Information Sciences 346–347 (2016) 261–274

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

An intelligent pattern recognition model for supporting investment decisions in stock market

Tai-liang Chen[a,*], Feng-yu Chen[b]

[a] *Department of Digital Content Application and Management, Wenzao Ursuline College of Languages, 900 Mintsu, 1st Road, Kaohsiung 807, Taiwan, ROC*

[b] *Computer and Network Center, National Kaohsiung University of Applied Science, Chien Kung Campus 415 Chien Kung Road, Kaohsiung 807, Taiwan, ROC*

ARTICLE INFO

Article history:
Received 10 December 2012
Revised 2 April 2015
Accepted 29 January 2016
Available online 10 February 2016

Keywords:
Stock market forecasting
Perceptually important point (PIP) identification matching
Template matching technique
Chart pattern
Technical indicator

ABSTRACT

For many years, how to make stock market predictions has been a prevalent research topic. To carry out accurate forecasting, stock analysts and academic researchers have tried various analysis techniques, algorithms, and models. For example, "technical analysis" is a popular approach used by common stock investors to analyze market trend, and Artificial Intelligence (AI) algorithms such as genetic algorithms (GAs), neural network (NN), and fuzzy time-series (FTS), were proposed by researchers to forecast the future stock index. Although the daily forecasts are very useful for professional investors who implement intraday trading, we argue that forecasting a bullish turning point is a more interesting issue than the future stock index for common investor because an accurate forecast will bring a huge amount of stock return. Therefore, this paper proposes an intelligent pattern recognition model, based on two new stock pattern recognition methods, “PIP bull-flag pattern matching” and the “floating-weighted bull-flag template,” to recognize a bull-flag stock pattern. The bull-flag pattern is a stock’s turning point with proper timing, which can enable a stock investor to profit. To promote recognition accuracy, the proposed model employs chart patterns and technical indicators, simultaneously, as pattern recognition factors. In the model verification, we evaluate the proposed model with stock returns by forecasting two stock databases (TAIEX and NASDAQ), and comparing the returns with other advanced algorithms. The experimental results indicate that the proposed model outperforms the published algorithms, such as rough set theory (RST), genetic algorithms (GAs) and their hybrid model, and gives a high-level of profitability. Additionally, the trading strategies, provided by the proposed model, also help investors to make beneficial investment decisions in the stock market.

1. Introduction

Stock market is a very complex and changeable system influenced by many factors such as economic environment, political policy, industrial development, and market news, etc. To make profit from the capital market, stock investor has been looking for right tools and techniques to analyze stock market. As we known, the technical analysis [9] is a popular approach used by investment professionals and common investors. The principles of technical analysis are derived from the

* Corresponding author: Tel.: +886 (0)920975168; fax: +886 (0)73105785.
E-mail address: 97007@mail.wzu.edu.tw, too.cool.chen@gmail.com (T.-l. Chen).

http://dx.doi.org/10.1016/j.ins.2016.01.079

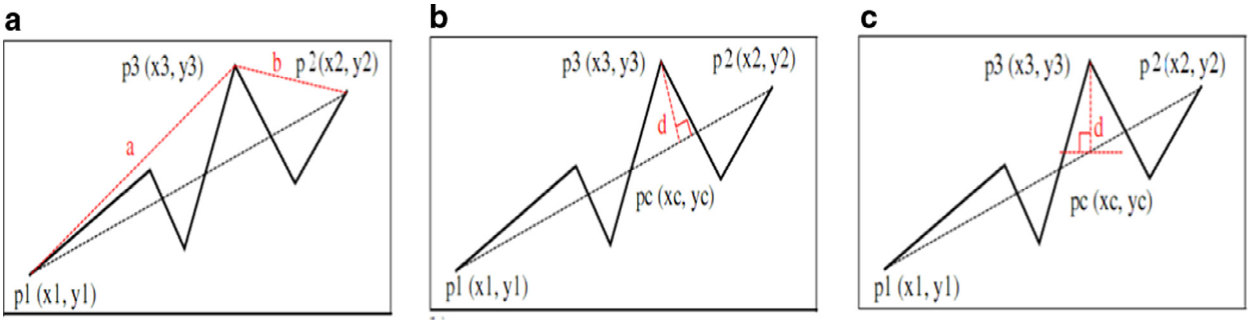

Fig. 1. Distance measure methods for PIP identification: (a) Euclidean distance based: PIP-ED, (b) perpendicular distance based: PIP-PD and (c) vertical distance based: PIP-VD [11].

observations of stock markets over hundreds of years and many evidences have shown that the technical analysis can help stock investor to make right judgment on stock market [2,3,19,21,22].

