MASTER: Market-Guided Stock Transformer for Stock Price Forecasting





Tong Li, Zhaoyang Liu, Yanyan Shen, Xue Wang, Haokun Chen, Sen Huang Shanghai Jiao Tong University

Alibaba Group

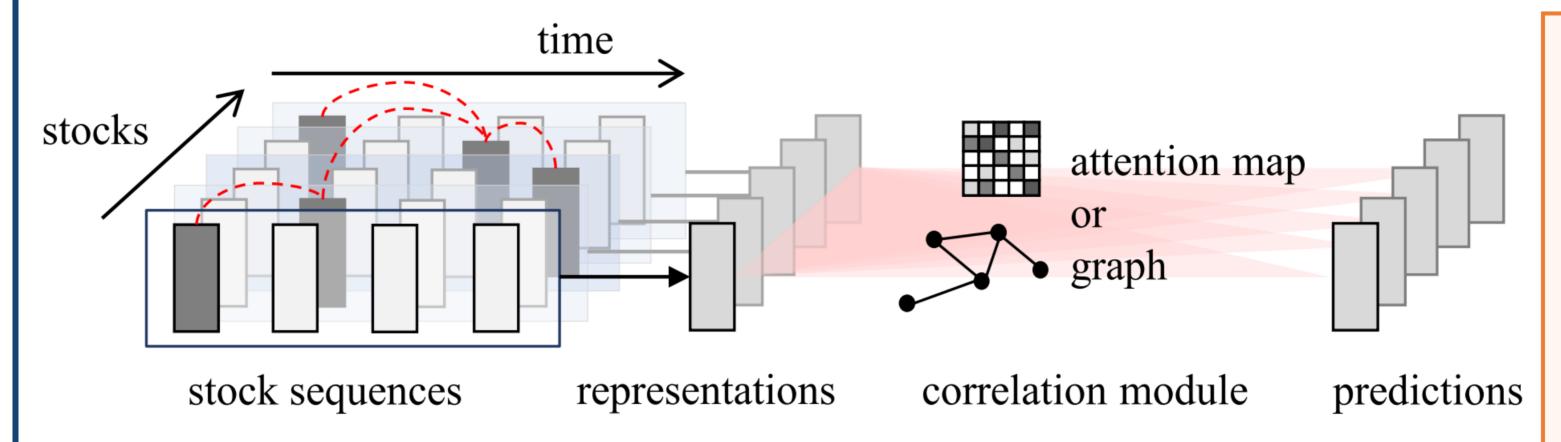


1. Background

- Stock price forecasting uses the historical data of stocks to predict their future trends, which is vital in profitable stock investment.
- Multiple factors, such as macroeconomic factors, capital flows and investor sentiments, interweave the stocks as a correlated network.
- Previous works model stock correlation
 - Static: Predefined relationships, concepts or rules. (e.g. industry graph). (1) relationship ≠ real-time correlation (2) Not generalizable when events such as company listing, delisting or change in main business happen.
 - **Dynamic:** Leverage attention mechanism to mine the latent correlation. (1) Data-driven, (2) More flexible.

2. Motivation

• Issue: Existing works share a framework which fails to model the realistic stock correlation.



