Fusing Multi-Granularity Data for Stock Trend Prediction with Contrastive Pre-training

Haonan Xu
xuhaonan2020@ia.ac.cn
Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, University of Chinese Academy of Sciences
Haidian Qu, Beijing Shi, China

Jiange Li
lijiange@csc.com.cn
Department of Information Technology
China Securities Co., Ltd.
Beijing Shi, China

Peng Wang
wangpengfund@csc.com.cn
Department of Index and Quantitative Investment
China Securities Co., Ltd.
Beijing Shi, China

XianChen Yin
yinxianchen@csc.com.cn
Department of Information Technology
China Securities Co., Ltd.
Beijing Shi, China

Wenfang Xue*
wenfang.xue@ia.ac.cn
Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, University of Chinese Academy of Sciences
Haidian Qu, Beijing Shi, China

ABSTRACT

The advances in financial theory and power of computer technology have driven the development of quantitative investing. Stock trend forecasting as an important topic in quantitative investment has attracted more and more researchers. The development of deep learning has also led to the emergence of many interesting models in the field. However, most of the existing models only use daily frequency data to predict the stock price fluctuations, lacking the consideration of high-frequency data and relational data. At the same time, it remains a challenge to incorporate multi-granularity data into model training. Based on this, we propose a contrastive pre-training model fusing high frequency data and daily frequency data(HRC), which construct pre-training tasks with a positive and negative sample design in terms of data granularity and stock relationships. In addition to this, we also constructed a label-based training task using the performance of returns to classify stocks. The experiments on real-world stock dataset show significant improvements of our approach over the baselines.

*To whom correspondence should be addressed.

ICMLC 2023, February 17–20, 2023, Zhuhai, China

ACM ISBN 978-1-4503-9841-1/23/02.
https://doi.org/10.1145/3587716.3587814

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; • **Information systems** → **Data mining**.

KEYWORDS

stock trend forecasting, contrastive learning, pre-training

ACM Reference Format:
Haonan Xu, Jiange Li, Peng Wang, XianChen Yin, and Wenfang Xue. 2023. Fusing Multi-Granularity Data for Stock Trend Prediction with Contrastive Pre-training . In *2023 15th International Conference on Machine Learning and Computing (ICMLC 2023), February 17–20, 2023, Zhuhai, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3587716.3587814

1 INTRODUCTION

Quantitative investment is an investment model based on a large amount of historical data and quantitative modeling for the purpose of obtaining stable returns[5]. With the improvement of people's living standard, people's investment style is also changing dramatically, and more and more people start to pay attention to and participate in the stock market investment[9, 12, 21]. The stock market is affected by many factors and the price is unpredictable. The study of stock price prediction is of great value, so the prediction of stock price has been a difficult problem for stockholders and scholars to explore relentlessly since the birth of the stock market[2, 14, 17].

In recent years, with the integration of artificial intelligence technology and financial investment business, deep learning has unique advantages in data feature screening and analysis[10]. The application of deep learning not only reduces the difficulty of stock trend analysis and forecasting, but also introduces new technical solutions and ideas for investors[4, 20, 22]. TRA[13] utilizes the optimal transportation planning as a shunt method, which mine different types of stock trading patterns and get more accurate predictions. HIST[23] puts forward a residual graph neural network model, which explores the potential relationship between stocks to make up for the deficiency of the explicit relationship. Co-CPC[18] integrates macroeconomic variables into model training so as to reduce the uncertainty of model prediction and make the model more robust. Wang et al.[19] design a hierarchical temporal relational network, which not only mine the multi-scale time information, but also construct the adaptive relation matrix from the point of view of space. There are also some models that enrich data by integrating data from different sources, such as event data[24], fund position data[3, 11], etc.

