Stock Market Movement Prediction by Gated Hierarchical Encoder

Peibin Chen and Ying Tan(✉)

Laboratory of Machine Perception (MOE), Department of Machine Intelligence,
School of Electronics Engineering and Computer Science,
Peking University, Beijing 100871, China
ytan@pku.edu.cn

Abstract. Stock movement prediction is an important but challenging topic in the stock market. Previous methods mainly focus on predicting up or down of one stock, ignoring the significant up and down of the whole stock market that is more related to the final return. In this paper, a novel framework Gated Hierarchical Encoder (GHE) is proposed, which consists of two components: hierarchical feature learning and dynamic gate. Hierarchical feature learning helps the model do prediction from coarse to fine, while dynamic gate dynamically ensembles results from different branches. Experiments show that compared with MLP and LSTM, GHE achieves higher return on multiple stock markets, and predicts more accurately on the significant up and down.

Keywords: Quantitative investment · End-to-end · Classification · Stock market movement prediction · Time series · Deep learning

1 Introduction

Financial activities play a key role in the modern economy, of which stock market is an important part. Stock market movement prediction is conducive to better macroeconomic strategies and monetary policies, and also conducive to more reasonable resource allocation and less investment losses. For the reason that the price of a stock is unpredictable [17,21], researchers mainly focus on predicting the stock price movement.

Stock data is highly chaotic, random and noisy, making it hard to generalize well with traditional machine learning methods, such as SVM [3] and ARIMA [1]. Recently, due to the great success of deep neural network [9,10,19], more and more researchers bring in the technology to stock movement prediction [8,14–16]. Some methods [5] apply adversarial learning to enhance the robustness of the model. Some methods [6] use multiple sources, like news data and stock time series data, to jointly predict the movement. And some methods [11] temp to explore the relationship between stocks. However, these methods are limited in the one-stock-binary-classification framework, failed to warn for the significant up and down of the stock market.

Y. Tan and Y. Shi (Eds.): ICSI 2021, LNCS 12690, pp. 511–521, 2021.
https://doi.org/10.1007/978-3-030-78811-7_48

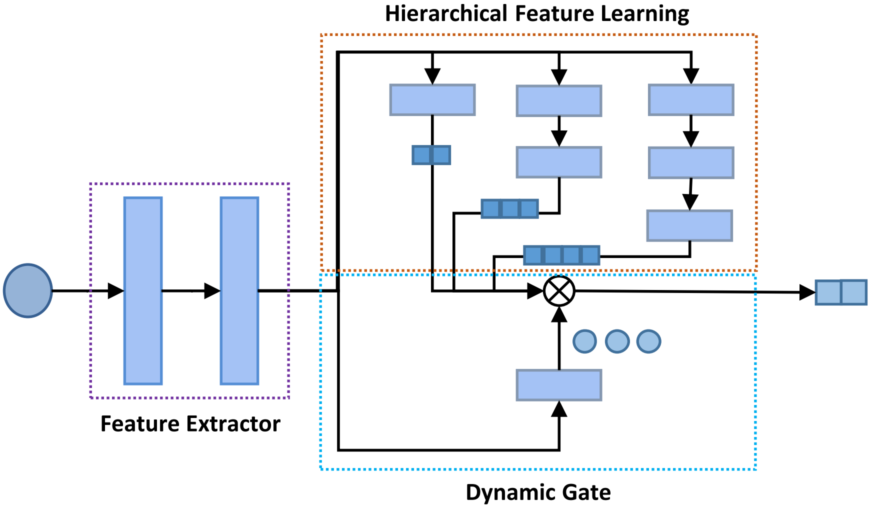


Fig. 1. A brief description of gated hierarchical encoder.

In this paper, the gated hierarchical encoder (GHE) is proposed, with which model is able to predict the up and down as well as the significant up and down. Different from previous methods [5,6], GHE decomposes the task of binary classification into multiple tasks, and splits the top of the model into multiple branches. Each task is still a classification, but the number of classes is positively correlated to the depth of the corresponding branch. Furthermore, a dynamic gate is introduced to ensemble results for different branches according to the stock input. Experiments on China A-shares market and U.S. stock market demonstrate that GHE surpasses baseline methods by an obvious margin, and is able to accurately predict the significant up and down. Ablation study proves each part of GHE contributes to the final strong performance.