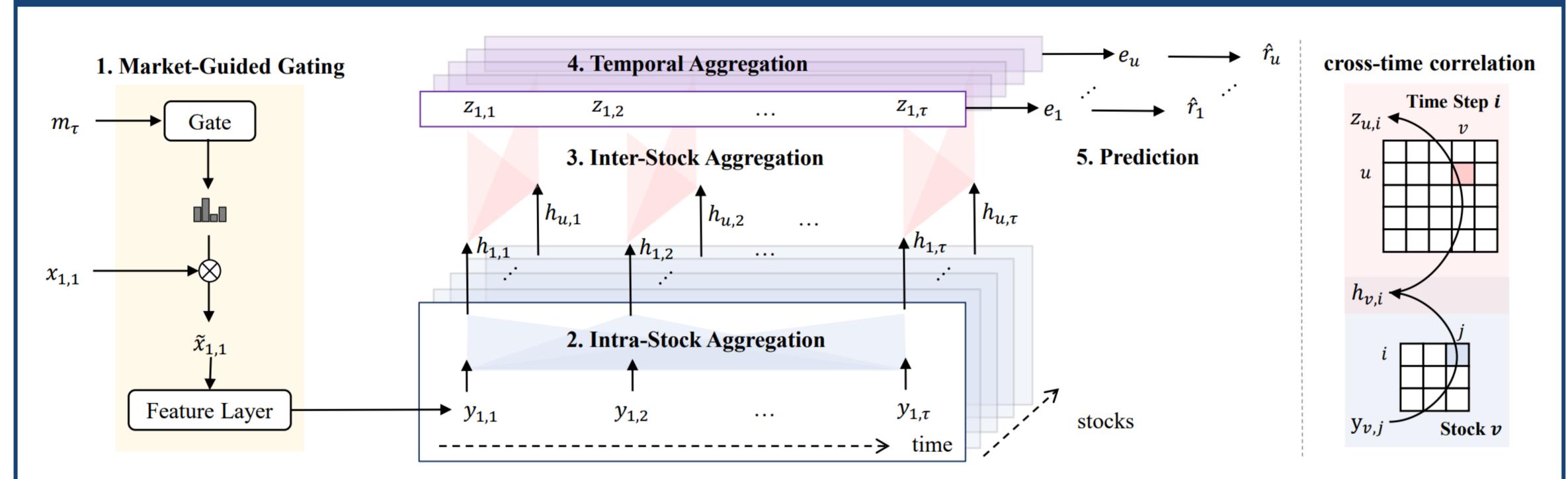
- 1. Momentary: The dominating factors of stock prices constantly change.
- 2. **Cross-time:** Different stocks may react to the same factor with different delays. (e.g. Upstream companies' stock prices may react faster to a shortage of raw materials than those of downstream companies.)

RankICIR

- Challenges:
 - 1. The large and complex attention field is challenging in learning in stock domain.
 - Our solution: Aggregate information from different time steps and other stocks alternatively.
 - 2. The stock correlation is different under varying market status. (e.g. in a bull market, the correlation are more significant due to investors' optimism.)
 - Our solution: Incorporate the market information to perform automatic feature selection.

