Heterogeneous Interactive Snapshot Network for Review-Enhanced Stock Profiling and Recommendation

Heyuan Wang[1,3], **Tengjiao Wang**[1,3], **Shun Li**[2*], **Shijie Guan**[1,3], **Jiayi Zheng**[1,3] and **Wei Chen**[1,3]
[1]School of Computer Science, National Engineering Laboratory for Big Data Analysis and Applications, Peking University, China
[2]University of International Relations
[3]Institute of Computational Social Science, Peking University(Qingdao)
{wangheyuan, tjwang, guanshijie, jiayizheng, pekingchenwei}@pku.edu.cn, lishunmail@foxmail.com

Abstract

Stock recommendation plays a critical role in modern quantitative trading. The large volumes of social media information such as investment reviews that delegate emotion-driven factors, together with price technical indicators formulate a "snapshot" of the evolving stock market profile. However, previous studies usually model the temporal trajectories of price and media modalities separately while losing their interrelated influences. Moreover, they mainly extract review semantics via sequential or attentive models, whereas the rich text associated knowledge is largely neglected. In this paper, we propose a novel heterogeneous interactive snapshot network for stock profiling and recommendation. We model investment reviews in each snapshot as a heterogeneous document graph, and develop a flexible hierarchical attentive propagation framework to capture fine-grained proximity features. Further, to learn stock embedding for ranking, we introduce a novel twins-GRU method, which tightly couples the media and price parallel sequences in a cross-interactive fashion thereby catching dynamic dependencies between successive snapshots. Our model excels state-of-the-arts over 7.6% in terms of cumulative and risk-adjusted returns in trading simulations on both English and Chinese benchmarks.

1 Introduction

Accurately predicting stock movements is a prerequisite for making profitable investment decisions. Conventional literatures from the mathematical and deep learning communities [Nayak *et al.*, 2015; Qin *et al.*, 2017] hammer at decomposing stock's volatility patterns from history technical indices. Nevertheless, unlike general time series, the dynamic of stock markets is highly stochastic and susceptible to non-stationary behaviors. With the growth of the Internet, massive online media reviews (such as financial news and tweets) provide great potential to reveal market environmental variations [Hu *et al.*, 2018; Sawhney *et al.*, 2021b]. However,

*Corresponding Author.

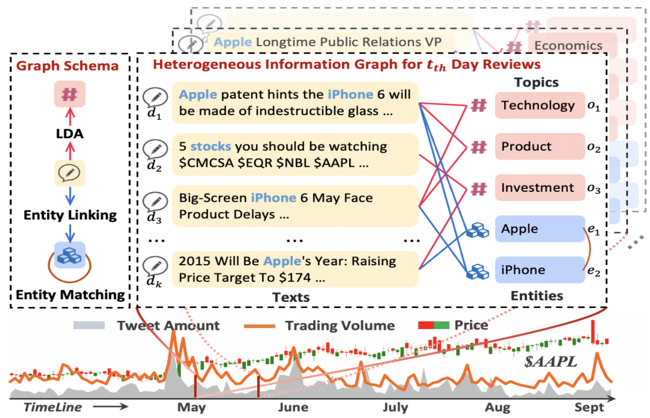


Figure 1: Dynamic snapshots of stock $AAPL and heterogeneous information graph for intra-day social media investment reviews.

previous studies on text-enhanced stock forecasting have limitations in terms of semantic representation and temporal dependency modeling. We elaborate our research insights from the following two perspectives.

First, existing stock prediction methods usually capture the semantic features of investment reviews via sequential or attentive models, whereas rich inter-text association knowledge is largely neglected. In fact, the consecutively delivered stock reviews are often context-relevant. Bridging them together through endogenous textual aspects such as topics and entities furnishes a broader scope of information for interpreting stock price oscillations. Fig. 1 illustrates the snapshots including daily technical signals and media reviews of $AAPL's evolving profile. On the t_{th} trading day, the first and third texts both refer to "technology" and "product", while the second and last texts are about "investment" opinions. These matched pairs portray complementary information toward specific topics. In addition, the first and last reviews mentioning the same entity transitively reveal the public's expectation of stock benefits resulted from "product launches". Moreover, the relatedness between different entities (e.g., "Apple" and "iPhone") is also an important indication to expound the unified context, which can be discerned by referring to additional knowledge like Wikipedia. This inspires us building a heterogeneous information graph for media signals of each stock snapshot, in order to make full use of semantic relevant clues to comprehensively assess underlying corporate status.

On another front, to synthesize the manifold market and media signals for stock profiling, prior studies generally follow two paradigms: (i) *Early-stage concatenation*, which treats the two modalities equally to constitute a compact input at each time point; and (ii) *Last-stage interaction*, which separately models each type of sequential data, then blends them together at the final layer. For instance, Xu and Cohen [2018] spliced every day's price and text vectors without discriminating the traits of different feature spaces. Wang *et al.* [2020] implemented case-by-case bullish/bearish stance classification of substantial stock reviews to expand discrete price indicators, whereas the pipelined processing is inevitably labor-intensive. In [Sawhney *et al.*, 2020], text and price sequences are independently modeled followed by a bilinear combination layer. However, fine-granular matching signals between the two-sided factors are lost since they cannot communicate until the final step of prediction. Indeed, real-life stock market is rapidly responsive to new messages, and conversely drastic price changes are apt to trigger extensive discussions on social media [Foucault *et al.*, 2016]. In Fig. 1, the dynamics of stock trading volume and the amount of relevant investment reviews are great synchronous especially when a mutation occurs. This phenomenon tallies with the Efficient Market Hypothesis (EMH) [Malkiel, 2003] and Behavioral Finance (BF) [Shiller, 2003], which state that stock prices can reflect all available information yet actual irrational markets are greatly affected by participant psychological factors. Hence a finer interaction method is required to better learn the interdependencies between market and media information.

