A Novel Differentiable Rank Learning Method Towards Stock Movement Quantile Forecasting

Chenyou Fan, South China Normal University
Hengyang Lu, Jiangnan University
Aimin Huang, Hangzhou Higgs Asset Management Co., Ltd.

Motivation

- We focus on Stock Movement Forecasting (SMF).
- Price movement forecasting applies to many financial assets such as futures and options, carbon credits, commodities, and more.
- An emerging area of interest within this field is the prediction of carbon credit prices, specifically within the leading carbon trading markets such as the European Union Emission Trading Scheme (EU ETS) and Chinese Emission Allowance (CEA).
- Previous deep-learning-based SMF techniques only considered binary up-or-down classification tasks, ignoring the importance of fine-grained categorization.

Motivation Cont’d

- Previous deep-learning-based SMF techniques only considered binary up-or-down classification tasks, ignoring the importance of fine-grained categorization.
- We present a novel end-to-end learning-to-rank framework that incorporates both market-level and stock-level dynamics.
- Our method learns to rank steps with the most significant movements in the temporal dimension.
- We conduct extensive evaluations over real-world market data and show 5-15% improvements in the Gain and Sharpe Ratio.

Overview of our approach

- Many recent studies have simplified SMF as a binary classification task that focuses solely on prominent rising and falling steps, ignoring neutral movements.
- However, the movement trends are highly imbalanced. E.g., tiny movements within ±0.1% can account for up to 40% of per-minute trading scenarios.
- A finer-grained prediction is necessary to depict the distribution of movements and make adaptive decisions, such as determining the optimal trading quantities.
- **We focus on modeling SMF as a fine-grained Dis- tributional Quantile Classification (DQC) task.**

Overview of our approach

- We focus on modeling SMF as fine-grained Distributional Quantile Classification (DQC).
- We partitioned the movement range into a series of finely divided quantile intervals.
- We forecast the future movement distribution across the designated domain.
- The probability assigned to each quantile interval reflects the probability of the corresponding movement range occurring.

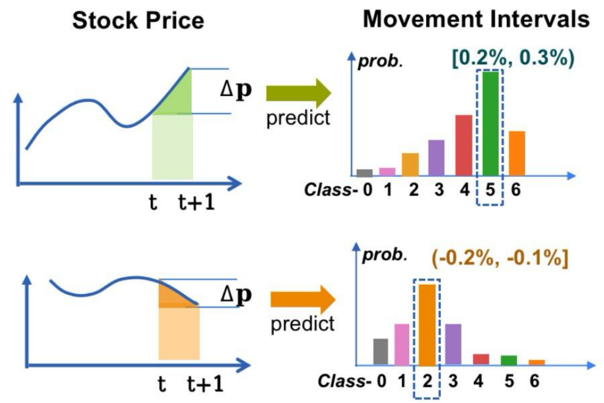


Figure 1: Illustration of price movement quantile prediction with rising case (top) and falling case (bottom). Our task is to categorize the movement at each step to its correct quantile interval.

Definition of the task

We consider a market of N common stocks and M stock market indices (e.g. S&P 500). We observe the past T_h steps $\mathcal{T}_h = \{1, 2, ..., T_h\}$ and predict the subsequent T_f future steps $\mathcal{T}_f = \{T_h + 1, ..., T_h + T_f\}$ as the SMF task. Depending on the data frequency, one step can be a day for daily trading data or a minute for per-minute trading data.

Task features

Table 1: Numerical inputs and features.

No.	Features	Formulation from raw data
1	v_close	e.g., $v\_close = close_t/close_{t\text{-}1} - 1$
2	$v\_open/high/low$	e.g., $v\_open = open_t/close_t - 1$
3	$v\_avg, k = 5, 10, \ldots$	e.g., $v\_avg = \frac{\sum_{i=1}^{k} close_{v\text{-}i+1}/k}{close_t} - 1$
4	$v\_trade/vol/amt$	e.g., $v\_vol = vol_t/vol_{t-1} - 1$

(e.g., trade number, volume, and amount) at each time step. We process the numeric data as previous works [7, 20] as Table 1 shows.