Dataset | Model

3. Method

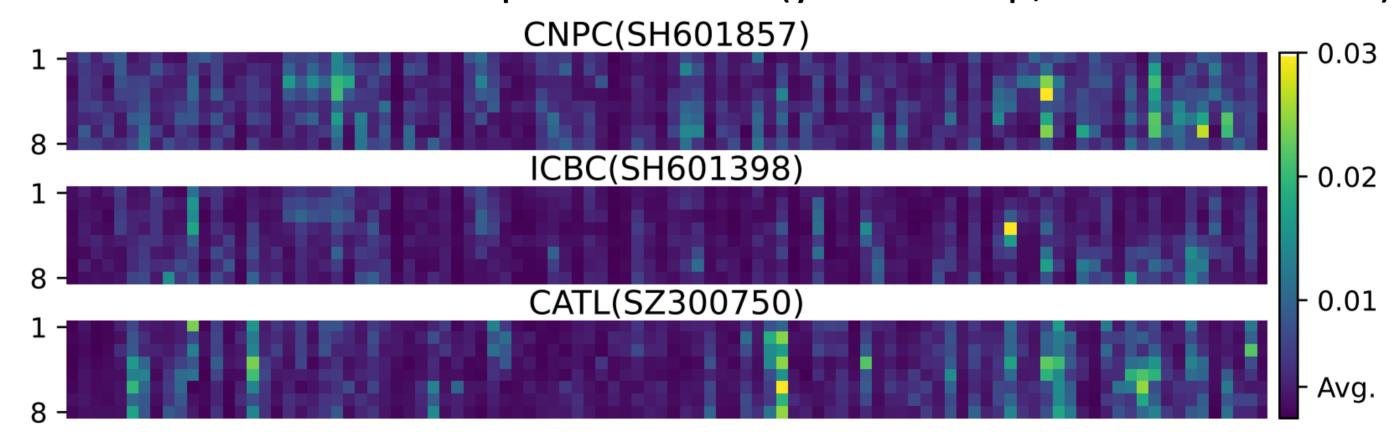


4. Experiments

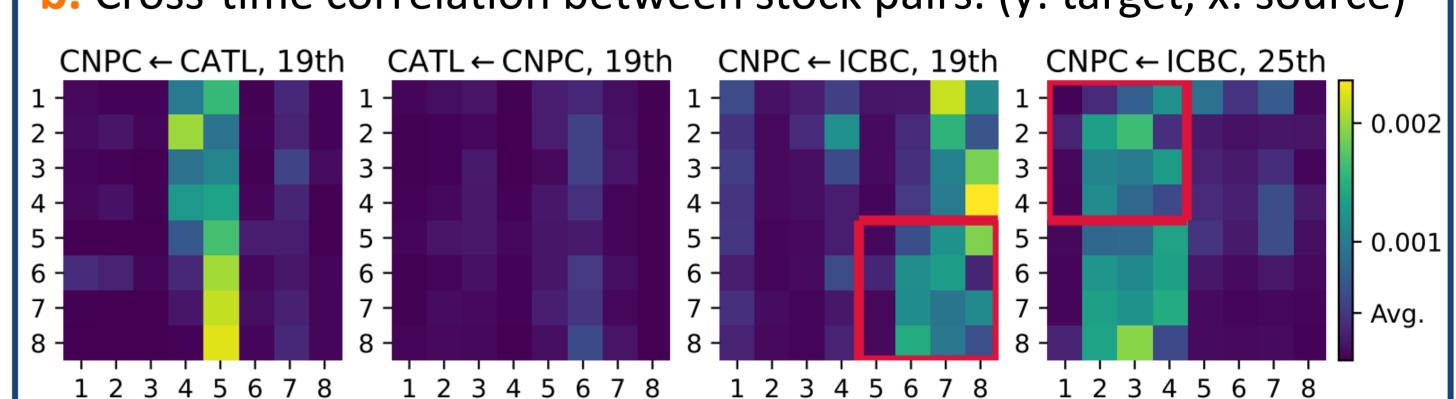
Performance. MASTER outperforms the second-best by 13% on ranking metrics, and 47% on portfolio-based metrics.