Along these lines, in this paper, we formulate stock forecast as a *learning to rank* task and propose a *Heterogeneous Interactive Snapshot Network* (HISN) for review-enhanced investment recommendation. Our main contributions include:

- We represent stock media reviews for each trading day as a heterogeneous document graph incorporating additional knowledge. Besides, a *hierarchical attentive propagation (HAP)* framework with flexible node- and view-level aggregators is developed to assimilate important proximity features for more accurate semantic modeling.
- We propose *twins-GRU*, which tightly couples market and media parallel sequences in a cross-interactive fashion to catch time-evolving mutual impacts between different modalities. Being aware of time attenuation, significant hidden states of stock dynamic snapshots are integrated to learn stock profile embeddings for expected profit ranking.
- We conduct extensive experiments on English and Chinese datasets spanning various stock exchange markets to verify HISN's applicability. The cumulative and risk-adjusted returns outperform state-of-the-arts by over 7.6% and 10.2%.

2 Related Work

Technical Analysis. Traditional stock prediction studies are based on analyses of price quote data. Many machine learning algorithms such as HMM, SVM and Random Forest were early used to capture volatility patterns [Nayak *et al.*, 2015; Khaidem *et al.*, 2016], whereas the hypothetical stochastic process may become stranded in handling highly non-stationary oscillations. Recently, deep neural networks have shown great prospects by formulating the stock forecast as a time series modeling problem, especially RNNs are mainstream technologies [Qin *et al.*, 2017; Zhang *et al.*, 2017; Sawhney *et al.*, 2021a]. For instance, Qin *et al.* [2017] applied an attentive recurrent network to identify driving input and hidden states of stock dynamics. Zhang *et al.* [2017] injected a state frequency memory in LSTM to discover multi-frequency trading patterns based on Discrete Fourier Transform. Modern researches resorted to other neural methods to embed stock time series, such as adopting adversarial training to improve model generalization [Feng *et al.*, 2019a], mining compositive volatilities based on Gaussian Transformer encoder [Ding *et al.*, 2020] and gating causal convolutions with the idea of WaveNet [Wang *et al.*, 2021].

Text Assisted Analysis. Except for numeric-based methods, another line of studies probes into exploring stock-related social media content such as financial news, tweets and earnings reports to learn stock environmental variables [Hu *et al.*, 2018; Xu and Cohen, 2018; Sawhney *et al.*, 2021b]. These works usually align the bullishness/bearishness of stock price movement with the sentiment of media messages. For instance, Hu *et al.* [2018] proposed bidirectional GRUs with hybrid attention on news and days for stock prediction. Wang *et al.* [2020] developed a stance detection and expert mining procedure to identify high-quality investment opinions from online discussion boards. Sawhney *et al.* [2020] adopted deep attentive learning for multipronged stock-related information. Sawhney *et al.* [2021b] applied BERT [Devlin *et al.*, 2019] to encode intraday reviews and then used a time-aware LSTM to aggregate them considering temporal irregularities. Despite their success, the dynamic interdependencies between media and market information have not been handled delicately. Besides, in these works stock reviews are generally represented as the set or sequence of texts, which may be not optimal since inter-text relatedness cannot be effectively captured. As the chaotic online content usually involves various information and is characterized with high degrees of obscurity, the intersection of Finance and heterogeneous semantic information embedding presents a promising research avenue.

3 Methodology

3.1 Problem Statement

The stock recommendation can be formulated as a *learning to rank* problem. Let $\mathcal{S} = \{s_1, \ldots, s_N\}$ denotes a set of N candidate stocks, where for stock s_i on trading day t, there is an associated closing price p_i^t and a 1-day return ratio $r_i^t = \frac{p_i^t - p_i^{t-1}}{p_i^{t-1}}$. To learn an optimal investment ranking $\mathcal{Y}^t = \{y_1^t > y_2^t \ldots > y_N^t\}$ of all stocks on t, each s_i entails a series of dynamic snapshots with lag size of ΔT as the input tensor $\{\mathcal{F}_i^{t-\Delta T}, \ldots, \mathcal{F}_i^{t-1}\}$, where $\mathcal{F}_i^\tau = (\mathcal{P}_i^\tau, \mathcal{D}_i^\tau)$ represents its price and media review signals at historical day τ. Briefly, the top-ranked stocks are expected to earn higher investment revenues in daily trading with such an ordering framework.

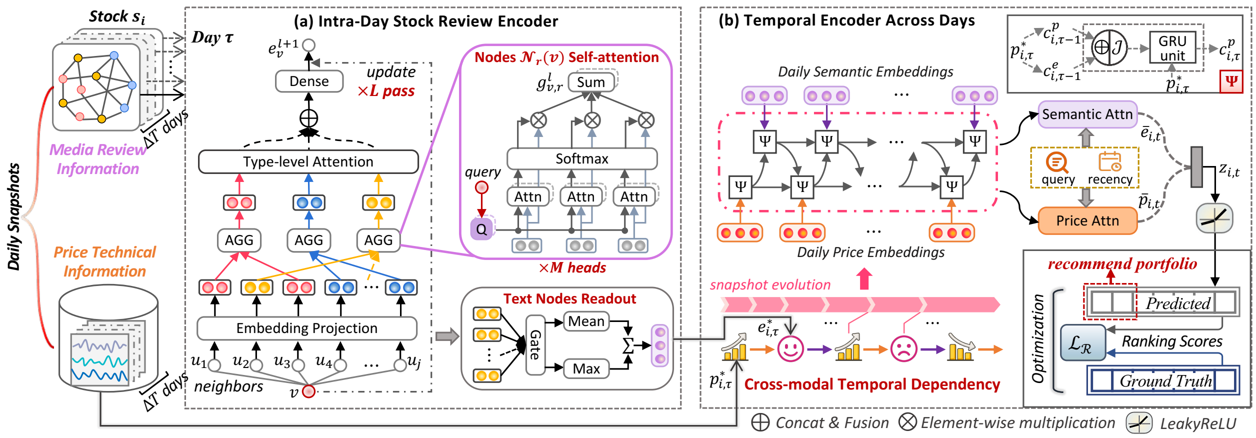


Figure 2: Overview of HISN, including representation and interaction of media & price signals across days, and stock ranking optimization.