In the financial field, high-frequency data often contain richer and more detailed information. However, many existing models tend to rely only on daily-frequency volume and price data to complete forecasts of future price fluctuations, lacking consideration of high-frequency data and relational data. When considering fusion of multi-granularity data, problems associated with cross-frequency processing and high noise can arise. Therefore we need to explore how to maximize the information gain instead of introducing interference and fluctuations. Contrast learning offers the possibility of achieving this goal. CMLF[7]attempts to effectively integrate high-frequency data by means of comparative learning and achieves a good result improvement. Similar work has been done in other areas, TCL[16] also using comparative learning to accomplish effective utilization of videos at different speeds. The advantage of contrast learning is that it can take advantage of the dependencies between data to construct training targets that capture the essence of data features and mine information most efficiently. Based on this, we consider the design of contrastive learning from three different perspectives and construct three contrastive loss functions in terms of data granularity, stock relationships, and trend changes.

We incorporate them into the model training by pre-training. We first used the contrastive loss function for training to obtain effective encoders of different fine-grained data. Then we add a prediction module to integrate information of different frequencies for prediction, and fine-tuned it through supervised training. Experiments on the real-world dataset of China's A-share market show that our model can achieve excellent performance. And the effectiveness of the contrastive pretraining is proved again through the backtest experiment

To sum up, the main contributions of this work are summarized as follows:

- HRC model is proposed. It not only considers daily frequency data, but also incorporates high frequency data and relational data into the model training, enriching the information source.
- The model considers the design of contrastive pretraining from three different perspectives and construct three contrastive loss functions in terms of data granularity, stock relationships and trend changes.
- The experiments on the real stock market dataset demonstrate that HRC achieves best performance than most of baselines and achieves excellent benefits in the backtest experiment.

The rest of the paper is organized as follows. In Section 2, we introduce the details of HRC. Then, we show the contents of the experiments and analyze the results in Section 3. Finally, in Section 4, we summarize the main content of the paper.

2 THE PROPOSED METHOD: HRC

In this section, we introduce the proposed framework: HRC. The overall structure of the model is shown in Figure 1. The whole model can be divided into two stages. In the first stage, we carry out pre-training through contrastive learning to complete the training of data feature encoders. We have designed three training tasks, and the details will be carried out later. In the second stage, we fine-tune the model parameters through supervised learning to complete stock trend prediction.

2.1 Notations and Problems Statement

In this paper, we use $X_d^t \in R^{N\times C}$ and $X_m^t \in R^{N\times K\times C}$ to represent the stock daily frequency features and stock high frequency features respectively, where C denotes the numbers of attribute features, K denotes the K time periods in a trading day and t denotes the time step. Let $G = (V, E, R)$ denotes a spatial network, where $V = \{s_1, s_2, ..., s_N\}$ is the set of vertices, R denotes the set of stock relations, $e_{ij}^r \in E$ denotes the correlated intensity between s_i and s_j w.r.t. the r^{th}-type relation.

The problem of stock trend forecasting can be described as: learning a mapping function which maps the historical network series $(X_d^i, X_m^i, G, i \in [t-T+1, t])$ into the future stock return rate $Y^{t+1} \in R^{N\times 1}$

2.2 Feature Encoder

We use the time series model($e.g., GRU$) to obtain the representation of the input time series features. Specially, for high frequency representation, we nest two-tier time series model, first input the timing features of the same trading day into the first-tier model, and then transfer the output to the second-tier model, so as to get the final embedding. For ease of description, we next use H_d and H_m to denote the embedding of daily frequency features and high frequency features, respectively.

2.3 Temporal Contrastive Learning

In this section, we present the design of temporal contrastive task.

