Retrieving Implicit Information for Stock Movement Prediction

Tsun-Hsien Tang
Department of Computer Science and Information Engineering,
National Taiwan University, Taiwan
thtang@nlg.csie.ntu.edu.tw

Chung-Chi Chen
Department of Computer Science and Information Engineering,
National Taiwan University, Taiwan
cjchen@nlg.csie.ntu.edu.tw

Hen-Hsen Huang
Department of Computer Science,
National Chengchi University, Taiwan
MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
hhhuang@nccu.edu.tw

Hsin-Hsi Chen
Department of Computer Science and Information Engineering,
National Taiwan University, Taiwan
MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
hhchen@ntu.edu.tw

ABSTRACT

Previous studies on the financial news focus mainly on the news articles explicitly mentioning the target financial instruments, and may suffer from data sparsity. As taking into consideration other related news, e.g., sector-related news, is a crucial part of real-world decision-making, we explore the use of news without explicit target mentions to enrich the information for the prediction model. We develop a neural network framework that jointly learns with a news selection mechanism to extract implicit information from the chaotic daily news pool. Our proposed model, called the news distilling network (NDN), takes advantage of neural representation learning and collaborative filtering to capture the relationship between stocks and news. With NDN, we learn latent stock and news representations to facilitate similarity measurements, and apply a gating mechanism to prevent noisy news representations from flowing to a higher level encoding stage, which encodes the selected news representation of each day. Extensive experiments on real-world stock market data demonstrate the effectiveness of our framework and show improvements over previous techniques.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; *Natural language processing*.

KEYWORDS

Stock Movement Prediction, News Sequence, Neural Network

ACM Reference Format:
Tsun-Hsien Tang, Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021. Retrieving Implicit Information for Stock Movement Prediction. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3404835.3462999

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

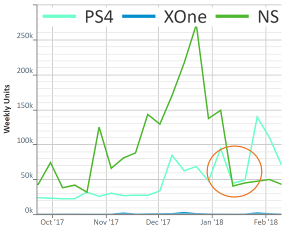
ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3462999

1 INTRODUCTION

Financial news is one of the important and focused textual resources in the financial domain. Lots of previous works present models to extract the information embedded in financial news, and evidence their models via stock market prediction task [2, 18]. However, using only news articles that make explicit mentions of specific stocks ignores topics potentially related to the target stock. That is, the presence of related information is not limited to news with explicit mentions of a stock. For example, the article "*Sales of PS4 Breaks 6 Million in Japan and Beats Switch First Time*", published on 2018/01/19, would be expected to influence the stock movements of supply chain vendors for that product. Figure 1 lists the related suppliers and their share prices corresponding to the date. Although the boosted sales could benefit the related suppliers, none of the suppliers' names are mentioned in the article. Therefore, when filtering out news articles that lack specific company names during preprocessing, crucial information is taken out of consideration. Based on this observation, implicitly related news articles could be integrated as a source of information. This calls for a model that effectively extracts relevant news articles from the chaotic daily news pool.

Investors bridge the implicit relation by domain knowledge and then make a financial decision. In previous works [11, 17, 23], models are fed by manually selected news articles that mention the target stock. They do not own the ability to distill the chaotic daily news into models automatically. To deal with this deficiency issue, we propose the news distilling network (NDN), a deep learning based framework integrated with collaborative filtering to predict stock trends. This framework ranks and selects the related news articles on the fly by the relevance score from collaborative filtering and fed them into the neural networks.

In addition to the implicit information issue, the sample temporal sequences for training are sparse because most companies are not mentioned daily. Thus the sparsity of training samples complicates



2018/01/19 news:
Sales of PS4 Breaks 6 Million and Beats Switch First Time

	Date	TAIEX	Foxconn	Delta Electronics
Price (NTD)	2018/1/17	11004.80	93.4	139.0
	2018/1/18	11071.57	93.8	143.0
	2018/1/19	11150.85	96.9	144.5
	2018/1/22	11231.46	96.6	149.0
Return (%)	2018/1/18	0.61%	0.43%	2.88%
	2018/1/19	0.72%	3.30%	1.05%
	2018/1/22	0.72%	-0.31%	3.11%
CAR_3			**1.38%**	**5.00%**

Figure 1: Example of implicit stock mentions impacting related companies.

the development of prediction model. By taking implicit information into consideration, i.e., widely leveraging information from daily news, we are able to generate a dense temporal sequence.