Fig. 2 presents an overview of our proposed *HISN*. In the following subsections, we first show how the heterogeneous document graph is built for investment reviews of each stock snapshot, and present the *hierarchical attentive propagation* method to coordinate multiple neighbor features (§**3.2**). We then describe *twins-GRU*, which couples duplex media and market signals across successive snapshots in the lookback period (§**3.3**). Lastly, we combine significant hidden states to learn stock embeddings for profit ranking in trading (§**3.4**).

3.2 Intra-Day Semantic Encoder

Heterogeneous Document Graph Construction

Given the set of media reviews $\mathcal{D}_i^\tau$ discussing news events or opinions from $\mathcal{F}_i^\tau$, we construct a heterogeneous graph which contains three kinds of nodes including texts $\bar{\mathcal{D}}$, topics $\bar{\mathcal{O}}$ and entities $\bar{\mathcal{E}}$ to learn the unified semantic context. As illustrated in Fig. 1, topics are important indications for revealing the nub of text content and public concerns. Therefore, we leverage LDA [Blei *et al.*, 2003] to mine latent topics of the entire daily review document as graph nodes, and connect each text to $|\bar{o}|$ topic nodes with the largest relevance probabilities. To be consistent with previous research [Sawhney *et al.*, 2021b], we generate the initial states of text nodes by averaging token level outputs from the final layer of BERT [Devlin *et al.*, 2019]. Each topic node is represented by weighing word embeddings based on its probability distribution over the vocabulary. Moreover, stock reviews delivered consecutively usually concentrate on a few specific entities which complementarily portray the dynamic corporate status. In this regard, we utilize an entity linking tool TAGME[1] to recognize the meaningful spots in each text, and map them to pertinent entities in Wikipedia. Whereas, the textual relationships formed only based on explicitly included entities may be sparse. Motivated by [Hu *et al.*, 2019], we further attach edges between similar entity nodes to promote information diffusion in semantic learning. Specifically, we empower entity embeddings using word2vec trained on Wikipedia annotation corpus. Then a pair of entities are connected if their cosine similarity is above a threshold δ to finalize the graph structure $\mathcal{G}_i^\tau$.

[1]https://sobigdata.d4science.org/group/tagme/

Hierarchical Attentive Propagation (HAP)

Once the graph topology is anchored, we perform node interactions to push forward semantic diffusion taking initial node states $[\boldsymbol{e}_v^0]_{v\in\bar{\mathcal{D}}\cup\bar{\mathcal{O}}\cup\bar{\mathcal{E}}}$ as input. Specifically, for target node v at l_{th} propagation layer, *HAP* convolves over its neighborhood following a hierarchical attention structure: $Attention_{node} \rightarrow Attention_{type} \rightarrow \boldsymbol{e}_v^{l+1}$. Features of each type of v's adjacent nodes are first transformed and fused together, then fed to the inter-type polymerization to enrich the representation of v.

Node-level Aggregator. We implement $Attention_{node}$ by a multi-head self-attention layer [Vaswani *et al.*, 2017]. The primary motivation is that multiple node neighbors from the same relation type may complement each other in characterizing specific corporate profiles (e.g., implicit semantic correlation between entities of "Apple" and "iPhone", and topics of "technological innovation" and "launching new products"). Specifically, by concatenating the r-type of v's neighborhood including itself (denoted as $\mathcal{N}_r(v)$) into a context matrix $\boldsymbol{E}_{v,r}^l \in \mathbb{R}^{|\mathcal{N}_r(v)|\times d_c}$ (where d_c is the dimension of node states), we transform the *Key-Value* tuples as $\boldsymbol{K}_{v,r,h}^l = \boldsymbol{V}_{v,r,h}^l = \boldsymbol{E}_{v,r}^l\boldsymbol{W}_h^Q \in \mathbb{R}^{|\mathcal{N}_r(v)|\times d_f}$, and take the current state of target node v as a *Query* matrix $\boldsymbol{Q}_{v,r,h}^l = \boldsymbol{e}_v^l\boldsymbol{W}_h^Q \in \mathbb{R}^{1\times d_f}$, where $h=1,\ldots,M$ are indices of different attention heads, $\boldsymbol{W}_h^Q \in \mathbb{R}^{d_c\times d_f}$ is a linear projection parameter. Here, we set $d_f = \frac{d_c}{M}$ in order to maintain the dimension of the node states. After projection, the self-attention operator exploits element-wise dependencies to enhance the representation of v (the query) by attending to the context nodes (keys):

$$\begin{aligned}\boldsymbol{g}_{v,r}^l &= \text{Multihead}(\boldsymbol{Q}_{v,r,h}^l, \boldsymbol{K}_{v,r,h}^l, \boldsymbol{V}_{v,r,h}^l) \\ &= ||_{h=1}^{h=M}\text{Softmax}(\boldsymbol{Q}_{v,r,h}^l\boldsymbol{K}_{v,r,h}^{l^\top}/\sqrt{d_f})\boldsymbol{V}_{v,r,h}^l\,,\end{aligned} \quad (1)$$

where $||$ denotes the concatenation of output vectors from all M heads, the scaled dot product measures the informativeness of different neighbors for contextual summarization.

