Predicting Stock Buy and Sell Points Based on Machine Learning through Technical Indicators

Roy

20241007

Motivation

- Current state of stock prediction: direct data input into deep learning models
- Common use of LSTM and RNN, but lacking feature engineering
- Feature engineering enhances model performance
- PCA can be combined with other techniques, potentially improving accuracy

Methodology

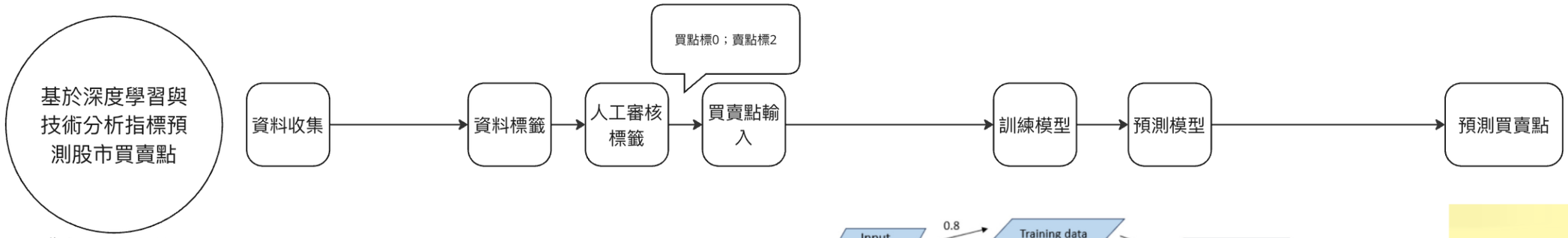
- Data source: Taiwan 50 Index
- TA-lib package for calculating technical indicators
- EDA, feature engineering, and PCA application
- Model selection: CNN, LSTM, combining different deep learning models
- Predicting stock buy/sell points

Goal

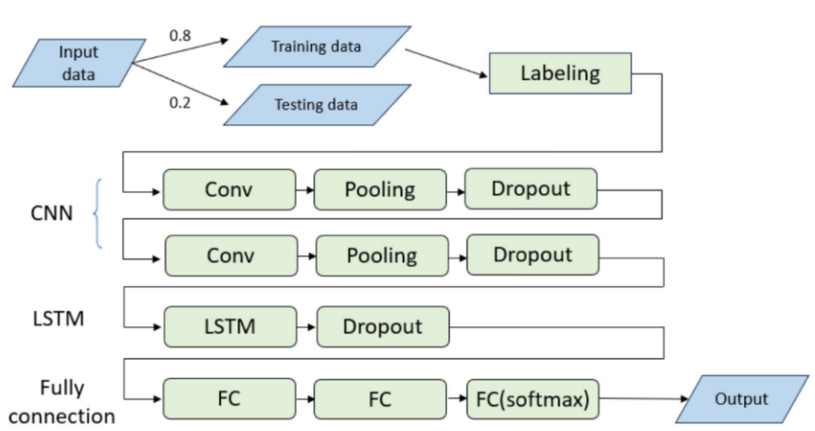
- Model evaluation metric: MSE (Mean Squared Error)Comparison of CNN and LSTM training
- Exploring the impact of feature engineering and PCA on model performance
- https://miro.com/welcomeonboard/YzhVZzZjUWZoVzl2dWlRczVEZ1ZUVnM5WGFhTzY0WGRFUFNEVzNpU3hxMjUxOWJYaEZjdFd2TUkxUzByVDVsY3wzNDU4NzY0NjAxMDcxMjQwNTExfDI=?share_link_id=11471725507

計算方式為
將原有54 一組的K 棒在細分成 9 個一組，使用9 個K 棒中的 DIF 正負值去判斷趨勢，DIF 和 K 值在 9 根 K 棒中最小值(買點)處標上 0，DIF 和 K 值在9 根K 棒中的最大值(賣點)處則標上2。

