Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 126 (2018) 87–96

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2018, 3-5 September 2018, Belgrade, Serbia

Trading Decision of Taiwan Stocks with the Help of United States Stock Market

Shih-Chan Huang, Chang-Biau Yang*, Hung-Hsin Chen

Department of Computer Science and Engineering
National Sun Yat-sen University
70 Lienhai Rd., Kaohsiung 80424, Taiwan

Abstract

The paper studies how to improve the trading decision of Taiwan stocks with the information of US stock market. Our method first aligns the trading days between Taiwan and US stock markets. Next, the similarity between the *portfolio index* (PI, constructed from 100 Taiwan stocks) and one of the US stock indices, the *Dow Jones Industrial Average* (DJIA), *NASDAQ composite index* (NASDAQ), or *Standard & Poor's 500* (S&P 500), is computed, respectively. The trading signals of PI or each US stock index are generated by the method of Lee *et al.* Finally, the consensus signals of PI are determined by the majority vote scheme with the weighted functions, calculated from the similarity.

The testing period of PI starts from 2000/1/4 to 2017/12/29, totally 4480 days. As the experimental results show, the index combination (PI, DJIA, NASDAQ) with the weighted function $W^{(4)}$ is considered to be the best combination for trading PI. Its average annualized return (cumulative return) achieves 15.03% (1170.42%), which is better than the method of Lee *et al.* 13.88% (947.65%), and the buy-and-hold strategy 9.85% (442.90%).

Keywords: stock investment; Taiwan stock; US stock indices; trading signal; gene expression programming.

1. Introduction

Many researches have devoted to the trend prediction of stock market by using machine learning methods, such as the *support vector machine* (SVM)[1], *support vector regression* (SVR)[2,3] and *artificial neural network* (ANN)[4,5]. To earn profit, a variety of methods and strategies has been utilized to trade stocks in the market. Because the fluctuation of a stock price is a nonstationary time series, it is a big challenge to predict the trend of a stock precisely.

Some *evolution-based* algorithms have also been applied to evolve profitable trading strategies, such as the *genetic algorithm* (GA)[6,7], *genetic programming* (GP)[8,5,9], *genetic expression algorithm* (GEP)[10,11,12], *particle swarm opti-*

* Corresponding author. Tel.: +886-7-5252000 ext. 4333; fax: +886-7-5254301.
E-mail address: cbyang@cse.nsysu.edu.tw

10.1016/j.procs.2018.07.212

mization (PSO)[13] and *firefly algorithm* (FA)[3]. In addition to the trend prediction of the stock market, some researches predicted the value of the stock market index by using the *fuzzy time series*[14].

In this paper, we hybridize the method of Lee *et al.*[11] for generating the consensus trading signals, and the weighted functions to weight the signals of *portfolio index* (PI) and US stock indices as auxiliary information for generating consensus signal of PI. PI, defined by Lee *et al.*[11], is calculated from the stocks with the top 100 market capital on 1995/1/5 in Taiwan stock market. The method of Lee *et al.* assumed that there is a mean-reversion property in a stock market, and the GEP is applied to evolve suitable trading strategies from historical data. The dynamic time warping (DTW) is first used to get three template intervals, which are similar to the leading interval, on the trading days. Then, some pre-trained trading strategies of these three template intervals are extracted from the strategy pool. These chosen strategies are validated in the validation interval, and consensus trading signals can be finally generated by the majority vote on the suitable trading strategies.

Due to strong dependence between Taiwan and US stock markets and disjoint trading hours, the information of the previous closed US market can be utilized to predict the next opened Taiwan market. In other words, the time-shift property is utilized to predict the trend of Taiwan stock market by the information of US stock markets[15]. Our method computes the similarity between PI and US indices, including the *Dow Jones industrial average* (DJIA), *National association of securities dealers automated quotations composite index* (NASDAQ), and *Standard & Poor's 500 index* (S&P 500). The weighted functions, inspired from the odd ratio[12,16,17,18], are calculated from the similarity on each trading day. Thus, the consensus trading signals for PI are obtained from the combination of PI and US stock indices with the weighted functions.

The data set of US stock indices were obtained from *Yahoo finance*, and the data set of PI was obtained from Lee *et al.*[11]. The training period of trading strategies for DJIA, NASDAQ and S&P 500 is from 1996/1/2 to 2017/12/29, totally 5539 days. The training period of trading strategies for PI is from 1996/1/4 to 2017/12/29, totally 5591 days. The testing period is from 2000/1/4 to 2017/12/29, totally 4480 days. The best average annualized return (cumulative return) of PI with our method achieves 15.03% (1170.42%), which is better than the method of Lee *et al.* 13.88% (947.65%), and the buy-and-hold strategy 9.85% (442.90%).

The rest of this paper is organized as follows. Some preliminaries are introduced in Section 2. Our method is presented in Section 3. The similarity analysis, experimental results and comparison are described in Section 4. Finally, the conclusion and the future work are given in Section 5.

2. Preliminaries

The concept of the survival-of-the fittest from the biology is used in the evolutionary algorithms. The evolutionary algorithms, such as *gene expression programming* (GEP)[10] and *genetic programming* (GP)[8], are used for searching the optimal solution of one problem. GEP improves GP in two aspects. GEP utilizes the linear string (genotype) to implement the program, and the tree structure (phenotype) to represent a gene. GEP implementation is easier than GP, and the performance of GEP is more efficient than GP. The procedure of GEP is described as follows.

Step 1: Population creation: The initial population is generated by random chromosomes.