The main contributions of this work are threefold: (1) We point out the limitation of market movement predictions based only on the news articles that contain companies' names, called explicit news. (2) We propose a novel model for imitating the real-world decision-making process by creating a dense temporal sequence of news. (3) The experimental results show that the proposed model performs the best on both accuracy and profitability.

2 RELATED WORK

Simulating the decision-making process of the investors has long been an attractive research topic [14]. Conventionally, researchers use a stock market prediction task to evaluate how much information is captured by their models. Some works simulate the decision-making process based on the technical analysis, and focus on mining the trading patterns from structured data such as market price [8, 26]. Some works simulate the process based on fundamental analysis with the textual data such as financial news [11, 17] and earnings call [14, 20, 24]. Although Liou et al. [16] explore the implicit information in the financial news, they do not apply it in the real-world application scenario, i.e., stock movement prediction. In this paper, we aim to enrich the model with the implicit information by simulating the investors' thoughts when reading financial news.

With advances in deep learning, models can induce useful hidden features from raw text and have achieved impressive performance in stock movement prediction. Ding et al. [6] propose an approach to learn dense vector representations for financial events extracted from news text to address the problem of feature sparsity. In addition, neural networks are further used to incorporate external knowledge to enrich event representation [7]. More recently, Hu et al. [11] mine news sequences directly from text with a hierarchical attention mechanism for stock trend prediction. Xu et al. [23] introduce neural variational inference into hierarchical attention networks to better incorporate stochastic factors. Liu et al. [17] consider the title and content of news separately and use a complementary attention mechanism to avoid capturing redundant information from the title and content. The major problem of the previous work is that they focus on the news with explicit stock (company) mentions, yielding a prediction model limited to a worm's eye view of the market environment. In contrast, we propose news distilling networks to utilize the dense temporal news sequence consisting of explicit news and implicitly related news, together with the concept of collaborative filtering [3, 10, 21, 25].

3 METHODOLOGY

3.1 Problem Formulation

We adopt the cumulative abnormal return, which is often used to evaluate the impact of news on a stock price, as our prediction target. Given a target stock s within an n-day event window, the cumulative abnormal return (CAR) is defined as

$$CAR_{s,n} = \sum_{d=1}^{n}(R_{s,d} - \hat{R}_d) \tag{1}$$

where $R_{s,d}$ denotes the return of a specific stock s on day d and $\hat{R}_d$ represents the return on an index such as the Dow Jones or the S&P 500 during the same period. We follow previous works [4, 5, 22] to set n as 3. Furthermore, we treat stock movement prediction as a binary classification problem. RISE and FALL classes are defined based on the polarity of cumulative abnormal return.

3.2 News Distilling Networks

To approximate real-world financial decision-making and take into account the sparsity of training samples, we propose the news distilling networks (NDN). NDN is a neural network-based framework that leverages related information and tames the chaos of unrelated news by combining a collaborative filtering mechanism (CFM) with hierarchical attention networks.

3.2.1 Collaborative Filtering Module. Let M and N denote the numbers of considered stocks and news in the collected dataset, respectively. An interaction between a stock s and a news article γ indicates if the stock is explicitly mentioned or labeled as implicitly related in the news article. A stock-news interaction matrix $A \in \{0, 1\}^{M \times N}$ is then defined as

$$a_{s,\gamma} = \begin{cases} 1 & \text{if interaction}(s, \gamma) \text{ is observed} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Given a target stock and a set of news articles, we seek to measure the relevance between them. The news articles with higher relevance are selected to flow into the next layer. Thus the news distilling problem is treated as the problem of estimating the relevance scores of entries in A, which are then used to rank the news articles. Here we apply generalized matrix factorization (GMF) [10] to learn the stock latent vector $p_s \in \mathbb{R}^{K_1}$ and news latent vector $q_\gamma \in \mathbb{R}^{K_1}$ with embedding size K_1. The estimation of relevance through a projection function consists of an element-wise product operation and a fully-connected layer:

$$\hat{a}_{s,\gamma} = \sigma\left(W_g^T \left(p_s \odot q_\gamma\right) + b_g\right) \tag{3}$$

where $\odot$ denotes the element-wise product of vectors, and W_g and b_g are the weights and bias in the fully-connected layer, respectively. The output score is mapped to the range from 0 to 1 using the sigmoid function $\sigma(x) = 1/(1 + e^{-x})$. The loss function for updating CFM is binary cross-entropy. We uniformly sample negative instances from unobserved interactions and control the sampling ratio w.r.t. the number of observed interactions. In addition, for simplicity, we denote the output $\hat{a}_{s,\gamma}$ as an estimated relevance score $r_{s,\gamma}$ hereafter.

3.2.2 News-level Encoding. To control memory costs, a news article is in terms of sentence-level representations by averaging pre-trained 100-dimensional word vectors for each sentence using FastText [13]. Given a target stock and a set of news articles, a news encoding layer encodes each sentence into a news vector $\tau_{d,\gamma} \in \mathbb{R}^{K_2}$ for the i-th news on day d using bidirectional gated recurrent units (Bi-GRU) [1, 9].

3.2.3 News Article Selection. Given the target stock s, the news representation $\tau_{d,\gamma} \in \mathbb{R}^{k_2}$ for the news article γ in day d is sorted by the relevance score $r_{s,\gamma}$ from the CFM; the top k news representations are allowed to flow to the next layer via a gating mechanism. Specifically, a binary mask m is used to control the gate and zero the representation vector of news ranked lower than the top k. In this work, we set k to 25 according to experimental results. Since the selected news articles may not equally contribute, we introduce an attention layer to focus on informative news.

3.2.4 Dense Temporal Encoding. The news articles published closer to the target date of stock prediction may have more impact than those already published for several days. Thus we model the dense temporal sequence using bi-directional GRUs and capture temporal patterns using an attention layer with a global context vector.

3.2.5 Prediction and Loss Function. Given a hidden representation for the whole news collection within the considered days, a discriminator, a network with a single fully connected layer, is used to generate a binary classification for the stock trend movement. The loss function of the hierarchical attention network is binary cross entropy as follows:

$$\mathcal{L}_{NDN} = -\sum_{\imath=1}^{N} y_\imath \log \hat{y}_\imath + (1 - y_\imath) \log (1 - \hat{y}_\imath) \tag{4}$$

where N is the number of training samples, and $y_\imath$ and $\hat{y}_\imath$ denote the ground truth and predicted movement, respectively.

3.2.6 Training and Inference Procedures. Based on the loss function and the proposed networks, NDN, the overall learning process can be described as Algorithm 1. There are two stages for pre-training CFM and jointly learning with NDN. To begin with, given the existing interaction dataset, we initialize the parameters in CFM by training with both positive and negative samples. Second, for training the main network NDN, we combine the loss from CFM and NDN itself to be $\mathcal{L}_{Joint}$, i.e., $\mathcal{L}_{Joint} = \mathcal{L}_{CFM} + \mathcal{L}_{NDN}$. The integrated loss is then propagated to CFM and NDN for jointly learning. In this way, the CFM is learned to select news articles based on both co-occurrence and content relatedness since the hidden features in NDN is encoded by news content.

As for the inference stage, we could utilize stock labels mapped from sector labels, which can be easily derived from most of the financial websites, to solve the cold start problem caused by unseen news articles and update the CFM within a few epochs.

Algorithm 1 Parameter updating algorithm of NDN

Input: Interaction dataset $\mathcal{V}$ and $\mathcal{V}^-$, negative sampling rate η,
News sequences dataset $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$,
Incoming interaction dataset $\mathcal{V}_{new}$
Collaborative Filtering Module (CFM) C_θ
News Distilling Network $\mathcal{D}_\Phi$
Output: Parameters of the network.