The contributions of this paper are summarized as:

- Gated hierarchical encoder is proposed to accurately predict the significant up and down of the market.
- A dynamic gate is proposed to dynamically ensemble results from different branches in GHE according to the stock input.
- Stock indexes are used as input rather than multiple stocks, making it easier to explore the characteristics of the whole stock market.

2 Related Work

Stock movement prediction is a highly challenging topic for academia and industry. Recently, due to the great success of deep learning, it has been possible to extract useful information from stock data automatically for final prediction.

Researches on stock prediction can be divided into two main directions: market-historic-data based analysis and outside market-historic-data based analysis.

Market-historic-data based analysis deals with historical market data, such as opening price and stock volumes. However, this kind of data is highly chaotic, random and noisy. Simply trying to fit the data can easily cause overfitting. Adv-ALSTM [5] introduces adversarial learning in the feature space, forcing the model to be robust against the noise. LSTM-RGCN [11] argues that the movement of one stock has connections with that of others. With this assumption, the authors model the stock relation with a graph neural network, with stocks as nodes and the relation between each other as edges. HMGTF [4] introduce the powerful Transformer [20] that has been widely used in NLP. The method enhances the locality, avoids redundant heads of the Transformer, and utilizes the modified Transformer to learn hierarchical features. With the model, extremely long-term dependencies are mined from financial time series.

Although historical market data is widely used, it is unable to reflect the current market trend in time. Based on this observation, many researchers bring in the outside market-historic-data and build models to predict the emotional signals from the data. HAN [6] uses a multi-level sequence model to extract information, and further uses an attention model to obtain the final signals from the news. DP-LSTM [12] extracts hidden information from the news and integrates multiple news sources through the differential privacy mechanism. CapTE [13] effectively encodes the rich semantics and relation for a certain stock with the powerful Transformer [20] and the capsule network [18].

Different from the methods above, this paper focuses on predicting the significant up and down of the whole stock market, rather than the up and down of a single stock. For this purpose, Gated Hierarchical Encoder is proposed, which significantly outperforms MLP and LSTM on China A-shares market and U.S. stock market.

3 Gated Hierarchical Encoder

Previous methods prefer to consider the stock market movement prediction as a single stock binary classification problem. This brings three problems: (i) Ignoring the change percentage that investors concern about. (ii) Ignoring the connection between China A-shares market and overseas markets. (iii) Single stock prediction is unable to represent the movement of the whole market. Based on these observations, Gated Hierarchical Encoder is proposed, which is formalized by three components: multi indexes data preprocessing, hierarchical feature learning and dynamic gate for ensembling. The details of the model are shown in Fig. 1.

3.1 Multi-indexes Data Preprocessing

Since stock indexes are able to indicate the movement of the stock market, multiple stock indexes are used as input rather than stocks. In order to introduce external information, stock indexes from overseas markets are also used. However, different kind of features and different markets have different scales, making it hard to catch the key clues for prediction. Inspired by the normalization operation in computer vision [7], we normalize the indexes along the data dimension. Assuming the training set $\mathcal{D}$ is consist of N_c China A-shares stock indexes examples x_i^c and N_o overseas markets indexes examples x_i^o, the normalization can be written as

$$\mu_{\mathcal{B}}^l = \frac{1}{N_l}\sum_{i=1}^{N_j} x_i^l, \tag{1}$$

$$\sigma_{\mathcal{B}}^l = \sqrt{\frac{1}{N_l}\sum_{i=1}^{N_l}\left(x_i^l - \mu_{\mathcal{B}}^l\right)^2 + \epsilon}, \tag{2}$$

$$x_i^l = \frac{x_i^l - \mu_{\mathcal{B}}^l}{\sigma_{\mathcal{B}}^l}, \tag{3}$$

where $l \in \{c, o\}$.