In academic research, many researchers also applied this approach in their forecasting models. For example, the study of charting patterns analysis for investment decision [19], investigating price charting patterns with kernel regression for the identification of ten patterns [16–18], and implementing a variation of the bull-flag stock chart with a template matching technique based on pattern recognition [31,33]. The stock chart patterns of these studies used for pattern recognition are established in fixed patterns given by experts and researchers, which are not very similar to the actual fluctuations in historical stock data, due to possible chart patterns contained historical stock data are not considered to forecasting.

Additionally, many advanced models and algorithms were proposed and these have achieved considerable results in forecasting accuracy. According to theories of model building, stock index forecasting models can be summarized in two categories [28]: (1) linear models based on statistical theories, such as General Autoregressive Conditional Heteroskedasticity (GARCH) and Stochastic Volatility model [12]; and (2) non-linear models based on artificial intelligence, such as the fuzzy time-series [5,29,30,34,35], Rough sets theory [5,14,23–26,30] artificial neural network [4,36,37], the support vector machine [32,38,39], and the particle swarm optimization [20,36,40].

After reviewing the related research, we argued that two issues are worth to be discussed further. Firstly, most researchers employed forecasting error as performance indicator, such as RMSE and MAPE, to evaluate their models [1,5,6, 28–30]. Although the daily forecasts are very useful for professional investors who implement intraday trading, we think that forecasting a bullish turning point is a more interesting issue than the future stock index because an accurate forecast for bullish stock pattern will bring a huge amount of stock return. Secondly, most stock pattern recognition models employed one matching template with many fixed weights assigned by researcher and it is not objective approach [8]; [16–18,31].

In this paper, we propose three new approaches to refine past methods: (1) a new definition for a bullish turning point with the probability of occurrence; (2) a new weighted method, based on the occurrence of observations, to produce dynamic weights for matching template; and (3) a new hybrid model based on PIP bull-flag pattern matching [11] and bull-flag template [18,31] for recognizing stock pattern.

To evaluate the profitability of the proposed model, we will give several trading criteria in experiments such as thresholds, stock holding period and stopping loss. In model verification, the NASDAQ composite index and Taiwan stock market weighted index (TAIEX) are taken as experimental datasets, and four advanced algorithms as benchmarks.

2. Related works

This section gives a brief review of methodology for two stock pattern identification methods: perceptually important point (PIP) identification matching and template matching.

2.1. Perceptually important point (PIP) identification matching

Reducing the dimension (i.e. the number of data point) by preserving the salient points is a promising method for time series representation [10]. These points are called as "*perceptually important points* (*PIP*)". The PIP identification process is introduced by Chung et al. [7] and used for matching of technical (analysis) patterns in financial applications.

In the PIP identification process, the vertical distance measure method is an automatic algorithm to recognize a specific pattern [11]. Three parameters are defined in the method: *fitting-window*, *holding-period*, and *distance-threshold*. The *fitting-window* is a window size of trading day for a stock pattern within a specific period of trading day such as 20-day, 40-day and 60-day. The *holding-period* is a certain period of trading day for investor to hold a stock. The *distance-threshold* is a specific price distance, *d*, for the method to define a stock pattern.