procedure PRE-TRAINING CFM
  while loop until convergence **do**
    Sample $\{B_p\}_{p=1,\ldots,u} \in \mathcal{V}$
    Sample $\{B_n\}_{n=1,\ldots,v} \in \mathcal{V}-$ ▷ $v = \eta u$
    $\hat{a}_i \leftarrow C_\theta(x_i),\ x_i \in B_p \cup B_n,\ i = 1, \ldots, u + v$
    Update θ via $\mathcal{L}_{CFM}$ with Stochastic Gradient Descent
  return θ
procedure JOINT LEARNING PROCESS FOR NDN
  while loop until convergence **do**
    Sample $\{B_s\}_{s=1,\ldots,N} \in \mathcal{X}$
    Sample $\{B_p\} \in \mathcal{V}_{new}$ with (news, stock) pairs in $\{B_s\}$
    $\hat{a}_i \leftarrow C_\theta(x_i),\ x_i \in B_p$
    Compute $\mathcal{L}_{CFM}$
    $\hat{y}_\imath \leftarrow \mathcal{D}_\Phi(x_\imath),\ x_\imath \in B_s$
    Compute $\mathcal{L}_{NDN}$
    $\mathcal{L}_{Joint} \leftarrow \mathcal{L}_{NDN} + \mathcal{L}_{CFM}$
    Update θ, Φ via $\mathcal{L}_{Joint}$ with SGD
  return θ and Φ

Table 1: Number of instances in the dataset. RISE and FALL denote the price movement.

	Train	Test
Period	2013/06/22-2018/01/01	2018/01/01-2018/06/20
RISE	29,067	2,906
FALL	26,778	2,678
Total	55,845	5,584

4 EXPERIMENTS

4.1 Dataset

We collected financial news articles and the manual labels from a well-known newsvendor[1] in Taiwan from 2013/06/22 to 2018/06/20 along with the stock price data and *Taiwan Stock Exchange Weighted Index* (TAIEX) data for the same period. Note that all annotations are provided by professional journalists when the articles are published. We considered 1,566 stocks and 19 sectors in the market. The statistics of training and test data are shown in Table 1.

4.2 Experimental Setup

We compare the performances of the above models with the following baseline models: (1) randomly guesses (Random Guess), (2) Bag-of-words (BoW) + Random Forest [12], (3) FastText + Random Forest [13, 19], and (4) Hybrid Attention Network (HAN) [11]. Because the previous works do not consider implicit information, only explicit news articles are used for training these models. We updated the networks with a batch size of 64 using the Adam optimizer [15]

[1] https://www.moneydj.com/

Table 2: Experimental results. Acc. denotes accuracy. The asterisk (*) denotes models that significantly outperform the HAN_{S1} at $p < 0.05$ (using McNemar's test).

Model	Acc. (%)	MCC
Random Guess	50.77	0.0147
BoW + Random Forest	50.97	0.0159
FastText + Random Forest	52.83	0.0485
HAN	54.37	0.0788
HAN_{s1}	53.85	0.0736
NDN_{s1} (Proposed)	56.75*	0.1302*
HAN_{s2}	56.77*	0.1300*
NDN_{s2} (Proposed)	**57.89***	**0.1536***

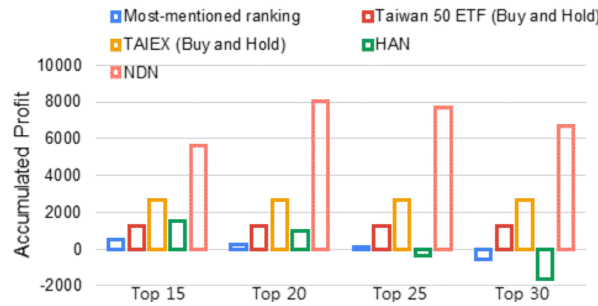


Figure 2: Market simulation with various k-best portfolio.

with a learning rate of 0.001. Furthermore, in our CFM, both stock and news embeddings were projected into a 50-dimensional latent space, i.e., K_1=50.

In order to show the usefulness of implicit information, we explore the following settings in the proposed methods and HAN model.

- **Setting 1 ($S1$)**: We fused all news regardless of the existence of manual stock labels to evaluate the robustness of the selection and attention mechanism.
- **Setting 2 ($S2$)**: We integrate manually labeled implicitly related news with explicit news to determine whether information from related news aids the prediction task. For NDN, all kinds of news were fed into the model. The results under this setting can be considered as the upper-bound of the models with implicit information.

