Stock Price Prediction via Discovering Multi-Frequency Trading Patterns

Liheng Zhang
University of Central Florida
4000 Central Florida Blvd.
Orlando, Florida 32816
lihengzhang1993@knights.ucf.edu

Charu Aggarwal
IBM T. J. Watson Research Center
New York 10598
charu@us.ibm.com

Guo-Jun Qi*
University of Central Florida
4000 Central Florida Blvd.
Orlando, Florida 32816
guojun.qi@ucf.edu

SRT Group

Zhang Chuheng

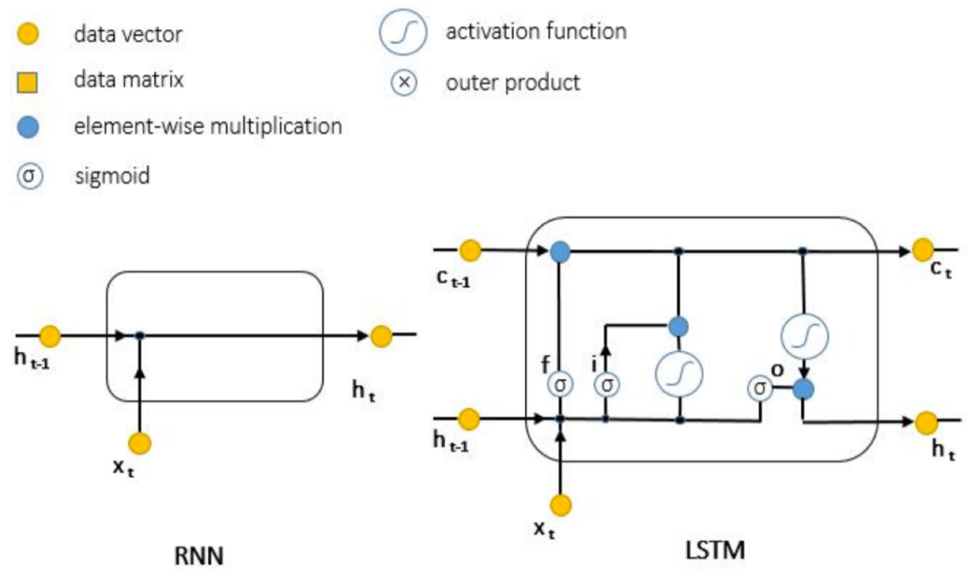
2018-05-12

Motivation

- Strategies are based on different patterns on different frequencies
- **DFT** transform time domain to frequency domain
- **RNN** can learn temporal patterns from nonlinear and non-stationary data
- **LSTM**, a variant of RNN, can capture long-term dependency of stock price

Recap: LSTM

$$\boldsymbol{i_t} = \text{sigmoid}(\boldsymbol{W_i x_t} + \boldsymbol{U_i h_{t-1}} + \boldsymbol{b_i})$$
$$\boldsymbol{f_t} = \text{sigmoid}(\boldsymbol{W_f x_t} + \boldsymbol{U_f h_{t-1}} + \boldsymbol{b_f})$$
$$\boldsymbol{\tilde{c}_t} = \tanh(\boldsymbol{W_c x_t} + \boldsymbol{U_c h_{t-1}} + \boldsymbol{b_c})$$
$$\boldsymbol{c_t} = \boldsymbol{i_t} \circ \boldsymbol{\tilde{c}_t} + \boldsymbol{f_t} \circ \boldsymbol{c_{t-1}}$$
$$\boldsymbol{o_t} = \text{sigmoid}(\boldsymbol{W_o x_t} + \boldsymbol{U_o h_{t-1}} + \boldsymbol{V_o c_t} + \boldsymbol{b_o})$$
$$\boldsymbol{h_t} = \boldsymbol{o_t} \circ \tanh(\boldsymbol{c_t})$$