圖 3- 5 標籤示意圖

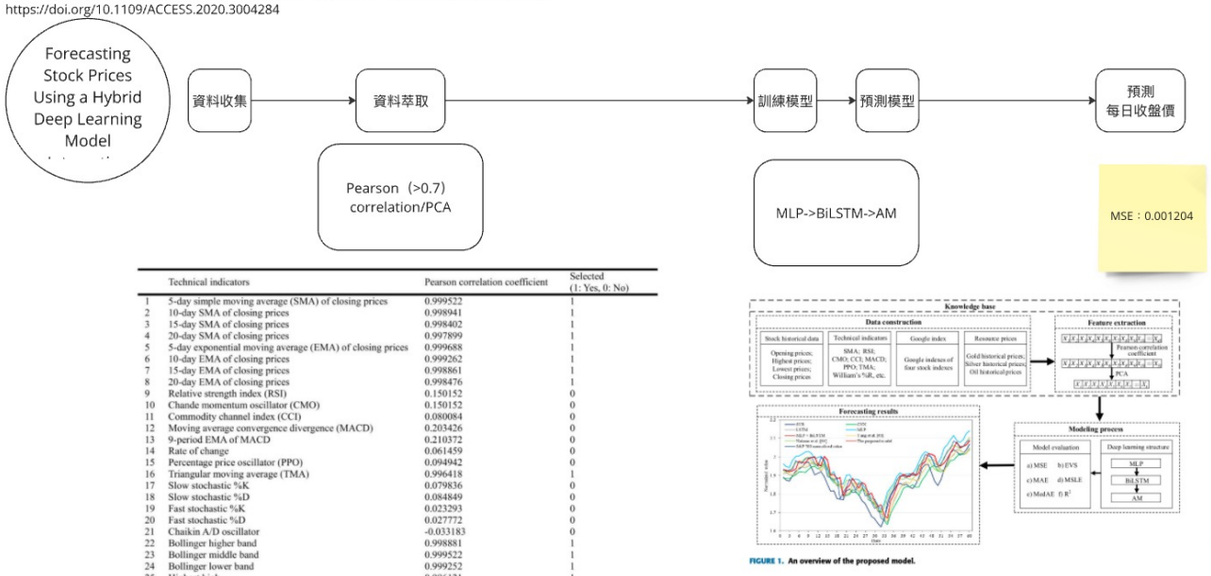


{鄭邦廷, 2023 #11}



MSE：0.1447

Chen, Q., Zhang, W., & Lou, Y. (2020). **Forecasting Stock Prices Using a Hybrid Deep Learning Model Integrating Attention Mechanism, Multi-Layer Perceptron, and Bidirectional Long-Short Term Memory Neural Network.** ***IEEE Access***, *8*, 117365-117376. https://doi.org/10.1109/ACCESS.2020.3004284

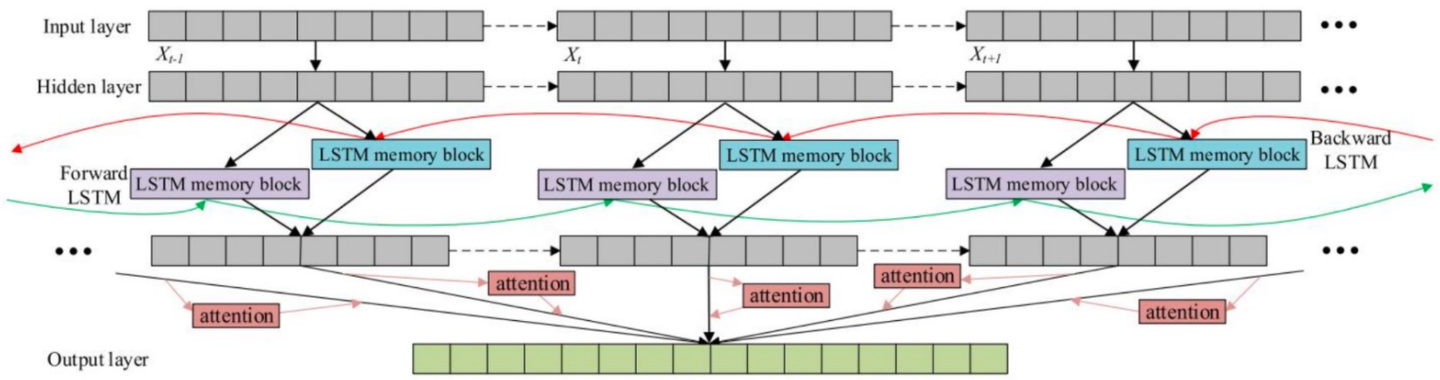


	Technical indicators	Pearson correlation coefficient	Selected (1: Yes, 0: No)
1	5-day simple moving average (SMA) of closing prices	0.999522	1
2	10-day SMA of closing prices	0.998941	1
3	15-day SMA of closing prices	0.998402	1
4	20-day SMA of closing prices	0.997899	1
5	5-day exponential moving average (EMA) of closing prices	0.999688	1
6	10-day EMA of closing prices	0.999262	1
7	15-day EMA of closing prices	0.998861	1
8	20-day EMA of closing prices	0.998476	1
9	Relative strength index (RSI)	0.150152	0
10	Chande momentum oscillator (CMO)	0.150152	0
11	Commodity channel index (CCI)	0.080084	0
12	Moving average convergence divergence (MACD)	0.203426	0
13	9-period EMA of MACD	0.210372	0
14	Rate of change	0.061459	0
15	Percentage price oscillator (PPO)	0.094942	0
16	Triangular moving average (TMA)	0.996418	1
17	Slow stochastic %K	0.079836	0
18	Slow stochastic %D	0.084849	0
19	Fast stochastic %K	0.023293	0
20	Fast stochastic %D	0.027772	0
21	Chaikin A/D oscillator	-0.033183	0
22	Bollinger higher band	0.998881	1
23	Bollinger middle band	0.999522	1
24	Bollinger lower band	0.999252	1
25	Highest high	0.996121	1
26	Lowest low	0.996121	1
27	William's %R	0.112234	0