Division of movement intervals as classes

We partition the range of movements into C = 7 intervals. Table 2 presents each interval along with its corresponding empirical data percentage and the approximate quantile of the median movement (M-Quantile) for that specific interval.

Table 2: Movement intervals and quantiles of CSI-21 dataset.

Class C^{qt}	0	1	2	**3**	4	5	6
Move Δp (%)	<-0.3	<-0.2	[-0.2,-0.1)	[-0.1,0.1]	>0.1	>0.2	>0.3
Percentage	10%	10%	10%	40%	10%	10%	10%
M-Quantile	0.05	0.15	0.25	0.5	0.75	0.85	0.95

Model Architecture

- **History Encoder**

$$\boldsymbol{O}^{St} \leftarrow \text{Self-Attention}(\boldsymbol{X}^h) \in \mathcal{R}^{T_h \times d_{\text{model}}}$$

- **Multi-Modality Fusion (MMF)**

$$\boldsymbol{O}^{Ind} \leftarrow \text{Concat}[\boldsymbol{O}_1^{Ind}; \ldots; \boldsymbol{O}_M^{Ind}] \in \mathcal{R}^{T_h \times (M \times d_{\text{model}})}$$

$$\boldsymbol{O}^{S\text{-}I} \leftarrow \boldsymbol{O}^{St} \oplus \text{Cross-Attention}(\boldsymbol{O}^{St}, \boldsymbol{O}^{Ind}, \boldsymbol{O}^{Ind})$$

- **Future Decoder**

$$\boldsymbol{H}^{dec} \leftarrow \text{Cross-Attention}(\boldsymbol{O}^{f}, \boldsymbol{O}^{S\text{-}I}, \boldsymbol{O}^{S\text{-}I})$$

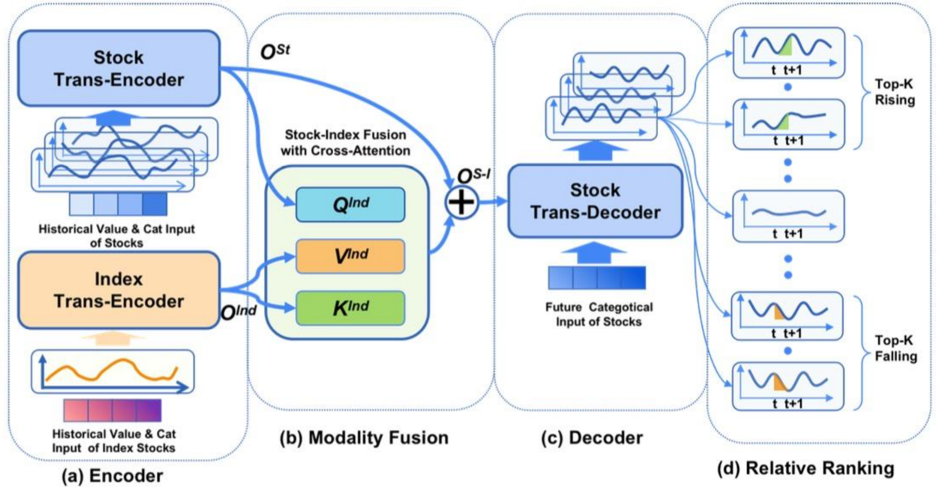


Figure 2: Overview of DQC-Rank learning framework with Encoder, Multi-Modality Fusion, Decoder and Rank Learning modules.

Distributional Quantile Classification (DQC)

Let the classifier f produces a C-way logits over movement intervals such that $\boldsymbol{\psi}_t \leftarrow f(\boldsymbol{H}^{dec}(t)) \in \mathcal{R}^C$ for each future step t. The **Distributional Quantile Classification Loss** is the weighted cross-entropy loss of C-quantile classification averaged over all T future steps such that:

$$L^{dqc}(\boldsymbol{\psi}, \boldsymbol{z}) = \frac{1}{T} \sum_{t=1}^{T} - \underbrace{(1 - p_t)^{\gamma}}_{\text{focal-term}} \cdot \underbrace{\log p_t}_{\text{ce-loss}} , \tag{5}$$

$$s.t. \;\; p_t = \sigma(\boldsymbol{\psi}_t[z_t]) \, ,$$

in which σ is softmax function, z_t is the true quantile interval as class label at time t. The focal-term inherits from focal loss [11] to balance class, which is critical as the movements are unevenly distributed as Table 2 shows.