Step 2: Fitness value evaluation: The expression of each chromosome is the genotype initially, and then fitness value of each chromosome is evaluated with its phenotype.

Step 3: The condition of termination: Check whether the condition of termination is satisfied or not. If the termination condition is satisfied, stop the evolution. Otherwise, go to Step 4.

Step 4: Evolution operation: Preserve the best chromosome, and some chromosomes are picked with the elitist selection. These selected chromosomes are applied to a series of evolution operations with probability, such as mutation, transposition, and recombination, to generate probabilistically better chromosomes.

Step 5: New population: After a series of evolution operations, the new population is formed. Then, go to Step 2 to evolve the next generation.

The definition of PI is based on the average of the cumulative daily return of the 100 listed companies with the largest market capitalization on 1995/1/5 in Taiwan stock market, described as follows[11].

$$R_i(t) = \frac{C_i(t) - C_i(t-1)}{C_i(t-1)}, \quad PR(t) = \frac{\sum_{i=1}^{100} R_i(t)}{100}, \quad PI(t) = PI(t-1) \times (1 + PR(t)), \quad PI(0) = 100, \tag{1}$$

where $C_i(t)$, $R_i(t)$, $PR(t)$, and $PI(t)$ are the close price of the ith stock, the return of the ith stock, the return of the portfolio, and portfolio index on day t, respectively. The initial value of PI is set to 100 on January 5, 1995.

Lee *et al.*[11] utilized GEP to obtain the profitable strategies on PI. The trading signal, BUY, SELL, or WAIT, of each trading strategy is generated, and the consensus trading signal is determined by the majority vote scheme. Their method is presented as follows.

Step 1: Strategy training
A *template interval* is made up of every L days of the historical trading data. Every two adjacent template intervals have a distance of 10 days for reducing the training time. The 10 best profitable strategies are preserved for each template interval from the GEP training.

Step 2: Similar Template Searching
Use DTW to search historical template intervals which are similar to the leading interval. The *leading interval* is defined as day $(t - L + 1)$ to day t, totally L days, where t denotes the current trading day. The similarity relationship between the template interval and leading interval is adopted the sliding window scheme.

Step 3: Strategy Validation
The *validation interval* with L_v days is from day $(t - L_v + 1)$ to day t. The trading strategies selected in Step 2 are verified whether they are suitable or not in the validation interval. When the return of buy-and-hold (BAH) strategy is positive, a strategy is suitable if its return $> 0.8 \times BAH$. When $BAH \leq 0$, a strategy is suitable if its return is not negative.

Step 4: Determination of the consensus trading signal
Each suitable strategy generates the trading signals, BUY, SELL, or WAIT, on day t. Accordingly, the buying vote V^B, selling vote V^S, and waiting vote V^W are counted. Then, the following ratios are calculated.

$$\tau_A = \frac{V^B + V^S}{V^B + V^S + V^W}, \quad \tau_B = \frac{V^B}{V^B + V^S}, \quad \tau_S = \frac{V^S}{V^B + V^S}. \tag{2}$$

The majority vote scheme utilizes the *available threshold* γ_A and *voting threshold* γ_V to get the consensus trading signal as follows.

- **Buying signal**: If $\tau_A > \gamma_A$ and $\tau_B > \gamma_V$, and we do not hold any share of PI, then all capital is invested to buy PI.
- **Selling signal**: If $\tau_A > \gamma_A$ and $\tau_S > \gamma_V$, and we hold some shares of PI in hand, then all the shares are sold.
- **Waiting signal**: All other conditions belong to the waiting signal. There is no action with this signal.

Step 5: Repetition
Steps 2 to 4 are repeated until the end of the testing period.

3. Our method

3.1. Overview

The main parameters of our method are the pair of available threshold and voting threshold (γ_A, γ_V), the weighted function, and index combination. Figure 1 shows the flowchart of our method, and the details are described as follows.

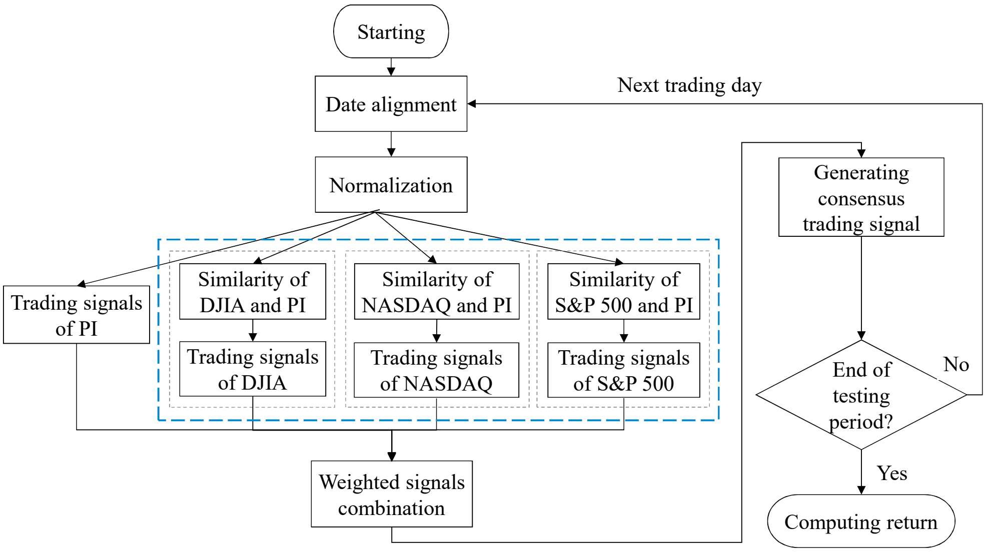


Fig. 1: The flowchart of our method. The blue dotted frame means that there are totally 7 index combinations, which are (PI, DJIA), (PI, NASDAQ), (PI, S&P 500), (PI, DJIA, NASDAQ), (PI, DJIA, S&P 500), (PI, NASDAQ, S&P 500), and (PI, DJIA, NASDAQ, S&P 500).