Fig. 1 illustrates the vertical distance measure method how to produce a price distance, *d*, for the "*head-and-shoulder*" pattern. In Fig. 1, three stock data, p_1, p_2 and p_3, are defined as three graphic points of two-dimension coordinates(x represents "a series of trading days" and y represents "stock price") contained in a specific period of *fitting-window*, where p_1

0.5	0	-1	-1	-1	-1	-1	-1	-1	0
1	0.5	0	-0.5	-1	-1	-1	-1	-0.5	0
1	1	0.5	0	-0.5	-0.5	-0.5	-0.5	0	0.5
0.5	1	1	0.5	0	-0.5	-0.5	-0.5	0	1
0	0.5	1	1	0.5	0	0	0	0.5	1
0	0	0.5	1	1	0.5	0	0	1	1
-0.5	0	0	0.5	1	1	0.5	0.5	1	1
-0.5	-1	0	0	0.5	1	1	1	1	0
-1	-1	-1	-0.5	0	0.5	1	1	0	-0.2
-1	-1	-1	1	-1	0	0.5	0.5	-2	-2.5
Consolidation							**Breakout**		

Fig. 2. A 10 × 10 grid of weights to represent a variation of the "bull-flag" charting pattern, which used price and trading volume as fitting values, advanced by Leigh et al. [17,18].

-.25	-.4	-.45	-.7	-15.	-1.6	-1.6	-1.6	-1.6	-.7
-.25	-.4	-.45	-.6	-.75	-1.4	-1.4	-1.4	-.8	1
-.25	-.4	-.45	-.55	-.5	-.75	-.75	-.5	-.5	.4
-.25	-.4	-.45	-.55	-.25	.9	.9	.9	-.15	-.35
-.25	-.5	-.6	-.25	.9	1	1	1	1	-.55
-.3	-.6	-.25	.8	1	.9	.9	.9	.8	-.45
.35	.1	.8	1	.65	.6	.6	.4	.75	-.15
.1	.8	1	.5	.3	.5	.5	.3	0	.1
.8	1	.5	.35	.15	0	0	0	.3	.35
1	.8	.35	0	0	0	0	.1	.25	.3

Fig. 3. A 10 × 10 grid of weights to represent a variation of the "*bull-flag*" charting pattern, which used price as fitting values, advanced by Wang and Chan [31].

and p_2 represent the first and last points respectively; and p_3 represents the point with the highest stock price between the period of *fitting-window* (*fitting-window* $= x_1 - x_2 + 1$). The price distance, d, is defined as the vertical distance from p_3 to the line connecting p_1 and p_2. The two lines, p_1, p_2 and p_3p_c, intersect vertically at point p_c. The formula to produce vertical distance, d, is defined as Eq. (1), where $x_c = x_3$.

$$d = VD(p_3, p_c) = y_c - y_3 = \left(y_1 + (y_2 - y_1) \times \frac{x_c - x_1}{x_2 - x_1}\right) - y_3 \tag{1}$$

With the given parameters, x and y, belong to p_1, p_2 and p_3, the above equation can be refined as Eq. (2).

$$d = VD(p_3, p_1, p_2) = y_c - y_3 = \left(y_1 + (y_2 - y_1) \times \frac{x_3 - x_1}{x_2 - x_1}\right) - y_3 \tag{2}$$

2.2. Template matching

As developing in computer technology and arising in cross-domain research between computer sciences and financial forecasting, many researchers have increasingly paid attention to using pattern analysis based on computer-based algorithms for investment decisions. For example, Lo et al. [19] tested price charting patterns using kernel regression for the identification of ten patterns; Leigh et al. [18] advanced a novel forecasting model with technical analysis, pattern recognizer, neural networks, and genetic algorithm to predict the NYSE composite index; Leigh et al. [17], discoveryed stock market trading rules using technical charting heuristics; Leigh et al. [16] provided a computational implementation of stock charting to produce a signal for movement in New York Stock Exchange Composite Index. These researches implemented a variation of the "*bull-flag*" charting pattern (shown in Fig. 2) and used a template matching technique [8] for pattern recognition. Although Wang and Chan [31] also used the template matching method to examine the potential profit for forecasting the Nasdaq Composite Index (NASDAQ) and Taiwan stock market weighted index (TAIEX), the two researchers concentrated on identifying an increasing price trend and proposed another version of template (shown in Fig. 3) for the bull-flag charting pattern.