Inter- / Intra- Stock Rank Learning

- We first define the inter-stock ranking, wherein we identify the top rising and declining stocks across the entire market to obtain market-level insights.
- Then we define the intra-stock ranking, wherein we identify the largest moving future time steps for each individual stock to uncover its internal patterns at the stock-level.

Inter-Stock Rank Learning (Inter-Rank)

- The objective of Inter-Rank is to identify the top-K stocks that are rising or falling, within a pool of N cross-sectional stocks in the market.

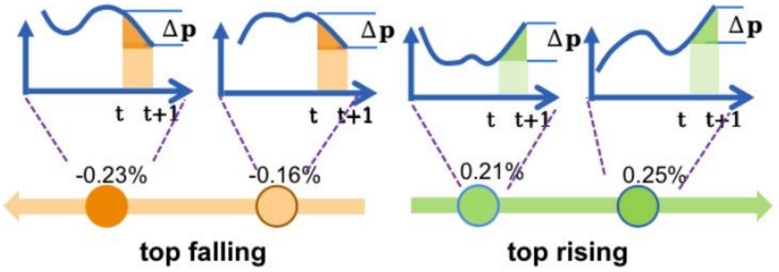


Figure 3: Demo of inter-stock ranking by finding the top-K ($K = 2$) rising and falling stocks over all stocks at step t.

- We average the inter-loss over all T_f future steps as the complete Inter-Stock Ranking Loss as:

$$L^{inter} = \frac{1}{2 \cdot T_f} \sum_{t=1}^{T_f} (\ell_t^r + \ell_t^f)$$

Intra-Stock Rank Learning (Intra-Rank)

- We further propose the Intra-Rank which aims to explore the internal volatility of each time-series by learning to identify its top-K′ rising and falling steps over all future steps.

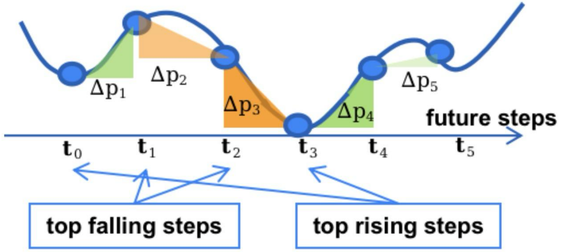


Figure 4: Demo of intra-stock ranking by finding the top-$K'(=2)$ rising and falling future steps of one stock.

- We average the intra loss for each of N stocks as the Intra-Stock Ranking Loss as

$$L^{intra} = \frac{1}{2 \cdot N} \sum_{i=1}^{N} (\mathcal{E}_i^r + \mathcal{E}_i^f)$$

Multi-task training objective

- Finally, we can optimize the model by jointly minimizing the distri- butional quantile loss Eq.(5), inter-rank loss (10), and intra-rank loss (15) in end-to-end fashion as:

Finally, we can optimize the model by jointly minimizing the distributional quantile loss Eq.(5), inter-rank loss (10), and intra-rank loss (15) in end-to-end fashion as:

$$L^{final} = L^{dqc} + \alpha_1 L^{inter} + \alpha_2 L^{intra} \ , \tag{16}$$

where factors α_1, α_2 are searched by cross-validation. We denote our learning framework as **DQC-Rank** with multi-task learning objective as in Eq. (16).

Experiments

- We evaluate DQC-Rank on 3 benchmark datasets described as follows.
- **KDD-17 [Zhang et al., KDD 2017]** has daily prices of 50 top performing US stocks from 10 sectors. **ACL-18 [Xu et al., ACL 2018]** contains daily prices of 88 US stocks with top capital sizes. We take 5 days as history and predict the closing prices at next 5 days.
- **CSI-21 (ours)** is self-collected from 800 China A-shares from 2018-2021. We collect the per-minute stock prices at 240 trading steps in a 4-hour trading day. We take 5 minutes as history and predict the closing prices at next 5 minutes.