Although there is a time fine-grained difference between the daily frequency feature and the high frequency feature, they both depict the changes of a certain period of time series of stocks. So the temporal contrastive task is designed to match the daily frequency representation with the high frequency representation. The high frequency representation and daily frequency representation of

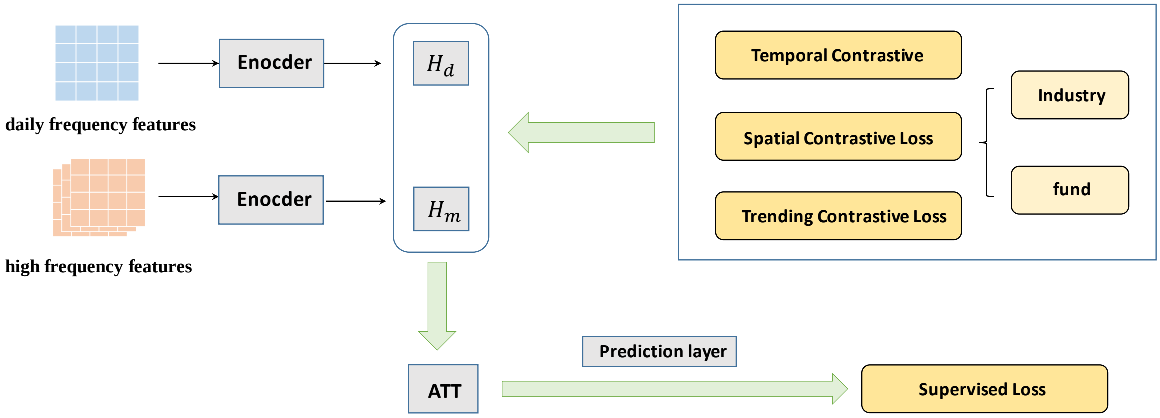

Figure 1: The Framework of the contrastive pre-training model HRC.

the same time step form a positive sample pair, while those with different time steps are regarded as a negative sample pair. The task of this paper in temporal contrast learning is shown in Formula 1.

$$\mathcal{L}_t(H^i_{d,t}, H^i_{m,t}) = -log\frac{h(H^i_{d,t}, H^i_{m,t})}{h(H^i_{d,t}, H^i_{m,t}) + \sum\limits_{\substack{p\in\{m,d\}\\ k\in S, q\neq t}} h(H^i_{d,t}, H^k_{p,q})} \quad (1)$$

where $h(u, v) = exp(\frac{u^T v}{\|u\|_2\|v\|_2})$ is the exponential of cosine similarity measure.

2.4 Spatial Contrastive Learning

we will introduce how to use construct positive sample pair and negative sample pair through stock relationship data. In real world, it is unreasonable to predict the stock as an independent individual. There are all kinds of relationships between stocks, which influence each other.

In this paper, we consider the industry relationship and fund position relationship to construct a positive sample pair. To put it simply, if two stocks belong to the same industry, they are positive sample pairs, otherwise they are negative sample pairs. After classifying the stocks by industry, we average the embedding of stocks belonging to the same class as the embedding expression for that industry. In the construction of contrastive loss function, as shown in Figure 2, we regard the expression of the same industry in different frequency data as a positive sample pair in terms of industry.

We use $R_{h,d}$ and $R_{h,m}$ to denote industry embedding of daily frequency features and high respectively. Based on the industry relationship, we can build the loss function as follows:

$$\mathcal{L}_h(R^i_{h,d}, R^i_{h,m}) = -log\frac{h(R^i_{h,d}, R^i_{h,m})}{h(R^i_{h,d}, R^i_{h,m}) + \sum\limits_{\substack{p\in\{m,d\}\\ k\neq i}} h(R^i_{h,d}, R^k_{h,p})} \quad (2)$$

In addition to industry relationships, there are other types of relationships that can also be used to help us mine more effective information from a spatial perspective. In this paper, we add the positive and negative sample pairs on the fund position relationship. We have a similar approach to fund positions relationship. So the formula for calculating the spatial contrast loss function is as follows:

$$\mathcal{L}_r(R^i_{r,d}, R^i_{r,m}) = \gamma_h * \mathcal{L}_h(R^i_{h,d}, R^i_{h,m}) + \gamma_f * \mathcal{L}_f(R^i_{f,d}, R^i_{f,m}) \quad (3)$$

2.5 Trending Contrastive Learning

In addition to temporal contrastive loss and spatial contrastive loss, we also use labels to construct positive and negative sample pairs. We divide the sample into three category by the yield range: rising, falling and unchanged. Take the rising category as an example.