3.2 Hierarchical Feature Learning

Previous methods ignore the change percentage that investors concern about. But directly predicting the change percentage is impossible because of the insufficient training data [2,17,21]. A novel framework is proposed for hierarchical feature learning, with which model is able to do prediction from coarse to fine.

Formally, given an example x_i, the feature extractor H outputs the embedding of x_i, after which the embedding is passed through K branches $B_1 \dots B_K$. With deeper the layer, the branches are more complex and are able to encode richer features. Each branch is assigned a classification task whose difficulty depends on the depth of branch. Besides, we assume that the task at lower layer can be solved if the task at higher layer is perfectly solved. For example, B_1 temps to predict the market up or down, and B_K is assigned a more complex task, predicting if market will go significant up, slight up, slight down or significant down. Since different branches share the same feature extractor, and the tasks between branches are highly correlative, the model is expected to learn better with the hierarchical supervise. Defining the classification layer after B_k as F_k, the model forward process can be defined by the following equations,

$$h_i = H(x_i), \tag{4}$$

$$b_i^k = B_k(h_i) \quad \text{for } k \text{ in } 1 \dots K, \tag{5}$$

$$f_i^k = F_k(b_i^k) \quad \text{for } k \text{ in } 1 \dots K, \tag{6}$$

$$p_{ij}^k = \frac{exp(f_{ij}^k)}{\sum_{j=1}^{Cls} exp(f_{ij}^k)} \quad \text{for } k \text{ in } 1 \dots K, \tag{7}$$

where p_{ij}^k represents the output score of class j from branch k for x_i.

Table 1. Comparison on China A-shares market and U.S. stock market. The results are shown as Acc/R.

Market	Methods	MLP	LSTM	Ours
China A-shares market	CSI300	0.592/1.645	0.558/1.752	**0.596/2.037**
	CSI500	0.539/1.387	0.562/1.739	**0.589/2.016**
	CYB	0.552/2.215	0.555/2.686	**0.573/2.917**
	SH	0.544/1.406	0.536/1.461	**0.608/1.691**
	SH50	0.542/1.381	0.546/1.561	**0.613/1.837**
	ZXB	0.546/1.502	0.555/2.248	**0.597/2.485**
U.S. stock market	IXIC	0.563/1.325	**0.587**/1.838	0.561/**2.076**
	SPX	0.523/0.924	**0.578**/1.623	0.558/**1.776**

3.3 Dynamic Gate for Ensembling

Due to the high correlation between tasks of branches, it is possible to ensemble results from different branches to enhance the performance of the model. Different from simply averaging the results, an ensembling gate is proposed to dynamically decide the weight of different branches. The output of the gate is depended on the input data, thus is more flexible and intelligent. Formally, give h_i from the feature extractor, the gate G encodes the feature and outputs K scores, with which the output of K branches are weighted summed to get f_i^e,

$$w_i = G(h_i), \tag{8}$$

$$f_i^e = \sum_{k=1}^{K} M(f_i^k) \times w_i^k, \tag{9}$$

$$p_{ij}^e = \frac{exp(f_{ij}^e)}{\sum_{j=1}^{Cls} exp(f_{ij}^e)}, \tag{10}$$

where p_{ij}^e represents the dynamically ensembling score of class j for x_i.

3.4 Loss Functions

Given multiple stock indexes, Gated Hierarchical Encoder outputs K results for K tasks, and the dynamic gate further ensembles K outputs to get a better prediction. We use cross entropy loss to guide the learning of the model. Furthermore, in order to encourage different branches to be close to the ensembling result, Kullback-Leibler divergence loss is added in all branches. The final loss function is

$$L = w_1 \times L_{cls} + w_2 \times L_{ens} + w_3 \times L_{kl}, \tag{11}$$

where w_1, w_2, w_3 are hyperparameters, and L_{cls}, L_{ens} and L_{kl} are defined as

$$L_{cls} = -\frac{1}{(N_c + N_o) \times K} \sum_{i=1}^{N_c+N_o} \sum_{k=1}^{K} y_i^k \log p_i^k, \tag{12}$$

Table 2. Ablation study on CSI300 and IXIC. The results are shown as Acc/R.