Baselines

- DQC-Plain is our proposed Quantile Classification framework with proposed Encoder-MMF-Decoder design and optimizes with DQC loss.
- DQC-Rank further improves DQC-Plain by incorporating our proposed Inter- / Intra- Rank learning framework.
- We compare our methods with LightGBM, ALSTM (Feng et al., IJCAI 2019), RSR (Feng et al., TOIS 2019), TFT (Lim et al., IJF 2021), and DTML [20] (Yoo et al., SIGKDDD 2021).

CSI-21 Results

- *DQC-Rank outperforms other baselines in Gain and SR,* leading the second place DQC-Plain by 14.4% in Gain (21.4 vs. 18.7) and 5.9% in SR (1.26 vs. 1.19), at a cost of a larger MDD (retraction).
- The overall trends of the three splits are downward, oscillating and upward, respectively. The Gain of each split is 7.8%, 23.4% and 33.0%, respectively, contributing to an averaged 21.4% in Gain (last row, col. Gain in Table 3).
- Due to the strong class-imbalance issue, LightGBM and ALSTM (shaded grey) had a seemingly high overal Acc (QAcc) (39.2% and 44.2%) while got extremely low in per-class (PAcc) (26.1% and 21.2%) and SR (0.16 and 0.88).

Table 3: Average result on CSI-21 with three rolling splits.

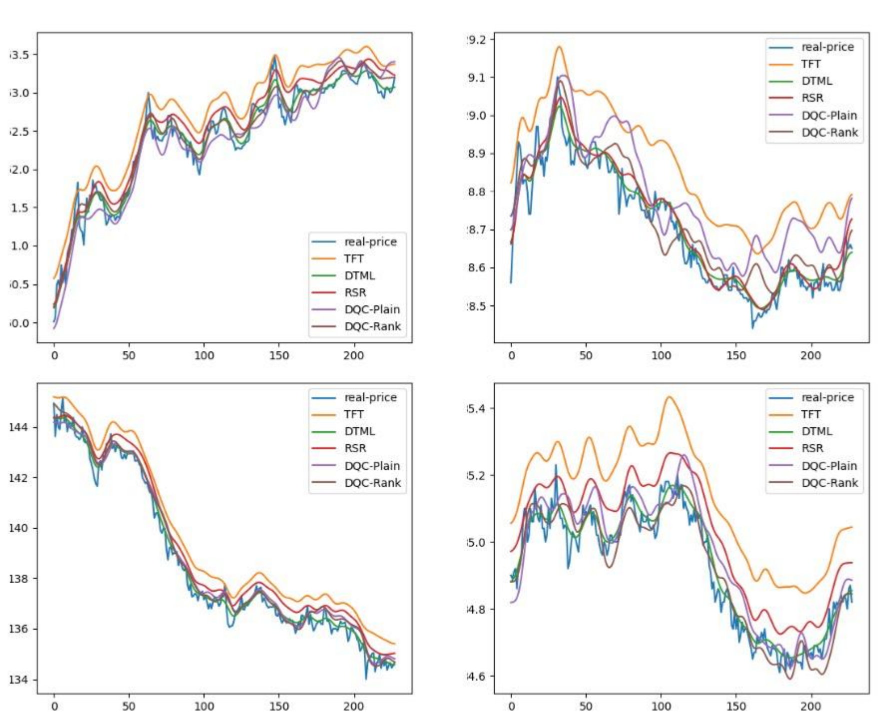
Method \ Setting	QAcc (%) ↑	PAcc (%) ↑	MCC (%) ↑	MDD (%) ↓	Gain (%) ↑	SR ↑
LightGBM [9]	39.2	26.1	-9.05	3.39	2.63	0.16
ALSTM [7]	44.2	21.2	-8.34	2.20	8.07	0.88
TFT [10]	31.4	32.0	-2.38	3.82	14.8	1.09
DTML [20]	33.6	32.7	**-1.76**	3.79	12.7	1.10
RSR [8]	34.5	32.6	-6.71	**2.98**	12.9	1.07
DQC-Plain (ours)	35.2	34.1	-6.22	3.32	18.1	1.19
DQC-Rank (ours)	**36.1**	**34.2**	-4.94	3.05	**21.4**	**1.26**