				Part of the Control o		
XGBoost LSTM GRU TCN Transformer GAT DTML	0.051 ± 0.001 0.049 ± 0.001 0.052 ± 0.004 0.050 ± 0.002 0.047 ± 0.007 0.054 ± 0.002 0.049 ± 0.006	0.37 ± 0.01 0.41 ± 0.01 0.35 ± 0.04 0.33 ± 0.04 0.39 ± 0.04 0.36 ± 0.02 0.33 ± 0.04	0.050 ± 0.001 0.051 ± 0.002 0.052 ± 0.005 0.049 ± 0.002 0.051 ± 0.002 0.041 ± 0.002 0.052 ± 0.005	0.36 ± 0.01 0.41 ± 0.03 0.34 ± 0.04 0.31 ± 0.04 0.42 ± 0.04 0.25 ± 0.02 0.33 ± 0.04	0.23 ± 0.03 0.20 ± 0.04 0.19 ± 0.04 0.18 ± 0.05 0.22 ± 0.06 0.19 ± 0.03 0.21 ± 0.03	$ \begin{array}{c} 1.9 \pm 0.3 \\ 2.0 \pm 0.4 \\ \hline 1.5 \pm 0.3 \\ 1.4 \pm 0.5 \\ 2.0 \pm 0.4 \\ 1.3 \pm 0.3 \\ 1.7 \pm 0.3 \end{array} $
The state of the s	100 100 100 100 100 100 100 100 100 100			And the second s	1 11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2.4 ± 0.4
XGBoost LSTM GRU TCN	0.040 ± 0.000 0.028 ± 0.002 0.039 ± 0.002 0.038 ± 0.002	0.37 ± 0.01 0.32 ± 0.02 0.36 ± 0.05 0.33 ± 0.04	0.047 ± 0.000 0.039 ± 0.002 0.044 ± 0.003 0.045 ± 0.002	0.42 ± 0.01 0.41 ± 0.03 0.39 ± 0.07 0.38 ± 0.05	0.08 ± 0.02 0.09 ± 0.02 0.07 ± 0.04 0.05 ± 0.04	0.6 ± 0.2 0.9 ± 0.2 0.6 ± 0.3 0.4 ± 0.3
Transformer GAT DTML MASTER	0.040 ± 0.003 0.043 ± 0.002 0.039 ± 0.004 $0.052^* \pm 0.006$	0.43 ± 0.03 0.39 ± 0.02 0.29 ± 0.03 0.40 ± 0.06	0.048 ± 0.003 0.042 ± 0.002 0.053 ± 0.008 $\mathbf{0.066 \pm 0.007}$	0.51 ± 0.05 0.35 ± 0.02 0.37 ± 0.06 0.48 ± 0.06	0.13 ± 0.04 0.10 ± 0.04 0.16 ± 0.03 $0.28^* \pm 0.02$	1.1 ± 0.3 0.7 ± 0.3 1.3 ± 0.2 $2.3^* \pm 0.3$
	LSTM GRU TCN Transformer GAT DTML MASTER XGBoost LSTM GRU TCN Transformer GAT DTML	$\begin{array}{ c c c } \textbf{LSTM} & 0.049 \pm 0.001 \\ \textbf{GRU} & 0.052 \pm 0.004 \\ \textbf{TCN} & 0.050 \pm 0.002 \\ \textbf{Transformer} & 0.047 \pm 0.007 \\ \textbf{GAT} & 0.054 \pm 0.002 \\ \textbf{DTML} & 0.049 \pm 0.006 \\ \textbf{MASTER} & \textbf{0.064*} \pm \textbf{0.000} \\ \textbf{LSTM} & 0.028 \pm 0.002 \\ \textbf{GRU} & 0.039 \pm 0.002 \\ \textbf{GRU} & 0.039 \pm 0.002 \\ \textbf{TCN} & 0.038 \pm 0.002 \\ \textbf{Transformer} & 0.040 \pm 0.003 \\ \textbf{GAT} & 0.043 \pm 0.002 \\ \textbf{DTML} & 0.039 \pm 0.004 \\ \textbf{DTML} & 0.039 \pm 0.004 \\ \textbf{DTML} & 0.039 \pm 0.004 \\ \textbf{DTML} & 0.0039 \pm 0.004 \\ \textbf{DTML} $	$\begin{array}{ c c c c } \textbf{LSTM} & 0.049 \pm 0.001 & 0.41 \pm 0.01 \\ \hline \textbf{GRU} & 0.052 \pm 0.004 & 0.35 \pm 0.04 \\ \hline \textbf{TCN} & 0.050 \pm 0.002 & 0.33 \pm 0.04 \\ \hline \textbf{Transformer} & 0.047 \pm 0.007 & 0.39 \pm 0.04 \\ \hline \textbf{GAT} & 0.054 \pm 0.002 & 0.36 \pm 0.02 \\ \hline \textbf{DTML} & 0.049 \pm 0.006 & 0.33 \pm 0.04 \\ \hline \textbf{MASTER} & \textbf{0.064*} \pm \textbf{0.006} & \textbf{0.33} \pm 0.04 \\ \hline \textbf{XGBoost} & 0.040 \pm 0.000 & 0.37 \pm 0.01 \\ \hline \textbf{LSTM} & 0.028 \pm 0.002 & 0.32 \pm 0.02 \\ \hline \textbf{GRU} & 0.039 \pm 0.002 & 0.36 \pm 0.05 \\ \hline \textbf{TCN} & 0.038 \pm 0.002 & 0.33 \pm 0.04 \\ \hline \textbf{Transformer} & 0.040 \pm 0.003 & \textbf{0.43} \pm \textbf{0.03} \\ \hline \textbf{GAT} & 0.043 \pm 0.002 & 0.39 \pm 0.02 \\ \hline \textbf{DTML} & 0.039 \pm 0.004 & 0.29 \pm 0.03 \\ \hline \end{array}$	$\begin{array}{ c c c c c } \textbf{LSTM} & 0.049 \pm 0.001 & 0.41 \pm 0.01 & 0.051 \pm 0.002 \\ \textbf{GRU} & 0.052 \pm 0.004 & 0.35 \pm 0.04 & 0.052 \pm 0.005 \\ \textbf{TCN} & 0.050 \pm 0.002 & 0.33 \pm 0.04 & 0.049 \pm 0.002 \\ \textbf{Transformer} & 0.047 \pm 0.007 & 0.39 \pm 0.04 & 0.051 \pm 0.002 \\ \textbf{GAT} & 0.054 \pm 0.002 & 0.36 \pm 0.02 & 0.041 \pm 0.002 \\ \textbf{DTML} & \overline{0.049} \pm 0.006 & 0.33 \pm 0.04 & 0.052 \pm 0.005 \\ \textbf{MASTER} & \textbf{0.064*} \pm \textbf{0.006} & \textbf{0.33} \pm 0.04 & 0.076* \pm \textbf{0.005} \\ \textbf{XGBoost} & 0.040 \pm 0.000 & 0.37 \pm 0.01 & 0.047 \pm 0.000 \\ \textbf{LSTM} & 0.028 \pm 0.002 & 0.32 \pm 0.02 & 0.039 \pm 0.002 \\ \textbf{GRU} & 0.039 \pm 0.002 & 0.36 \pm 0.05 & 0.044 \pm 0.003 \\ \textbf{TCN} & 0.038 \pm 0.002 & 0.33 \pm 0.04 & 0.045 \pm 0.002 \\ \textbf{Transformer} & 0.040 \pm 0.003 & \textbf{0.43} \pm \textbf{0.03} & 0.048 \pm 0.003 \\ \textbf{GAT} & 0.043 \pm 0.002 & 0.39 \pm 0.02 & 0.042 \pm 0.002 \\ \textbf{DTML} & \overline{0.039} \pm 0.004 & 0.29 \pm 0.03 & 0.053 \pm 0.008 \\ \end{array}$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{ c c c c c c c c } \textbf{LSTM} & 0.049 \pm 0.001 & 0.41 \pm 0.01 & 0.051 \pm 0.002 & 0.41 \pm 0.03 & 0.20 \pm 0.04 \\ \hline \textbf{GRU} & 0.052 \pm 0.004 & 0.35 \pm 0.04 & 0.052 \pm 0.005 & 0.34 \pm 0.04 & 0.19 \pm 0.04 \\ \hline \textbf{TCN} & 0.050 \pm 0.002 & 0.33 \pm 0.04 & 0.049 \pm 0.002 & 0.31 \pm 0.04 & 0.18 \pm 0.05 \\ \hline \textbf{Transformer} & 0.047 \pm 0.007 & 0.39 \pm 0.04 & 0.051 \pm 0.002 & 0.42 \pm 0.04 & 0.22 \pm 0.06 \\ \hline \textbf{GAT} & 0.054 \pm 0.002 & 0.36 \pm 0.02 & 0.041 \pm 0.002 & 0.25 \pm 0.02 & 0.19 \pm 0.03 \\ \hline \textbf{DTML} & 0.049 \pm 0.006 & 0.33 \pm 0.04 & 0.052 \pm 0.005 & 0.33 \pm 0.04 & 0.21 \pm 0.03 \\ \hline \textbf{MASTER} & \textbf{0.064}^* \pm \textbf{0.006} & \textbf{0.42} \pm \textbf{0.04} & \textbf{0.076}^* \pm \textbf{0.005} & \textbf{0.49} \pm \textbf{0.04} & \textbf{0.27} \pm \textbf{0.05} \\ \hline \textbf{LSTM} & 0.028 \pm 0.002 & 0.37 \pm 0.01 & 0.047 \pm 0.000 & 0.42 \pm 0.01 & 0.08 \pm 0.02 \\ \hline \textbf{LSTM} & 0.039 \pm 0.002 & 0.32 \pm 0.02 & 0.039 \pm 0.002 & 0.41 \pm 0.03 & 0.09 \pm 0.02 \\ \hline \textbf{GRU} & 0.039 \pm 0.002 & 0.36 \pm 0.05 & 0.044 \pm 0.003 & 0.39 \pm 0.07 & 0.07 \pm 0.04 \\ \hline \textbf{TCN} & 0.038 \pm 0.002 & 0.33 \pm 0.04 & 0.045 \pm 0.002 & 0.38 \pm 0.05 & 0.05 \pm 0.04 \\ \hline \textbf{Transformer} & 0.040 \pm 0.003 & \textbf{0.43} \pm \textbf{0.03} & 0.048 \pm 0.003 & \textbf{0.51} \pm \textbf{0.05} & 0.13 \pm 0.04 \\ \hline \textbf{GAT} & 0.043 \pm 0.002 & 0.39 \pm 0.02 & 0.042 \pm 0.002 & 0.35 \pm 0.02 & 0.10 \pm 0.04 \\ \hline \textbf{DTML} & 0.039 \pm 0.004 & 0.29 \pm 0.03 & 0.053 \pm 0.008 & 0.37 \pm 0.06 & 0.16 \pm 0.03 \\ \hline \end{array}$