Step 1: Date alignment

The previous trading day of the US stock market is aligned with the current trading day of Taiwan stock market, with the time-shift property (or the next-day property)[15].

Step 2: Similarity computation

The daily return and 13 technical indicators are used as the features of PI and the US stock indices. These values form a 14-dimensional feature vector for each index on the current trading day t. Each element in the vector is normalized to locate between -1 and 1. Then, the *similarity* between PI and one of the US stock indices is computed. Therefore, there are three similarity values for each trading day, between PI and DJIA, between PI and NASDAQ, and between PI and S&P 500. The details will be described in Section 3.2.

Step 3: Generating trading signals

The method of Lee *et al.* is utilized to generate profitable strategies for each of the PI and US indices. For each index, DTW is applied to obtain three template intervals, which are similar to the leading interval, and the trading strategies of the chosen template intervals are extracted. $3 \times 10 = 30$ strategies are got, since each template interval has 10 pretrained strategies. Then the validation interval is applied to verify whether the trading strategies are suitable or not. Finally, the suitable strategies generate trading signals, BUY, SELL or WAIT, for each index on day t.

Step 4: Weighting signals in each index combination

There are 7 index combinations for aggregating the trading signals of PI and the US indices. For each index combination, the weighted functions employing the similarity values are utilized to combine the trading signals of PI and the US indices on day t. Five weighted functions are used and they will be described in Section 3.3.

Step 5: Generating consensus trading signal

After combining the weighed trading signals from various index combinations, the consensus trading signal is determined by the majority vote scheme, described in Section 2 .

Step 6: Repetition

Repeat Steps 2 to 5 until the testing period ends.

3.2. Similarity computation

The 13 technical indicators which are significantly related to the one-day-ahead returns of PI are MOM^5, MOM^{10}, MOM^{20}, OSC^5, OSC^{10}, OSC^{20}, RSI, CMO^{14}, $BIAS^5$, $BIAS^{10}$, $BIAS^{20}$, $MACD$, $TAPI$. The daily return and the 13 indicators form a 14-dimensional feature vector on a trading day. When the new data of a stock index comes on trading day t, each element in the feature vector is normalized between -1 to 1 from the starting day of simulation to day t. The RMSD between PI and one of US indices is first calculated with the normalized vector. And the similarity value is defined as the reciprocal of the RMSD, calculated as follows.

$$RMSD_{t,index} = \sqrt{\frac{\Sigma_{i=1}^{n}(y_{t,index}^{(i)} - y_{t,PI}^{(i)})^2}{n}}, \quad Similarity_{t,index} = \frac{1}{RMSD_{t,index}}, \tag{3}$$

where $y_{t,index}^{(i)}$ is the ith element of a 14-normalized feature vector of the $index \in$ {DJIA, NASDAQ, S&P 500} on day t, n is the number of element in the feature vector (here $n = 14$.), $Similarity_{t,index}$ is the similarity value between PI and $index$ on day t, and $Similarity_{t,PI}$ is equal to 1 for all t.

3.3. Signal weights in each index combination

There are totally 7 index combinations, including (PI, DJIA), (PI, NASDAQ), (PI, S&P 500), (PI, DJIA, NASDAQ), (PI, DJIA, S&P 500), (PI, NASDAQ, S&P 500), and (PI, DJIA, NASDAQ, S&P 500). An index combination is denoted as a variable **I** in the following. Given an index combination **I**, we adopt five weighted functions for the combination as shown in Equations (4), (5), (6), (7), and (8).

$$W_{t,index}^{(0)} = 1, \tag{4}$$

where $W_{t,index}^{(i)}$ denotes the ith weighted function of stock $index$ on day t, and $W_{t,index}^{(0)}$ means that the weight of each stock $index$ is equal to 1. In other words, all indices are of the same importance in $W_{t,index}^{(0)}$.

$$W_{t,index}^{(1)} = Similarity_{t,index}, \tag{5}$$

where $Similarity_{t,index}$ is the similarity value between PI and $index$ on day t. $W_{t,index}^{(1)}$ represents that similar indices are more important than those are not similar.

$$W_{t,index}^{(2)} = \frac{Similarity_{t,index}}{\sum\limits_{index \in \mathbf{I}} Similarity_{t,index}}. \tag{6}$$

$W_{t,index}^{(2)}$ has the similar meaning as $W_{t,index}^{(1)}$, but it represents the degree of normalized similarity.

$$W_{t,index}^{(3)} = \frac{\frac{Similarity_{t,index}}{C - Similarity_{t,index}}}{\sum\limits_{index \in \mathbf{I}} \frac{Similarity_{t,index}}{C - Similarity_{t,index}}}, \tag{7}$$

where C is a constant greater than 1. In $W_{t,index}^{(3)}$, the weight of high similarity is increased. In this paper, C is set 2, or $\frac{4}{3}$ because our primary experimental results shows that the weighted function does not perform so well with other values of C, such as $C = \frac{6}{5}, \frac{8}{7}$, or $\frac{10}{9}$.

$$W_{t,index}^{(4)} = \frac{(Similarity_{t,index})^2}{\sum\limits_{index \in \mathbf{I}} (Similarity_{t,index})^2}. \tag{8}$$

With the square computation, $W_{t,index}^{(4)}$ enhances the weight of high similarity.

