock prices [18,44]. Moreover, research on event-driven trading shows that material news can significantly influence stock return volatility [45,46], while the market volatility also impacts the cognition of market participants [47]. Excessive media attention and negative news would cause abnormal stock price volatility [48,49]. Several existing research works combine technical indicators with financial news to predict stock price, demonstrating that financial news is essential information for effective stock price prediction [21,40,50]. Besides, with the developments of the market economy, listed companies are connected through various relations, such as supply chain [25], shareholding chain [26] and industry competitor relations [27]. Research on public opinion crisis management argues that there are spillover effects among interrelated firms. The public judgment of a company will be affected by the information about its related companies [51].

Recently, several works have attempted to use stock relations to aid stock price prediction [46,52]. Graph-based models were used to capture the impacts of related stocks [28,53,54]. To incorporate stock relation information, Wu et al. [54] used historical price correlations among stocks to construct a correlation graph, which improved the stock price prediction in the Chinese stock markets. Chen et al. [26] proposed a GCN-based framework to improve the performance of stock price movement prediction by aggregating the data of relevant stocks from a shareholding graph. Hsu et al. [8] built a hierarchical industry graph by analyzing the relation between industries for stock ranking prediction and profitable stock recommendations. Although the above models have attempted to capture the effects of related stocks, these studies only fuse the effects from a single stock relation, such as shareholding [26] or industry [8,10].

In practice, listed companies are connected through various relationships [25–27]. Stock prices are also affected by the spillover effects of various relations [15,29]. It may result in sub-optimal prediction results due to insufficient valid information if only a single type of related stock is considered. Ma et al. [15] proposed a multi-source aggregated classification model based on GCN for stock price prediction, which can aggregate the effects of related stocks in the supply chain, shareholding chain, and industry competitors. They organized the stock relation graph with three relations as a homogeneous graph and indiscriminately integrated information from related stocks with equal weights. However, the impact of each relation on a specific stock node is not uniform [5,29]. For example, related stocks in shareholding chains and industry competitors possess distinct influence levels on the target stocks [30–32]. Therefore, it is critical to consolidate the spillover impacts of related stocks across multiple relations and proficiently discern the variations in the degrees of these effects in different relations.

To sum up, there are two shortcomings in the aforementioned related works. On the one hand, the research on stock ranking prediction for portfolio optimization obtained the stock ranking results only based on higher expected returns while the stock price volatility risk is disregarded. As a crucial aspect of financial investments, risk-oriented criteria should be considered and introduced into the stock ranking method for stock selection.

On the other hand, most existing research preliminarily attempted to capture the effects of related stocks from a singular relation, disregarding the multiple spillover effects from related stocks with different relationships. Moreover, existing research ignored the distinctions in effects from various relations. The effects of each relation on the target stock are different. To effectively capture the effects of related stocks, the effects of related stocks from various relations should be aggregated, and the distinctions in relation strength should be learned. Table 1 shows the difference between the most relevant works and our proposed method.

3. Methodology

Following previous studies [8–10], stock portfolio optimization is formulated as a stock ranking problem to select superior stocks in this work. We propose an HGA-MT method for stock ranking and portfolio optimization. The model introduces an HGA mechanism to fuse the textual and quantitative features of multiple related stocks and employs a multi-task learning paradigm to concurrently learn both the stock risk and return. Given $S = \{s_1, s_2, \ldots, s_N\}$ is the set of N stocks, our task is to predict the stock ranking results of the N stocks on the next trading day, and then accordingly select Top-K ranked stocks to achieve a higher and sound investment revenue. As shown in Fig. 1, the proposed HGA-MT method includes four technical components. First (Section 3.1), we obtain the quantitative indicators and news features of N stocks in the set S and use a BiLSTM encoder to capture the temporal information. The input feature (F) of each training step aggregates the features of the N stocks over a time window of T days. T represents the count of trading days preceding the day of stock ranking prediction. The BiLSTM yields the post-encoding feature matrix (H) of the N stocks, subsequently. Second (Section 3.2.1), the stock relationship data of the N stocks are collected and presented via a heterogeneous graph (G), including the supply chain, shareholding chain, and industry competitor relationships. To distinguish the effects of related stocks in different relation types, we construct three relation-specific subgraphs that are defined as G_{su}, G_{sh} and G_{co}, corresponding to supply, shareholding, and industry competitor relationships, respectively. Third (Section 3.2.2), GCN is utilized to capture the stock relations, fuse the effects of related stocks in each relationship, and an attention mechanism is utilized to assign different weights to different relation types. We aggregate the overall effects of related stocks to form a post-fusion matrix (X) by summing up the weighted information extracted from the supply chain (X_{su}), shareholding chain (X_{sh}), and industry competitor (X_{co}) relationships. Finally (Section 3.3), the post-fused matrix X is learned in a multi-task learning paradigm to predict the stock return ($\hat{y}^{Return}_{t+1}$) and volatility risk ($\hat{y}^{Risk}_{t+1}$) on the next trading day $t+1$. We obtain the stock ranking results ($\hat{Rank}^{ROR}_{t+1}$) by considering both the risk and return in a score function. The Top-K-ranked stocks are selected as an investment portfolio to achieve sound investment profits. The description of symbols that are introduced in this paper is shown in Table 2.

3.1. Stock feature extraction

3.1.1. Quantitative indicator and news feature extraction

The extraction of quantitative indicators. Extensive research [2,6,18] has proved that fundamental[2] and technical indicators[3] are effective features for stock price prediction. Following previous studies [15,

[2] The fundamental features are used to show the intrinsic value of stocks from several aspects, such as profitability, operating ability, capital structure, and liquidity analysis.

[3] The technical indicators are used to capture the stock price patterns by the mathematical operation based on the given historical transaction data, e.g., price, trading volume or time span.

Table 1
The comparison of related works with our proposed method. QI denotes the quantitative indicators; News denotes the financial news features. RSL denotes relation strength learning.

Studies	Features		Modeled relationships	RSL	Selected by	
	QI	News			Return	Risk
Sawhney et al. [10]	✓	–	Industry	—	✓	—
Hsu et al. [8]	✓	–	Industry	—	✓	—
Ma et al. [5]	✓	–	Industry, Concept, Fund-hold.	—	✓	—
Feng et al. [12]	✓	–	Price trend	—	✓	—
Wu et al. [54]	✓	–	Price trend	—	✓	—
Chen et al. [26]	✓	–	Sharehold.	—	✓	—
Ma et al. [15]	✓	✓	Supply, Sharehold., Industry competitor	—	✓	—
Ours	✓	✓	Supply, Sharehold., Industry competitor	✓	✓	✓

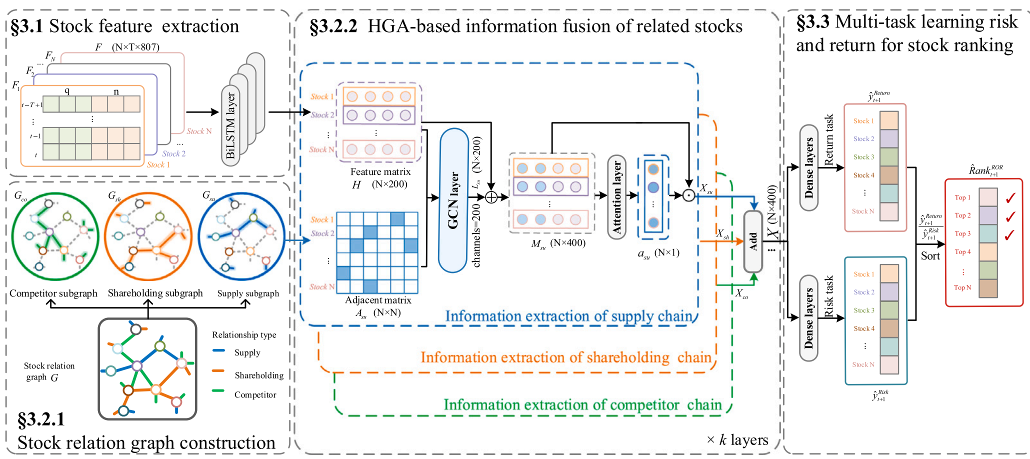


Fig. 1. Overview of the HGA-MT framework.

20,43,44,55] with superior prediction results, 39 widely used indicators are used in this work, including 8 fundamental features that show the intrinsic values of stocks and 31 technical indicators that capture the stock price patterns. These fundamental and technical indicators are denoted as quantitative indicators. Table 3 shows all quantitative indicators used in this work. The original stock data were gathered from the RESSET platform.[4] Then, Min-Max normalization is employed to standardize each indicator [43], owing to their distinct value ranges. The quantitative indicators of Stock i in the trading Day t are described as:

$$\mathbf{q}_{i,t} = [ROA_{i,t}, DAR_{i,t}, \dots, LowerBand_{i,t}], \tag{1}$$

where $\mathbf{q}_{i,t} \in \mathbb{R}^{1\times39}$. 39 is the vector dimension of quantitative indicators.

The extraction of news features. With the advancement of natural language processing techniques, financial news has been wildly incorporated with quantitative indicators to improve stock price prediction accuracy. Our previous work [15] proves that a pre-trained Market-driven Sentiment Classifier (MSC) can effectively generate market-driven news sentiment features. Therefore, the MSC module from the previous work [15] is also applied to generate news embedding features in this study. MSC is a pre-trained sentiment classifier based on Chinese-
RoBERTa [56] and news titles. The pre-training process encompasses training a Chinese-RoBERTa classifier to learn market-driven sentiment polarities. The polarity labels of market-driven sentiment are established based on the variation in stock prices on the day after the news release.

[4] RESSET (http://www.resset.com/) is a Chinese platform, providing financial and economic information.

If the stock price goes down, the sentiment label of the news title is negative; otherwise, positive. In our application, the pre-trained MSC from the work of Ma et al. [15] is used to generate embeddings for news titles. In practical scenarios, a stock may receive more than one news article in a day. Therefore, the news embeddings of the stock in a day are averaged, thereby obtaining a daily news feature ($\mathbf{n} \in \mathbb{R}^{1\times768}$, where 768 is the vector dimension of news feature) to represent all news titles in a day. The process of generating a news embedding of the Stock i in Day t can be defined as:

$$\mathbf{n}_{i,t} = MSC(News_1, News_2, \dots, News_m), \tag{2}$$

where $\mathbf{n}_{i,t} \in \mathbb{R}^{1\times768}$ denotes the daily news feature of Stock i on Day t. $MSC(\cdot)$ refers to the pre-trained MSC-based news embedding generator. $News_1, \dots, News_m$ are m financial news titles on Day t related to Stock i.

Then, the quantitative indicators $\mathbf{q}_{i,t}$ and news features $\mathbf{n}_{i,t}$ of Stock i on Day t are concatenated ($\oplus$), yielding its daily feature ($\mathbf{f}_{i,t}$) by

$$\mathbf{f}_{i,t} = \left[\mathbf{q}_{i,t} \oplus \mathbf{n}_{i,t}\right], \tag{3}$$

where $\mathbf{f}_{i,t} \in \mathbb{R}^{1\times807}$ denotes the daily feature of Stock i on Day t. 807 is the vector dimension of concatenated quantitative indicators and news features.

3.1.2. Temporal feature extraction

The stock market data are typical time series data. BiLSTM performs well in processing sequential data in stock price prediction [1–3]. Thus, BiLSTM is used to learn the time series features. For a training step that learns the stock risk and return at Day $t+1$, the features in lookback window T are used to capture the time series information. The lookback window T denotes T trading days before the prediction day (Day $t+1$), i.e., T trading day from $t-T+1$ to t.

Table 2
Symbol description.