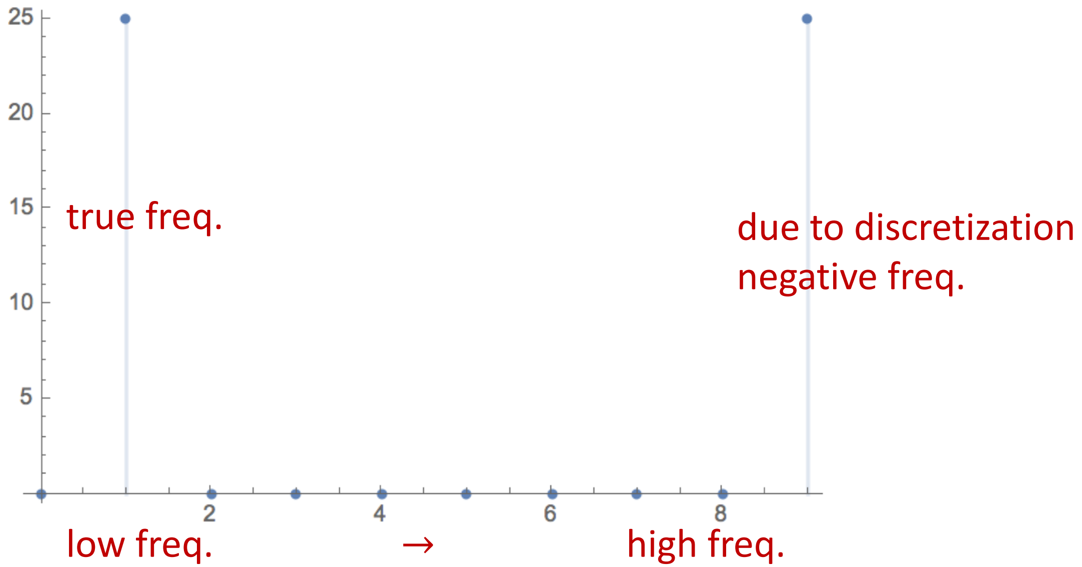


- **Input gate** regulates the allowed amount of new information flowing into the memory cell
- **Forget gate** controls how much information should be kept in the cell
- **Output gate** defines the amount of information that can be output

Recap: DFT

- $X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn}$, e.g. when $x_n = \sin(\frac{2\pi}{10}n)$, with $T = 10$, $f = \frac{1}{10}$, base frequencies are $\{0(const.), \frac{1}{10}, \frac{2}{10}, \cdots, \frac{9}{10}\}$

```
DiscretePlot[(Σ_{n=0}^{9} Sin[2π/10 n] Cos[2π/10 k n])^2 + (Σ_{n=0}^{9} Sin[2π/10 n] Sin[2π/10 k n])^2, {k, 0, 9}]
```



Recap: DFT

连续Fourier变换

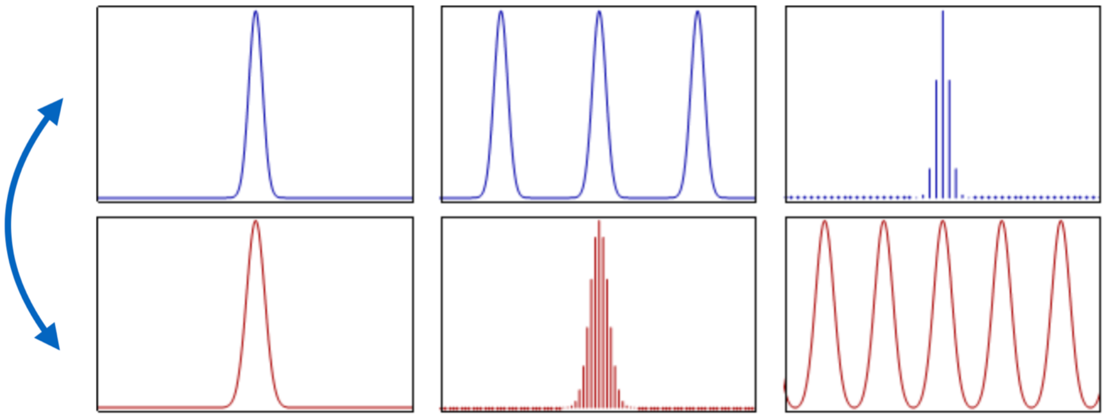
$$F(k)=\int_{-\infty}^{+\infty} f(x)\exp(-ikx)\,dx$$