Model	CSI300	IXIC
LSTM	0.558/1.752	0.587/1.838
LSTM+HF	0.573/1.878	0.634/2.039
LSTM +HF + DG (Ours)	0.596/2.037	0.561/2.076

$$L_{ens} = -\frac{1}{(N_c + N_o)} \sum_{i=1}^{N_c+N_o} y_i^e \log p_i^e, \tag{13}$$

$$L_{kl} = \frac{1}{(N_c + N_o) \times K} \sum_{i=1}^{N_c+N_o} \sum_{k=1}^{K} KL(p_i^k || p_i^e), \tag{14}$$

where $KL(\cdot)$ represents the Kullback-Leibler divergence loss.

4 Experiments

In this section, a set of experiments are conducted to verify the ability of Gated Hierarchical Encoder. Firstly, our method is compared with traditional methods on multiple indexes from China A-shares market and U.S. stock market, showing the advantages of our method. Secondly, the dynamic gate is removed from the model for proving the effectiveness of the gate. Finally, the output of the gate is visualized to observe how the gate ensembles branches on different indexes.

4.1 Dataset

We select 6 indexes, including CSI300, CSI500, CYB, SH, SH50 and ZXB, in China A-shares market and 2 indexes, including IXIC and SPX, in U.S. stock market. All data are from 2005-01-01 to 2020-12-14 and split into three parts: training set (from 2005-01-01 to 2017-12-31), validation set (from 2018-01-01 to 2018-12-31) and test set (from 2019-01-01 to 2020-12-14). Accuracy at significant up/down Acc_S and cumulative return R_S of index S are defined as

$$\mathrm{Acc}_S = \frac{\sum_{i \in S} \mathcal{I}\left(y_i = \tilde{y}_i\right) \times \mathcal{I}(abs(chg_i) > 0.5)}{\sum_{i \in S} \mathcal{I}(abs(chg_i) > 0.5)}, \tag{15}$$

$$\mathrm{R}_S = \sum_{i \in S} (1 + \mathcal{I}(y_i = \tilde{y}_i) \times chg_i), \tag{16}$$

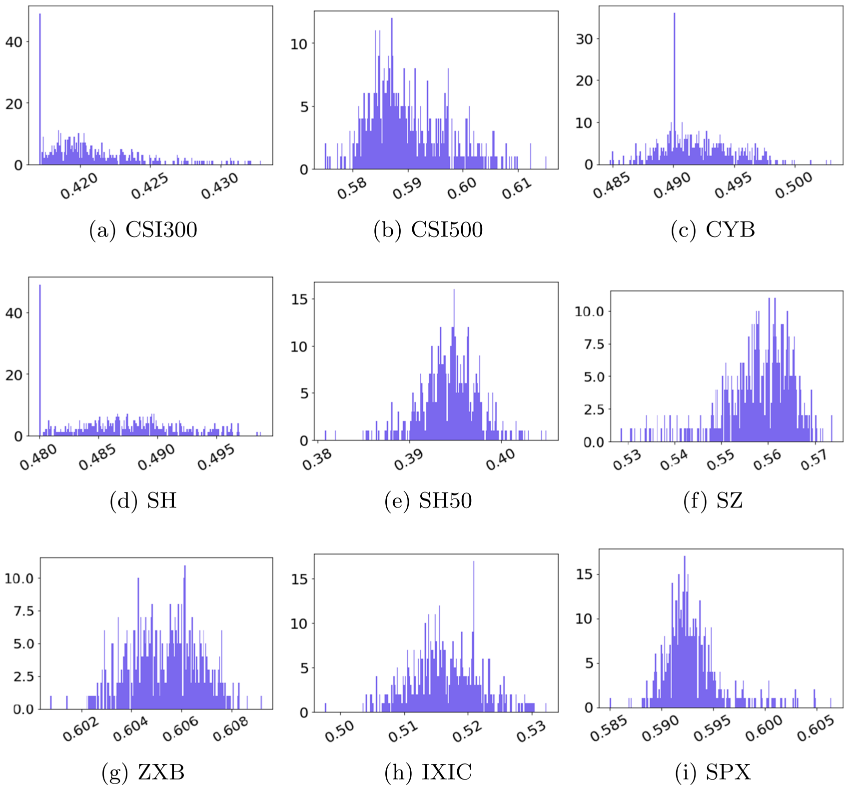
where $\mathcal{I}(.)$ is a indicator function that takes on a value of 1 if its augment is true, and 0 otherwise. chg_i is the change percentage of x_i.