Type-level Aggregator. With the aggregated semantic-specific node messages, we engage *Attention*$_{type}$, a type-level attention mechanism to capture the importance of different information aspects for updating a target node. Specifically, we compute the weighting coefficient of type r to v combining two terms – the matching similarity between $(\boldsymbol{g}^l_{v,r}, \boldsymbol{e}^l_v)$ determined by *SUBMULT+NN* function [Mou *et al.*, 2016], and an adaptive type-level attention tensor $\boldsymbol{T}_{1:|\mathcal{R}|} \in \mathbb{R}^{|\mathcal{R}|\times 4d_c}$ which serves to indicate the general semantic significance:

$$\mu^l_{v,r} = \boldsymbol{T}_r \left[\, \boldsymbol{g}^l_{v,r} \,||\, \boldsymbol{e}^l_v \,||\, (\boldsymbol{g}^l_{v,r} - \boldsymbol{e}^l_v) \,||\, (\boldsymbol{g}^l_{v,r} \circ \boldsymbol{e}^l_v) \,\right]^\top, \quad (2)$$

$$\beta^l_{v,r} = \frac{\exp(\mu^l_{v,r})}{\sum_{r'=1}^{|\mathcal{R}|} \exp(\mu^l_{v,r'})}, \quad (3)$$

where $-$ and $\circ$ mean the difference and element-wise product. The information of all types is fused as $\bar{\boldsymbol{g}}^l_v = \sum_{r=1}^{|\mathcal{R}|} \beta^l_{v,r}\boldsymbol{g}^l_{v,r}$. Additionally, we place each *HAP* layer inside a residual connection block for updating current states of all graph nodes, which helps retain low-order features and facilitate gradient back-propagation. By this means, we obtain $\boldsymbol{e}^{l+1}_v = \bar{\boldsymbol{g}}^l_v + \boldsymbol{e}^l_v$.

After substantiating the graph correlations for L rounds, we apply a soft gating transformation based on two multilayer perceptions (denoted as f_1, f_2) to integrate the features of all text nodes. Since each review plays a role in the day's public voice while the salient ones contribute more explicitly, we jointly apply max-pooling and average functions to produce the final sentiment embedding of $\mathcal{F}^\tau_i$, defined as:

$$\hat{\boldsymbol{e}}_v = \sigma(f_1(\boldsymbol{e}^L_v)) \circ \tanh(f_2(\boldsymbol{e}^L_v)) \quad \forall v \in \bar{\mathcal{D}}, \quad (4)$$

$$\boldsymbol{e}^*_{i,\tau} = \frac{1}{|\bar{\mathcal{D}}|}\sum_{v\in\bar{\mathcal{D}}} \hat{\boldsymbol{e}}_v + \mathrm{Maxpooling}(\hat{\boldsymbol{e}}_1 \cdots \hat{\boldsymbol{e}}_{|\bar{\mathcal{D}}|}). \quad (5)$$

3.3 Inter-day Temporal Encoder

Temporal Dependency Mining

How to effectively integrate the manifold information across dynamic snapshots in a lookback period is another key issue to discover the stock evolution regularities for future movement prediction. In light of EMH and BF literatures [Malkiel, 2003; Shiller, 2003], stock prices can timely react to the emotional signals from social media, and in turn can drive public's expectation of new market situations. It is essential to learn evolving stock profiles by synthetically modeling such highly correlated dual-pronged factors. To this end, unlike most methods separately modeling or simply splicing the price and media information sequences of all snapshots (denoted as $\boldsymbol{P}^*_{i,t} = [\boldsymbol{p}^*_{i,t-\Delta T}, \ldots, \boldsymbol{p}^*_{i,t-1}]$, $\boldsymbol{E}^*_{i,t} = [\boldsymbol{e}^*_{i,t-\Delta T}, \ldots, \boldsymbol{e}^*_{i,t-1}]$), we align them more finely along with the temporal trajectory of stock evolution in a cross-interactive fashion. Specifically, we design *twins-GRU*, which takes the advantage of two parallel GRU networks that not only focus on the sequence of each modality but interact with each other progressively.

As shown in Fig. 2, we represent the neural unit in *twins-GRU* cell as Ψ. Take price embedding for example, at every time-step τ, the new coming input vector $\boldsymbol{p}^*_{i,\tau}$ and two predecessor hidden states from both price and media sequences $\boldsymbol{c}^p_{i,\tau-1}, \boldsymbol{c}^e_{i,\tau-1}$ are fed into Ψ, such that the duplex information is simultaneously incorporated to characterize the dynamic context. Concerning different impacts of semantic and price factors, the $\oplus$ operator in Ψ engages an attentive summation:

$$\boldsymbol{u}^p_{i,\tau-1} = \phi\boldsymbol{c}^p_{i,\tau-1} + (1-\phi)\boldsymbol{c}^e_{i,\tau-1}, \quad (6)$$

where ϕ is a coefficient learned to rectify the fusing proportion of each kind of hidden signals, which is calculated by:

$$\phi = \frac{\exp(\eta^p_{i,\tau-1})}{\exp(\eta^p_{i,\tau-1}) + \exp(\eta^e_{i,\tau-1})}, \quad (7)$$

$$\eta^p_{i,\tau-1} = \boldsymbol{\nu}_c[\boldsymbol{c}^p_{i,\tau-1} \,||\, \boldsymbol{p}^*_{i,\tau}], \;\; \eta^e_{i,\tau-1} = \boldsymbol{\nu}_c[\boldsymbol{c}^e_{i,\tau-1} \,||\, \boldsymbol{p}^*_{i,\tau}], \quad (8)$$

where $\boldsymbol{\nu}_c$ is the attention vector. The module $\mathcal{J}$ in Ψ further transforms the unified hidden state by a dense layer:

$$\tilde{\mathbf{c}}^p_{i,\tau-1} = \mathrm{ReLU}(\boldsymbol{W}_c\boldsymbol{u}^p_{i,\tau-1} + \boldsymbol{b}_c), \quad (9)$$

where $\boldsymbol{W}_c$, $\boldsymbol{b}_c$ are weight and bias parameters. Then $\tilde{\mathbf{c}}^p_{i,\tau-1}$ and $\boldsymbol{p}^*_{i,\tau}$ are fed into a canonical GRU unit to update the price memory $\boldsymbol{c}^p_{i,\tau}$. Analogously we can acquire new memory $\boldsymbol{c}^e_{i,\tau}$.