For an index combination **I**, the formulas for the weighed signal combination are shown as follows.

$$V_t^B = \sum_{index \in \mathbf{I}} \Upsilon^B_{t,index} \times W^{(i)}_{t,index}, \quad V_t^S = \sum_{index \in \mathbf{I}} \Upsilon^S_{t,index} \times W^{(i)}_{t,index}, \quad V_t^W = \sum_{index \in \mathbf{I}} \Upsilon^W_{t,index} \times W^{(i)}_{t,index}, \tag{9}$$

where V_t^B, V_t^S and V_t^W denotes the number of weighted BUY, SELL and WAIT signals on day t, respectively, and $\Upsilon^B_{t,index}$, $\Upsilon^S_{t,index}$, $\Upsilon^W_{t,index}$ are the number of unweighted BUY, SELL, and WAIT signals of the stock *index* on day t, respectively.

After calculating the aggregating weighted votes V_t^B, V_t^S, and V_t^W, the majority vote scheme is utilized for determining the consensus trading signal of PI on day t.

4. Experimental results

4.1. Data sets and experimental environment

The data sets of the US stock indices were obtained from *Yahoo finance*, including DJIA, NASDAQ, and S&P 500. The adjusted close prices are adopted to eliminate the effect of stock split. Figure 2 shows the close price for PI, DJIA, NASDAQ, and S&P 500 from 2000 to 2017.

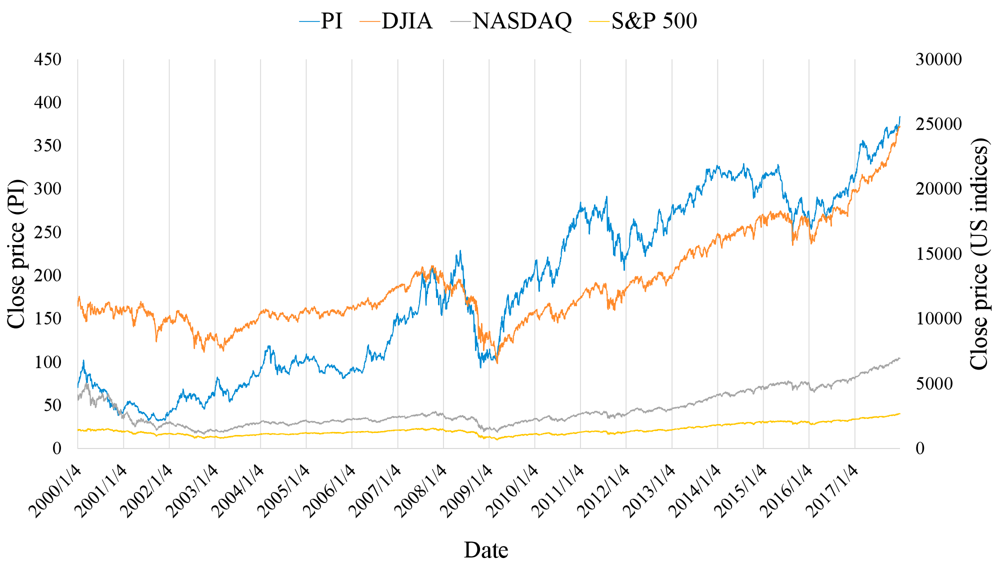


Fig. 2: The trend for PI, DJIA, NASDAQ, and S&P 500 from 2000 to 2017.

The parameters for generating trading signals are the same as the method of Lee *et al.*, template interval length $L = 90$, validation interval length $L_v = 20$. GEP is utilized to train the trading strategies of PI, DJIA, NASDAQ, and S&P 500, respectively. The training period of PI starts from 1996/1/4 through 2017/12/29, totally 5591 days. The testing period of PI starts from 2000/1/4 through 2017/12/29, totally 4480 days. In addition, the training period of DJIA, NASDAQ, and S&P 500 starts from 1996/1/2 through 2017/12/29, totally 5539 days. The testing period of DJIA, NASDAQ, and S&P 500 starts from 2000/1/3 through 2017/12/29, totally 4528 days. In Taiwan stock market, the transaction fee is 0.1425% per share when we buy or sell stocks, and the capital gains tax is 0.3% of the sold amount when we sell the stocks.

4.2. Similarity calculation

The average similarity of each year between PI and a US index is shown in Figure 3. The similarity is lower in 2000 and 2008 than other years, because *information technology bubble (dot-com) crisis* in 2000, and the *global financial* *crisis* in 2008 make the similarity lower. When the similarity is low between PI and the US stock indices, it means that the stock markets may be chaotic. In other words, although the overall trend of PI may be similar to that of the US stock indices, the short term trend is chaotic between the two indices. Therefore, it is difficult to improve the trading profit of PI by combining the trading signals of the US stock indices with low similarity.