4.2 Comparisons

Gated Hierarchical Encoder is compared with MLP and LSTM to show the strength of Gated Hierarchical Encoder. MLP consists of simple interconnected neurons or nodes, learning the nonlinear mapping between input and output vectors. LSTM is a special type of recurrent neural networks which are able to learn long-term dependencies, especially in sequence prediction problems. For fairness, all methods use the same features as input, and are tested using the model whose performance is the best in the validation set. Table 1 shows the results. From Table 1, it is obvious that our method achieves the best results in most of the indexes. The cumulative return earned by different methods in the test set are shown in Fig. 3 and Fig. 4. Gated Hierachical Encoder is able to predict the time of high risk or high returns, thus getting more returns on the stock markets.

4.3 Ablation Study

To clarify the effectiveness of hierarchical feature learning and dynamic gate, we measure the effect of (i) LSTM: Original LSTM model, (ii)LSTM + HF: Adding hierarchical feature learning to LSTM and (iii) LSTM +HF + DG: adding dynamic gate and hierarchical feature learning to LSTM. Experiments are carried out on CSI300 and IXIC. As shown in Table 2, each component improves the origin LSTM and boosts the returns by an obvious margin.

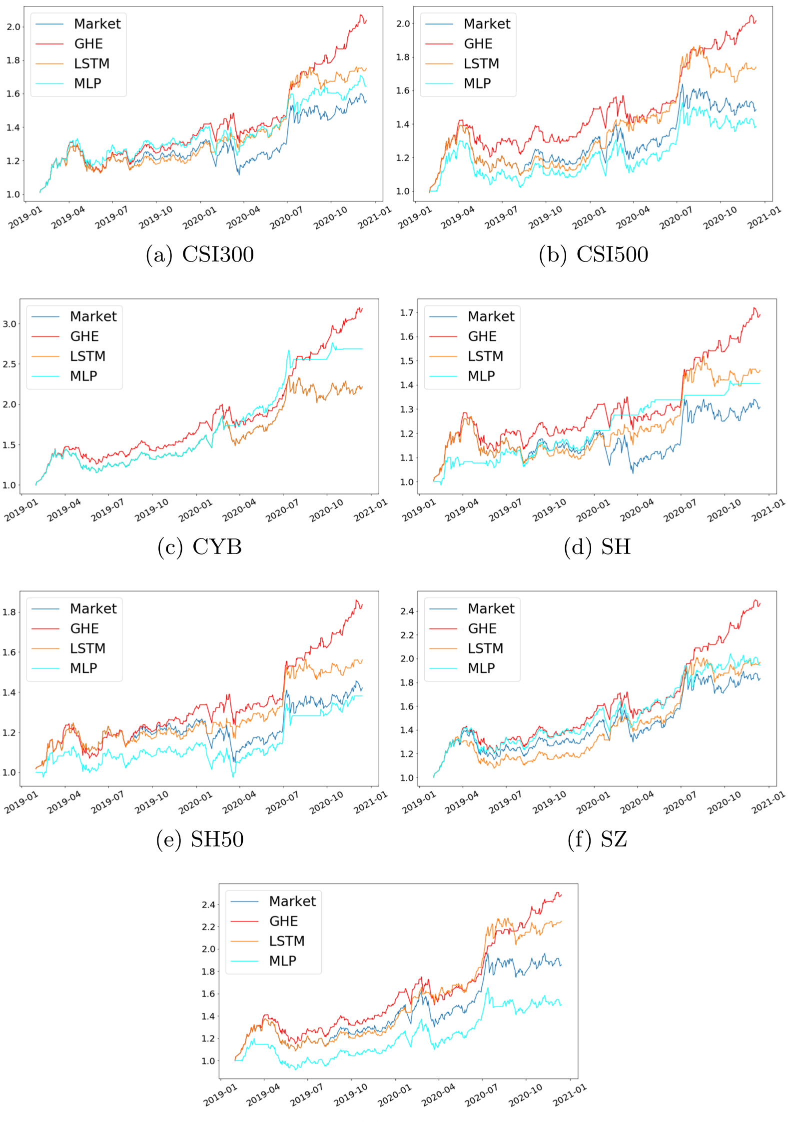


(a) CSI300 (b) CSI500 (c) CYB

(d) SH (e) SH50 (f) SZ

(g) ZXB (h) IXIC (i) SPX

Fig. 2. Visualizations of the dynamic gate for multiple indexes.



(a) CSI300

(b) CSI500

(c) CYB

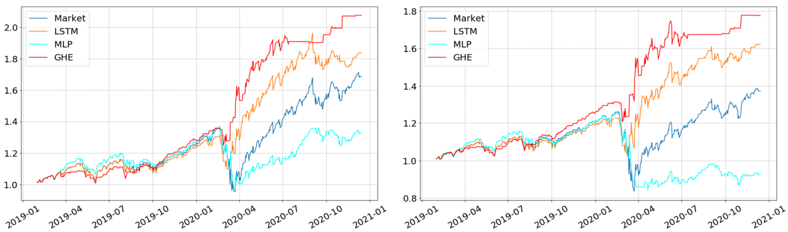
(d) SH

(e) SH50

(f) SZ

(g) ZXB

Fig. 3. Comparison of different methods on China A-shares market.



(a) IXIC (b) SPX

Fig. 4. Comparison of different methods on U.S. stock market.

4.4 Visualizations