Notation	Description
S	The set of stocks.
N	The number of stocks in the stock set S.
K	The number of top-ranked stocks in an investment portfolio.
T	The number of trading days before the prediction day.
$\mathbf{q}_{i,t}$	$\mathbf{q}_{i,t} \in \mathbb{R}^{1\times 39}$, the quantitative indicators of Stock i on Day t.
$\mathbf{n}_{i,t}$	$\mathbf{n}_{i,t} \in \mathbb{R}^{1\times 768}$, the news features of Stock i on Day t.
$\mathbf{f}_{i,t}$	$\mathbf{f}_{i,t} \in \mathbb{R}^{1\times 807}$, the daily features of Stock i on Day t.
F_i	$F_i \in \mathbb{R}^{T\times 807}$, the feature matrix of Stock i within a time window T.
F	$F \in \mathbb{R}^{N\times T\times 807}$, the feature matrix of N stocks within a time window T.
H	$H \in \mathbb{R}^{N\times 200}$, the hidden states of BiLSTM of N stocks, representing the feature matrix of N stocks.
$G = (V, E, R)$	G is a heterogeneous graph of stock relation. V is the set of nodes (stocks). E is the set of edges (relations). R is the set of relation types, including supply chain (su), shareholding chain (sh) and industry competitor (co) relation.
G_{su}, G_{sh}, G_{co}	G_{su}, G_{su}, G_{su} are the subgraphs of G only containing the su, sh and co relation, respectively.
A_{su}, A_{sh}, A_{co}	A_{su}, A_{sh}, $A_{co} \in \mathbb{R}^{N\times N}$, the adjacency matrix associated with the subgraphs G_{su}, G_{sh} and G_{co}, respectively.
L_{su}, L_{sh}, L_{co}	L_{su}, L_{sh}, $L_{co} \in \mathbb{R}^{N\times 200}$, the hidden states of GCN, fusing the features of related stocks in the su, sh and co relations, respectively.
M_{su}, M_{sh}, M_{co}	M_{su}, M_{sh}, $M_{co} \in \mathbb{R}^{N\times 400}$, the concatenation of L_{su}, L_{sh} and L_{co} with H, respectively.
a_{su}, a_{sh}, a_{co}	a_{su}, a_{sh}, $a_{co} \in \mathbb{R}^{N\times 1}$, the importance (attention) weights of related stocks in su, sh and co relations, respectively.
X_{su}, X_{sh}, X_{co}	X_{su}, X_{sh}, $X_{co} \in \mathbb{R}^{N\times 400}$, the effects of related stocks from the su, sh and co relations, respectively, given by HGA.
X	$X \in \mathbb{R}^{N\times 400}$, a post-fusion matrix that aggregates the effects of related stocks from the su, sh and co relations.
k	The hyperparameter for GCN layers.
α	The hyperparameter for balancing the losses of the two subtasks.

Table 3
The employed fundamental and technical indicators of stocks.

Feature name	Feature name
Return on Asset (ROA)	Debt Asset Ratio (DAR)
Earnings Per Share (EPS)	Total Assets Growth Rate (TAGR)
Operational Profit (OProfit)	Revenue Growth Rate (RGR)
Asset Turnover Rate (ATR)	Gross profit growth rate (GPGR)
Opening/Closing Price (OP/CP)	Moving Average (MA10/MA20/MA30)
Highest/Lowest Price (HP/LP)	Simple/Exponential Moving Average (SMA/EMA)
Volume of Trade (Vol)	Relative Strength Index (RSI6/RSI12/RSI24)
Turnover Rate (TurR)	Money Flow Index (MFI)
Daily Return (DR)	Daily Amplitude (DA)
Price-to-Earnings Ratio (P/E)	Stochastic Indicator KDJ (Slowk/Slowd)
Price-to-Book Ratio (P/B)	True/Average True Range (TR/ATR)
Price-to-Sales Ratio (P/S)	Williams Indicator (WR)
Moving Average Convergence and Divergence (DIFF/DEA/MACD)	Bollinger Bands (UpperBand/ MiddleBand/ LowerBand)

Then, the input features of Stock i in a training step are defined as

$$F_i = \left[\mathbf{f}_{i,t-T+1}, \dots, \mathbf{f}_{i,t-1}, \mathbf{f}_{i,t}\right], \tag{4}$$

where $F_i \in \mathbb{R}^{T\times 807}$ is the feature matrix of the Stock i during the preceding T trading days. We analyze the sensitivity of T in Section 5.3.

Next, F_i is fed into a single BiLSTM layer with a dropout ($\mathcal{D}(\cdot)$) to discover temporal patterns of Stock i:

$$h_{i,t+1} = \mathcal{D}(BiLSTM(F_i)), \tag{5}$$

where $h_{i,t+1} \in \mathbb{R}^{1\times 200}$ is the final hidden states of $BiLSTM(\cdot)$ outputs of Stock i with the dropout. The hidden state size of the BiLSTM layer is 200. Since there are N input stocks at a training step, the overall input feature set is $F \in \mathbb{R}^{N\times T\times 807}$. After BiLSTM encoding F and dropout, the temporal features (H) of all stocks are defined as:

$$H = \left[h_{1,t+1}, h_{2,t+1}, \dots, h_{i,t+1}, \dots, h_{N,t+1}\right], \tag{6}$$

where $H \in \mathbb{R}^{N\times 200}$ denotes the feature matrix of N stocks.

3.2. Information fusion of related stocks

In this section, we first develop the stock relation graphs that are used to present the relations between stocks (Section 3.2.1). Then, we present the process for the information fusion of related stocks based on the proposed HGA module (Section 3.2.2).

3.2.1. Stock relation graph construction

The relationships among listed companies are diverse and interconnected. Stock prices are affected by multiple spillover effects from related stocks. The relation graph is an appropriate method for representing the stock relationship information. In this study, we collect the relationship data for N stocks from Qichacha,[5] including the supply chain (suppliers and customers), the shareholding chain (holding and being held), and the industry competitor relationships. Qichacha is a leading information service in China that focuses on companies' enterprise credit and relationship information. They objectively determine the related companies of each company based on official reports and announcements. More details regarding the data collection and stock relations extraction by Qichacha are shown in Appendix A. In this work, the interactive relations among stocks are obtained from Qichacha.

Then, the relationship data of N stocks is presented by a heterogeneous graph (G). In the heterogeneous graph $G = (V, E, R)$, V is the set of N nodes, where each node denotes a stock. In this study, the node (stock) set V includes 86 nodes (stocks) (N=86, details shown in Section 4.1). E is the set of edges, where each edge represents that two connected stocks have a type of business relationship. In this study, the set E includes 336 edges, indicating 336 pairwise relationships among 86 stocks. R is the set of edge types (i.e., relation types). In our work, R includes three types: the supply chain (su), the shareholding chain (sh), and the industry competitor (co) relationships. Each edge in set E belongs to one of the three edge types.

In the stock relation heterogeneous graph, a stock node is usually linked with other stock nodes by multiple relations. Related nodes in different relation types have different influences on the target stock node because the spillover effects on different relation types are different. Therefore, the information aggregation of related stock nodes in different relation types should be distinguished. According to the relation types in the heterogeneous graph (G), we construct three relation-specific subgraphs, namely G_{su}, G_{sh} and G_{co}. Fig. 2 shows the process of developing the stock relation graphs and the adjacent matrices. $G_{su} = (V, E_{su})$ denotes the subgraph of G only containing the supply chain relation, where all nodes are linked by supply chain relations. Similarly, G_{sh} and G_{co} refer to the subgraphs only connecting stock nodes by shareholding and industry competitor relations, respectively.

[5] https://www.qcc.com/

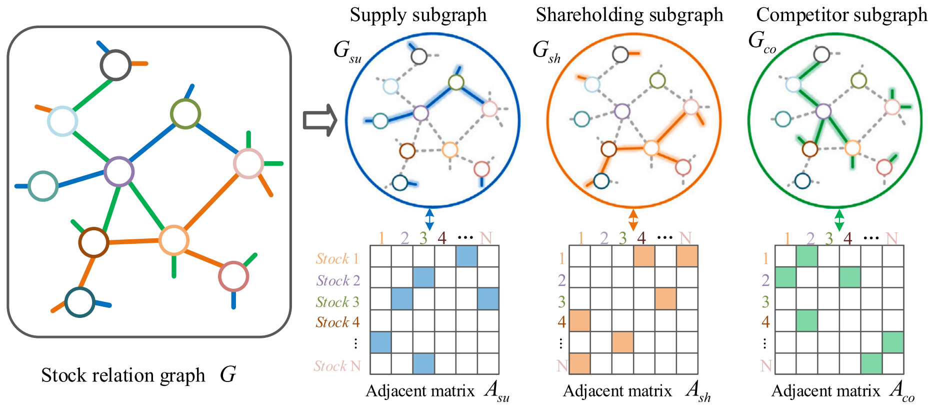


Fig. 2. The illustration of stock relation graph and adjacent matrix construction.

According to the subgraphs G_{su}, we construct the relation-specific adjacency matrix $A_{su} \in \mathbb{R}^{N\times N}$ which illustrates the supply relationships between the N stocks. The adjacency matrix A_{su} is defined as Eq. (7). $a_{i,j}$ is the element of Matrix A_{su}. If Stocks i and j share a supply chain relation (the nodes are connected), the conjunction $a_{i,j}$ is 1, otherwise is 0.

$$A_{su} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N,1} & a_{N,2} & \cdots & a_{N,N} \end{bmatrix} \tag{7}$$

Similarly, we can obtain the adjacency matrix $A_{sh} \in \mathbb{R}^{N\times N}$ and $A_{co} \in \mathbb{R}^{N\times N}$, representing the shareholding chain and industry competitor relationships among the N stocks, respectively.

3.2.2. HGA-based information fusion of related stocks

Since the stock relationships are represented by graphs, to combine the impacts of associated stocks, we adopt a GCN network [34] to learn the graph. The GCN has demonstrated the ability to efficiently acquire the structural information present in graphs [57–59]. The inputs of the GCN model consist of a feature matrix, including each node's features, and an adjacency matrix, presenting the relationships between nodes. The GCN model can update the feature representation of each node according to the adjacency matrix and fuse the features of its neighbor nodes by the convolutional operation. In practice, the effect extent of every relation to the nodes is different. For example, the effects of related stocks in the shareholding chain differ from those of industry competitors. Thus, assigning different importance weights to different relations is essential. We propose the HGA module to aggregate the effects of related stocks in multiple relationships by GCN networks and assign different importance weights to different relation types by an attention mechanism.

The process of the HGA module for the information fusion of related stocks is shown in Fig. 3. To learn the impacts of related stocks in supply chain relation, first, the feature matrix (H) and the adjacency matrix (A_{su}) containing the supply chain relations between these nodes are input into a GCN layer for convolutional processing (see Eq. (8)). The feature vector of every node is enhanced by integrating the information from neighboring nodes in GCN. Consequently, the updated hidden states of each stock node in L_{su} encompasses the information about its interrelated stocks in the supply chain relation.

$$L_{su} = ReLU(\widetilde{D}^{-1/2} A_{su} \widetilde{D}^{-1/2} H W), \tag{8}$$

where $\widetilde{D}^{-1/2} A_{su} \widetilde{D}^{-1/2} \in \mathbb{R}^{N\times N}$ is Laplacian matrix. $\widetilde{D} \in \mathbb{R}^{N\times N}$ denotes a diagonal degree matrix, where $\widetilde{D} = \Sigma_j A_{i,j}$. $A_{su} \in \mathbb{R}^{N\times N}$ is an adjacency matrix. $W \in \mathbb{R}^{200\times 200}$ denotes trainable parameters in the GCN layer. We use a $ReLU(\cdot)$ activation function. $L_{su} \in \mathbb{R}^{N\times 200}$ is the activated GCN output, whose hidden state size is 200. L_{su} represents the updated representation of each stock node in the graph by leveraging the information of its related stocks in the supply chain relationship. Then, the GCN output matrix L_{su} and the feature matrix H are concatenated.

$$M_{su} = L_{su} \oplus H, \tag{9}$$

where $M_{su} \in \mathbb{R}^{N\times 400}$. 400 is the matrix dimension of concatenated H and L_{su}.

Next, considering the fact that related stocks in different relation types have different impacts and importance on the target stock, the model should learn the importance of each relation. The matrix M_{su} is fed into a dense layer and sigmoid to calculate the importance weight ($a_{su} \in \mathbb{R}^{N\times 1}$) of related stocks in supply chain relation

$$a_{su} = sigmoid(Dense(M_{su})). \tag{10}$$

This process is defined as the attention layer in Fig. 3.

With the weight coefficient, the information of related stocks in the supply chain relation ($X_{su} \in \mathbb{R}^{N\times 400}$) can be captured by Eq. (11), where $\odot$ indicates dot product

$$X_{su} = a_{su} \odot M_{su}. \tag{11}$$

Similarly, we can obtain the feature matrices M_{sh} and M_{co}, and the importance weights a_{sh} and a_{co}, thereby capturing the effects of the related stocks in shareholding (X_{sh}) and industry competitor (X_{co}) relationships.

The final representation of a stock node can be obtained by aggregating the information of related stocks in the supply chain, shareholding chain, and industry competitor relationships:

$$X_t = X_{su} + X_{sh} + X_{co}. \tag{12}$$

This process of HGA can be repeated k times for more comprehensive feature fusion. We examine the GCN layers with different k values in Section 5.3.