Temporal Aggregation

The last part of the framework fuses the entire information stream to learn stock's compact representation for profit ranking. Intuitively, historical transaction days have varied predictive influences, where days with profound reviews and drastic price changes usually excite investors' behaviors more intensively. Meanwhile, the study of Hawkes process in finance shows that the impact of emerging market variables dwindles over time [Bacry *et al.*, 2015]. Thus we conduct day-level aggregation by taking both feature salience and time attenuation into account. The weight of the day τ is specified as:

$$\theta_\tau = \frac{\exp(\boldsymbol{c}^p_{i,\tau}\tilde{\boldsymbol{W}}\boldsymbol{c}^{p\top}_{i,t-1})}{\sum_k \exp(\boldsymbol{c}^p_{i,k}\tilde{\boldsymbol{W}}\boldsymbol{c}^{p\top}_{i,t-1})} \times \Big(1 + \epsilon\exp(-\gamma\Delta o_\tau)\Big), \quad (10)$$

where $\tilde{\boldsymbol{W}}$ is a transformation matrix, ϵ is excitation coefficient, γ is a decay scaler and Δo_τ is the lag of τ to latest time. The overall price sequence is then summed as $\bar{\boldsymbol{p}}_{i,t} = \sum_\tau \theta_\tau\boldsymbol{c}^p_{i,\tau}$. The condensed semantic embedding can be distilled in a very similar way. Finally, we tailor the profile of stock s_i as the concatenation of abstracts $\boldsymbol{z}_{i,t} = [\bar{\boldsymbol{p}}_{i,t} \,||\, \bar{\boldsymbol{e}}_{i,t}]$, and all candidate stocks form the tensor $\boldsymbol{Z}_t \in \mathbb{R}^{N\times d_o}$.

3.4 Model Training

Following [Sawhney *et al.*, 2021b], we pass $\boldsymbol{Z}_t$ to a dense layer with the activation of LeakyReLU to acquire stock expected return ratios $\hat{r}^t_{[1:N]}$ on day t. Based on joint optimization of point-wise regression and pairwise ranking loss, we minimize the discrepancy between $\hat{r}^t_{[1:N]}$ and ground-truth $r^t_{[1:N]}$ meanwhile maintaining the relative order of stocks:

$$\mathcal{L}_\mathcal{R} = \sum_{i=1}^N \left\|\hat{r}^t_i - r^t_i\right\|^2 + \lambda\sum_{i=1}^N\sum_{j=1}^N \max\left(0, -\left(\hat{r}^t_i - \hat{r}^t_j\right)\left(r^t_i - r^t_j\right)\right), \quad (11)$$

where λ is a weighting coefficient. Stocks at the top of the forecast order are selected for investment recommendations.

	Methods		S&P 500		Ashare&HK	
			CIR	*SR*	*CIR*	*SR*
CLF	**TSLDA** [Nguyen and Shirai, 2015]	LDA generative model jointly exploiting topics and sentiments in texts	0.40	0.39	0.50	0.51
CLF	**StockNet** [Xu and Cohen, 2018]	Variational autoencoder (HedgeFundAnalyst) with semantic attentions	1.09	0.81	1.12	0.93
CLF	**HAN** [Hu *et al.*, 2018]	Bidirectional GRU encoder with news and days attentions	1.07	0.80	1.20	1.01
CLF	**Adv-ALSTM** [Feng *et al.*, 2019a]	Simulate price dynamic stochasticity supported by adversarial training	0.97	0.75	1.05	0.83
CLF	**StockEmb** [Du and Tanaka-Ishii, 2020]	Stock embedding based on price and dual-vector textual representation	0.70	0.51	0.89	0.74
CLF	**MFN** [Wang *et al.*, 2020]	Discretize emotion indicators by refining lexical attributes of opinion tuples	1.16	0.87	1.27	1.09
CLF	**MAN-SF** [Sawhney *et al.*, 2020]	Attentive fusion learning of stock prices, tweets and correlations	1.24	0.92	1.30	1.14
CLF	**HMG-TF** [Ding *et al.*, 2020]	Gaussian Transformer + Orthogonal regularization on price sequence	1.11	0.82	1.25	1.04
CLF	**HATR** [Wang *et al.*, 2021]	Gated causal convolution to capture long term fluctuation patterns	1.21	0.84	1.28	1.10
REG	**DA-RNN** [Qin *et al.*, 2017]	RNNs + Dual-stage attention mechanism to weigh input and hidden states	0.71	0.60	0.83	0.72
REG	**SFM** [Zhang *et al.*, 2017]	RNNs + Discrete Fourier Transform to discover multi-frequency dynamics	0.47	0.42	0.54	0.50
RL	**S-Reward** [Yang *et al.*, 2018]	Gaussian Inverse-RL to model relevance between sentiments and returns	0.93	0.73	1.33	1.08
RL	**iRDPG** [Liu *et al.*, 2020]	Simulate RDPG model to exploit temporal data for reward of Sharpe Ratio	1.05	0.79	1.19	1.03
RL	**RAT** [Xu *et al.*, 2020]	Relation-aware Transformer under RL framework for portfolio learning	1.26	0.95	1.31	1.07
RAN	**RankNet** [Song *et al.*, 2017]	Stock ranking using sentiment-based shock and trend indicators	1.16	0.87	1.14	0.95
RAN	**RSR** [Feng *et al.*, 2019b]	Train temporal GCN on 5, 10, 20, 30-day's average and close prices	1.01	0.78	1.17	0.96
RAN	**STHAN-SR** [Sawhney *et al.*, 2021a]	Attentive LSTM-based price encoder + hypergraph relation modeling	1.29	<u>*1.08*</u>	1.42	<u>*1.27*</u>
RAN	**FAST** [Sawhney *et al.*, 2021b]	BERT + time-aware LSTM to encode reviews during and across days	<u>*1.34*</u>	0.96	<u>*1.44*</u>	1.19
	HISN (Ours)	Interactive modeling of stock price and heterogeneous review semantics	**1.53***	**1.21***	**1.55***	**1.40***

Table 1: Profitability comparison with *Classification (CLF), Regression (REG), Reinforcement Learning (RL)*, and *Ranking (RAN)* methods. Bold & underlines depict the best & second-best results. $\star$ means the improvement over SOTA is statistically significant ($p < 0.01$).

