StockMixer: A Simple Yet Strong MLP-Based Architecture for Stock Price Forecasting

Jinyong Fan, Yanyan Shen
Shanghai Jiao Tong University

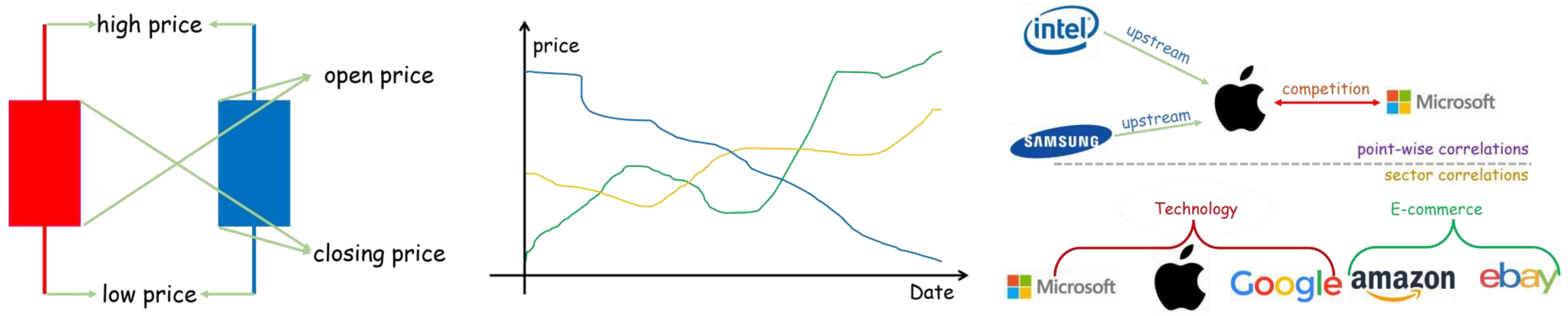
上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

The AAAI Conference on Artificial Intelligence

Paper, code, and data are available:
https://github.com/SJTU-Quant/StockMixer

Introduction

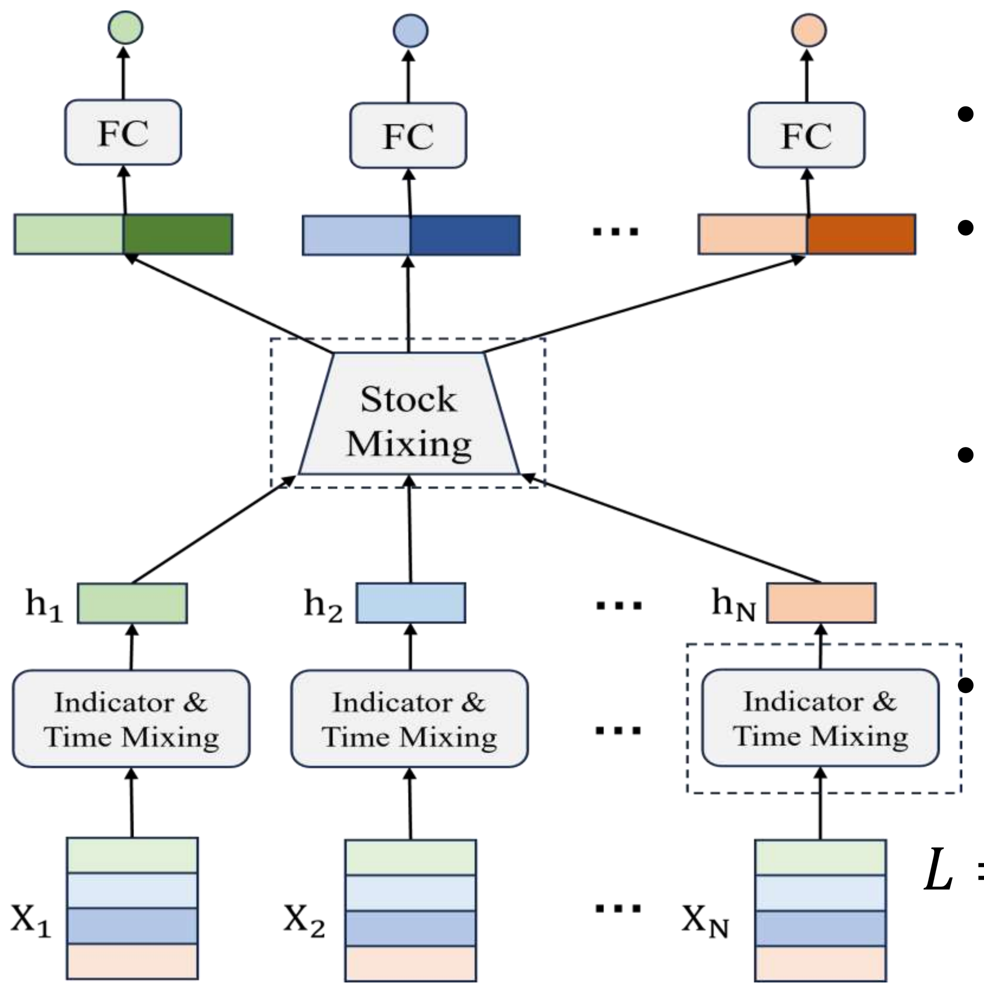
- **Stock price forecasting** is a fundamental task in the field of quantitative investment.
- As the stock market is highly *volatile* and *chaotic*, achieving high forecasting accuracy remains an open question.
- There exist three correlations in stock price data:



Could we develop a simple neural architecture that is easy to optimize and enjoys strong predictive performance by modeling the above-mentioned correlations effectively?

Our contributions:
- We propose a lightweight and effective MLP-based architecture for stock price forecasting.
- We introduce patch-based multi-scale time mixing and market-aware stock mixing that exploits the characteristics of stock patterns.

Overview of the proposed StockMixer



- **Inputs:** $X = \{X_1, \dots, X_N\},\ X_i \in R^{T\times F}$
- **Indicator & Time Mixing:**

$$h_i = \text{Indicator\&TimeMixing}(X_i)$$

- **StockMixing:** $H = \{h_1, \dots, h_N\}$

$$\hat{H} = \text{concat}(H, \text{StockMixing}(H))$$

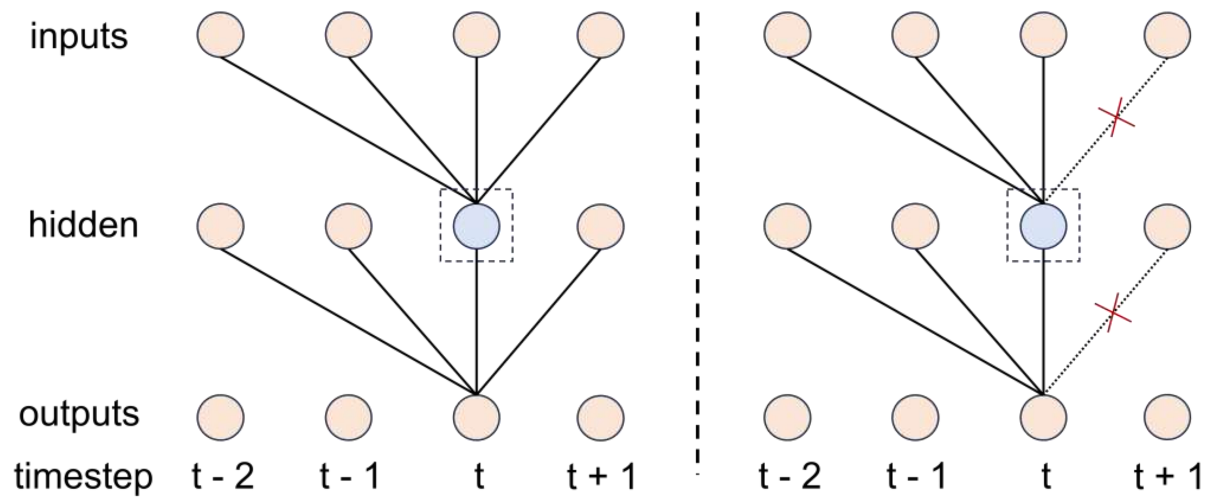
- **Loss function:**

$$L = L_{MSE} + \alpha \sum_{i=1}^{N}\sum_{j=1}^{N} \max(0, -(\hat{r}_i^t - \hat{r}_j^t)(r_i^t - r_j^t))$$