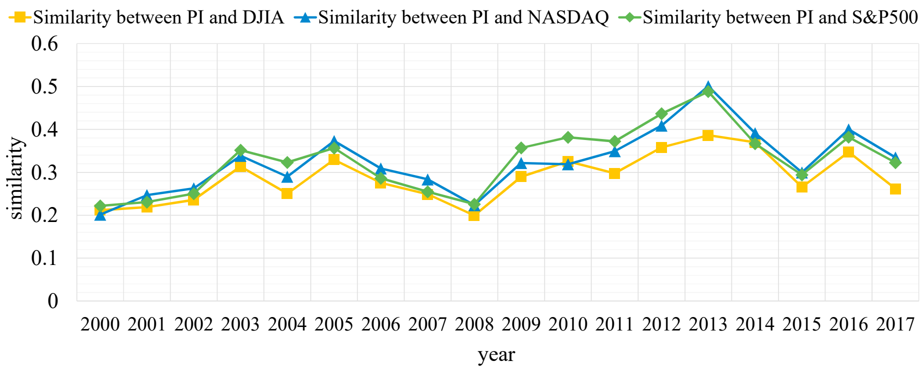


Fig. 3: The average similarity of PI and US stock indices.

4.3. *Performance analysis and comparison*

The consensus trading signals are obtained by using the majority vote scheme with the available threshold γ_A and the voting threshold γ_V. In all experiments, we test each value $\gamma_A \in \{0.00, 0.01, \cdots, 0.99\}$ and $\gamma_V \in \{0.50, 0.51, \cdots, 0.99\}$. Because the number of realizations of γ_A, γ_V, $\mathbf{I}$, and $W^{(i)}$ are 100, 50, 7, and (5+1), respectively, the total number of parameter combinations is $100 \times 50 \times 7 \times 6 = 210000$. The threshold pairs (γ_A, γ_V) with $\gamma_A \in \{0.00, 0.01, \cdots, 0.59\}$ and $\gamma_V \in \{0.50, 0.51, \cdots, 0.89\}$ are denoted as the *whole section*, and the average annualized returns in the whole section are usually greater than the average ones of the method of Lee *et al.* in our primary experiments.

To simplify the result representation, 25 threshold pairs of consecutive five γ_A values and consecutive five γ_V values are averaged to form a block. For example, the resulting value in block ($\gamma_A = \{0.30 \sim 0.34\}$, $\gamma_V = \{0.80 \sim 0.84\}$) is obtained by averaging the values of the 25 threshold pairs. To confirm the parameter stability, we select 100 threshold pairs (4 blocks), instead of 1 threshold pair, with the highest returns as the *optimal section*. The average annualized return of the method of Lee *et al.* is shown in Figure 4. The optimal and whole sections are marked with a yellow frame and a blue frame, respectively. The average annualized returns of various index combinations are shown in Table 1.

As seen in the average annualized returns of the optimal sections with various index combinations of the right side in Table 1, the most helpful index combination for combing one US index to PI for earning profits is (PI, NASDAQ), and the most helpful index combination for combining two US indices to PI for earning profits is (PI, DJIA, NASDAQ). Obviously, S&P 500 is the least helpful. When combining all the three US indices, the performance of (PI, DJIA, NASDAQ, S&P 500) is not the best. In Table 1, we select the top 5 returns (shown in the red bold font). Among these 5 results, (PI, DJIA, NASDAQ) is the best index combination in the optimal section, and $W^{(4)}$ is the most stable function since the $\frac{\mu}{\sigma}$ value of $W^{(4)}$ on the seven index combinations is the highest in the whole section. Obviously, the results of using the weighted functions are better than the unweighted result. Our suggestion is (PI, DJIA, NASDAQ) with $W^{(4)}$, whose average annualized return is 15.03%.

In Table 1, the index combinations $W^{(3)}(C = 2)$, $W^{(3)}(C = \frac{4}{3})$ and $W^{(4)}$ may also be good candidates because of their good performance in the optimal section. On the contrary, $W^{(0)}$ is the worst combination because its $\frac{\mu}{\sigma}$ value is the lowest in the optimal and whole section, which means the combination sensitive to the voting threshold. Therefore, the following three parameters, the (optimal section, (PI, NASDAQ), $W^{(3)}(C = \frac{4}{3})$), the (optimal section, (PI, DJIA, NASDAQ), $W^{(4)}$), and the (optimal section, (PI, DJIA, NASDAQ, S&P 500), $W^{(3)}(C = 2)$), are suggested in the investment.

Fig. 4: The average annualized return (%) by the method of Lee *et al.*, only based on Taiwan stock market. The whole section is marked with a blue frame, and the optimal section is marked with a yellow frame.