离散Fourier变换

$$F(k_m)=\sum_{n=0}^{N-1} f(x_n)\exp(-i2\pi\frac{m}{N dx} n\,dx)$$

k　x

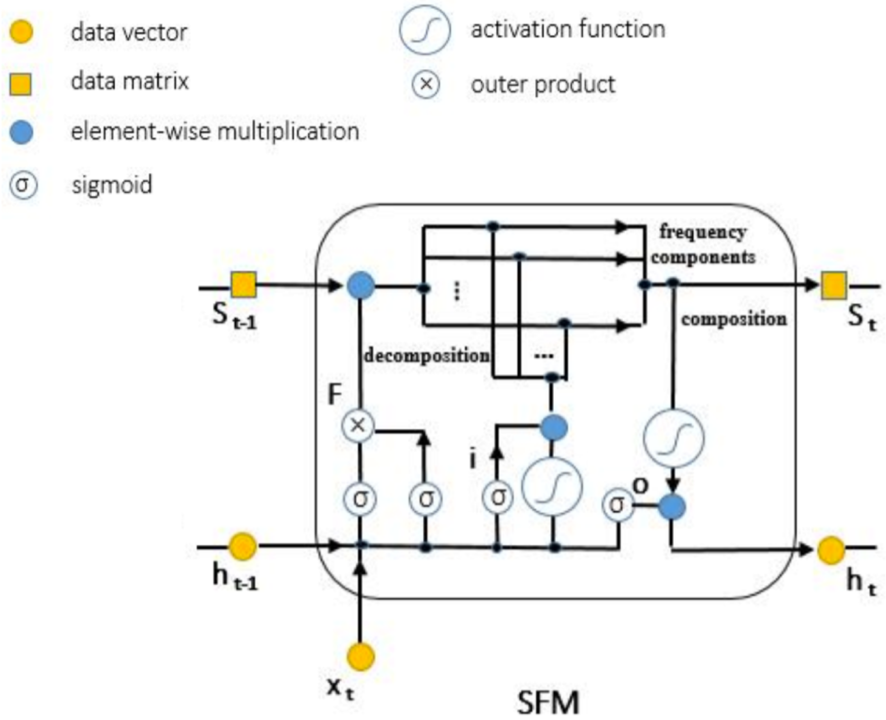
- x空间周期性 ⟺ k空间离散
- x空间离散 ⟺ k空间周期性



SFM

$$
\begin{aligned}
\boldsymbol{i_t} &= \text{sigmoid}(\boldsymbol{W_i x_t} + \boldsymbol{U_i h_{t-1}} + \boldsymbol{b_i}) \\
\boldsymbol{f_t} &= \text{sigmoid}(\boldsymbol{W_f x_t} + \boldsymbol{U_f h_{t-1}} + \boldsymbol{b_f}) \\
\boldsymbol{\tilde{c}_t} &= \tanh(\boldsymbol{W_c x_t} + \boldsymbol{U_c h_{t-1}} + \boldsymbol{b_c}) \\
\boldsymbol{c_t} &= \boldsymbol{i_t} \circ \boldsymbol{\tilde{c}_t} + \boldsymbol{f_t} \circ \boldsymbol{c_{t-1}} \\
\boldsymbol{o_t} &= \text{sigmoid}(\boldsymbol{W_o x_t} + \boldsymbol{U_o h_{t-1}} + \boldsymbol{V_o c_t} + \boldsymbol{b_o}) \\
\boldsymbol{h_t} &= \boldsymbol{o_t} \circ \tanh(\boldsymbol{c_t})
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{i_t} &= \text{sigmoid}(\boldsymbol{W_i x_t} + \boldsymbol{U_i h_{t-1}} + \boldsymbol{b_i}) \\
\boldsymbol{f_t^{ste}} &= \text{sigmoid}(\boldsymbol{W_{ste} x_t} + \boldsymbol{U_{ste} h_{t-1}} + \boldsymbol{b_{ste}}) \in \mathbb{R}^D \\
\boldsymbol{f_t^{fre}} &= \text{sigmoid}(\boldsymbol{W_{fre} x_t} + \boldsymbol{U_{fre} h_{t-1}} + \boldsymbol{b_{fre}}) \in \mathbb{R}^K \\
\boldsymbol{F_t} &= \boldsymbol{f_t^{ste}} \otimes \boldsymbol{f_t^{fre}} \in \mathbb{R}^{D \times K} \\
\boldsymbol{\tilde{c}_t} &= \tanh(\boldsymbol{W_c x_t} + \boldsymbol{U_c h_{t-1}} + \boldsymbol{b_c}) \\
\boldsymbol{S_t} &= \boldsymbol{F_t} \circ \boldsymbol{S_{t-1}} + (\boldsymbol{i_t} \circ \boldsymbol{\tilde{c}_t}) \begin{bmatrix} e^{j\omega_1 t} \\ e^{j\omega_2 t} \\ \dots \\ e^{j\omega_K t} \end{bmatrix}^T \in \mathbb{C}^{D \times K} \\
\boldsymbol{c_t} &= \tanh(\boldsymbol{A_t u_a} + \boldsymbol{b_a}) \\
\boldsymbol{o_t} &= \text{sigmoid}(\boldsymbol{W_o x_t} + \boldsymbol{U_o h_{t-1}} + \boldsymbol{V_o c_t} + \boldsymbol{b_o}) \\
\boldsymbol{h_t} &= \boldsymbol{o_t} \circ \boldsymbol{c_t}
\end{aligned}
$$