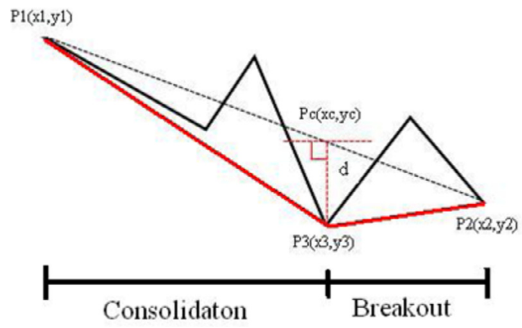

Fig. 4. A bull-flag pattern based on PIP-VD. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

The template matching is a pattern recognition technique used to match a given stock pattern with a pictographic image for object identification [8,31,18]. In the matching process, a selected period of stock pattern (a period of price time series defined as a "fitting window") is converted to a matrix based on the percentile order of stock price or volume and matched with a given grid template, such as a "bull flag" stock chart, to produce a "fit" value. The fit value is computed by a matrix multiplication operation with the selected period of stock pattern and the given grid template.

The matrix operation is called a cross correlation computation, in which the percentile value that falls in each cell of a column is multiplied by the weight in the corresponding cell of the given temple. If the amount of the fit values [8,31,18] is high, it represents that the selected period of stock pattern is close to "bull flag" stock pattern and denotes a right time to invest in the stock market.

Take Fig. 2 as example, the template used for the bull flag pattern is a 10-by-10 grid with weights ranging from -2.5 to $+1.0$ in the cells. The weighting is used to define areas in the template for the "descending" consolidation and for the "upward-tilting" breakout portions of this bull flag heuristic pattern [18]. For example, "0.5" means that the stock pattern falling in the grid has a better similarity to the shape of the full flag than "0" and less than "1". And "-0.5" means that the stock pattern falling in the grid has a worse similarity to the shape of the full flag than "0" and better than "1".

The 10-by-10 grid is applied to the time series of price or volume data one trading day at a time, with the leftmost time series data point being the values for the trading day which precedes the current day by 59 trading days, and the rightmost time series data point being the trading day which is currently being analyzed. Values for the earliest 10% of the trading days (6 days of the 60 in the rolling window) are mapped to the first column of the grid, values for the next-to earliest 10% of the trading days are mapped to the second column of the grid, and so on, until the most recent 10% of the trading days are mapped to the rightmost column. For example, there will be 6 trading days represented in each column of a single 60 trading day window. If all 6 of these trading days have price values which are in the lowest decile of the 60 price values for the day, then 100% (6 values out of a total of 6 in the column) will be the value in the lowest cell of the 10 cells in the column. If this column is the leftmost of the columns in the window, then this 100% will be multiplied by the value in the corresponding cell in the bull flag template (which is the one in the lowest left-hand corner), which has the value of "-1.0" (see Fig. 2), to result in a cell fit value of "-1.0" $\times$ 100% ="-1.0". This is done for the 10 cells in the column and summed, resulting in a fit value for the column of "-1.0", since there will be 0.0% in the other nine cells of the column [18].

In this way, 10-column fit values for price and 10-column fit values for volume are computed for each trading day. The summation of all 20 values for a trading day represents a total fit for the trading day [8,18].

3. Proposed model

In this paper, we proposed an intelligent recognition model, which integrated two new methods in identification process. The first method is the "*PIP bull-flag pattern matching*". It is refined out of two past pattern matching methods: (1) template matching method [8,18]; and (2) PIP identification matching method [7,10,11]. We crystallize the advantages of the methods into a new stock pattern recognition method which can provide an intelligent approach to recognize possible bull-flag patterns. The proposed concept to explain how they work simultaneously is introduced in the next paragraphs.