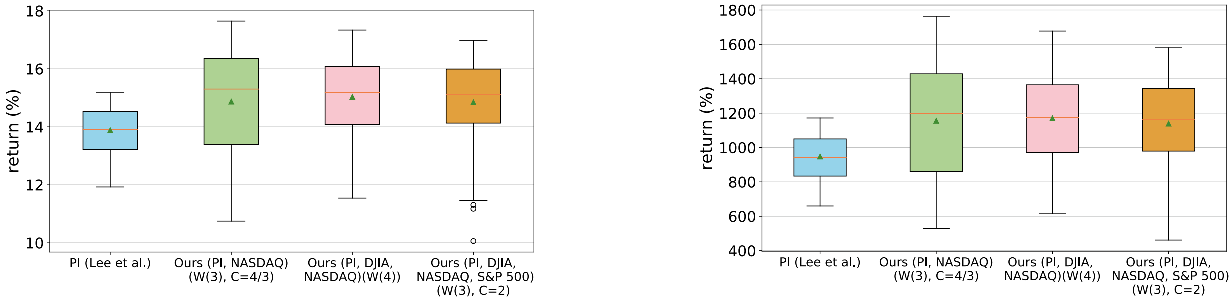
$\gamma A \backslash \gamma V$	0.5~0.54	0.55~0.59	0.60~0.64	0.65~0.69	0.70~0.74	0.75~0.79	0.80~0.84	0.85~0.89	0.90~0.94	0.95~0.99
0.00~0.04	-1.23	-0.20	2.48	5.45	7.11	8.06	7.18	8.50	8.63	7.68
0.05~0.09	-1.59	-0.58	2.24	5.23	6.84	7.76	5.24	5.71	5.45	4.51
0.10~0.14	-1.15	-0.12	2.69	5.06	6.44	6.63	4.35	4.92	4.05	3.09
0.15~0.19	-0.14	0.85	3.94	6.00	7.58	8.01	5.46	5.34	3.82	2.97
0.20~0.24	-0.23	0.46	3.17	6.37	9.03	8.63	5.78	6.64	6.66	5.47
0.25~0.29	1.14	1.84	4.86	7.52	10.74	10.82	7.99	9.06	8.99	7.80
0.30~0.34	2.50	3.34	6.02	7.66	10.51	10.50	9.56	10.65	8.90	8.42
0.35~0.39	1.97	2.23	4.54	8.51	10.33	12.38	12.50	12.65	11.31	10.35
0.40~0.44	2.45	3.79	6.77	11.85	13.09	14.07	14.11	13.06	11.18	11.26
0.45~0.49	3.46	3.83	6.47	11.04	13.44	13.92	12.02	11.29	11.85	14.49
0.50~0.54	7.81	7.40	7.31	9.93	12.28	11.17	11.59	11.41	10.93	10.07
0.55~0.59	7.57	7.42	8.36	9.25	10.63	11.38	11.33	8.29	6.09	3.00
0.60~0.64	12.34	12.33	11.10	10.05	10.70	9.74	7.39	6.47	5.84	3.62
0.65~0.69	11.52	11.87	10.35	7.97	8.00	7.58	7.58	7.58	5.52	3.62
0.70~0.74	6.48	6.15	5.93	5.93	4.73	2.93	2.93	2.93	2.91	2.90
0.75~0.79	1.09	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.80~0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.85~0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.90~0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.95~0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 1: The average annualized returns (%) with various index combinations from 2000 through 2017. The annualized return of BAH strategy is 9.85 %. The top 5 returns are shown in the red bold font, and the underlined red bold one is our suggested parameter.

Index combination	**Parameter section**	$\mathbf{W^{(0)}}$	$\mathbf{W^{(1)}}$	$\mathbf{W^{(2)}}$	$\mathbf{W^{(3)}}$ **(C = 2)**	$\mathbf{W^{(3)}}$ **(C =** $\mathbf{\frac{4}{3}}$**)**	$\mathbf{W^{(4)}}$	**Avg**	**Rank**	
									Opt.	Whole
PI (Lee *et al.*)	Optimal	13.88	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Whole	6.89	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
PI, DJIA	Optimal	14.08	14.51	13.85	14.08	14.09	13.83	14.07	**2**	N/A
	Whole	9.35	7.83	8.08	7.57	7.42	7.18	7.91	N/A	**2**
PI, NASDAQ	Optimal	13.15	13.99	14.42	13.92	**14.87**	14.76	14.19	**1**	N/A
	Whole	8.38	7.29	7.17	7.50	7.39	7.26	7.50	N/A	**3**
PI, S&P 500	Optimal	14.16	13.12	13.13	14.46	14.73	14.61	14.04	**3**	N/A
	Whole	9.08	8.56	8.14	8.25	7.74	7.59	8.23	N/A	**1**
PI, DJIA, NASDAQ	Optimal	**15.71**	14.03	12.86	14.12	13.83	**15.03**	**14.26**	**1**	N/A
	Whole	10.06	8.09	8.25	7.68	7.72	7.36	8.19	N/A	**2**
PI, DJIA, S&P 500	Optimal	**15.34**	13.41	13.96	14.38	13.78	13.42	14.05	**2**	N/A
	Whole	10.43	8.86	8.92	8.79	7.96	7.35	8.72	N/A	**1**
PI, NASDAQ, S&P 500	Optimal	12.87	13.13	12.83	13.36	12.70	13.79	13.11	**3**	N/A
	Whole	8.00	7.91	7.99	7.92	7.65	7.54	7.84	N/A	**3**
PI, DJIA ,NASDAQ, S&P 500	Optimal	13.65	13.90	14.82	**14.85**	14.60	14.03	14.31	N/A	N/A
	Whole	9.61	8.58	9.54	8.84	8.12	7.46	8.69	N/A	N/A
Avg. of optimal section (μ)		14.14	13.73	13.70	14.17	14.09	14.21			
Std. dev. of optimal section (σ)		1.06	0.52	0.78	0.47	0.75	0.59			
$\frac{\mu}{\sigma}$ of optimal section		13.34	26.34	17.56	30.21	18.78	23.95			
Avg. of whole section (μ)		9.27	8.16	8.30	8.08	7.71	7.39			
Std. dev. of whole section (σ)		0.87	0.54	0.75	0.56	0.27	0.15			
$\frac{\mu}{\sigma}$ of whole section		10.67	15.07	11.07	14.38	29.10	50.04			

Then, the box plots of average annualized and cumulative returns on the optimal section for the three suggested parameters and the method of Lee *et al.* are shown in Figure 5. As one can see, the results of the three suggested parameters are better than the method of Lee *et al.*.