4 Experiments

4.1 Dataset and Experimental Setting

We validate *HISN* on two real-world stock forecast datasets: **US S&P 500** [Xu and Cohen, 2018] contains 109,915 English Twitter data between Jan. 2014 and Jan. 2016, related to 88 high-capital-size stocks from popular S&P-500 Composite Index spanning 9 industries in *NASDAQ* and *NYSE* markets. Tweets and stocks are associated using the regex query of ticker symbols (e.g., $AAPL for Apple). Samples of the first 19 months are split for training, those of the last 3 months for testing, and the rest for validation in chronological order. **Ashare&HK** [Huang *et al.*, 2018] collects 90,361 news headlines from major financial websites in Chinese during Jan. and Dec. 2015. It targets at 99 top-traded A-share and HK stocks spanning *Shanghai*, *Shenzhen* and *Hong Kong* Exchange markets. Samples are chronologically divided, leaving us with a date range of the first 8 months for training, the last 3 months for testing and the others for validation.

We collect split-adjusted open, high, low, close and volume indicators from professional *Wind-Financial Terminal*[2] to constitute daily price vectors of all stocks. We keep the same setting as in previous works [Sawhney *et al.*, 2021b] and leverage a consecutive 5-day lookback trading window (i.e., 5 daily snapshots) to generate each sample. To build heterogeneous document graphs, we set the number of topic nodes $|\bar{\mathcal{O}}|$, the per-text related topics $\bar{o}$ and the similarity threshold between entities δ to $\{15, 2, 0.5\}$. Hence, 15,448 and 6,696 entity nodes are exploited for S&P 500 and Ashare&HK datasets, respectively. The word embeddings to initialize graph nodes are 300-dimensional. We set $L = 2$ graph layers and $M = 4$ attention heads in *HAP*. The hidden state size of *twins-GRU* is 64. The loss weighing factor $\lambda = 4$. We apply dropout [Srivastava *et al.*, 2014] with the ratio of 0.3 at the end of each layer to mitigate overfitting. Parameters are tuned using Adam optimizer [Kingma and Ba, 2015] on a GeForce RTX 3090 GPU for 50 epochs, the batch size is 16 and learning rate is 1e-3. Each experiment is repeated 5 times. We report average MRR, nDCG@5 to assess the model's ranking ability. To compare the practical revenue of stock recommendation, we follow [Feng *et al.*, 2019b; Sawhney *et al.*, 2021b] and calculate the Sharpe ratio (SR) and cumulative investment return ratio (CIR) by simulating a daily *buy-hold-sell* trading strategy. That is, when the market closes on day $t-1$, the trader buys κ stocks with the highest expected returns, then sells the bought shares on next day's close market. Specifically, CIR$^t = \sum_{i \in \hat{\mathcal{S}}^t} \frac{p_i^t - p_i^{t-1}}{p_i^{t-1}}$, where $\hat{\mathcal{S}}^t$ denotes selected stocks in portfolio. SR characterizes how well the earned return R_p compensates a trader for the borne risk, i.e., SR$= \frac{E[R_p]-R_f}{std[R_p]}$ where R_f is the risk-free return.

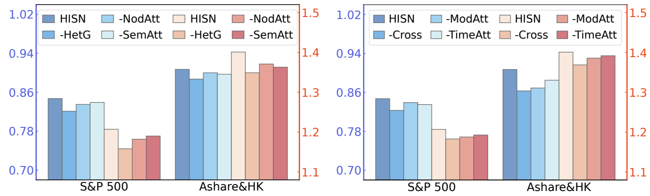
[2]https://www.wind.com.cn/en/wft.html

	S&P 500				Ashare&HK			
	MRR	*CIR*	*nDCG*	*SR*	*MRR*	*CIR*	*nDCG*	*SR*
P (Price)	0.067	1.307	0.762	0.896	0.105	1.339	0.796	1.142
P+T (Text)	0.086	1.478	0.821	1.158	0.129	1.496	0.887	1.349
P+T+Topic	0.096	1.490	0.834	1.179	0.136	1.515	0.896	1.370
P+T+Entity	0.104	1.518	0.838	1.191	0.142	1.538	0.903	1.386
P+T+Topic+Entity	**0.110**	**1.529**	**0.847**	**1.207**	**0.146**	**1.546**	**0.907**	**1.401**

Table 2: Effect of incorporating different information sources.

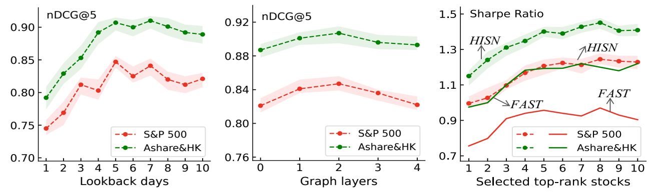
4.2 Performance Evaluation

We compare *HISN*'s profitability with different lines of stock prediction methods. The evaluation results are shown in Table 1, which reveal several findings: 1) Overall, RL and ranking approaches excel most traditional price regression and classification methods, illustrating the efficacy of formulating stock forecast as a learning-to-rank paradigm that directly hammers at optimizing investment profits. 2) Exploiting web



(a) Media review encoder (b) Temporal encoder

Figure 3: Ablation of components in review and temporal modeling. Blue- and red-shade show results on nDCG@5 and SR respectively.