4.3 Model Evaluation

In Table 2, we list the performances of the baseline models and the proposed model. Firstly, HAN, which only uses explicit news, performs the best among the baseline models, and the proposed model (NDN_{S1}) significantly outperforms all baseline models. Secondly, when under Setting 1, the performance of HAN drops slightly. That means the extra news articles cause the noise to the HAN model. However, the proposed model under the same setting performs well. That means the proposed NDN model can distilling useful information from the given articles. Thirdly, the experimental results under Setting 2 (HAN_{S2} and NDN_{S2} indicate that the proposed model achieves a higher upper-bound in stock movement prediction when using manually-labeled information.

Apart from the accuracy and Matthews correlation coefficient (MCC), testing the models on historical data is also an important factor for evaluating the performance of models in the real-world market. We further use the same testing period for an experiment. We follow the portfolio construction strategies used in previous works [6, 11] in the daily frequency to mimic the trader's behavior. At the beginning of each trading session, our model calculates a prediction score given a news sequence for each considered stock (1,566 stocks here). This score indicates the *RISE* probability relative to the whole market index (TAIEX here). We then rank the stocks based on this score and select the top k with which we construct the portfolio for the following trading day. We evenly invest in the selected stocks according to their *OpenPrices*. After purchasing, the stocks are held until we can sell them for a profit of 2%. Otherwise, the stocks are sold at market prices at the end of the day. Also, to approximate real-world trading practices, a transaction cost of 0.3% for each trading is included. Note that we purchase $10,000 worth of stocks equally for different trading portfolios.

To compare the performance of the proposed model, we not only consider HAN as baseline but also use *buy and hold* strategies for comparison. There are three *buy and hold* strategies, (1) we long the most-mentioned stock, i.e., the portfolio construction is based on the number of news items mentioning a specific stock, (2) we long Taiwan 50 ETF, i.e., representative ETF, and (3) we long TAIEX, i.e., stock market index. The simulation results are shown in Figure 2. The NDN achieves the best performance with accumulated profits of $8,053 when k is 20. Moreover, since the transaction cost increases once we purchase more stocks, the profit does not cover the costs and begins to drop when k is 25 and 30. HAN and the naive baseline (1) thus begin losing money when k is 25. When k is larger, NDN tend to converge on a similar profit since their portfolios become similar. Among the naive baselines, Taiwan 50 ETF with our trading strategy even outperform the previous method, HAN. In conclusion, this back-testing result is consistent with the classification results discussed in the previous section: the proposed methods make the best profit in comparison with the previous method.

5 CONCLUSIONS

In this work, we discuss the problem of sparse temporal news sequences as well as the shortcoming of previous works that are based on explicit news only. We propose a method to create dense temporal news sequences by extensively leveraging implicit information. To discover implicitly related news given limited labels, we develop news distilling networks. The experiments on the real-world data show that the proposed methods significantly improve the accuracy of stock movement prediction and yield greater profits than the previous methods.

ACKNOWLEDGMENTS

This research was partially supported by Ministry of Science and Technology, Taiwan, under grants MOST 109-2218-E-009-014, MOST 109-2634-F-002-040, MOST 109-2634-F-002-034, and MOST 110-2634-F-002-028.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[2] Ananth Balashankar, Sunandan Chakraborty, Samuel Fraiberger, and Lakshminarayanan Subramanian. 2019. Identifying Predictive Causal Factors from News Streams. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 2338–2348. https://doi.org/10.18653/v1/D19-1238

[3] Immanuel Bayer, Xiangnan He, Bhargav Kanagal, and Steffen Rendle. 2017. A Generic Coordinate Descent Framework for Learning from Implicit Feedback. In *Proceedings of the 26th International Conference on World Wide Web* (Perth, Australia) *(WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1341–1350. https://doi.org/10.1145/3038912.3052694

[4] Ching-Yun Chang, Yue Zhang, Zhiyang Teng, Zahn Bozanic, and Bin Ke. 2016. Measuring the Information Content of Financial News. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. The COLING 2016 Organizing Committee, Osaka, Japan, 3216–3225. https://www.aclweb.org/anthology/C16-1303

[5] Angela K Davis, Jeremy M Piger, and Lisa M Sedor. 2012. Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research* 29, 3 (2012), 845–868.

[6] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*. 2327–2333.