$$B^r_d(H_d) = \frac{1}{N}\sum_{y_i>c} H^i_d \quad (4)$$

$$B^r_m(H_m) = \frac{1}{N}\sum_{y_i>c} H^i_m \quad (5)$$

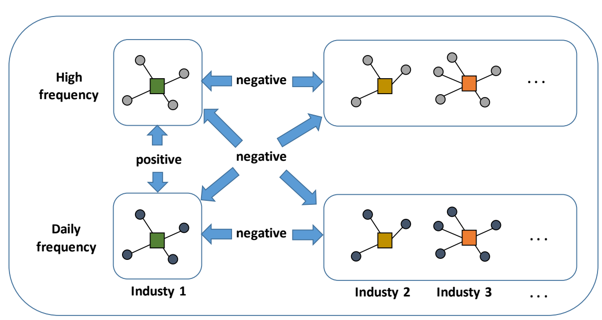

Figure 2: Spatial contrastive loss over industry

where N represents the number of samples that meet the conditions and c is the hyperparameter for the interval.

In the same way, we can calculate falling categories and constant categories. So we construct the trending contrastive loss.

$$\mathcal{L}_b(B_d^i, B_m^i) = -log \frac{h(B_d^i, B_m^i)}{h(B_d^i, B_m^i) + \sum_{\substack{p\in\{m,d\}\\ k\neq i}} h(B_d^i, B_p^i)} \tag{6}$$

2.6 Pretraining and Finetuning

We divide the training process of the model into two parts, the pretraining part and the fine-tuning part. In the pre-training part, we adopt the following loss function:

$$\mathcal{L} = \mathcal{L}_{sup} + \alpha_t * \mathcal{L}_t + \alpha_r * \mathcal{L}_r + \alpha_b * \mathcal{L}_b \tag{7}$$

where $\mathcal{L}_{sup}$ represents the mean square error loss function of the supervised regression task and α_t, α_r, α_b are the hyper-parameters to balance different losses.

The pretraining phase can assist us in learning effective Feature Encoder. After that, we remove the loss function of comparative learning, supervise the training according to the label, and fine-tune the parameters of the model. In order to effectively fuse high-frequency data and daily-frequency data to predict future return rates more accurately, we designed an attention module.The method of calculation is as follows:

$$Y = MLP(\sum_{r\in(d,m)} \beta_r H_r) = MLP(\beta_d H_d + \beta_m H_m) \tag{8}$$

The attention coefficient β_r is calculated as follows:

$$\beta_r = \frac{\exp\left(Sigmoid\left(H_r W_{att} + b_{att}\right)\right)}{\sum_{p\in\{d,m\}} \exp\left(Sigmoid\left(H_p W_{att} + b_{att}\right)\right)} \tag{9}$$

3 EXPERIMENT

In this section, we aim to answer the following two questions:

RQ1. Whether the proposed model is more effective than the benchmark model?

RQ2. Whether the proposed model can achieve higher returns in the backtest experiment?

3.1 Experiment Settings

Dataset. We validate our model on real-world stock dataset of China's A-share market and collect data from 08/01/2015 to 08/01/2020. We have constructed the dataset CSI300 which comprise the largest 300 stocks in the China's A-share market. We use the data of the 15-minute line as high frequency data and collect 11 features for stocks, such as opening price, closing price, volume, low price and so on. In terms of relationship data, we collect the industry relationship between stocks and fund position relationship.

Hyperparameter Setup. We split all datasets with ratio 8 : 1 : 1 into training sets, validation sets and test sets. We use 30 days historical data to predict future return rate. The loss function of the model is optimized using the Adam optimizer of which the initial learning rate is set to 0.0001, and the decay rate of the learning rate is set to 0.00001.