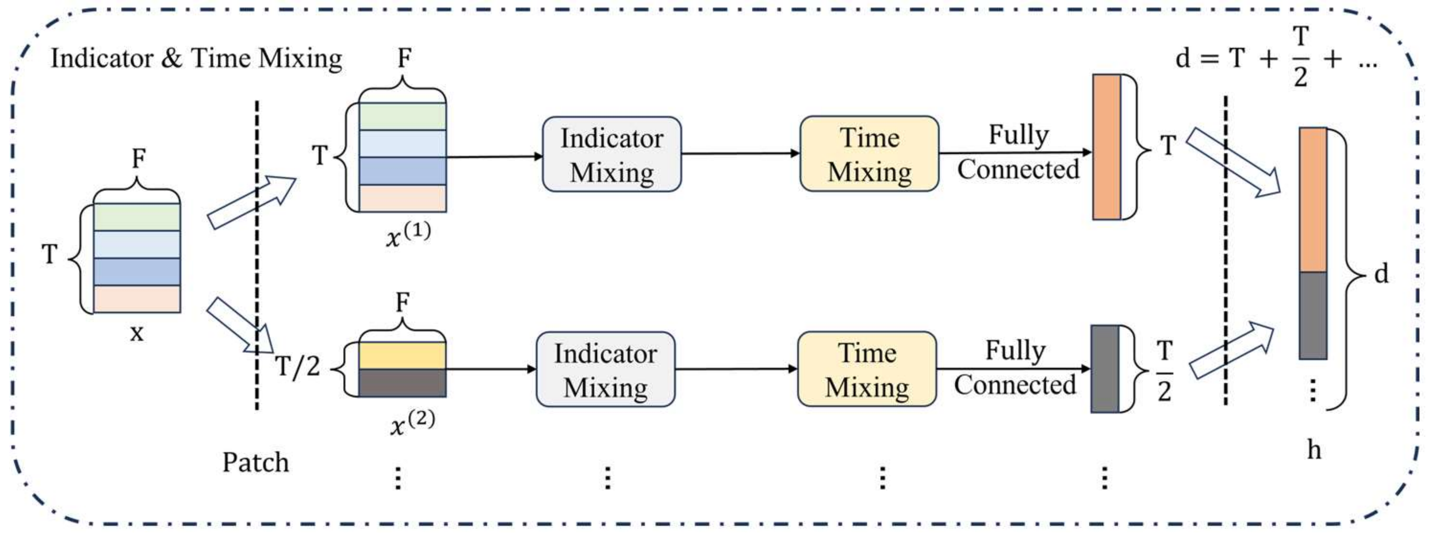
Indicator & Time Mixing

- **Indicator Mixing** is consistent with standard MLP-based mixing:

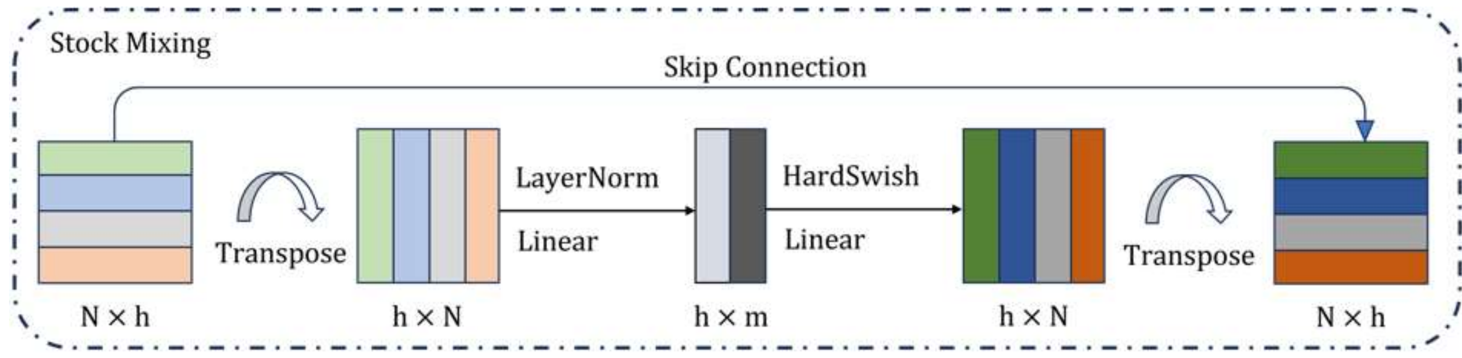
$$\hat{x}^T = x^T + W_2\sigma(W_1\text{LayerNorm}(x^T))$$



- Replace the weights with upper triangular matrix U_1, U_2 : $h = \hat{x} + \mathbf{U}_2\sigma(\mathbf{U}_1\text{LayerNorm}(\hat{x}))$,



Stock Mixing



- **Stock Mixing** replaces the hidden dimension of standard mixing related to stocks with a hyperparameter m :

$$\hat{H} = H + \mathbf{M}_2\sigma(\mathbf{M}_1\text{LayerNorm}(H))$$

- **Time Mixing** segments time sequence into subsequence-level patches and mix features at k scales.