FIGURE 1. An overview of the proposed model.

Methods	EVS	MAE	MSE	MSLE	MedAE	R^2
SVR	**0.943496**	0.049373	0.003009	0.000359	0.046726	0.716351
LSTM	0.853031	0.039465	0.002207	0.000266	0.033488	0.760591
CNN	0.703118	0.072582	0.006661	0.000792	0.073571	0.006610
MLP	0.901269	0.041324	0.002882	0.000338	0.038708	0.719088
MLP + BiLSTM	0.869217	0.025827	0.001230	0.000151	0.018968	0.867302
Yang et al. [63]	0.828717	0.039052	0.002147	0.000257	0.034421	0.752297
Nakano et al. [64]	0.903007	0.028608	0.001320	0.000161	0.024855	0.845055
The proposed model	0.869727	**0.025393**	**0.001204**	**0.000148**	**0.017290**	**0.869326**

Note: Significant values are boldfaced.



Bukhari, A. H., Raja, M. A. Z., Sulaiman, M., Islam, S., Shoaib, M., & Kumam, P. (2020). **Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting.** ***IEEE***

Logo and Links to Quantinar Courselets

- Use Quantinar icon and name as source

Logo and Links to Quantlet/GitHub

- Use Quantlet icon and name as source
- Hyperlink both to GitHub repository  Styleguide
- Change the presentation logo in the master slide (see View/Edit Master Slide, shortcut: Shift-Command-E)

LvB notations 1

- Use the formula creator within keynote ‘Insert/Equation’
- All operators are to be defined by \operatorname{}
  - without operatorname: ~~$argmax_i f(x_i)$~~
  - with operatorname: $\operatorname{argmax}_i f(x_i)$
- Equations covering multiple lines may be written aligned
- Use bracket sequence $[\{(\dots)\}]$
- Conventional bracket rules represent and exemption of the rule above. For example: $Y \sim \mathcal{N}(\mu(X), \sigma(X))$

LvB notations 2

- Use ^{\top} to write the transpose symbol: $x^\top x = \| x \|$
- Use \ldots to write the three dots symbol: $x \in \{1, \ldots, n\}$ $\hat{}$ $\tilde{}$
- Use \widehat{} and \widetilde{} rather than \hat{}, \tilde{}: $\widehat{Y}, \widetilde{Y}$
- Write norms via \|: $\| x \|$

LvB notations 3

- The for convergence may be written with \mathcal{O}: $\mathcal{O}$
- The operator for exponential terms with Euler's number as the base is defined by \exp: $\exp(1) \approx 2.718$
- Use \overset{\mathcal{L}}{\rightarrow} to write the symbol for convergence in distribution and denote the normal distribution by \mathcal{N}, this produces $X \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$
- Use \overset{\operatorname{as.}}{\sim} to write the symbol for asymptotic distribution $X \overset{\text{as.}}{\sim} \chi^2$
- To define a function, variable etc. use \overset{\operatorname{def}}{=} $f(x) \overset{\text{def}}{=} ax + b$

LvB notations 4

- Use \log for the natural logarithm: $\log\{\exp(1)\} = 1$
- Use \mathsf{E} for expectation: $\mathsf{E}[X] = \mu$
- Use \operatorname{P} to write the symbol for probability: P
- Use \operatorname{\mathbf{I}} for the indicator function: $\mathbf{I}\{x < 1\}$
- Use \varepsilon instead of epsilon: $\not\epsilon \rightarrow \varepsilon$

Tables

- Follow the Cambridge University Press Style
- Round appropriately (as much information as necessary, as little as possible)
- Align decimal points

d	10	11	12
10%	2.2886	2.4966	2.6862
5%	2.5268	2.7444	2.9490
1%	3.0339	3.2680	3.4911

Figures

- Give informative axis labels
- If x- and y-axis are on the same domain, the plot should be square
- Use same color scheme for multiple plots if they show the same content.

TEN Template

Your Name

Your affiliation

[Your Webpage]

Repeat on last slide the lead picture