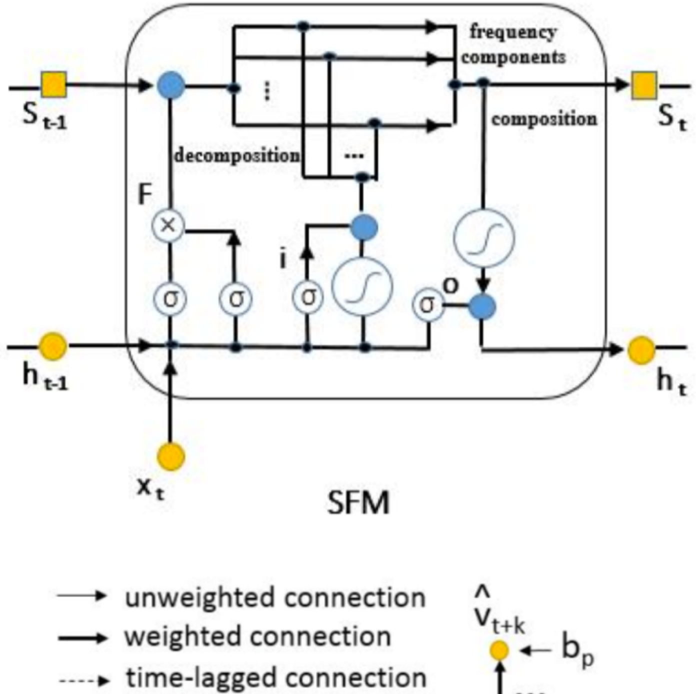


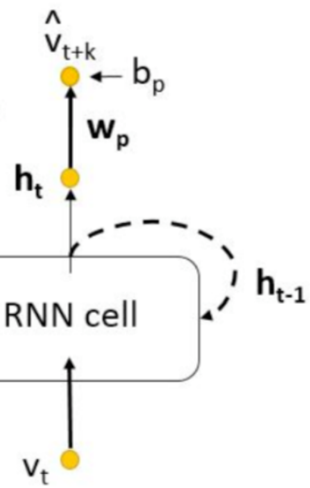
$$
\boldsymbol{A_t} = |\boldsymbol{S_t}| = \sqrt{(Re\boldsymbol{S_t})^2 + (Im\boldsymbol{S_t})^2} \in \mathbb{R}^{D \times K}
$$

$$
\angle \boldsymbol{S_t} = \arctan\left(\frac{Im\boldsymbol{S_t}}{Re\boldsymbol{S_t}}\right) \in [-\frac{\pi}{2}, \frac{\pi}{2}]^{D \times K}
$$