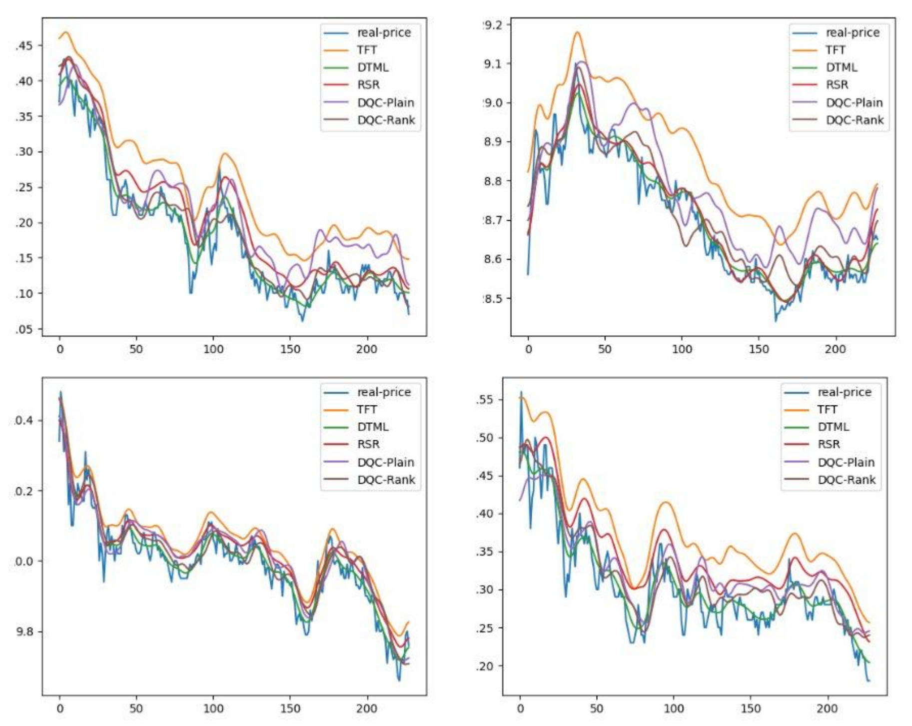
Results of CSI-21. We show the results on CSI-21 in Table 3, which collects *per-minute* high-frequency trading data in China market. We summarize the results as follows.

KDD-17 and ACL-18

- *On both datasets, DQC-Rank yields a highest Sharpe Ratio (SR)* with its better profit-to-volatility feature. E.g., on ACL-18 DQC-Rank has a 5% increase of SR compared with best baseline RSR (2.33 vs. 2.21); on KDD-17, DQC-Rank has a 8.2% better SR over RSR (1.59 vs. 1.47) as well.
- *DQC-Rank yields the best Gain*, leading the RSR and DTML by more than 5% relatively on ACL-18. DQC-Rank has a 28% higher SR than DQC-Plain (2.33 v.s. 1.82), as its rank learning better regularizes model training and reduces volatility. Similar trends are also with KDD dataset.
- *LightGBM yields the lowest Gains* due to their lacked capacity of performing complex temporal learning.

Visualizations





Conclusion

- We study stock movement forecasting as a fine-grained quantile classification task, with designed learning-to-rank tasks to explore global context of the market and internal moving patterns of an individual stock.
- Our model achieves significant improvement on realistic datasets with various evaluation metrics.
- In future work, we can apply our work to other financial assets such as futures and options, carbon credits, commodities, and more.
- This study can potentially benefit individual investors by helping them anticipate market risks and minimize losses, as well as policy makers who can take early action based on the market prices of agricultural products to promote social welfare.

THANKS

QA