Baselines and Evaluation Metrics. We evaluate the models on five baselines. They are LSTM[15], Transformer[6], HAN[8], TCN[1] and HRC-N (HRC without contrastive pre-training). For the evaluation system, in addition to the MSE, which is commonly used for regression tasks, we adopt the financial indicators IC and ICIR for the evaluation of the model. The IC is calculated as follows.

$$IC = \frac{1}{N} \frac{(\hat{y} - mean(\hat{y}))^T (y - mean(y))}{std(\hat{y}) \cdot std(y)} \tag{10}$$

where $\hat{y}$ and y are the predicted rankings and actual rankings respectively. In practice, IC is calculated for different trading days and their average will be reported. To show the stability of IC, we also report the information ratio of IC, i.e., $ICIR$, which is calculated by dividing the average by the standard deviation of IC.

3.2 Experiment Results(RQ1)

The results are shown in table 1. In the experimental results, it can be found that the proposed model achieves the best results compared with the benchmark model. We also carried out experiments on the model without pretraining, and found that contrastive pre-training can effectively improve the performance of the model.

Table 1: The performance of models

Model	*MSE*	*IC*	*ICIR*
LSTM	0.753	0.0168	0.1280
HAN	0.760	0.0143	0.1642
Transformer	0.748	0.0237	0.2449
TCN	0.749	0.0244	0.1594
HRC-N	0.750	0.0310	0.2880
HRC	**0.748**	**0.0506**	**0.3117**

3.3 Backtest experiment(RQ2)

In order to further verify the effectiveness of the model, we have carried out a simple simulation of the market and carried out back-test experiments. In this paper, we adopt the Topk-Dropout strategy. For each trading day, sell d number of stocks with the worst prediction score and buy the same number of unheld stocks with the best prediction score. The results of the backtest experiment are shown in Figure 3. Through the backtest on the test set, we found that HRC achieved the highest rate of return.

4 CONCLUSION

In this work, we propose a contrastive pre-training model fusing high frequency data and daily frequency data(HRC), which try to integrate high-frequency data into model training and makes full use of time attributes, relational data and future trends to establish the connections between different frequency data. Through pre-training and fine-tuning, we obtain two good encoders and make the prediction more accurate. The experiments on real-world stock dataset show that our model achieves the best results compared to the benchmark models. And our model obtains the highest benefit in the backtest experiment. We believe this work has important

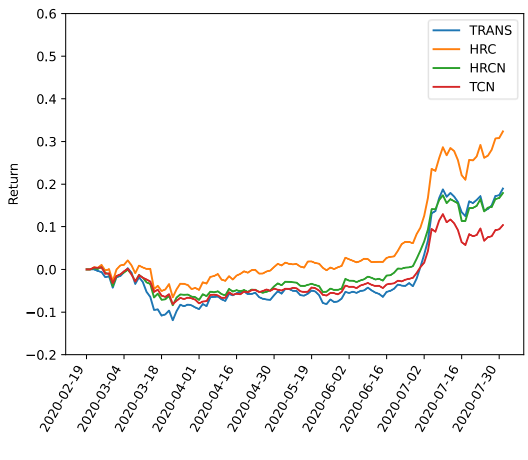

Figure 3: The return of backtest experiment

application value and good prospect for reducing investment risk and becomes an effective reference.

REFERENCES

[1] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).

[2] A. Brim and N. S. Flann. 2022. Deep reinforcement learning stock market trading, utilizing a CNN with candlestick images. *PLoS ONE* 17, 2 (2022), e0263181–.

[3] Chi Chen, Li Zhao, Jiang Bian, Chunxiao Xing, and Tie-Yan Liu. 2019. Investment behaviors can tell what inside: Exploring stock intrinsic properties for stock trend prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2376–2384.

[4] D. Cheng, F. Yang, S. Xiang, and J. Liu. 2022. Financial time series forecasting with multi-modality graph neural network. *Pattern Recognition: The Journal of the Pattern Recognition Society* 121- (2022), 121.

[5] Yujie Fang, Juan Chen, and Zhengxuan Xue. 2019. Research on quantitative investment strategies based on deep learning. *Algorithms* 12, 2 (2019), 35.