3.3. Multi-task learning risk and return for stock ranking

3.3.1. Multi-task learning risk and return

Both risk and return drive investment decision-making. This inspires us to use a multi-task learning paradigm to jointly learn risk and return for stock ranking. The multi-task learning module can improve return and risk predictability because these learning tasks can leverage shared knowledge and utilize the correlation between them to enhance their generalization performance. More importantly, by taking into account both the predicted stock return ratio and stock volatility risk, we can attain more appropriate stock ranking results rather than solely prioritizing them based on their return ratios.

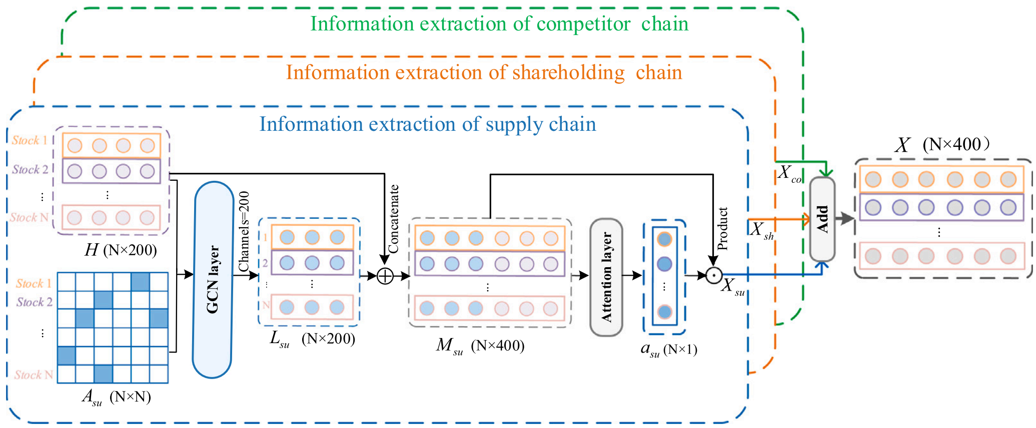


Fig. 3. The illustration of the proposed HGA module for the information fusion of related stocks.

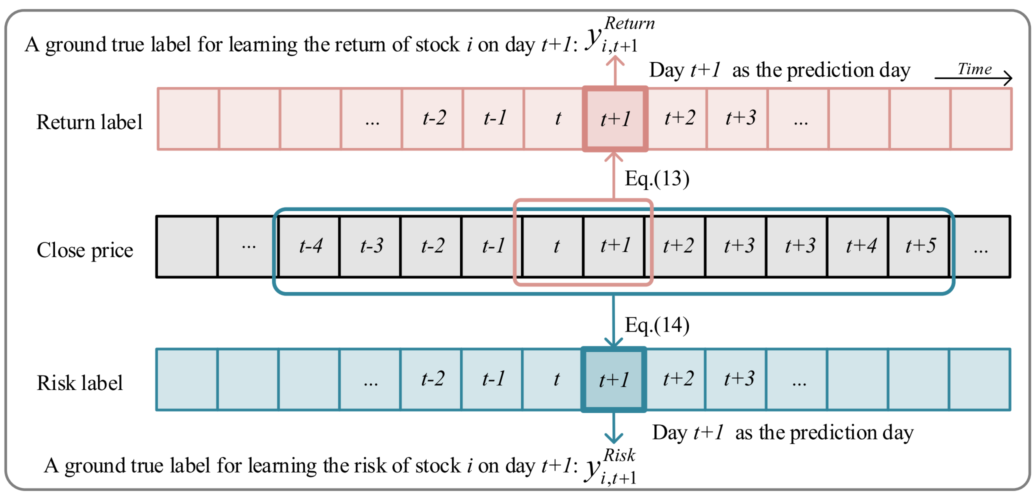


Fig. 4. The method for computing ground truth risk ($y^{Risk}_{i,t+1}$) and return ($y^{Return}_{i,t+1}$) labels for Stock i at trading Day $t+1$.

The stock return ratio is used as a training target for learning the “return” subtask. The ground truth return ratio of Stock i on the trading Day $t+1$ is denoted as

$$y^{Return}_{i,t+1} = \frac{CP_{i,t+1} - CP_{i,t}}{CP_{i,t}}, \tag{13}$$

where the $CP_{i,t+1}$ and $CP_{i,t}$ are the close price of Stock i on the trading Day t and Day $t+1$, respectively.

The stock price volatility during a period of time is used as another training target for learning the “risk” subtask. Wild volatility means higher risk. We define the standard deviation of 10-day close prices from Day $t-4$ to Day $t+5$ as the ground truth risk of Stock i on Day $t+1$:

$$y^{Risk}_{i,t+1} = \frac{\sqrt{\sum_{j=t-4}^{t+5}(CP_{i,j} - \overline{CP_i})}}{10}, \tag{14}$$

where $j \in \{t-4, t-3, \ldots, t+5\}$. $CP_{i,j}$ is the close price of Stock i on the trading Day j. $\overline{CP_i}$ is the mean value of $CP_{i,j}$ during 10 trading days from Day $t-4$ to Day $t+5$. We heuristically used the close prices during 10 trading days (trading days in two weeks) to calculate the risk label. Noticeably, the close prices from $t-4$ to $t+5$ are only used to calculate the risk label on Day $t+1$. The risk label ($y^{Risk}_{i,t+1}$) and return label ($y^{Return}_{i,t+1}$) are our training targets rather than model input. Thus, the proposed method does not involve the issue of label leakage. The calculation process for $y^{Return}_{i,t+1}$ and $y^{Risk}_{i,t+1}$ is shown in Fig. 4.

The process of stock selection by multi-task learning risk and return is shown in Fig. 5. The matrix X that aggregates the information from related stocks is fed into task-specific layers to learn features related to each subtask. Task-specific layers are composed of two dense layers. The predictions of return and risk measures for N stocks on the trading Day $t+1$ can be yielded from their task-specific layers

$$\begin{aligned} \hat{y}^{Return}_{t+1} &= Dense(ReLU(Dense(ReLU(Dense(X))))), \\ \hat{y}^{Risk}_{t+1} &= Dense(ReLU(Dense(ReLU(Dense(X))))), \end{aligned} \tag{15}$$

where $\hat{y}^{Return}_{t+1} = \left[\hat{y}^{Return}_{1,t+1}, \hat{y}^{Return}_{2,t+1}, \ldots, \hat{y}^{Return}_{N,t+1}\right]$, $\hat{y}^{Risk}_{t+1} = \left[\hat{y}^{Risk}_{1,t+1}, \hat{y}^{Risk}_{2,t+1}, \ldots, \hat{y}^{Risk}_{N,t+1}\right]$. $\hat{y}^{Return}_{t+1} \in \mathbb{R}^{N\times 1}$ and $\hat{y}^{Risk}_{t+1} \in \mathbb{R}^{N\times 1}$ are the predicted return ratio and risk of N stocks on the trading Day $t+1$, respectively.

The multi-task learning objective is optimized using a joint loss function, which simultaneously predicts the return ratio and risk of stocks. It updates network parameters by combining the loss values of both subtasks. We use the mean absolute error (MAE) as the loss function for the two subtasks. The final joint loss function of the proposed multi-task learning module ($\mathcal{L}_{RR}$) is denoted as

$$\begin{aligned} \mathcal{L}_{RR} &= (1-\alpha)\mathcal{L}_{Return} + \alpha\mathcal{L}_{Risk}, \\ \mathcal{L}_{Return} &= MAE(\hat{y}^{Return}_{t+1}, y^{Return}_{t+1}), \\ \mathcal{L}_{Risk} &= MAE(\hat{y}^{Risk}_{t+1}, y^{Risk}_{t+1}), \end{aligned} \tag{16}$$

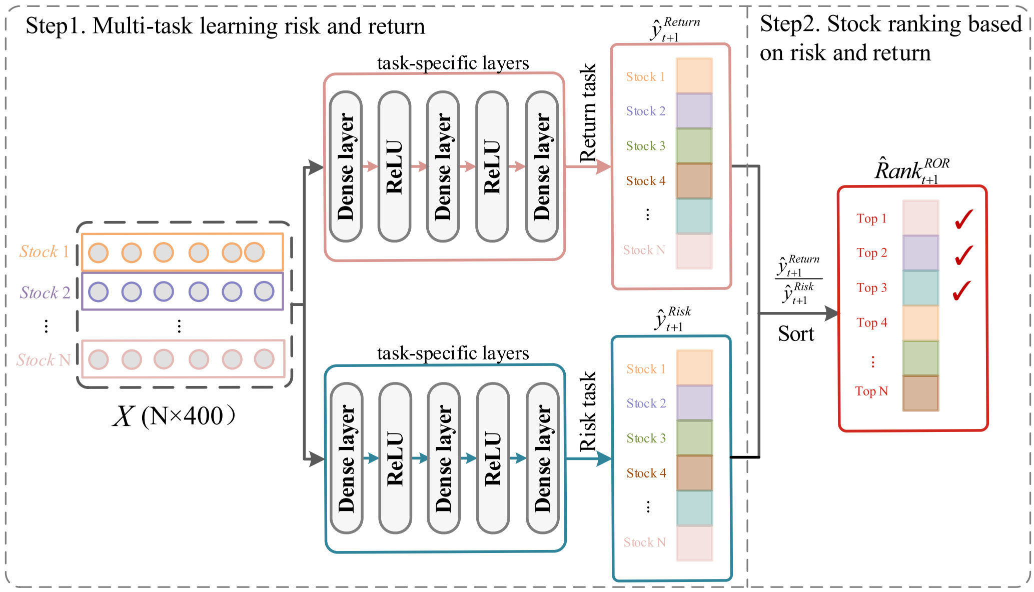


Fig. 5. Stock selection with multi-task learning return and risk.

where y_{t+1}^{Return} and y_{t+1}^{Risk} are the ground truth return ratio and risk of N stocks on trading Day $t+1$. α is a hyperparameter for balancing the losses of the two subtasks. Various values of α are evaluated in Section 5.3.

3.3.2. Trading-oriented stock selection based on risk and return

We aim to obtain a more sound ranking result for stock investment by considering both risk and return. The return per unit of risk (ROR) is used as a measure for stock ranking. It is the ratio of return to volatility (risk), which means how much return is yielded for each unit of risk. ROR of Stock i on the trading Day $t+1$ is defined as

$$ROR_{i,t+1} = \frac{y_{i,t+1}^{Return}}{y_{i,t+1}^{Risk}}. \tag{17}$$

We compute the ROR for all stocks and rank the stocks in descending order ($D\_sort(\cdot)$). The predicted ranking result based on ROR on Day $t+1$ is given by

$$\hat{Rank}_{t+1}^{ROR} = D\_sort([\hat{ROR}_{1,t+1}, \dots, \hat{ROR}_{i,t+1}, \dots, \hat{ROR}_{N,t+1}]). \tag{18}$$

Thus, the ranking result takes both the stock risk and return factors into account. Finally, the Top-K ranked stocks out of the N stock candidates are selected to invest as a portfolio, according to the ranking results.

4. Experiments

4.1. Datasets

As an emerging market and an essential part of the world economy, the Chinese stock markets are attractive to explore [2,3]. This work focuses on Chinese stock markets and executes the experiments using CSI100 index constituent stocks. CSI100 is the typical Index of Chinese markets. The stocks comprising the CSI100 Index are considered highly representative, characterized by significant market capitalization and robust liquidity in the Chinese market. As such, they serve as widely applicable and pertinent choices for financial investments. To ensure the validity of the experiment and minimize the impact of missing or abnormal data on the comparative analysis, we exclude stocks with limited trading days and financial news. In total, there are 86 stocks ($N = 86$) in our stock set.

The features of each stock in our stock set include 39 quantitative indicators and financial news features (shown in Section 3.1.1). The 39 quantitative indicators from January 2, 2018 to June 2, 2021 were collected from the RESSET platform[4]. Moreover, a total of 134,520 news titles concerning 86 stocks were gathered from Hundsun Electronics,[6] spanning from January 2, 2018 to June 2, 2021. Details of the stock candidates and the number of news titles of each stock are shown in Appendix B. Besides, the relations (edges) among 86 stocks (nodes) were gathered from Qichacha[5], including three relation types: supply chain, shareholding chain, and industry competitor relations. We obtained 336 relations among 86 stocks.

Table 4
Dataset statistics.