(a) Lookback period (b) Graph layer (c) Selected stocks

Figure 4: Influence of hyper-parameters ΔT, L and κ.

information beyond numeric prices (e.g., *MFN*, *MAN-SF*, *FAST*) can usually impose positive effect on judging the stock tendency. It indicates the desirability of mining dependencies of multifaceted affecting signals confronting the high non-stationarity of financial markets. 3) The test periods of S&P 500 and Ashare&HK datasets cover standard and turbulence (bearish) market conditions respectively. Facing the diverse scenarios, our proposed *HISN* can consistently outperform compared baselines in terms of both cumulative and risk-adjusted returns. We attribute the performance gains to two major reasons. First, *HISN* learns the joint effect of market and media signals in a cross-interactive way being aware of their temporal interdependencies. Further, by enriching the representation of chaotic reviews with additional knowledge including topics, entities and Wikipedia descriptions, latent semantic clews can be absorbed for more comprehensive stock profiling. Next, we move to explore how different components and hyper-parameters affect the capability of *HISN*.

4.3 Ablation Study

Effect of Absorbed Information. We first evaluate the effectiveness of different information sources on stock profiling. From Table 2, significant improvements are noted by blending the textual signals beyond unimodal price features. In addition, we find that both topic and entity nodes contribute to the performance by facilitating the semantic learning of online media texts that usually have high degrees of obscurity and insufficiency. This is mainly because topics can depict the nub of review content, while entities and their neighbors portray more detailed meanings of key elements in each text.

Effect of Model Variants. We then investigate ablation impacts of different components in *HISN*. Fig. 3a shows the results of applying plain GCN [Kipf and Welling, 2017] to convolve entire review document graph by concatenating feature spaces of all types of nodes (*w/o HetG*), and discarding node- and semantic-level attentions (*w/o NodAtt, SemAtt*). We find

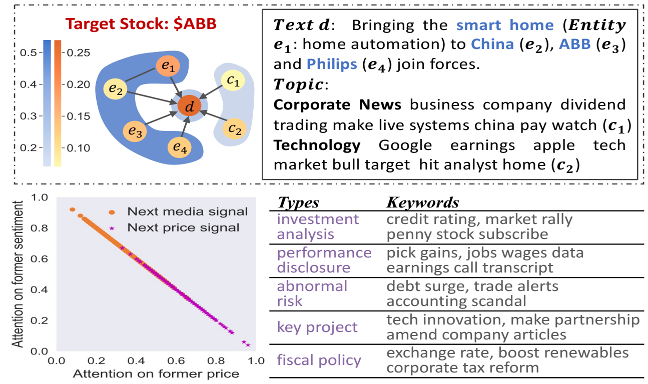


Figure 5: Visualization of semantic and temporal attentions.

that ignoring the heterogeneity of node relationships seriously hurts the effect of graph embedding. Meanwhile, the hierarchical attention mechanism is conducive to the model's performance. This proves the advantage of *HAP* in learning the significance of different nodes and semantics to a specific target thereby reducing the propagation of noise information.

We further conduct ablation research on the inter-day temporal encoder, including decoupling parallel GRUs to model price and semantic sequences in isolation (*w/o Cross*), discarding the attention on modalities in Ψ (*w/o ModAtt*) and directly feeding the last-day's hidden state for ranking (*w/o TimeAtt*). From Fig. 3b, interactively modeling the transition of media and market signals clearly benefits stock profiling and the estimation of expected investment revenues. Besides, weighing different kinds of information and historical time points is helpful to the modeling and integration of stock dynamic trajectory, which may be because it highlights salient influential variables and reduces false overreactions.

4.4 Parameter Analysis

We examine stock ranking ability regarding different lengths of lookback days. From Fig. 4a, the performance first increases with larger ΔT, then begins to decline due to the inclusion of outdated media and market signals. Fig. 4b shows the impact of applying varied numbers of graph layers for semantic embedding, where the optimal setting is $L = 2$. This is reasonable because as nodes incrementally absorb information from high-order neighbors, the graph may become oversmooth thus may deteriorate the accuracy of distilled text signals. We further explore the change of profitability w.r.t. the number of selected stocks κ. Fig. 4c shows that *HISN* consistently performs better than *FAST* (the SOTA stock ranking model exploiting media reviews). This demonstrates *HISN*'s adaptability to trading strategies with different risk appetites.

4.5 Case Study

We look closer to interpreting the attention mechanism of our model in semantic and temporal modeling. The upper part of Fig. 5 visualizes the node- and type-level weights *HAP* assigned to a randomly sampled review of stock $ABB. The text itself and entity relationship receive higher attentions. It means that intrinsic text content and associated knowledge of included entities (eg., *smart home*, *Philips* which reveal fine-grained clues of corporate business innovation and partnership) are most informative to characterize the review semantics. The lower left of Fig. 5 elaborates the attentions paid to diverse modalities (price and media memory states) given the next input. It can be seen that the duplex information act together during the modeling of stock evolution profile. In general, previous semantic state plays a major role when factoring in next media input, while the situation reverses when reconciling new price signals. Moreover, we summarize the top-5 types of reviews that have attracted the most attentions. As shown in the lower right of Fig. 5, *equity analysis* and *release of corporate operating conditions* would likely induce public's sentiment toward underlying stock variations more significantly, and thereby driving investment reactions.

5 Conclusion

We propose HISN, a heterogeneous interactive snapshot network for stock profiling and recommendation. We first devise a flexible HAP framework for modeling media investment reviews of each stock's dynamic snapshot, which incorporates additional information to capture implicit semantic relatedness. We also propose a twins-GRU that couples market and media signals in a cross-interactive fashion so as to catch their fine-granular interdependencies along the temporal trajectory. Extensive experiments demonstrate that HISN can identify evolving stock status to benefit the profitability of investment recommendation. In the future, we shall explore the correlation of enterprises and cold start problem of new stocks.

Acknowledgements

This research is supported by the National Key Research and Development Program of China (No. 2021ZD0111202), the National Natural Science Foundation of China (No.62176005, 91846303), PKU-Haier "Smart Home" Qingdao Joint Innovation Lab, Project 2020BD002 supported by PKU-Baidu Fund, and Research Funds for NSD Construction, University of International Relations.