[6] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019. Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems (TOIS)* 37, 2 (2019), 1–30.

[7] M. Hou, C. Xu, Y. Liu, W. Liu, J. Bian, L. Wu, Z. Li, E. Chen, and T. Y. Liu. 2021. Stock Trend Prediction with Multi-granularity Data: A Contrastive Learning Approach with Adaptive Fusion. (2021).

[8] 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*. 261–269.

[9] Z. Hu, Y. Zhao, and M. Khushi. 2021. A Survey of Forex and Stock Price Prediction Using Deep Learning. *Applied System Innovation* 4, 1 (2021), 9.

[10] X. Li, P. Wu, and W. Wang. 2020. Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management* (2020), 102212.

[11] Zhige Li, Derek Yang, Li Zhao, Jiang Bian, Tao Qin, and Tie-Yan Liu. 2019. Individualized indicator for all: Stock-wise technical indicator optimization with stock embedding. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 894–902.

[12] Peng Liang, Hai-Dong Yang, Wen-Si Chen, Si-Yuan Xiao, and Zhao-Ze Lan. 2018. Transfer learning for aluminium extrusion electricity consumption anomaly detection via deep neural networks. *International Journal of Computer Integrated Manufacturing* 31, 4-5 (2018), 396–405.

[13] Hengxu Lin, Dong Zhou, Weiqing Liu, and Jiang Bian. 2021. Learning multiple stock trading patterns with temporal routing adaptor and optimal transport. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1017–1026.

[14] V. Polepally, Nsn Reddy, M. Sindhuja, N. Anjali, and K. J. Reddy. 2021. A Deep Learning Approach for Prediction of Stock Price Based on Neural Network Models: LSTM and GRU. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*.

[15] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* 28 (2015).

[16] A. Singh, O. Chakraborty, A. Varshney, R. Panda, and A. Das. 2021. Semi-Supervised Action Recognition with Temporal Contrastive Learning.

[17] D. Song, Amc Baek, and N. Kim. 2021. Forecasting Stock Market Indices Using Padding-Based Fourier Transform Denoising and Time Series Deep Learning Models. *IEEE Access* PP, 99 (2021), 1–1.

[18] G. Wang, L. Cao, H. Zhao, Q. Liu, and E. Chen. 2021. Coupling Macro-Sector-Micro Financial Indicators for Learning Stock Representations with Less Uncertainty.. In *National Conference on Artificial Intelligence*.

[19] H. Wang, S. Li, T. Wang, and J. Zheng. 2021. Hierarchical Adaptive Temporal-Relational Modeling for Stock Trend Prediction.. In *International Joint Conference on Artificial Intelligence*.

[20] H. Wang, J. Wang, L. Cao, Y. Li, Q. Sun, and J. Wang. 2021. A Stock Closing Price Prediction Model Based on CNN-BiSLSTM. *Complexity* 2021 (2021).

[21] Jenq Haur Wang and Hsin Wen Liu. 2022. A Deep Fusion Model Combining News Content and Historical Prices for Stock Trend Prediction. *Springer, Cham* (2022).

[22] H. Xu, L. Chai, Z. Luo, and S. Li. 2022. Stock movement prediction via gated recurrent unit network based on reinforcement learning with incorporated attention mechanisms. *Neurocomputing* Jan.7 (2022), 467.

[23] Wentao Xu, Weiqing Liu, Lewen Wang, Yingce Xia, Jiang Bian, Jian Yin, and Tie-Yan Liu. 2021. HIST: A Graph-based Framework for Stock Trend Forecasting via Mining Concept-Oriented Shared Information. *arXiv preprint arXiv:2110.13716* (2021).

[24] Wentao Xu, Weiqing Liu, Chang Xu, Jiang Bian, Jian Yin, and Tie-Yan Liu. 2021. REST: Relational Event-driven Stock Trend Forecasting. In *Proceedings of the Web Conference 2021*. 1–10.