- Visualization. The stock correlation captured by MASTER is momentary, cross-time, asymmetric and evolve in time.
- a. Correlation towards specific stock. (y: time step, x: source stocks)



b. Cross-time correlation between stock pairs. (y: target, x: source)



- Market-Guided Gating generates one scaling coefficient for each feature.
- Feature vector $x_{u,t}$ is the indicators of stock $u \in S$ at time step $t \in [1, \tau]$, $F = |x_{u,t}|$.
- ullet Market status vector $m_{ au}$ contains (1) market index price, (2) market index trading volume.
- Coefficient $\alpha(m_{\tau}) = F \cdot \operatorname{softmax}_{\beta}(W_{\alpha}m_{\tau} + b_{\alpha}), \beta$ is the temperature, $|\alpha(\cdot)| = F$.
- Enlarge or shrink the magnitude of each feature by $ilde{x}_{u,t} = lpha(m_{ au}) \circ x_{u,t}$.
- Feature layer: $Y_u = \prod_{t \in [1,\tau]} \text{LayerNorm}(f(\tilde{x}_{u,t}) + p_t), \ f(\cdot)$ is a linear layer and p is positional code.
- Intra-stock and inter-stock aggregation to model cross-time stock correlation.
- For stock u, transform Y_u into Q_u^1, K_u^1, V_u^1 . $H_u^1 = ||_{t \in [1,\tau]} h_{u,t} = \text{FFN}^1(\text{MHA}^1(Q_u^1, K_u^1, V_u^1) + Y_u)$.
- At time t, transform $H_t^2 = ||_{u \in S} h_{u,t}$ into Q_t^2 , K_t^2 , V_t^2 . $Z_t = ||_{u \in S} z_{u,t} = \text{FFN}^2(\text{MHA}^2(Q_t^2, K_t^2, V_t^2) + H_u^2)$.
- Temporal aggregation summarizes temporal embeddings to gain one stock embedding.
- $e_u = \sum_{t \in [1,\tau]} \lambda_{u,t} z_{u,t}$. The last embedding queries from others for weights $\lambda_{u,t} = \frac{\exp(z_{u,t}^T W_{\lambda} z_{u,\tau})}{\sum_{i \in [1,\tau]} \exp(z_{u,i}^T W_{\lambda} z_{u,\tau})}$

Prediction

- $\hat{r}_u = g(e_u), g(\cdot)$ is a linear layer for regression.
- $L = \sum_{u \in S} MSE(r_u, \hat{r}_u)$. r_u is the normalized return ratio in d days, encoded with ranking information.

More Information

- (1) Data & Code: github.com/SJTU-Quant/MASTER
- Contact us: 2017lt@sjtu.edu.cn