Dataset	Start	End	#days	#news
Train	02/01/2018	25/05/2020	581	42 515
Val.	26/05/2020	21/09/2020	83	20 095
Test	22/09/2020	02/06/2021	166	71 910
All	02/01/2018	02/06/2021	830	134 520

We split the dataset 70% for training, 10% for validation, and 20% for testing. The statistics of the employed dataset are shown in Table 4. Following previous research [20,60], the "walk forward testing" method [60] is conducted for maximally utilizing the available data.

4.2. Baselines

To verify the validity of the proposed HGA-MT method, we compare it with the following baselines, namely MAC, FinGAT, LFM, TRAN, and TGC.

- MAC [15] is a GCN-based model that effectively aggregates the news sentiments of related stocks for stock price movement prediction. The original settings of MAC are adopted. We change the learning target of MAC from the stock price movement prediction (a classification task) to the stock return ratio prediction (a regression task). Then, we rank stocks by the predicted return ratios of MAC.

[6] Hundsun Electronics (https://www.hundsun.com) is a prominent Chinese financial technology company that offers comprehensive financial technology services.

- FinGAT [8] is a graph attention network-based model that incorporates a hierarchical correlation of sector and stocks to recommend the Top-K profitable stocks. It devises a multi-task objective that recommends profitable stocks and predicts stock movements. The original settings of FinGAT are adopted. The sector relationship information of the stocks in this study is collected from RESSET due to the different datasets.
- LFM [61] is a multi-task learning model that integrates LSTM and random forest (RF) to simultaneously predict stock return and return movement directions. LFM is a non-graph-based method that does not incorporate related stocks' information. The original settings of LFM are adopted. We first obtain the predicted stock return ratios, then accordingly rank stocks.
- TRAN [11] is a graph-based model considering the industry relationships between stocks to rank the return ratios for stock recommendations. We adopt the model structure and original settings of TRAN. The sector relationship between stocks is collected from RESSET. Our quantitative indicators and news features are employed in this model.
- TGC [9] is a classic model that treats stock prediction as a return ratio ranking task for stock recommendations. It captures the stock relations by a graph-based model. The quantitative indicators and model are employed from the original paper without modification, where the stock relation graph is constructed by our stock relation data.

4.3. Setups

We adopt Keras and TensorFlow frameworks to perform the proposed model. The batch size is set to 32. The training process is limited to a maximum of 500 epochs and equipped with an early-stop mechanism. The Adam optimizer [62] is used to optimize the model parameters with an initial learning rate of 0.001. To regularize the weight matrix of the kernel, we apply activity regularization L2 (0.001). The experiments are performed on GPU machines (Tesla V100). In particular, we carry out several experiments to explore the optimal sliding window length T within $\{1,3,5,7,9,11,13,15\}$, the number of GCN layers (k) within $\{1,2,3,4\}$ and the loss weighing factor (α) of risk specific-task within $\{0.001,0.01,0.1,0.2,0.3,0.4,0.5,0.7\}$, based on the validation sets. A grid search process is adopted to optimize the model performance, where all the possible combinations of 3 hyperparameters are tested to select the optimal values. Section 5.3 shows more details on verifying these hyperparameters. The testing results are reported by using the optimal time step ($T=7$), 2 GCN layers, and the loss weighing factor $\alpha=0.3$.

4.4. Evaluation metrics

The proposed model is evaluated in terms of ranking performance and financial evaluation. Ranking performance is an intuitive evaluation of the data science model for stock ranking. Financial evaluation is a crucial aspect of investments.

To evaluate the ranking performance, Precision and Mean Reciprocal Rank (MRR) are employed as evaluation metrics, following previous studies [8,12,13]. A larger value of Precision or MRR indicates more accurate ranking performance. Specifically, Precision reflects the hit percentage of the predicted ranking results.

$$Precision = \frac{\left|L@K(\hat{Rank}^{Return}) \cap L@K(Rank^{Return})\right|}{K}, \tag{19}$$

where the ranking result of the ground truth return ratios of N stocks is denoted as $Rank^{Return}$, which orders the ground truth return ratios in y^{Return} in descending; The ranking result of predicted return ratios of all stocks is denoted as $\hat{Rank}^{Return}$, which orders the predicted return ratios in $\hat{y}^{Return}$ in descending.

The lists of ground truth and predicted Top-K stocks are denoted as $L@K(Rank^{Return})$ and $L@K(\hat{Rank}^{Return})$, respectively. MRR reflects the ranking order of the correct results among the predicted ranking results.

$$MRR = \frac{1}{K} \sum_{i \in L@K(\hat{Rank}^{Return})} \frac{1}{position(Rank_i^{Return})}, \tag{20}$$

where $position(Rank_i^{Return})$ returns the ground truth ranking position of Stock i. In the following evaluation that uses Precision and MRR metrics in Section 5, we do not use ROR that takes both risks and returns into account for the stock ranking ((see Section 3.3.2)). This is because the ground truth ranking results are ordered by return ratios only. Using $\hat{Rank}^{Return}$ along is more in line with the evaluation criteria of the stock ranking task setup.

To evaluate the financial performance of a model, we conduct a simulation of stock trading based on the predicted rankings and the Top-K stocks recommended by the model. Following previous studies [9–11,13,15], the metrics including cumulative Investment Return Ratio (IRR) [9], Sharpe ratio (SP) [63], Total money (TMoney) [28] and Maximum Drawdown (MDD) [64] are adopted in this study for financial evaluation in terms of return and risk. A larger value of IRR, SP, or TMoney generally indicates better performance in profitability; A lower value of MDD indicates lower downside risk and a smoother performance trajectory.

The IRR is adopted as the main metric to evaluate profitability, which is calculated by summing over the mean return ratios of the K selected stocks on each testing day.

$$\begin{aligned} IRR &= \sum_{q=1}^{Q} IRR_q, \\ IRR_q &= \frac{\sum_{i \in L@K(\hat{Rank}_q)} \frac{CP_{i,q}-CP_{i,q-1}}{CP_{i,q-1}}}{K}, \end{aligned} \tag{21}$$

where IRR_q is the real return ratio on the qth testing day, Q is the number of testing days; $L@K(\hat{Rank}_q)$ is the set of Top-K stocks according to the predicted ranking results on the qth testing day; $CP_{i,q}$ is the close price of Stock i on the qth testing day.

TMoney [28] is adopted as an essential metric to evaluate profitability intuitively. TMoney is the summation of the funds for each stock in the portfolio, which represents the real funds statement of the portfolio. The concept of TMoney is inspired by the work of Chou et al. [36] and basically uses a similar idea of its funds standardization. The daily TMoney in a portfolio can be calculated by

$$TMoney_q = \sum_{i \in L@K(\hat{Rank}_q)} CP_{i,q} \times Share_{i,q}, \tag{22}$$

where $Share_{i,q}$ is the number of holing shares of Stock i on the qth testing day. SP is the measure of risk-adjusted return, which is the average return earned in excess of the risk-free rate per unit of volatility (risk). A higher SP means a higher return relative to the amount of risk taken. SP is one of the most common financial evaluation metrics that has been widely used in previous studies [10,13,15].[7]

$$\begin{aligned} SP &= \frac{E(R_p) - R_f}{std(R_p)}, \\ &= \frac{\frac{1}{n}\sum_{q=1}^{n} R_{p,q} - R_f}{\sqrt{\frac{\sum_{q=1}^{n}(R_{p,q} - \frac{1}{n}\sum_{q=1}^{n} R_{p,q})^2}{n}}} \end{aligned} \tag{23}$$

[7] Trend Ratio [35] is also a useful metric. Compared with the Sharpe ratio, the risk in the trend ratio is the deviation between the portfolio trend and the fund standardization, instead of the deviation between the average of fund standardization and the fund standardization. The trend ratio takes the trend of the portfolio into consideration. This study uses SP instead of Trend Ratio because of their alignment with investors' intuitive comprehension of risk-adjusted returns. Besides, SP is widely used in related works and our baseline works.

Table 5
The ranking performance of different models over Top-K stocks. $K \in \{3, 5, 10\}$. "Prec." denotes the Precision.

	Top3		Top5		Top10		Avg.	
	Prec.	MRR	Prec.	MRR	Prec.	MRR	Prec.	MRR
TGC	0.0695	0.1687	0.0761	0.2066	0.0869	0.2969	0.0775	0.2241
TRAN	0.0675	0.1731	0.0845	0.2199	0.1104	0.3586	0.0875	0.2505
LFM	0.0577	0.1686	0.0757	0.2117	0.0859	0.2864	0.0731	0.2222
FinGAT	0.0634	0.1644	0.0712	0.2275	0.1372	0.3920	0.0906	0.2613
MAC	0.0670	0.1736	0.0911	0.2206	0.1612	0.4054	0.1064	0.2665
Ours	**0.0859**	**0.1820**	**0.1006**	**0.2252**	**0.1681**	**0.4587**	**0.1182**	**0.2886**

where

$$R_{p,q} = \frac{TMoney_q - TMoney_{q-1}}{TMoney_{q-1}}, \tag{24}$$

R_f is the annual risk-free rate. In our evaluation, the yearly risk-free rate stands at 2.5%, aligning with established literature [28] and comparable to various residual maturities of China Government Bonds, spanning from 1 year to 10 years. R_p is the return rate of the portfolio p over a testing period. $R_{p,q}$ denotes the return rate of the qth day within the period. n is the number of trading days during the testing period. The daily return rates are calculated by the daily TMoney (Eq. (24)) that represents the real fund statement of the portfolio. Thus, the change in return rates given by TMoney can truly reflect the return and risk varieties in a portfolio.

MDD also serves as an instinctive and widely used financial metric, effectively capturing the portfolio's downside risk. In portfolio analysis, MDD represents the largest percentage decline in value that the investment experiences from its peak value to its trough value before reaching a new peak. A lower MDD denotes less severe losses during an investment period, which is favored by investors.

$$MDD = \max_{q,l:q<l} \left(\frac{TMoney_q - TMoney_l}{TMoney_q} \right), \tag{25}$$

where l denotes the lth testing day that is later than qth testing day within the testing period.

The stock ranking results of our model are based on ROR whose score function considers both risk and return factors. The stock ranking of baselines is only based on their predicted return ratios because they did not consider the risk factor in their original works.

5. Results

5.1. Main results

5.1.1. Ranking performance of stock selection

In this section, we compare the ranking performance of the proposed HGA-MT model against the state-of-the-art baselines. The results of different models are shown in Table 5. We report the Precision and MRR of Top-K ranking prediction, where the values of K are 3, 5, and 10. The results reveal that our proposed HGA-MT method achieves the best ranking performance than all baselines across the Top3, Top5, and Top10 predictions in terms of both the Precision and MRR metrics.

We observe that FLM, which does not capture related stock information, falls behind the other graph-based models. It indicates that integrating the effects of related stocks is important for stock ranking. Compared with FinGAT and TRAN, which only use the industry relation graphs, the improvements of HGA-MT demonstrate that our method can better aggregate the effects of related stocks from multiple relationships based on the heterogeneous stock relation graphs. Besides, compared with TGC and TRAN which generate the results by a single return ranking factor, the proposed method can achieve better performance since our multi-task learning paradigm can extract more robust features and reduce overfitting by jointly learning two complementary tasks [65–67]. The improvements verify the advantage of using the multi-task learning paradigm to simultaneously learn stock return and risk.

Table 6
The profitability comparison between different methods over Top-K stocks. NB: B&H strategy buys the CSI 100 Index, which does not involve the trading of Top-K stocks.

Strategy	Models	IRR	SP	MDD
Top3	TGC	0.2985	1.0827	0.1862
	TRAN	0.4010	1.5502	0.1278
	LFM	0.3320	1.3978	0.1053
	FinGAT	0.4143	1.9031	0.0920
	MAC	0.4420	2.0960	0.0898
	Ours	**0.5123**	**2.4092**	**0.0792**
Top5	TGC	0.2339	1.0844	0.1655
	TRAN	0.3104	1.3059	0.1545
	LFM	0.2463	1.2349	0.1199
	FinGAT	0.2983	1.6560	0.1004
	MAC	0.3297	1.8679	0.0867
	Ours	**0.3817**	**1.9511**	**0.0806**
Top10	TGC	0.2081	1.0144	0.1427
	TRAN	0.2611	1.2684	0.1200
	LFM	0.2186	1.1495	0.1191
	FinGAT	0.2443	1.4669	0.1081
	MAC	0.2521	1.5198	0.0877
	Ours	**0.2828**	**1.7400**	**0.0664**
Avg.	B&H	0.1665	0.9316	0.1579
	TGC	0.2468	1.0605	0.1648
	TRAN	0.3242	1.3748	0.1341
	LFM	0.2656	1.2607	0.1148
	FinGAT	0.3190	1.6753	0.1002
	MAC	0.3413	1.8279	0.0881
	Ours	**0.3923**	**2.0334**	**0.0754**

In particular, our proposed HGA-MT method outperforms MAC (the strongest baseline) with 0.0118 (11.09%) and 0.0221 (8.29%) gains in Precision and MRR on average, respectively. MAC also aggregates the financial news information of related stocks from multiple relationships. However, it yields the same weight for different relationships and delivers the results by a single return learning task. The improvements in our results demonstrate that the proposed method can better aggregate the effects of related stocks and optimize the model by multi-task learning.