[7] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2016. Knowledge-Driven Event Embedding for Stock Prediction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. The COLING 2016 Organizing Committee, Osaka, Japan, 2133–2142. https://www.aclweb.org/anthology/C16-1201

[8] Fuli Feng, Huimin Chen, Xiangnan He, Ji Ding, Maosong Sun, and Tat-Seng Chua. 2019. Enhancing stock movement prediction with adversarial training. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 5843–5849. https://doi.org/10.24963/ijcai.2019/810

[9] Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* 18, 5-6 (2005), 602–610.

[10] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *Proceedings of the 26th International Conference on World Wide Web* (Perth, Australia) *(WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 173–182. https://doi.org/10.1145/3038912.3052569

[11] Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. 2018. Listening to Chaotic Whispers: A Deep Learning Framework for News-Oriented Stock Trend Prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA) *(WSDM '18)*. Association for Computing Machinery, New York, NY, USA, 261–269. https://doi.org/10.1145/3159652.3159690

[12] Thorsten Joachims. 1998. Text categorization with Support Vector Machines: Learning with many relevant features. In *Machine Learning: ECML-98*, Claire Nédellec and Céline Rouveirol (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 137–142.

[13] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In *EACL*. Association for Computational Linguistics, Valencia, Spain, 427–431. http://aclweb.org/anthology/E17-2068

[14] Katherine Keith and Amanda Stent. 2019. Modeling Financial Analysts' Decision Making via the Pragmatics and Semantics of Earnings Calls. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 493–503. https://doi.org/10.18653/v1/P19-1047

[15] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014).

[16] Yi-Ting Liou, Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021. Dynamic Graph Transformer for Implicit Tag Recognition. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Association for Computational Linguistics, Online, 1426–1431. https://www.aclweb.org/anthology/2021.eacl-main.122

[17] Qikai Liu, Xiang Cheng, Sen Su, and Shuguang Zhu. 2018. Hierarchical Complementary Attention Network for Predicting Stock Price Movements with News. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) *(CIKM '18)*. Association for Computing Machinery, New York, NY, USA, 1603–1606. https://doi.org/10.1145/3269206.3269286

[18] Ye Ma, Lu Zong, Yikang Yang, and Jionglong Su. 2019. News2vec: News Network Embedding with Subnode Information. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 4843–4852. https://doi.org/10.18653/v1/D19-1490

[19] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi. 2016. Sentiment analysis of Twitter data for predicting stock market movements. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*. 1345–1350.

[20] Yu Qin and Yi Yang. 2019. What You Say and How You Say It Matters: Predicting Stock Volatility Using Verbal and Vocal Cues. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 390–401. https://doi.org/10.18653/v1/P19-1038

[21] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-Based Collaborative Filtering Recommendation Algorithms. In *Proceedings of the 10th International Conference on World Wide Web* (Hong Kong, Hong Kong) *(WWW '01)*. Association for Computing Machinery, New York, NY, USA, 285–295. https://doi.org/10.1145/371920.372071

[22] Paul C Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance* 62, 3 (2007), 1139–1168.

[23] Yumo Xu and Shay B. Cohen. 2018. Stock Movement Prediction from Tweets and Historical Prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, 1970–1979. https://doi.org/10.18653/v1/P18-1183

[24] Linyi Yang, Tin Lok James Ng, Barry Smyth, and Riuhai Dong. 2020. HTML: Hierarchical Transformer-Based Multi-Task Learning for Volatility Prediction. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 441–451. https://doi.org/10.1145/3366423.3380128

[25] Hanwang Zhang, Fumin Shen, Wei Liu, Xiangnan He, Huanbo Luan, and Tat-Seng Chua. 2016. Discrete Collaborative Filtering. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Pisa, Italy) *(SIGIR '16)*. Association for Computing Machinery, New York, NY, USA, 325–334. https://doi.org/10.1145/2911451.2911502

[26] Liheng Zhang, Charu Aggarwal, and Guo-Jun Qi. 2017. Stock Price Prediction via Discovering Multi-Frequency Trading Patterns. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS, Canada) *(KDD '17)*. Association for Computing Machinery, New York, NY, USA, 2141–2149. https://doi.org/10.1145/3097983.3098117