For ensuring that our method has the ability to make profit, but not to earn in occasional events, the annual returns of the three suggested parameters are compared with the method of Lee *et al.* and the BAH strategy, and the results are shown in Table 2. In the table, the average annual returns of our suggested parameters are distributed evenly and they are usually better than the ones of the method of Lee *et al.*. During the bear markets, such as years 2000, 2001,



(a) (b)

Fig. 5: The box plots for the returns of the three suggested parameters and the method of Lee *et al.* on the optimal section. (a) Annualized returns (%). (b) Cumulative returns (%).

2005, 2008, 2011, 2014 and 2015, our suggested parameters usually suffer less losses than the method of Lee *et al.* and the BAH strategy.

In Table 2, R_{BAH} denotes the ratio of the method of Lee *et al.* or our suggested parameters on the optimal section which get higher return than the BAH strategy. R_{Lee} denotes the ratio of out suggested parameters which get higher return than the method of Lee *et al.* on the optimal section. During the bear markets (negative BAH), R_{BAH} of our suggested parameters is usually high, meaning that the parameters are able to avoid loss. However, when the bear market turned into the bull market (positive BAH), R_{BAH} of our suggested parameters are sometimes lower because the weights are computed according to historical information, which may cause conservative trading decision.

Next, we compare our suggested parameters with the method of Lee *et al.* The returns of our suggested parameters are slightly higher than that of Lee *et al.* When the translation between bull and bear market alternatively, R_{Lee} is sometimes lower because the translation may make the weighted signals chaotic. More specifically, when the bull market turned into the bear market, the weights for signal combination were computed optimistically and suffered the loss. In contrast, when bear market turned into bull market, the weights for signal combination are computed conservatively and decreased the profits.

Table 2: The annual returns (%) for the three suggested parameters and the method of Lee *et al.* on the optimal section. σ denotes the standard deviation (%). R_{BAH} and R_{Lee} denote the ratio compared with BAH strategy or the method of Lee *et al.*, respectively.

	BAH	**Lee *et al.* (PI)**			**$W^{(3)}$, C = $\frac{4}{3}$ (PI, NASDAQ)**				**$W^{(4)}$ (PI, DJIA, NASDAQ)**				**$W^{(3)}$, C = 2 (PI, DJIA, NASDAQ, SP500)**			
	Annual return (%)	**Avg (%)**	σ(%)	R_{BAH}	**Avg (%)**	σ(%)	R_{BAH}	R_{Lee}	**Avg (%)**	σ(%)	R_{BAH}	R_{Lee}	**Avg (%)**	σ(%)	R_{BAH}	R_{Lee}
2000	**-40.71**	**6.79**	7.68	1	**9.30**	13.45	1	0.72	**17.83**	0.63	1	1	**4.70**	6.02	1	0.37
2001	**-7.40**	**24.13**	8.81	1	**24.12**	3.93	1	0.15	**20.74**	6.79	1	0.05	**23.40**	8.99	1	0.38
2002	**47.95**	**45.59**	8.05	0.72	**47.17**	9.02	0.7	0.38	**53.70**	6.44	0.9	0.73	**39.73**	6.07	0.21	0.2
2003	**39.56**	**16.57**	8.96	0.03	**33.40**	12.25	0.52	0.99	**39.56**	6.43	0.65	0.97	**47.44**	6.05	0.94	1
2004	**17.59**	**12.67**	8.49	0.39	**16.19**	5.36	0.59	0.66	**14.58**	3.40	0.25	0.49	**9.76**	3.59	0	0.28
2005	**-15.24**	**-2.86**	5.83	1	**-9.02**	4.99	1	0	**-12.58**	5.43	0.82	0	**6.93**	7.87	1	0.97
2006	**71.59**	**63.39**	4.72	0.03	**63.63**	2.95	0	0.35	**59.09**	15.46	0	0.3	**60.98**	3.48	0	0.12
2007	**10.01**	**8.34**	1.47	0	**6.00**	5.22	0.03	0.26	**7.88**	6.86	0.14	0.51	**16.82**	4.11	0.88	1
2008	**-35.39**	**-2.01**	1.83	1	**-0.52**	2.01	1	0.66	**0.56**	2.15	1	0.91	**6.80**	11.23	1	0.91
2009	**85.11**	**55.29**	8.31	0	**55.37**	12.39	0	0.42	**51.18**	11.60	0	0.08	**21.68**	4.84	0	0
2010	**32.96**	**24.22**	7.39	0.27	**23.12**	4.95	0	0.53	**26.09**	6.84	0.2	0.65	**29.95**	6.77	0.7	0.93
2011	**-22.09**	**-17.62**	4.81	0.8	**-17.45**	2.93	1	0.66	**-20.56**	4.22	0.75	0	**-19.04**	3.42	0.9	0.17
2012	**24.84**	**17.70**	7.52	0.4	**21.04**	6.33	0.12	0.7	**16.45**	5.36	0	0.45	**12.69**	7.16	0	0.13
2013	**18.01**	**16.13**	0.77	0	**16.71**	1.54	0.35	0.36	**18.24**	1.05	0.9	0.92	**17.53**	1.56	0.65	0.85
2014	**-3.22**	**-3.59**	2.26	0.35	**-3.40**	1.85	0.38	0.64	**-3.76**	2.07	0.16	0.59	**-1.14**	3.48	0.67	0.8
2015	**-13.02**	**-17.58**	1.18	0	**-15.03**	6.86	0.15	0.41	**-11.00**	8.97	0.44	0.86	**-18.78**	4.06	0.17	0.17
2016	**16.09**	**10.81**	4.88	0.36	**9.76**	7.04	0.45	0.25	**5.33**	4.94	0	0	**10.92**	4.54	0.06	0.56
2017	**22.74**	**19.17**	2.38	0	**17.11**	4.61	0	0.17	**17.26**	4.94	0	0.36	**19.53**	0.43	0	0.38
Avg	**13.85**	**15.40**	0.93	0.41	**16.53**	1.80	0.46	0.46	**16.70**	1.58	0.46	0.49	**16.11**	1.48	0.51	0.51

As a summary, taking both profit and risk into consideration, index combination (PI, DJIA, NASDAQ) with $W^{(4)}$ is considered to be the best for trading PI from 2000 through 2017. Its average annualized return (cumulative return) is 15.03% (1170.42%), which is better than the method of Lee *et al.* 13.88% (947.65%), and the BAH strategy 9.85% (442.90%).