5.1.2. Profitability analysis of stock selection

For financial evaluation, we perform trading simulations (backtesting) based on the predicted ranking results of different models. Following previous studies [9–11], the daily buy-hold-sell trading strategy is adopted for simulation trading. Specifically, the model is applied to obtain a predicted ranking list of N stocks on the trading Day t. Then, we buy the Top-K ranked stocks with the highest predicted ranking scores at the closing price on the trading Day t and sell it at the closing price on Day $t+1$. The budget is equally split to trade the Top-K stocks.

Here, the K is set to 3, 5, and 10. The backtesting is conducted on the data of the test set. The initial investment capital is set to 10,000 RMB. The transaction cost is assumed as zero, which is in line with previous works [9,11,43]. Besides, we add an additional benchmark for financial evaluation. We buy the CSI 100 Index (the index that represents the weighted average price of 100 indexed stocks) on the first trading day of the testing period with all of the initial capital and hold it until the end of the testing period, which is defined as the Buy & Hold (B&H) strategy.

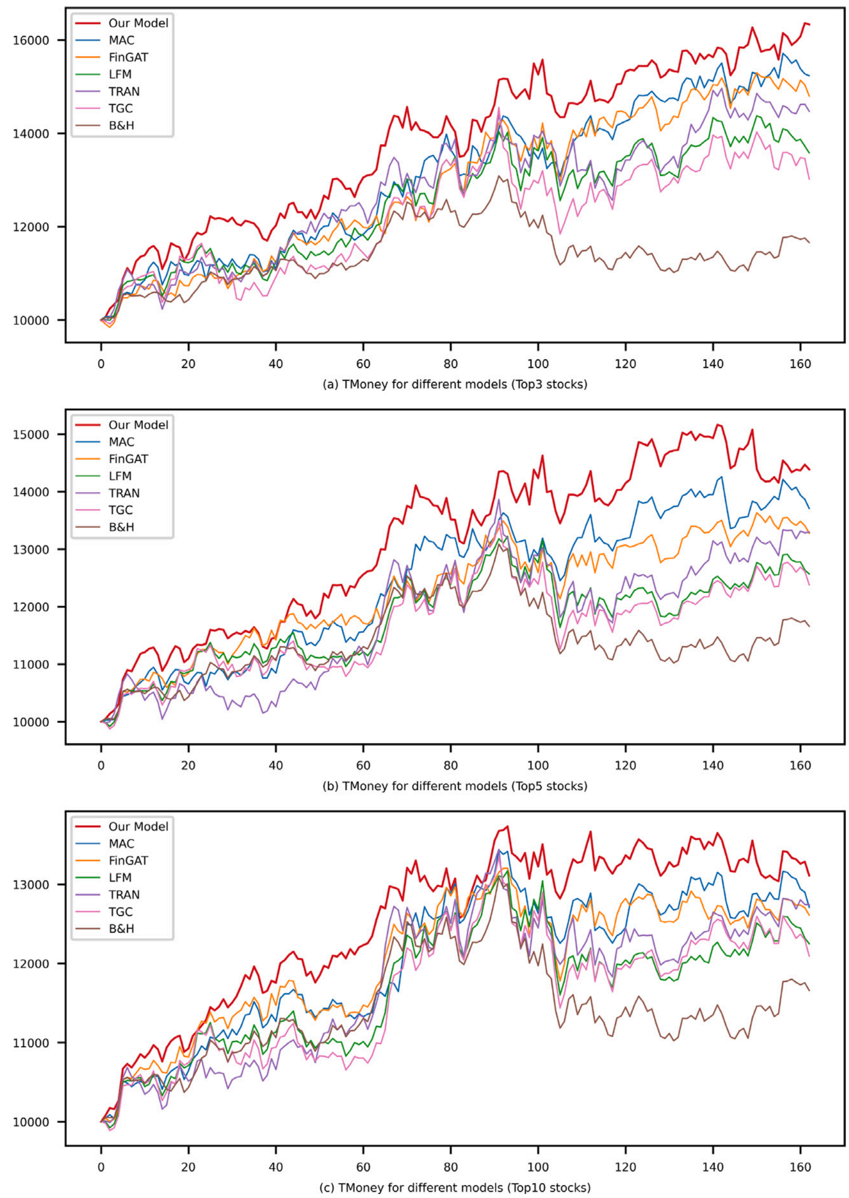


Fig. 6. TMoney of different models.

Table 6 presents the profitability evaluation results for IRR, SP and MDD of different methods. Fig. 6 visualizes the detailed changes in TMoney with different methods during the backtesting period. The evaluation results reveal the following findings:

(1) As shown in Table 6, our proposed HGA-MT method achieves the best performance across the Top3, Top5, and Top10 prediction tasks in terms of IRR and SP, yielding at least 0.0510 (14.94%) gains in IRR and 0.2055 (11.24%) gains in SP on average. HGA-MT also presents the lowest MDD (7.54%) on average, which is 14.42% lower than the strongest baseline. Moreover, Fig. 6 shows that our model can gain significantly higher TMoney across the Top3, Top5, and Top10 prediction tasks than all baselines during backtesting. These results demonstrate that our HGA-MT method can produce higher profits with lower risk.

(2) As shown in Table 6, the proposed method can achieve the best performance of SP with 2.0334 on average, which is higher than those of MAC (+0.2055), FinGAT (+0.3581), LFM (+0.7727), TRAN (+0.6586) and TGC (+0.9729). The higher SP means that HGA-MT can achieve a higher return on the investment relative to the same risk level. These results demonstrate the superiority of our HGA-MT method for stock ranking to optimize portfolios by considering return and risk simultaneously.

(3) All models achieve better evaluation results in terms of IRR and SP compared to directly buying the stock of the CSI 100 Index (B&H), which yields an IRR of 0.1665 and an SP of 0.9316. Such an observation indicates that effectively capturing valuable information by promising models to assist investment can obtain more profits. Surprisingly, B&H has a lower MDD compared to TGC.

Moreover, we conducted experiments to show the performance of different methods under the Top1-ranked stock selection setup. The results are shown in Appendix C.

5.2. Ablation studies

In this section, several ablation experiments are performed to examine the effectiveness of each technical component of HGA-MT, including the stock relation graphs, numerical quantitative indicator input, textual news feature input, the HGA module, the multi-task learning paradigm, and the adjusted ranking based on predicted risk values. Several variants of the model are conducted as follows:

Table 7
The ranking performance of the ablation analysis over Top-K stocks. $K \in \{3,5,10\}$.

	Top3		Top5		Top10		Avg.	
	Prec	MRR	Prec	MRR	Prec	MRR	Prec	MRR
w/o Graph	0.0675	0.1728	0.0896	0.2135	0.1166	0.3625	0.0912	0.2496
w/o News	0.0745	0.1741	0.0903	0.2147	0.1287	0.3985	0.0978	0.2624
w/o QI	0.0777	0.1703	0.0920	0.2157	0.1478	0.4168	0.1058	0.2676
w/o HGA	0.0754	0.1759	0.0936	0.2178	0.1552	0.4274	0.1081	0.2737
w/o MT	0.0781	0.1754	0.0929	0.2228	0.1577	0.4485	0.1096	0.2822
Ours	**0.0859**	**0.1820**	**0.1006**	**0.2252**	**0.1681**	**0.4587**	**0.1182**	**0.2886**

- **w/o stock relation graphs (w/o Graph)**: The model excludes the stock relation graphs and GCN layers, which is used to verify the effectiveness of aggregating information from related stocks.
- **w/o news sentiment features (w/o News)**: The model excludes the news sentiment features of each stock. The information aggregated from related stocks only involves their quantitative indicators as well. It is used to verify the effectiveness of utilizing financial news features.
- **w/o quantitative indicators (w/o QI)**: The model excludes the quantitative indicators of each stock. The information aggregated from related stocks only involves their financial news features as well. It is used to verify the effectiveness of utilizing quantitative indicators.
- **w/o HGA module (w/o HGA)**: This model organizes the stock relation graph with three relations as a homogeneous graph and integrates information from related stocks with equal weights. This is used to verify the effectiveness of our HGA module.
- **w/o multi-task learning (w/o MT)**: This model is optimized solely on stock return loss, which is used to examine the effect of the multi-task learning paradigm.
- **w/o risk adjustment (w/o Risk)**: The model is still trained with multi-task learning. However, the ranking results are only sorted based on the predicted stock return ratios ($\hat{Rank}^{Return}$) without considering the risk values. It is used to verify the effectiveness of taking both stock risk and return into account at the scoring stage.

For the ranking performance, Table 7 shows the evaluation results of the ablation experiments. As seen, the model without graph (w/o Graph) obtains the worst performance, which indicates that integrating the information of related stocks is critical for target stock ranking due to the complicated relationship between stocks. The proposed HGA-MT model performs better than the model without financial news sentiment features (w/o News) or quantitative indicators (w/o QI). The results demonstrate that news sentiment features and quantitative indicators are effective features for stock ranking prediction. The model excluding news sentiment features (w/o News) performs worse than the model excluding quantitative indicators (w/o QI), showing that financial news sentiment features have a higher utility.

Besides, our HGA-MT model outperforms the model without the HGA module (w/o HGA), demonstrating that the proposed HGA method can capture more useful information than naively integrating information from multiple related stocks without distinguishing the weights of different relationship types. Moreover, the results of the model without multi-task learning (w/o MT) show that the performance is degraded when removing the subtask of risk prediction. This proves that our proposed method with a multi-task paradigm can improve return-based ranking prediction performance. We do not include the performance of the ablation model w/o Risk in Table 7, because our full model in the ranking evaluation task does not take the risk factor into account in the scoring function (see Section 4.4), which is the same as the w/o Risk setup.

For the financial evaluation, Table 8 shows the evaluation results of IRR, SP, and MDD. As seen, the proposed HGA-MT model achieves the best performance among all the evaluation metrics. The performance is degraded by removing any component. Fig. 7 shows the TMoney of our proposed HGA-MT model and its submodels across Top3, Top5, and Top10 prediction tasks. The solid line represents the results of our full model. The results of the ablation models, which remove a particular component, are represented by the dashed lines. It can also be observed that subtracting a component (i.e., w/o Graph, w/o News, w/o QI, w/o HGA, w/o MT and w/o Risk) leads to lower TMoney across all tasks. Compared with the ablation model w/o HGA, the proposed model can achieve higher and more stable profits, demonstrating the effectiveness of integrating information by the HGA module. Moreover, the ablation model w/o Risk performs better than the ablation model w/o MT, showing the effect of multi-task learning in optimizing model parameters, thereby achieving better profitability.

Table 8
The profitability results of the ablation analysis over Top-K stocks. $K \in \{3,5,10\}$.