Dynamic gate is designed to ensemble multiple branches with dynamic weights. In order to show whether the gate is able to output different scores for different indexes, we count the outputs of dynamic gate and plot the histogram as shown in Fig. 2. For CSI300, CYB, SH and SH50, dynamic gate gives more weights on the multiple classification branch. For the others, dynamic gate gives more weights on the binary classification branch. Therefore, the gate is proved to have the ability to ensemble branches according to different situations.

5 Conclusion

In this paper, Gated Hierarchical Encoder is proposed to predict the significant up and down of the stock market. Stock indexes are used as input for the reason that stock indexes are able to indicate the movement of stock market. Then, the stock movement prediction are divided into multiple highly related classification tasks via hierarchical feature learning. To further boost the performance of the model, a dynamic gate is introduced to ensemble different results from different branches. The ensembling weights are dynamically determined according to the input index. Experiments on China A-shares market and U.S. stock market show that GHE surpasses the baselines by an obvious margin.

Acknowledgements. This work is supported by Science and Technology Innovation 2030 - "New Generation Artificial Intelligence" Major Project (Grant Nos.: 2018AAA0102301, 2018AAA0100302), and the National Natural Science Foundation of China (Grant No. 62076010).

References

1. Ariyo, A.A., Adewumi, A.O., Ayo, C.K.: Stock price prediction using the ARIMA model. In: Al-Dabass, D., Orsoni, A., Cant, R.J., Yunus, J., Ibrahim, Z., Saad, I. (eds.) UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014, Cambridge, United Kingdom, 26–28 March 2014, pp. 106–112. IEEE (2014). https://doi.org/10.1109/UKSim.2014.67