Price Prediction

$$
\begin{aligned}
\boldsymbol{i_t} &= \mathrm{sigmoid}(\boldsymbol{W_i x_t} + \boldsymbol{U_i h_{t-1}} + \boldsymbol{b_i}) \\
\boldsymbol{f_t^{ste}} &= \mathrm{sigmoid}(\boldsymbol{W_{ste} x_t} + \boldsymbol{U_{ste} h_{t-1}} + \boldsymbol{b_{ste}}) \in \mathbb{R}^D \\
\boldsymbol{f_t^{fre}} &= \mathrm{sigmoid}(\boldsymbol{W_{fre} x_t} + \boldsymbol{U_{fre} h_{t-1}} + \boldsymbol{b_{fre}}) \in \mathbb{R}^K \\
\boldsymbol{F_t} &= \boldsymbol{f_t^{ste}} \otimes \boldsymbol{f_t^{fre}} \in \mathbb{R}^{D \times K} \\
\boldsymbol{\tilde{c}_t} &= \tanh(\boldsymbol{W_c x_t} + \boldsymbol{U_c h_{t-1}} + \boldsymbol{b_c}) \\
\boldsymbol{S_t} &= \boldsymbol{F_t} \circ \boldsymbol{S_{t-1}} + (\boldsymbol{i_t} \circ \boldsymbol{\tilde{c}_t}) \begin{bmatrix} e^{j\omega_1 t} \\ e^{j\omega_2 t} \\ \cdots \\ e^{j\omega_K t} \end{bmatrix}^T \in \mathbb{C}^{D \times K} \\
\boldsymbol{c_t} &= \tanh(\boldsymbol{A_t u_a} + \boldsymbol{b_a}) \\
\boldsymbol{o_t} &= \mathrm{sigmoid}(\boldsymbol{W_o x_t} + \boldsymbol{U_o h_{t-1}} + \boldsymbol{V_o c_t} + \boldsymbol{b_o}) \\
\boldsymbol{h_t} &= \boldsymbol{o_t} \circ \boldsymbol{c_t}
\end{aligned}
$$





Price prediction is a simple linear transform of latest hidden state $\hat{v}_{t+n} = \boldsymbol{w_p h_t} + b_p$

Loss is the standard mean squared loss

$$\mathcal{L} = \sum_{m=1}^{M} \sum_{t=1}^{T} (v_{t+n}^m - \hat{v}_{t+n}^m)^2$$

v is normalized price

Experiment

- Dataset: top five stocks with largest market capitalization in each sector in US market

basic materials	cyclicals [1]	energy	financials	healthcare	industrials	non-cyclicals [2]	technology	telecommunications [3]	utilities
BHP	AMZN	CVX	BAC	JNJ	BA	KO	AAPL	CHL	D
DOW	CMCSA	PTR	BRK-B	MRK	GE	MO	GOOGL	DCM	DUK
RIO	DIS	RDS-B	JPM	NVS	MA	PEP	INTC	NTT	EXC
SYT	HD	TOT	SPY	PFE	MMM	PG	MSFT	T	NGG
VALE	TM	XOM	WFC	UNH	UPS	WMT	ORCL	VZ	SO

- Result: smaller squared error compared with AR and LSTM

	1-step	3-step	5-step
AR	6.01	18.58	30.74
LSTM	5.93	18.38	30.02
SFM	**5.57**	**17.00**	**28.90**

Analysis

$$\boldsymbol{S}_t = \boldsymbol{F}_t \circ \boldsymbol{S}_{t-1} + (\boldsymbol{i}_t \circ \tilde{\boldsymbol{c}}_t) \begin{bmatrix} e^{j\omega_1 t} \\ e^{j\omega_2 t} \\ \cdots \\ e^{j\omega_K t} \end{bmatrix}^T \in \mathbb{C}^{D \times K}$$

- Hidden state dimension: #patterns expected to explore on each frequency

# of states	10	20	40	50
1-step	**5.57**	5.79	6.15	5.91
3-step	18.48	19.20	17.25	**17.00**
5-step	29.48	29.84	31.30	**28.90**

- #frequencies: how many frequency levels expected to explore

# of frequencies	5	10	15	20
1-step	6.69	5.91	5.91	**5.88**
3-step	18.39	**17.00**	19.15	19.52
5-step	30.95	**28.9**	30.57	31.22

注

- 关于 $\boldsymbol{S_t} = \boldsymbol{F_t} \circ \boldsymbol{S_{t-1}} + (\boldsymbol{i_t} \circ \boldsymbol{\tilde{c}_t}) \begin{bmatrix} e^{j\omega_1 t} \\ e^{j\omega_2 t} \\ \dots \\ e^{j\omega_K t} \end{bmatrix}^T \in \mathbb{C}^{D \times K}$
  - S的储存一直是复数，直到output的时候才变成幅值
  - 如果认为F都是1（不遗忘），那么S可以展开为$S_t = \sum_{t'=1}^{t} x_t{}' \ \Omega_{t'}$，$x_t = i_t \circ \widetilde{c_t}$，**Ω是傅里叶基底**，那么一行是对于$d \in [D]$**这个**hidden state的时序上面的傅里叶变换