Strategy	Model	IRR	SP	MDD
Top3	w/o Graph	0.4365	1.6565	0.1327
	w/o News	0.4428	1.8482	0.0843
	w/o QI	0.4455	1.9956	0.0801
	w/o HGA	0.4687	2.0126	0.0907
	w/o MT	0.4716	2.1384	0.1101
	w/o Risk	0.5071	2.2669	0.1360
	Ours	**0.5123**	**2.4092**	**0.0792**
Top5	w/o Graph	0.3429	1.5336	0.1601
	w/o News	0.3436	1.6483	0.0816
	w/o QI	0.3430	1.7174	0.0876
	w/o HGA	0.3437	1.8973	0.0720
	w/o MT	0.3471	1.8592	0.0822
	w/o Risk	0.3792	1.8668	0.1141
	Ours	**0.3817**	**1.9511**	**0.0806**
Top10	w/o Graph	0.2529	1.2987	0.1274
	w/o News	0.2600	1.5028	0.0650
	w/o QI	0.2667	1.5052	0.0595
	w/o HGA	0.2742	1.5971	0.0668
	w/o MT	0.2720	1.6135	0.0900
	w/o Risk	0.2824	1.6446	0.1173
	Ours	**0.2828**	**1.7400**	**0.0664**
Avg.	w/o Graph	0.3441	0.3441	0.1401
	w/o News	0.3488	0.3488	0.0770
	w/o QI	0.3517	0.3517	0.0757
	w/o HGA	0.3622	0.3622	0.0765
	w/o MT	0.3636	0.3636	0.0941
	w/o Risk	0.3896	0.3896	0.1225
	Ours	**0.3923**	**0.3923**	**0.0754**

In particular, the HGA-MT approach, as proposed in this study, demonstrates superior performance compared to the ablation model w/o Risk. In the model w/o Risk, ranking results are solely based on the descending order of stock returns, without taking into account the predicted risk values to adjust the ranking outcome. The improvements consolidate our argument that financial investment should consider risk and return simultaneously, thereby achieving higher and more stable profits. Although IRR results are similar by comparing HGA-MT and w/o Risk in Table 8, SP has achieved a significant increase (+0.1073, i.e., 5.57% in averaged SP). As mentioned in Section 4.4, a higher SP means a higher return relative to the amount of risk taken. In other words, a higher SP means more stable profits. The MDD metric also confirms that the full model can yield portfolios with lower investment risk, compare to other ablation models.

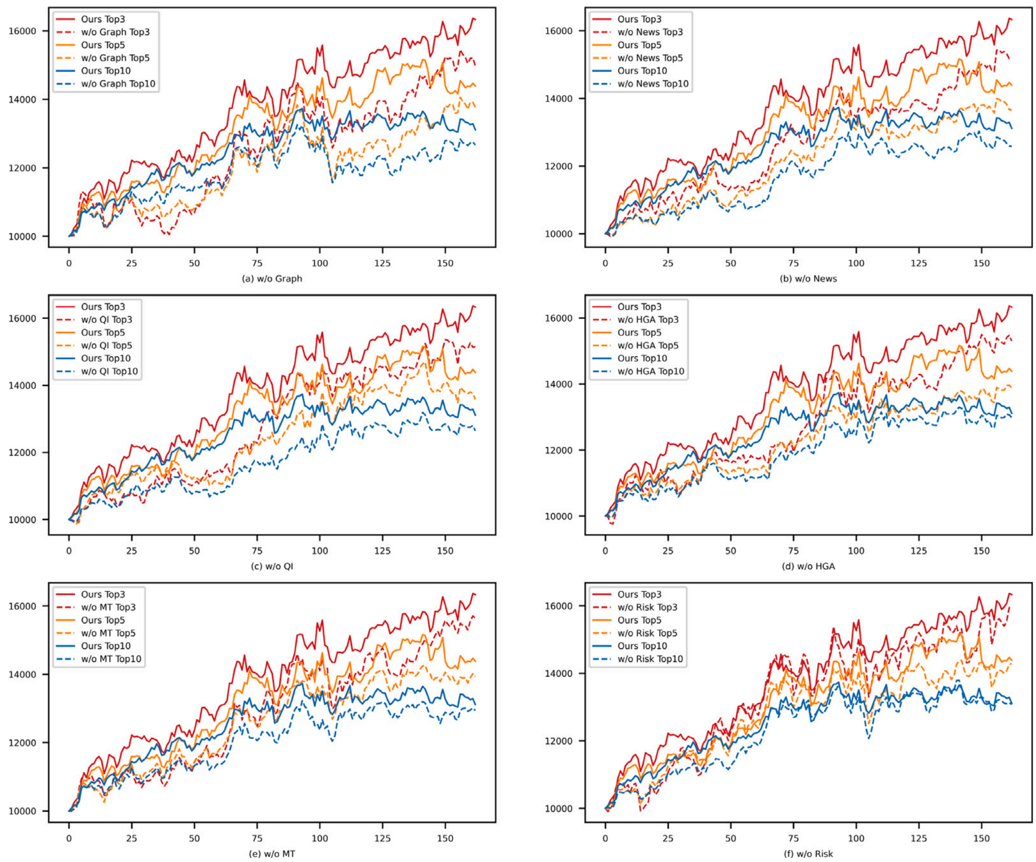


Fig. 7. TMoney of the proposed HGA-MT model and its ablation models over various selected stocks K. $K \in \{3, 5, 10\}$.

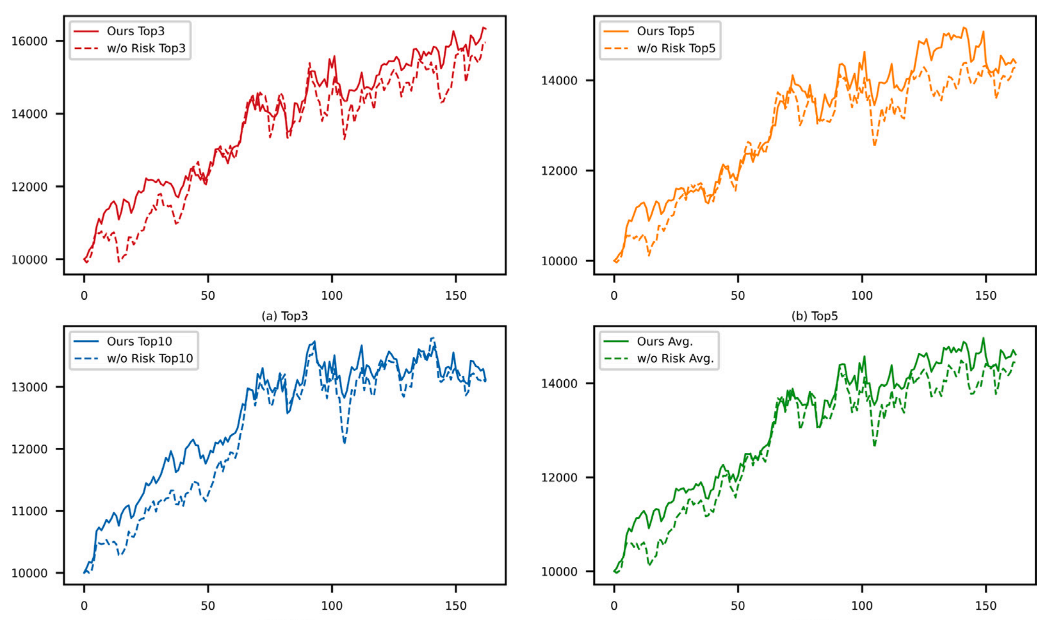
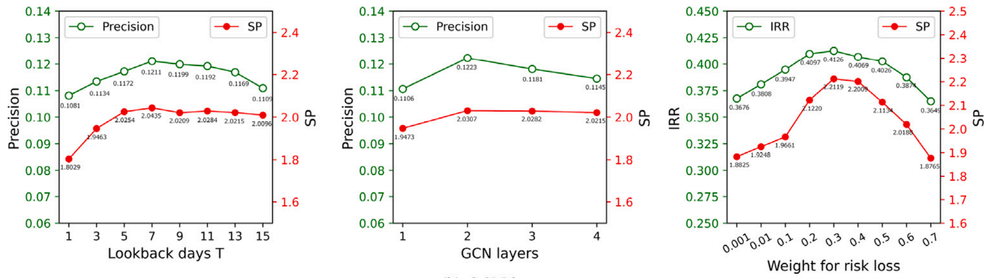


Fig. 8. TMoney of the proposed HGA-MT model and the ablation model w/o Risk over different number of selected stocks K. $K \in \{3, 5, 10\}$.

Fig. 8 shows the TMoney of the proposed HGA-MT model and the ablation model w/o Risk across Top3, Top5, Top10, and on average. The TMoney lines of HGA-MT and that of w/o Risk are very close in many cases. However, the TMoney lines of w/o Risk present greater volatility, which shows consistent findings from IRR, SP, and MDD metrics. In summary, the improved performance of our proposed model in terms of IRR, SP, MDD, and TMoney demonstrates that HGA-MT can achieve higher and more stable profits by considering both risk and return for investment-oriented stock ranking.

5.3. Hyperparameter analysis

As mentioned in Section 4.3, a grid search process is first adopted to explore the optimal values of different hyperparameters, including



(a) Lookback window size T (b) GCN layer (c) Loss weight

Fig. 9. Sensitivity analysis to (a) lookback window sizes, (b) the number of GCN layers, and (c) the weights for risk loss. The results are averaged over different portfolios including Top3, Top5 and Top10 stocks.

the sliding window length (T), the number of GCN layers, and the weight parameter (α) that determines the contribution of the loss of the risk subtask to the overall loss. All the possible combinations of 3 hyperparameters are tested to select the optimal values. The results show that the optimal values are time step $T = 7$, 2 GCN layers, and the loss weighing factor by $\alpha = 0.3$. To clearly visualize the result variation of changing each hyperparameter, we show the results that when a hyperparameter is varying, the others are fixed by the optimal values. The Precision and SP with different windows (T), GCN layers, and loss weights (α) are shown in Fig. 9.

Fig. 9(a) shows that our model obtains the best Precision and SP by T=7, which can capture enough financial temporal information as well as avoid including outdated media and market signals. Moreover, Fig. 9(b) shows the performance by using varied numbers of GCN layers in the proposed model, where the optimal setting is given by 2 GCN layers. Besides, we show the performance by varying α in Fig. 9(c). Specifically, the α is first tested within $\{0.001, 0.01, 0.1, 0.5, 1\}$, following an existing study about multi-task learning [8].

The preliminary results show that our model can achieve better performance, given α ranging from 0.1 to 0.5. Then, additional tests with finer-grained interval $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$ are tested. Finally, the results by varying $\alpha \in \{0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7\}$ are shown in Fig. 9(c). A lower value of α indicates a decreased influence of stock risk loss and an increased influence of stock return loss towards the overall loss. The results demonstrate the inclination of an initial increase followed by a subsequent decline, where the best performance is achieved by α=0.3. The results imply that properly fusing the risk prediction task can improve stock ranking performance and profitability. Moreover, the performance is slightly worse when the α is higher than 0.4. This is likely because too much focus on risk control undermines profits.

6. Conclusion

In order to select profitable stocks for portfolio optimization, we proposed a model to predict the stock ranking results on the next trading day by fusing the effects of related stocks and taking into account the stock return and volatility risk simultaneously. We designed a heterogeneous graph attention module that combines GCN and attention to capture various effects of related stocks from diverse relationships. The attention assigns distinct weights to scrutinize the distinction in effects among different relationships. Moreover, our model leverages a multi-task learning paradigm that jointly learns and predicts the stock return ratios and volatility risk. Thus, we can obtain stock ranking results by considering both the risk and return.

We examined our method on CSI 100 dataset in Chinese stock markets. The results showed that our proposed HGA-MT model significantly outperforms the state-of-the-art baselines, yielding at least 11.09% and 8.29% gains in terms of Precision and MRR on average, respectively. Moreover, the backtesting demonstrated the superiority of our method in financial evaluation.

Our method exceeds the baselines at least by 14.94% in IRR 11.24% in SP, and 14.42% in MDD, achieving higher and more stable profits. In addition, the results of the ablation analysis indicate that jointly learning the stock risk and return significantly improves the stock ranking performance and financial profits. The designed HGA module is more effective for capturing the effects of related stocks than a homogeneous graph without weighing different relation types. These results demonstrate the effectiveness and applicability of the proposed HGA-MT method. In summary, our proposed HGA-MT method achieves better outcomes for stock ranking than baselines, resulting in higher and more stable profits.

On the other hand, there are several limitations that can be improved in future work. First, more relations between stocks should be explored to improve the robustness of neural networks. We will continue to explore the latent interactions between stocks and time-evolving stock relationship information to improve the performance further. Enhancing the explainability of neural networks is also a promising direction to improve the trustworthiness of trading signal predictions in the investment domain [68–70]. Second, we observed that many figurative languages exist in financial news, which results in a misunderstanding of the semantics [71,72] of the input text. Thus, the metaphor understanding should be leveraged in future financial news- and natural language processing-based quantitative investments. Third, the risk measure of individual stocks is an intuitive way of ranking stocks, whereas it is sub-optimal for measuring the risk of a portfolio. We will extend our proposed method with other portfolio risk measures.

CRediT authorship contribution statement

Yu Ma: Conceptualization, Methodology, Writing – original draft. **Rui Mao:** Conceptualization, Methodology, Writing – review & editing. **Qika Lin:** Methodology, Software. **Peng Wu:** Funding acquisition, Writing – review & editing. **Erik Cambria:** Funding acquisition, Project administration, Supervision, Writing – review & editing.