5. Conclusion

In this paper, we improve the trading strategy for Taiwan stock market with the help of US stock markets. First, the trading days between Taiwan and the US stock markets are aligned. Next, the similarity reflecting the daily dependence between PI and one of the US stock indices is calculated. The trading signals of PI or each US stock index are generated by the method of Lee *et al.* Then, the weight of each US index is calculated according to the similarity. Finally, the consensus trading signal is determined by the majority vote scheme with the weighted functions.

As the experimental results show, DJIA and NASDAQ are the most helpful indices for the trading PI with weight $W^{(4)}$. From 2000 through 2017, for the index combination (PI, DJIA, NASDAQ) with $W^{(4)}$, and the average annualized return (cumulative return) achieves 15.03% (1170.42%), which is better than the method of Lee *et al.* 13.88% (947.65%), and the BAH strategy 9.85 (442.90%). Besides, the results of the weighted functions are better than the unweighted results.

For the future work, the portfolio selection may consider the Taiwan stocks that are more dependent on US stock markets. The index combination may utilize the dynamically adjusting method to pick the most suitable combination as time going on. In addition, the method of bidirectional recognition may be utilized to select the features.

References

1. Ni, L.P., Ni, Z.W., Gao, Y.Z.. Stock trend prediction based on fractal feature selection and support vector machine. *Expert Systems with Applications* 2011;**38**:5569–5576.
2. Huang, C.F.. A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing* 2012; **12**:807–818.
3. Ahmad Kazem, E.S., Hussain, F.K., Saberi, M., Hussain, O.K.. Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied Soft Computing 13* 2013;**13**:947–958.
4. Chang, P.C., Fan, C.Y., Lin, J.J.. Integrating a piecewise linear representation method with dynamic time warping system for stock trading decision making. In: *Proc. of Fourth International Conference on Natural Computation*; vol. 2. Jinan, China; 2008, p. 434–438.
5. Hsu, C.M.. A hybrid precedure for stock price prediction by integrating self-organizing map and genetic programming. *Expert Systems with Applications* 2011;**38**:14026–14036.
6. Huang, C.F., Chang, B.R., Cheng, D.W., Chang, C.H.. Feature selection and parameter optimization of a fuzzy-based stock selection model using genetic algorithms. *International Journal of Fuzzy Systems* 2012;**14**(1):65–75.
7. Tsai, T.J., Yang, C.B., Peng, Y.H.. Genetic algorithms for the investment of the mutual fund with global trend indicator. *Expert Systems with Applications* 2011;**38(3)**:1697–1701.
8. Koza, J.R.. *Genetic Programming: On the programming of computers by means of natural selection*; vol. 1. Cambridge, USA, MA: MIT press; 1992.
9. Jhou, S.M., Yang, C.B., Chen, H.H.. Taiwan stock forecasting with the genetic programming. In: *Proc. of the 16th Conference on Artificial Intelligence and Application (Domestic Track)*. Chungli, Taiwan; 2011, p. 151–157.
10. Ferreira, C.. Gene expression programming: A new adaptive algorithm for solving problems. *Complex Systems* 2001;**13**:87–129.
11. Lee, C.H., Yang, C.B., Chen, H.H.. Taiwan stock investment with gene expression programming. *Procedia Computer Science* 2014; **35**:137–146.
12. Chou, C.Y., Yang, C.B., Chen, H.H.. Portfolio investment based on gene expression programming. In: *Proc. of the 32nd International Conference on Computers and Their Applications*. Honolulu, Hawaii, USA; 2017, p. 225–230.
13. Wang, T.L., Wang, M.. Features extraction based on particle swarm optimization for high frequency financial data. In: *Proc. of IEEE International Conference on Granular Computing (GrC)*. Kaohsiung, Taiwan; 2011, p. 728–733.
14. Chen, S.M., Chu, H.P., Sheu, T.W.. Taiex forecasting using fuzzy time series and automatically generated weights of multiple factors. *IEEE Transactions on Systems, Man, and Cybernetics part A: Systems and Humans* 2012;**42**(6):1485–1494.
15. Aityana, S.K., Ivanov-Schitz, A.K., Izotov, S.S.. Time-shift asymmetric correlation analysis of global stock markets. *International Financial Markets, Institutions and Money* 2010;**20**:590–605.
16. Edwards, A.W.. The measure of association in a 2× 2 table. *Journal of the Royal Statistical Society Series A (General)* 1963;:109–114.
17. Mosteller, F.. Association and estimation in contingency tables. *Journal of the American Statistical Association* 1968;**63**(321):1–28.
18. Cornfield, J.. A method of estimating comparative rates from clinical data. applications to cancer of the lung, breast, and cervix. *Journal of the National Cancer Institute* 1951;**11**(6):1269–1275.