Learning Multiple Stock Trading Patterns with Temporal Routing Adaptor and Optimal Transport

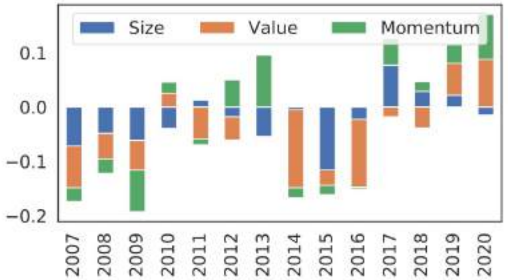
Hengxu Lin*, Dong Zhou*, Weiqing Liu, Jiang Bian

Microsoft Research · Sun Yat-sen University

KDD 2021 · 27th SIGKDD Conference on Knowledge Discovery and Data Mining · Online · August 14–18, 2021

Problem

- Accurate stock prediction is of vital importance for successful investment.
- However, as investors usually hold various beliefs and strategies to guide their trading, there will be multiple patterns in the stock market data.
- A single model often fails to capture such diverse or even contrary trading patterns.



Highlights

- We are the first to design stock prediction solutions to address the challenge of multiple trading patterns.
- We propose *Temporal Routing Adaptor (TRA)* to empower existing stock prediction models with the ability to model multiple trading patterns.
- We design an *optimal transport (OT)* based learning algorithm to further guarantee the discovery of multiple trading patterns.
- Our model outperforms the state-of-the-art by 11.3% in terms of IC and achieves higher annualized return and Sharpe ratio in real-world experiments.

Temporal Routing Adaptor

Temporal Routing Adaptor: Predictors + Router

Predictors:

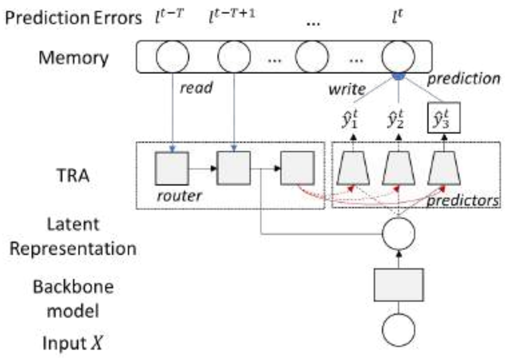
- Responsible for representing multiple patterns.

Router:

- Responsible for detecting a sample's underlying pattern and assign to the appropriate predictor.

Router can utilize two types of information to "predict" $p(y_t|\mathbf{x}_t)$:

- Latent representation: $p(\hat{y}_t|\mathbf{x_t})$.
- Temporal prediction errors: $p(y_{<t}|\mathbf{x}_{<t})$.



Optimal Transport

Optimal Transport: Objective + Constraint

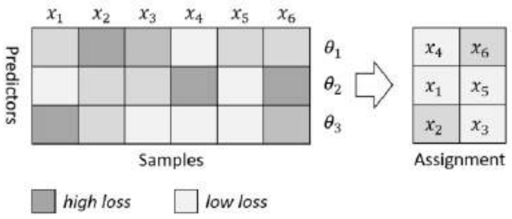
- **Constraint**: keep the assigned samples to different predictors almost balanced
- **Objective**: assign samples to predictors to minimize the overall loss

$$\min_{\mathrm{P}} \langle \mathrm{P}, \mathrm{L} \rangle,$$

$$s.t. \sum_{i=1}^{N} \mathrm{P}_{ik} = \nu_k * N, \ \forall k = 1 \ldots K$$

$$\sum_{k=1}^{K} \mathrm{P}_{ik} = 1, \ \forall i = 1 \ldots N$$

$$\mathrm{P}_{ik} \in \{0, 1\}, \ \forall i = 1 \ldots N, k = 1 \ldots K,$$



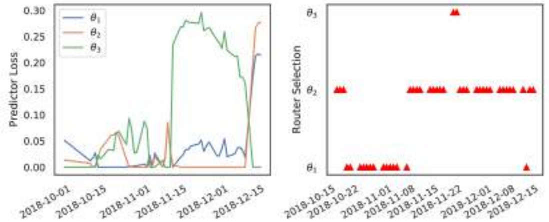
$$\min_{\Theta,\pi,\psi} \mathbb{E}_{(\mathbf{x}_i,\mathbf{y}_i)\in\mathcal{D}^{\text{train}}} \left[\ell(\mathbf{x}_i, \mathbf{y}_i; \Theta, \pi, \psi) \boxed{- \lambda \sum_{k=1}^{K} \mathrm{P}_{ik} \log(q_{ik})}\right]$$

Results

TRA beats all state-of-the-art stock prediction methods:

Method	Ranking Metrics				Portfolio Metrics			
	MSE (↓)	MAE (↓)	IC (↑)	ICIR (↑)	AR (↑)	AV (↓)	SR (↑)	MDD (↓)
Linear	0.163	0.327	0.020	0.132	-3.2%	16.8%	-0.191	32.1%
LightGBM	0.160(0.000)	0.323(0.000)	0.041	0.292	7.8%	15.5%	0.503	25.7%
MLP	0.160 (0.002)	0.323 (0.003)	0.037	0.273	3.7%	15.3%	0.264	26.2%
SFM	0.159 (0.001)	0.321 (0.001)	0.047	0.381	7.1%	14.3%	0.497	22.9%
ALSTM	0.158 (0.001)	0.320 (0.001)	0.053	0.419	12.3%	**13.7%**	0.897	**20.2%**
Trans.	0.158 (0.001)	0.322 (0.001)	0.051	0.400	14.5%	14.2%	1.028	22.5%
ALSTM+TS	0.160 (0.002)	0.321 (0.002)	0.039	0.291	6.7%	14.6%	0.480	22.3%
Trans.+TS	0.160 (0.004)	0.324 (0.005)	0.037	0.278	10.4%	14.7%	0.722	23.7%
ALSTM+TRA (Ours)	**0.157 (0.000)**	**0.318 (0.000)**	**0.059**	**0.460**	12.4%	14.0%	0.885	20.4%
Trans.+TRA (Ours)	**0.157 (0.000)**	0.320 (0.000)	0.056	0.442	**16.1%**	14.2%	**1.133**	23.1%

TRA can adaptively select predictor with smaller loss:



TRA benefits from both the latent representation and temporal prediction errors:

Information	MSE (↓)	MAE (↓)	IC (↑)	ICIR (↑)
Random	0.159 (0.001)	0.321 (0.002)	0.048	0.362
LR	0.158 (0.001)	0.320 (0.001)	0.053	0.409
TPE	0.158 (0.001)	0.321 (0.001)	0.049	0.381
LR+TPE	**0.157 (0.000)**	**0.318 (0.000)**	**0.059**	**0.460**

A moderate selection of the number of predictors could give desirable performance gains