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.

Data availability

The authors do not have permission to share data.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (project numbers are 72274096, 72174087, 71774084 and 71874082), the National Social Science Fund of China (project number is 17ZDA291), Foreign Cultural and Educational Expert Program of the Ministry of Science and Technology of China (G202218-2009L).

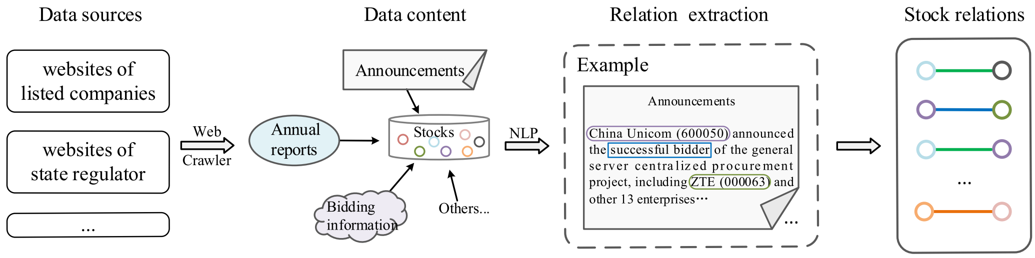


Fig. A.10. Illustration of the process of stock relation extraction by Qichacha.

Table B.9
A list of stocks that are used for ranking in this work. Code denotes the stock code. #news denotes the number of news of the stock.

Code	#news	Code	#news	Code	#news	Code	#news
000001	1718	002594	6264	600519	4460	601360	562
000002	924	002714	1564	600585	870	601390	648
000063	3543	002736	2927	600588	481	601398	3055
000166	2237	300015	689	600606	762	601601	1918
000333	2124	300059	1162	600690	954	601628	2206
000538	624	300122	839	600809	600	601668	184
000568	798	300498	1026	600837	4302	601688	4023
000651	2610	600000	1954	600887	646	601766	688
000661	792	600009	606	600900	475	601800	368
000725	996	600015	742	600999	3592	601818	1603
000776	2316	600018	240	601006	178	601857	586
000858	1462	600019	421	601012	2097	601888	640
000876	1621	600028	892	601088	758	601899	981
000895	334	600030	11 049	601169	698	601933	876
001979	1200	600031	869	601186	424	601988	4159
002024	877	600036	3257	601211	1970	601989	134
002027	532	600048	656	601225	249	601998	2150
002142	853	600050	4039	601229	688	603160	519
002304	576	600104	2022	601288	1913	603288	654
002352	1608	600276	1394	601318	4007	603501	561
002415	1219	600346	526	601328	2642	——	——
002475	1574	600406	126	601336	1437	——	——

Appendix A

More details regarding stock relations provided by Qichacha are as follows:

- **Data sources**: (1) the official website of listed companies, (2) the official website of the state regulator, such as the National Enterprise Credit Information Publicity System,[8] China Tendering & Bidding Public Service Platform,[9]Ministry of Industry and Information Technology[10] and Cninfo.[11]
- **Technology**: (1) web crawler for raw information collection, (2) natural language processing technology for key information extraction.
- **Data form**: text data for key information extraction.
- **Data content**: the public information, including listed companies' announcements, annual reports, bidding information, qualification certificates, prospectus, listing announcement, etc.

The process of stock relation extraction by Qichacha is illustrated in Fig. A.10. Qichacha primarily collects textual data from the official websites of listed companies and the state regulator through web crawlers, including various sources such as announcements, annual reports, bidding information, qualification certificates, prospectuses, listing announcements, etc. Then, key information is extracted and analyzed using natural language processing technology. Qichacha obtains relation information among stocks, including supply chain, shareholding chain, and industry competitors. Finally, the related stocks in three relation types of each stock are provided by Qichacha.

Table C.10
The ranking performance and profitability of different methods over Top1-ranked stocks.

Model	Ranking performance		Profitability performance		
	Precision	MRR	IRR	SP	MDD
TGC	0.0122	0.0122	0.2832	1.0115	0.1997
TRAN	0.0184	0.0184	0.3730	1.5065	0.1732
LFM	0.0123	0.0123	0.2718	1.3912	0.1323
FinGAT	0.0366	0.0366	0.3909	1.8825	0.0939
MAC	0.0429	0.0429	0.4195	1.9227	0.0934
Ours	**0.0491**	**0.0491**	**0.4869**	**2.3633**	**0.0875**

Appendix B

The pool of stocks for ranking can be seen in Table B.9.

Appendix C

As seen in Table C.10 and Fig. C.11, our proposed HGA-MT method achieves the best performance on the Top1 prediction task in terms of ranking prediction and profitability, which demonstrates the effectiveness of (1) simultaneously acquiring knowledge regarding the stock risk and return and (2) fusing the effects of multiple related stocks.

On the other hand, the evaluation outcomes of Table C.10, in conjunction with the findings presented in Tables 5 and 6 for Top3/5/10, illustrate a pattern that Top3 outperforms Top5 and Top10, and Top1 ranks lower than Top3. The observed performance trend, such as Top3

[8] https://shiming.gsxt.gov.cn/index.html
[9] http://www.cebpubservice.com/
[10] https://www.miit.gov.cn/
[11] http://www.cninfo.com.cn/new/index

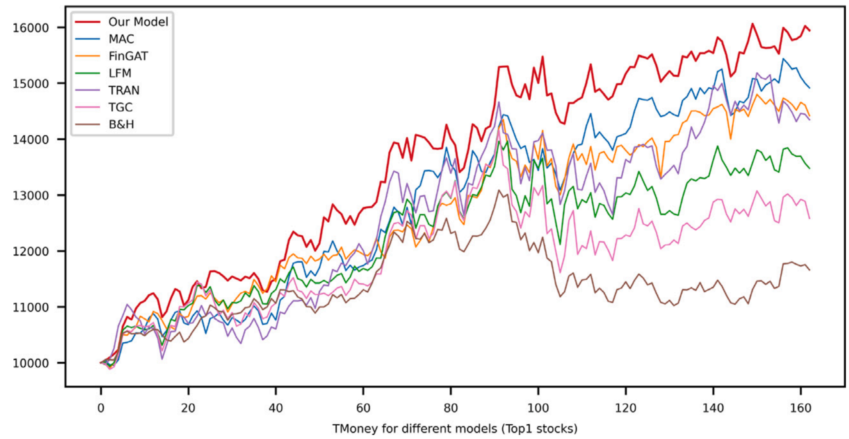


Fig. C.11. TMoney of different methods over Top1-ranked stocks.

surpassing Top5 and Top10, can be attributed to the potential for heightened portfolio profitability when selecting a very limited number of stocks with superior rates of return, assuming a highly accurate model. At the extreme end of this spectrum is the notion of investing solely in the most profitable stock (Top1) every day, which would be the case if a ranking model is 100% accurate. However, acknowledging the impracticality of such a hypothetical scenario, we note the performance disparity where Top1 ranks lower than Top3.

This discrepancy underscores the significance of a risk diversification strategy, exemplified by investing in multiple stocks, e.g., Top3. This approach, driven by portfolio optimization, enables the model to achieve superior profitability compared to a strategy exclusively focused on the Top1-ranked stocks. In practice, given the inherent uncertainty surrounding the accuracy of predictions over a specified period, the risk-diversified approach of investing in multiple stocks offers investors a means to attain expected returns with reduced exposure to risks.

References

[1] W. Jiang, Applications of deep learning in stock market prediction: Recent progress, Expert Syst. Appl. 184 (2021) 115537, http://dx.doi.org/10.1016/j.eswa.2021.115537.

[2] M.M. Kumbure, C. Lohrmann, P. Luukka, J. Porras, Machine learning techniques and data for stock market forecasting: A literature review, Expert Syst. Appl. 197 (2022) 116659, http://dx.doi.org/10.1016/j.eswa.2022.116659.

[3] T. Kehinde, F.T. Chan, S. Chung, Scientometric review and analysis of recent approaches to stock market forecasting: Two decades survey, Expert Syst. Appl. 213 (2023) 119299, http://dx.doi.org/10.1016/j.eswa.2022.119299.

[4] H. Markowitz, Portfolio selection, J. Finance 7 (1) (1952) 77–91, http://dx.doi.org/10.2307/2975974.

[5] X. Ma, T. Zhao, Q. Guo, X. Li, C. Zhang, Fuzzy hypergraph network for recommending top-K profitable stocks, Inform. Sci. 613 (2022) 239–255, http://dx.doi.org/10.1016/j.ins.2022.09.010.

[6] A. Thakkar, K. Chaudhari, Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions, Inf. Fusion 65 (2021) 95–107, http://dx.doi.org/10.1016/j.inffus.2020.08.019.

[7] K. Adam, A. Marcet, J.P. Nicolini, Stock market volatility and learning, J. Finance 71 (1) (2016) 33–82, http://dx.doi.org/10.1111/jofi.12364.

[8] Y.-L. Hsu, Y.-C. Tsai, C.-T. Li, FinGAT: Financial graph attention networks for recommending top-K profitable stocks, IEEE Trans. Knowl. Data Eng. 35 (1) (2023) 469–481, http://dx.doi.org/10.1109/TKDE.2021.3079496.

[9] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, T.-S. Chua, Temporal relational ranking for stock prediction, ACM Trans. Inf. Syst. 37 (2) (2019) 30, http://dx.doi.org/10.1145/3309547.

[10] R. Sawhney, S. Agarwal, A. Wadhwa, T. Derr, R.R. Shah, Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 497–504.

[11] J. Gao, X. Ying, C. Xu, J. Wang, S. Zhang, Z. Li, Graph-based stock recommendation by time-aware relational attention network, ACM Trans. Knowl. Discov. Data 16 (1) (2021) 21, http://dx.doi.org/10.1145/3451397.

[12] S. Feng, C. Xu, Y. Zuo, G. Chen, F. Lin, J. XiaHou, Relation-aware dynamic attributed graph attention network for stocks recommendation, Pattern Recognit. 121 (2022) 108119, http://dx.doi.org/10.1016/j.patcog.2021.108119.

[13] H. Wang, T. Wang, S. Li, S. Guan, J. Zheng, W. Chen, Heterogeneous interactive snapshot network for review-enhanced stock profiling and recommendation, in: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, International Joint Conferences on Artificial Intelligence Organization, 2022, pp. 3962–3969, http://dx.doi.org/10.24963/ijcai.2022/550, Main Track.

[14] B. Li, Q. Wang, Y. Yu, M.-Z. Sun, L.-X. Chen, Z.-L. Xiang, F. Zhao, Q.-C. Lv, Z.-Y. An, A novel risk-control model for the online portfolio selection of high-frequency transactions, Knowl.-Based Syst. 240 (2022) 108176, http://dx.doi.org/10.1016/j.knosys.2022.108176.

[15] Y. Ma, R. Mao, Q. Lin, P. Wu, E. Cambria, Multi-source aggregated classification for stock price movement prediction, Inf. Fusion 91 (2023) 515–528, http://dx.doi.org/10.1016/j.inffus.2022.10.025.

[16] Q. Li, Y. Chen, L.L. Jiang, P. Li, H. Chen, A tensor-based information framework for predicting the stock market, ACM Trans. Inf. Syst. 34 (2) (2016) http://dx.doi.org/10.1145/2838731.

[17] O. Bustos, A. Pomares-Quimbaya, Stock market movement forecast: A systematic review, Expert Syst. Appl. 156 (2020) 113464, http://dx.doi.org/10.1016/j.eswa.2020.113464.

[18] I.K. Nti, A.F. Adekoya, B.A. Weyori, A systematic review of fundamental and technical analysis of stock market predictions, Artif. Intell. Rev. 53 (4) (2020) 3007–3057, http://dx.doi.org/10.1007/s10462-019-09754-z.

[19] L. Malandri, F.Z. Xing, C. Orsenigo, C. Vercellis, E. Cambria, Public mood–driven asset allocation: The importance of financial sentiment in portfolio management, Cogn. Comput. 10 (2018) 1167–1176, http://dx.doi.org/10.1007/s12559-018-9609-2.

[20] X. Li, P. Wu, W. Wang, Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong, Inf. Process Manag. 57 (5) (2020) 102212, http://dx.doi.org/10.1016/j.ipm.2020.102212.

[21] W.-C. Lin, C.-F. Tsai, H. Chen, Factors affecting text mining based stock prediction: Text feature representations, machine learning models, and news platforms, Appl. Soft Comput. 130 (2022) 109673, http://dx.doi.org/10.1016/j.asoc.2022.109673.