2. Cogswell, M., Ahmed, F., Girshick, R.B., Zitnick, L., Batra, D.: Reducing overfitting in deep networks by decorrelating representations. In: Bengio, Y., LeCun, Y. (eds.) 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, 2–4 May 2016. Conference Track Proceedings (2016). http://arxiv.org/abs/1511.06068
3. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995). https://doi.org/10.1007/BF00994018
4. Ding, Q., Wu, S., Sun, H., Guo, J., Guo, J.: Hierarchical multi-scale gaussian transformer for stock movement prediction. In: Bessiere, C. (ed.) Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, pp. 4640–4646. ijcai.org (2020). https://doi.org/10.24963/ijcai.2020/640
5. Feng, F., Chen, H., He, X., Ding, J., Sun, M., Chua, T.: Enhancing stock movement prediction with adversarial training. In: Kraus, S. (ed.) Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, 10–16 August 2019, pp. 5843–5849. ijcai.org (2019). https://doi.org/10.24963/ijcai.2019/810
6. Hu, Z., Liu, W., Bian, J., Liu, X., Liu, T.: Listening to chaotic whispers: a deep learning framework for news-oriented stock trend prediction. In: Chang, Y., Zhai, C., Liu, Y., Maarek, Y. (eds.) Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, 5–9 February 2018, pp. 261–269. ACM (2018). https://doi.org/10.1145/3159652.3159690
7. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Bach, F.R., Blei, D.M. (eds.) Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015. JMLR Workshop and Conference Proceedings, vol. 37, pp. 448–456. JMLR.org (2015). http://proceedings.mlr.press/v37/ioffe15.html
8. Kim, R., So, C.H., Jeong, M., Lee, S., Kim, J., Kang, J.: HATS: a hierarchical graph attention network for stock movement prediction. CoRR abs/1908.07999 (2019). http://arxiv.org/abs/1908.07999
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Bartlett, P.L., Pereira, F.C.N., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a Meeting Held Lake Tahoe, Nevada, United States, 3–6 December 2012, pp. 1106–1114 (2012). http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks
10. LeCun, Y., Bengio, Y., Hinton, G.E.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539
11. Li, W., Bao, R., Harimoto, K., Chen, D., Xu, J., Su, Q.: Modeling the stock relation with graph network for overnight stock movement prediction. In: Bessiere, C. (ed.) Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, pp. 4541–4547. ijcai.org (2020). https://doi.org/10.24963/ijcai.2020/626
12. Li, X., Li, Y., Yang, H., Yang, L., Liu, X.: DP-LSTM: differential privacy-inspired LSTM for stock prediction using financial news. CoRR abs/1912.10806 (2019). http://arxiv.org/abs/1912.10806
13. Liu, J., et al.: Transformer-based capsule network for stock movement prediction. In: Proceedings of the First Workshop on Financial Technology and Natural Language Processing, pp. 66–73 (2019)

14. Long, W., Lu, Z., Cui, L.: Deep learning-based feature engineering for stock price movement prediction. Knowl. Based Syst. **164**, 163–173 (2019). https://doi.org/10.1016/j.knosys.2018.10.034
15. Ma, T., Tan, Y.: Multiple stock time series jointly forecasting with multi-task learning. In: 2020 International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, United Kingdom, 19–24 July 2020, pp. 1–8. IEEE (2020). https://doi.org/10.1109/IJCNN48605.2020.9207543
16. Nelson, D.M.Q., Pereira, A.C.M., de Oliveira, R.A.: Stock market's price movement prediction with LSTM neural networks. In: 2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA, 14–19 May 2017, pp. 1419–1426. IEEE (2017). https://doi.org/10.1109/IJCNN.2017.7966019
17. Nguyen, T.H., Shirai, K., Velcin, J.: Sentiment analysis on social media for stock movement prediction. Expert Syst. Appl. **42**(24), 9603–9611 (2015). https://doi.org/10.1016/j.eswa.2015.07.052
18. Sabour, S., Frosst, N., Hinton, G.E.: Dynamic routing between capsules. In: Guyon, I., et al. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017, pp. 3856–3866 (2017). http://papers.nips.cc/paper/6975-dynamic-routing-between-capsules
19. Szegedy, C., et al.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, 7–12 June 2015, pp. 1–9. IEEE Computer Society (2015). https://doi.org/10.1109/CVPR.2015.7298594
20. Vaswani, A., et al.: Attention is all you need. In: Guyon, I., et al. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017, pp. 5998–6008 (2017). http://papers.nips.cc/paper/7181-attention-is-all-you-need
21. Walczak, S.: An empirical analysis of data requirements for financial forecasting with neural networks. J. Manag. Inf. Syst. 17(4), 203–222 (2001). http://www.jmis-web.org/articles/997