$$x^{(k)} = \text{Avgpool}(x)_{\text{kernel}=k}, k \in \{\frac{T}{2}, \frac{T}{4}, \dots, 1\},$$

$$h^{(k)} = \text{TimeMixing}(\text{IndicatorMixing}(x^{(k)})),$$

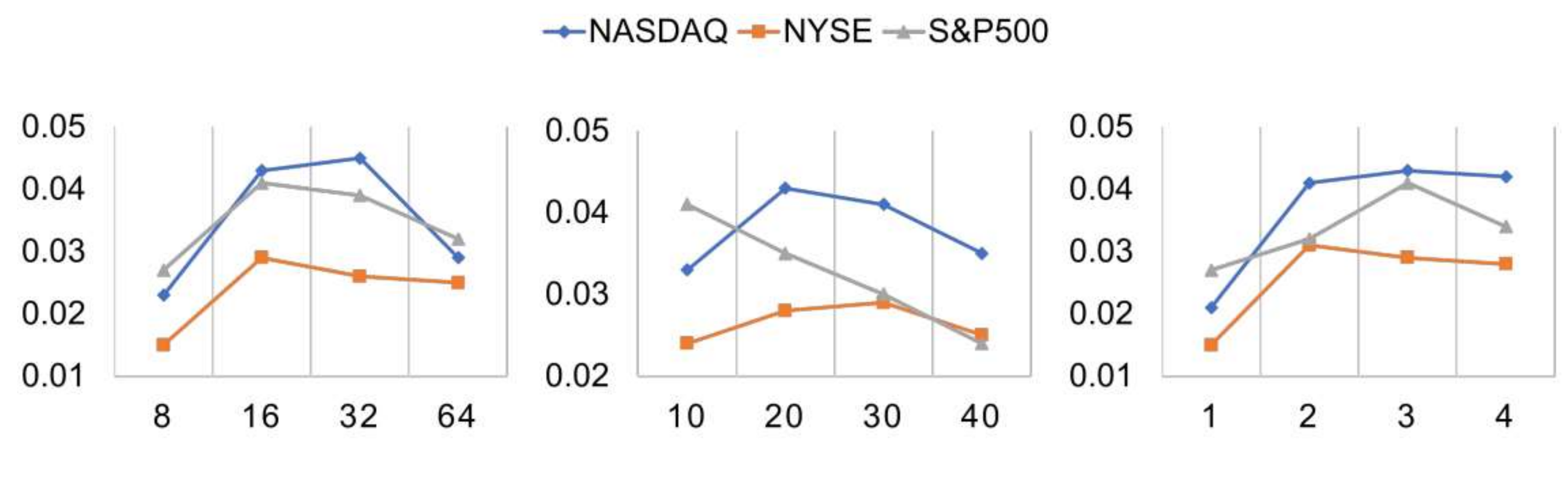
$$h = \text{FC}(\text{concat}(h^{(k)})), k \in \{\frac{T}{2}, \frac{T}{4}, \dots, 1\}.$$

Experiments

Model	NASDAQ				NYSE				S&P500			
	IC	RIC	prec@N	SR	IC	RIC	prec@N	SR	IC	RIC	prec@N	SR
LSTM	0.032	0.354	0.514	0.892	0.024	0.256	0.512	0.857	0.031	0.186	0.531	1.332
ALSTM	0.035	0.371	0.522	0.941	0.023	0.276	0.519	0.764	0.029	0.181	0.532	1.298
RGCN	0.034	0.382	0.516	1.054	0.025	0.275	0.517	0.932	0.028	0.175	0.528	1.359
GAT	0.035	0.377	0.530	1.233	0.025	0.297	0.521	1.070	0.034	0.191	0.541	1.484
RSR-I	0.038	0.398	0.531	1.238	0.026	0.284	0.519	0.098	0.033	0.200	0.542	1.437
STHAN-SR	0.039	0.451	0.543	1.416	0.029	0.344	**0.542**	1.228	0.037	0.227	0.549	1.533
ESTIMATE	0.037	0.444	0.539	1.307	**0.030**	0.327	0.536	1.115	0.035	0.241	**0.553**	1.547
Linear	0.019	0.188	0.505	0.517	0.015	0.163	0.497	0.625	0.016	0.156	0.520	0.674
StockMixer	**0.043**	**0.501**	**0.545**	**1.465**	0.029	**0.351**	0.539	**1.454**	**0.041**	**0.262**	0.551	**1.586**

Model	# of params	training time(s)
LSTM	14.28K	3.011
LSTM-RGCN	8464K	8.976
GAT	1039K	7.388
RSR-I	9493K	6.195
STHAN-SR	56.72K	7.085
ESTIMATE	185.3K	6.678
StockMixer	45.08K	4.242

Ablation	NASDAQ		NYSE	
Model Component	**IC**	**RIC**	**IC**	**RIC**
LSTM	0.032	0.354	0.024	0.256
w.o.Indicator Mixing	0.040	0.465	0.027	0.291
w.o.Time Mixing	0.018	0.164	0.016	0.161
w.o.Stock Mixing	0.037	0.376	0.026	0.285
LSTM + Stock Mixing	0.041	0.476	**0.030**	0.307
STHAN-SR	0.039	0.451	0.029	0.344
StockMixer	**0.043**	**0.501**	0.029	**0.351**



(a) Window length (b) Market (c) Scale factor