[22] E. Cambria, Q. Liu, S. Decherchi, F. Xing, K. Kwok, SenticNet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022, pp. 3829–3839.

[23] K. Du, F. Xing, R. Mao, E. Cambria, FinSenticNet: A concept-level lexicon for financial sentiment analysis, in: 2023 IEEE Symposium Series on Computational Intelligence, SSCI, 2023, pp. 109–114.

[24] T. Yu, M. Sengul, R.H. Lester, Misery Loves company: The spread of negative impacts resulting from an organizational crisis, Acad. Manag. Rev. 33 (2) (2008) 452–472, http://dx.doi.org/10.5465/amr.2008.31193499.

[25] S. Pandit, C.E. Wasley, T. Zach, Information externalities along the supply chain: The economic determinants of suppliers' stock price reaction to their customers' earnings announcements, Contemp. Account. Res. 28 (4) (2011) 1304–1343, http://dx.doi.org/10.1111/j.1911-3846.2011.01092.x.

[26] Y. Chen, Z. Wei, X. Huang, Incorporating corporation relationship via graph convolutional neural networks for stock price prediction, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM '18, Association for Computing Machinery, 2018, pp. 1655–1658, http://dx.doi.org/10.1145/3269206.3269269.

[27] S. He, H. Rui, A.B. Whinston, Social media strategies in product-harm crises, Inf. Syst. Res. 29 (2) (2018) 362–380, http://dx.doi.org/10.1287/isre.2017.0707.

[28] W. Chen, M. Jiang, W.-G. Zhang, Z. Chen, A novel graph convolutional feature based convolutional neural network for stock trend prediction, Inform. Sci. 556 (2021) 67–94, http://dx.doi.org/10.1016/j.ins.2020.12.068.

[29] J. Tan, Q. Li, J. Wang, J. Chen, FinHGNN: A conditional heterogeneous graph learning to address relational attributes for stock predictions, Inform. Sci. 618 (2022) 317–335, http://dx.doi.org/10.1016/j.ins.2022.11.013.

[30] S. He, H. Rui, A.B. Whinston, Social media strategies in product-harm crises, Inf. Syst. Res. 29 (2) (2018) 362–380, http://dx.doi.org/10.1287/isre.2017.0707.

[31] P. Zou, G. Li, How emerging market investors' value competitors' customer equity: Brand crisis spillover in China, J. Bus. Res. 69 (9) (2016) 3765–3771, http://dx.doi.org/10.1016/j.jbusres.2015.12.068.

[32] B. Sun, H. Li, P. An, Z. Wang, Dynamic energy stock selection based on shareholders' coholding network, Physica A 542 (2020) 122243, http://dx.doi.org/10.1016/j.physa.2019.122243.

[33] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, Neural Netw. 18 (5–6) (2005) 602–610, http://dx.doi.org/10.1016/j.neunet.2005.06.042.

[34] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, 2017, pp. 24–26, http://dx.doi.org/10.48550/ARXIV.1609.02907.

[35] Y.-H. Chou, S.-Y. Kuo, Y.-C. Jiang, A novel portfolio optimization model based on trend ratio and evolutionary computation, IEEE Trans. Emerg. Top. Comput. Intell. 3 (4) (2019) 337–350, http://dx.doi.org/10.1109/TETCI.2018.2868939.

[36] Y.-H. Chou, S.-Y. Kuo, Y.-T. Lo, Portfolio optimization based on funds standardization and genetic algorithm, IEEE Access 5 (2017) 21885–21900, http://dx.doi.org/10.1109/ACCESS.2017.2756842.

[37] Y.-H. Chou, Y.-C. Jiang, Y.-R. Hsu, S.-Y. Kuo, S.-Y. Kuo, A weighted portfolio optimization model based on the trend ratio, emotion index, and ANGQTS, IEEE Trans. Emerg. Top. Comput. Intell. 6 (4) (2022) 867–882, http://dx.doi.org/10.1109/TETCI.2021.3118041.

[38] W. Chen, H. Zhang, M.K. Mehlawat, L. Jia, Mean–variance portfolio optimization using machine learning-based stock price prediction, Appl. Soft Comput. 100 (2021) 106943, http://dx.doi.org/10.1016/j.asoc.2020.106943.

[39] F. Yang, Z. Chen, J. Li, L. Tang, A novel hybrid stock selection method with stock prediction, Appl. Soft Comput. 80 (2019) 820–831, http://dx.doi.org/10.1016/j.asoc.2019.03.028.

[40] Q. Song, A. Liu, S.Y. Yang, Stock portfolio selection using learning-to-rank algorithms with news sentiment, Neurocomputing 264 (2017) 20–28, http://dx.doi.org/10.1016/j.neucom.2017.02.097, Machine learning in finance.

[41] T. Ma, Y. Tan, Stock ranking with multi-task learning, Expert Syst. Appl. 199 (2022) 116886, http://dx.doi.org/10.1016/j.eswa.2022.116886.

[42] H. Liu, T. Zhao, S. Wang, X. Li, A stock rank prediction method combining industry attributes and price data of stocks, Inf. Process. Manage. 60 (4) (2023) 103358, http://dx.doi.org/10.1016/j.ipm.2023.103358.

[43] A. Picasso, S. Merello, Y. Ma, L. Oneto, E. Cambria, Technical analysis and sentiment embeddings for market trend prediction, Expert Syst. Appl. 135 (2019) 60–70, http://dx.doi.org/10.1016/j.eswa.2019.06.014.

[44] K.-Y. Shen, G.-H. Tzeng, Combined soft computing model for value stock selection based on fundamental analysis, Appl. Soft Comput. 37 (2015) 142–155, http://dx.doi.org/10.1016/j.asoc.2015.07.030.

[45] Y. Jeon, T.H. McCurdy, X. Zhao, News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies, J. Financ. Econ. 145 (2) (2022) 1–17, http://dx.doi.org/10.1016/j.jfineco.2021.08.002.

[46] Q. Li, J. Tan, J. Wang, H. Chen, A multimodal event-driven LSTM model for stock prediction using online news, IEEE Trans. Knowl. Data Eng. 33 (10) (2021) 3323–3337, http://dx.doi.org/10.1109/TKDE.2020.2968894.

[47] R. Mao, K. Du, Y. Ma, L. Zhu, E. Cambria, Discovering the cognition behind language: Financial metaphor analysis with MetaPro, in: 2023 IEEE International Conference on Data Mining, ICDM, IEEE, 2023.

[48] Y. Shi, K.Y. Ho, News sentiment and states of stock return volatility: Evidence from long memory and discrete choice models, Finance Res. Lett. 38 (2021) 101446, http://dx.doi.org/10.1016/j.frl.2020.101446.

[49] N. Strauß, R. Vliegenthart, P. Verhoeven, Lagging behind? Emotions in newspaper articles and stock market prices in the netherlands, Public Relat. Rev. 42 (4) (2016) 548–555, http://dx.doi.org/10.1016/j.pubrev.2016.03.010.

[50] X. Chen, X. Ma, H. Wang, X. Li, C. Zhang, A hierarchical attention network for stock prediction based on attentive multi-view news learning, Neurocomputing 504 (2022) 1–15, http://dx.doi.org/10.1016/j.neucom.2022.06.106.

[51] J. Lei, N. Dawar, J. Lemmink, Negative spillover in brand portfolios: Exploring the antecedents of asymmetric effects, J. Market. 72 (3) (2008) 111–123, http://dx.doi.org/10.1509/JMKG.72.3.111.

[52] J. Long, Z. Chen, W. He, T. Wu, J. Ren, An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market, Appl. Soft Comput. 91 (2020) 106205, http://dx.doi.org/10.1016/j.asoc.2020.106205.

[53] X. Hou, K. Wang, C. Zhong, Z. Wei, ST-Trader: A spatial-temporal deep neural network for modeling stock market movement, IEEE/CAA J. Autom. Sin. 8 (5) (2021) 1015–1024, http://dx.doi.org/10.1109/JAS.2021.1003976.

[54] J. Wu, K. Xu, X. Chen, S. Li, J. Zhao, Price graphs: Utilizing the structural information of financial time series for stock prediction, Inform. Sci. 588 (2022) 405–424, http://dx.doi.org/10.1016/j.ins.2021.12.089.

[55] Y.-M. Li, L.-F. Lin, C.-Y. Hsieh, B.-S. Huang, A social investing approach for portfolio recommendation, Inf. Manag. 58 (8) (2021) 103536, http://dx.doi.org/10.1016/j.im.2021.103536.

[56] Z. Zhao, H. Chen, J. Zhang, X. Zhao, T. Liu, W. Lu, X. Chen, H. Deng, Q. Ju, X. Du, UER: An open-source toolkit for pre-training models, 2019, http://dx.doi.org/10.48550/ARXIV.1909.05658.

[57] Q. Lin, Y. Zhu, H. Lu, K. Shi, Z. Niu, Improving university faculty evaluations via multi-view knowledge graph, Future Gener. Comput. Syst. 117 (2021) 181–192, http://dx.doi.org/10.1016/j.future.2020.11.021.

[58] Q. Lin, J. Liu, L. Zhang, Y. Pan, X. Hu, F. Xu, H. Zeng, Contrastive graph representations for logical formulas embedding, IEEE Trans. Knowl. Data Eng. (2021) 1, http://dx.doi.org/10.1109/TKDE.2021.3139333.

[59] W. Li, L. Zhu, R. Mao, E. Cambria, SKIER: A symbolic knowledge integrated model for conversational emotion recognition, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, (no. 11) 2023, pp. 13121–13129, http://dx.doi.org/10.1609/aaai.v37i11.26541.

[60] A. Picasso, S. Merello, L. Oneto, Y. Ma, L. Malandri, E. Cambria, Ensemble of technical analysis and machine learning for market trend prediction, in: 2018 IEEE Symposium Series on Computational Intelligence, SSCI, 2018, pp. 2090–2096, http://dx.doi.org/10.1109/SSCI.2018.8628795.

[61] H.J. Park, Y. Kim, H.Y. Kim, Stock market forecasting using a multi-task approach integrating long short-term memory and the random forest framework, Appl. Soft Comput. 114 (2022) 108106, http://dx.doi.org/10.1016/j.asoc.2021.108106.

[62] D. Kingma, J. Ba, Adam: A method for stochastic optimization, in: International Conference on Learning Representations, ICLR, 2015, pp. 1–15.

[63] S.A. Sharpe, Financial market imperfections, firm leverage, and the cyclicality of employment, Am. Econ. Rev. 84 (4) (1994) 1060–1074.

[64] M. Magdon-Ismail, A.F. Atiya, Maximum drawdown, Risk Mag. 17 (10) (2004) 99–102.

[65] R. Mao, X. Li, Bridging towers of multi-task learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, (no. 15) 2021, pp. 13534–13542, http://dx.doi.org/10.1609/aaai.v35i15.17596.

[66] M. Ge, R. Mao, E. Cambria, Explainable metaphor identification inspired by conceptual metaphor theory, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, (no. 10) 2022, pp. 10681–10689, http://dx.doi.org/10.1609/aaai.v36i10.21313.

[67] R. Liu, G. Chen, R. Mao, E. Cambria, A multi-task learning model for gold-two-mention co-reference resolution, in: 2023 International Joint Conference on Neural Networks, IJCNN, Queensland, Australia, 2023, pp. 1–9.

[68] W.J. Yeo, W. van der Heever, R. Mao, E. Cambria, R. Satapathy, G. Mengaldo, A comprehensive review on financial explainable AI, 2023, arXiv preprint arXiv:2309.11960.

[69] H. Turbé, M. Bjelogrlic, C. Lovis, G. Mengaldo, Evaluation of post-hoc interpretability methods in time-series classification, Nat. Mach. Intell. 5 (3) (2023) 250–260.

[70] E. Cambria, R. Mao, M. Chen, Z. Wang, S.-B. Ho, Seven pillars for the future of artificial intelligence, IEEE Intell. Syst. 38 (6), pp. 62-69 (2023).

[71] R. Mao, C. Lin, F. Guerin, Word embedding and WordNet based metaphor identification and interpretation, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 1222–1231, http://dx.doi.org/10.18653/v1/P18-1113.

[72] R. Mao, X. Li, K. He, M. Ge, E. Cambria, MetaPro online: A computational metaphor processing online system, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (System Demonstrations), Vol. 3, Association for Computational Linguistics, 2023, pp. 127–135.