References

[Bacry *et al.*, 2015] Emmanuel Bacry, Iacopo Mastromatteo, and Jean Fran?Ois Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01), 2015.

[Blei *et al.*, 2003] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[Devlin *et al.*, 2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, pages 4171–4186, 2019.

[Ding *et al.*, 2020] Qianggang Ding, Sifan Wu, Hao Sun, Jiadong Guo, and Jian Guo. Hierarchical multi-scale gaussian transformer for stock movement prediction. In *IJCAI*, pages 4640–4646, 2020.

[Du and Tanaka-Ishii, 2020] Xin Du and Kumiko Tanaka-Ishii. Stock embeddings acquired from news articles and price history, and an application to portfolio optimization. In *ACL*, pages 3353–3363, 2020.

[Feng *et al.*, 2019a] Fuli Feng, Huimin Chen, Xiangnan He, Ji Ding, Maosong Sun, and Tat-Seng Chua. Enhancing stock movement prediction with adversarial training. In *IJCAI*, pages 5843–5849, 2019.

[Feng *et al.*, 2019b] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. Temporal relational ranking for stock prediction. *TOIS*, 37(2):1–30, 2019.

[Foucault *et al.*, 2016] Thierry Foucault, Johan Hombert, and Ioanid Roşu. News trading and speed. *The Journal of Finance*, 71(1):335–382, 2016.

[Hu *et al.*, 2018] Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *WSDM*, pages 261–269, 2018.

[Hu *et al.*, 2019] Linmei Hu, Tianchi Yang, Chuan Shi, Houye Ji, and Xiaoli Li. Heterogeneous graph attention networks for semi-supervised short text classification. In *EMNLP-IJCNLP*, pages 4821–4830, 2019.

[Huang *et al.*, 2018] Jieyun Huang, Yunjia Zhang, Jialai Zhang, and Xi Zhang. A tensor-based sub-mode coordinate algorithm for stock prediction. In *DSC*, pages 716–721, 2018.

[Khaidem *et al.*, 2016] Luckyson Khaidem, Snehanshu Saha, and Sudeepa Roy Dey. Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*, 2016.

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[Liu *et al.*, 2020] Yang Liu, Qi Liu, Hongke Zhao, Zhen Pan, and Chuanren Liu. Adaptive quantitative trading: An imitative deep reinforcement learning approach. In *AAAI*, pages 2128–2135, 2020.

[Malkiel, 2003] Burton G Malkiel. The efficient market hypothesis and its critics. *Journal of economic perspectives*, 17(1):59–82, 2003.

[Mou *et al.*, 2016] Lili Mou, Rui Men, Ge Li, Yan Xu, Lu Zhang, Rui Yan, and Zhi Jin. Natural language inference by tree-based convolution and heuristic matching. In *ACL*, 2016.

[Nayak *et al.*, 2015] Rudra Kalyan Nayak, Debahuti Mishra, and Amiya Kumar Rath. A naïve svm-knn based stock market trend reversal analysis for indian benchmark indices. *Applied Soft Computing*, 35:670–680, 2015.

[Nguyen and Shirai, 2015] Thien Hai Nguyen and Kiyoaki Shirai. Topic modeling based sentiment analysis on social media for stock market prediction. In *ACL*, pages 1354–1364, 2015.

[Qin *et al.*, 2017] Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison W. Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. In *IJCAI*, pages 2627–2633, 2017.

[Sawhney *et al.*, 2020] Ramit Sawhney, Shivam Agarwal, Arnav Wadhwa, and Rajiv Ratn Shah. Deep attentive learning for stock movement prediction from social media text and company correlations. In *EMNLP*, pages 8415–8426, 2020.

[Sawhney *et al.*, 2021a] Ramit Sawhney, Shivam Agarwal, Arnav Wadhwa, Tyler Derr, and Rajiv Ratn Shah. Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach. In *AAAI*, pages 497–504, 2021.

[Sawhney *et al.*, 2021b] Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Shah. Fast: Financial news and tweet based time aware network for stock trading. In *EACL*, pages 2164–2175, 2021.

[Shiller, 2003] Robert J Shiller. From efficient markets theory to behavioral finance. *Journal of economic perspectives*, 17(1):83–104, 2003.

[Song *et al.*, 2017] Qiang Song, Anqi Liu, and Steve Y. Yang. Stock portfolio selection using learning-to-rank algorithms with news sentiment. *Neurocomputing*, 264:20–28, 2017.

[Srivastava *et al.*, 2014] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *JMLR*, 15(1):1929–1958, 2014.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, pages 5998–6008, 2017.

[Wang *et al.*, 2020] Heyuan Wang, Tengjiao Wang, and Yi Li. Incorporating expert-based investment opinion signals in stock prediction: A deep learning framework. In *AAAI*, pages 971–978, 2020.

[Wang *et al.*, 2021] Heyuan Wang, Shun Li, Tengjiao Wang, and Jiayi Zheng. Hierarchical adaptive temporal-relational modeling for stock trend prediction. In *IJCAI*, pages 3691–3698, 2021.

[Xu and Cohen, 2018] Yumo Xu and Shay B Cohen. Stock movement prediction from tweets and historical prices. In *ACL*, pages 1970–1979, 2018.

[Xu *et al.*, 2020] Ke Xu, Yifan Zhang, Deheng Ye, Peilin Zhao, and Mingkui Tan. Relation-aware transformer for portfolio policy learning. In *IJCAI*, pages 4647–4653, 2020.

[Yang *et al.*, 2018] Steve Y Yang, Yangyang Yu, and Saud Almahdi. An investor sentiment reward-based trading system using gaussian inverse reinforcement learning algorithm. *Expert Systems with Applications*, 114:388–401, 2018.

[Zhang *et al.*, 2017] Liheng Zhang, Charu Aggarwal, and Guo-Jun Qi. Stock price prediction via discovering multi-frequency trading patterns. In *KDD*, pages 2141–2149, 2017.