As Leigh et al. [17,18] defined, a bull-flag pattern is a horizontal or sloping flag of "*consolidation*" followed by a sharp rise in the positive direction, the "*breakout*" (see Fig. 2). Based on the definition, the bull-flag pattern can also be defined as a PIP pattern (Fig. 4 demonstrates the PIP patterns), where the black line denotes an actual stock pattern; the red line between P_1 and P_3 denotes consolidation; and the red line between P_3 and P_2 denotes breakout. Therefore, a new approach to recognize the bull-flag pattern from stock market is proposed in this paper and it is named as "PIP bull-flag pattern matching". To detail the method, the pseudo codes for it are given and shown in Fig. 5 (where the input variable *P* is a set that contains candidate stock time-series patterns; the output variable, *SP*, is a set that contains selected "*bull-flag*" patterns from *P*; the "*distance-threshold*" is a given value to define the height of the "*bull-flag*" pattern; the "*fitting-window*" is a given

```
Procedure Bull_Pattern _Search (P, distance-threshold, fitting-window, holding-period)
    Input: sequence P[1…m], distance-threshold, fitting-window, holding-period.
    Output: pattern SP.
        For each data point p[i] in P
        p[j] = The data point with maximum distance between p[i] and p[i + fitting-window-1]
        d = VD(p[j], p[i], p[i + fitting-window-1])/**defined in Eq. (2)**/
        If   d ≥ distance-threshold   And
             Closing_Index ( p[j + holding-period ]) – Closing_Index ( p[j] ) > 0
           Then Add p[i] to SP
        End If
    End For
    Return SP
End
```

Fig. 5. Pseudo codes for PIP *bull-flag* pattern matching method.

value to define the window size of stock time-series pattern in *P*; the "*holding-period*" is a given period of time for holding the bought stock).

The second method is "*floating weighted*" method to build a "*bull-flag template*" for pattern recognition. From the first pattern recognition method, the stock patterns with bull-flag shape are selected and they can be used as the basis for recognizing the bull-flag pattern. In this method, a bull-flag template is built based on the selected bull-flag patterns and each cell of the template is assigned a "*floating*" weight according to its occurrence. Additionally, to extract the characteristics of the bull-flag pattern as much as possible by technical analysis approaches, two types of templates, charting pattern and technical indicator, are proposed to create the bull-flag template.

In the pattern recognition process of the proposed model, the template matching technique is applied to calculate a "*fitting value*" for the present stock pattern with the bull-flag template. The value can be used to judge whether it is a good timing (i.e. bull-flag pattern) to invest in stock market or not. For example, if the fitting value for the present stock pattern reaches a certain high-level, then there is a good chance to make profit on investing the stock because the stock pattern is highly similar to the bull-flag pattern.

Lastly, to examine model performance, five experimental parameters are taken to evaluate the proposed model: (1) pattern fitting window; (2) PIP-distance, *d*, for stock pattern; (3) holding period; and (4) stock return indicator.

3.1. The research framework of the proposed model

The research framework for the proposed model contains two phases, model training (Fig. 6) and testing (Fig. 7), and each phase contains four steps. In the first phase, the "*bull-flag template*" and "*threshold*" are produced from historical stock data and defined as the criteria for "*bull-flag pattern*". In the second phase, the criteria are used to forecast testing stock data with various factors to examine model performance. Each step in the proposed model is described in detail as follows.

3.2. Data preparation

The initial experimental stock database contains six basic attributes: date, opening index, closing index, highest index, lowest index, and trading volume. Because two types of template grids, charting patterns and technical indicators, are employed for representing the "*bull-flag template*", stock patterns and technical indicators should be produced in advance as forecasting attributes. Stock pattern is composed of daily closing index. Nine popular technical indicators are generated by the six basic attributes and they are MA, RSI, STOD, OBV, ROC, VR, PSY, AR, and DIS. These indicators are selected because they are highly related to the future stock index [5,15]. Additionally, to reduce computing complexity, cumulative probability distribution approach (CDPA) [29,30] is employed to granulate the value of each technical indicator into linguistic values (the initial linguistic values are set as three values : low (L_1), medium (L_2) and high (L_3)).

Phase I: Model training

Step 1: Bull pattern recognizing

This step uses the "*PIP-bull pattern matching*" (the pseudo codes are listed in Fig. 5) method to recognize the "*bull-flag pattern*" from training stock data. Three parameters are used as experimental factors and those are *fitting-window* (a window

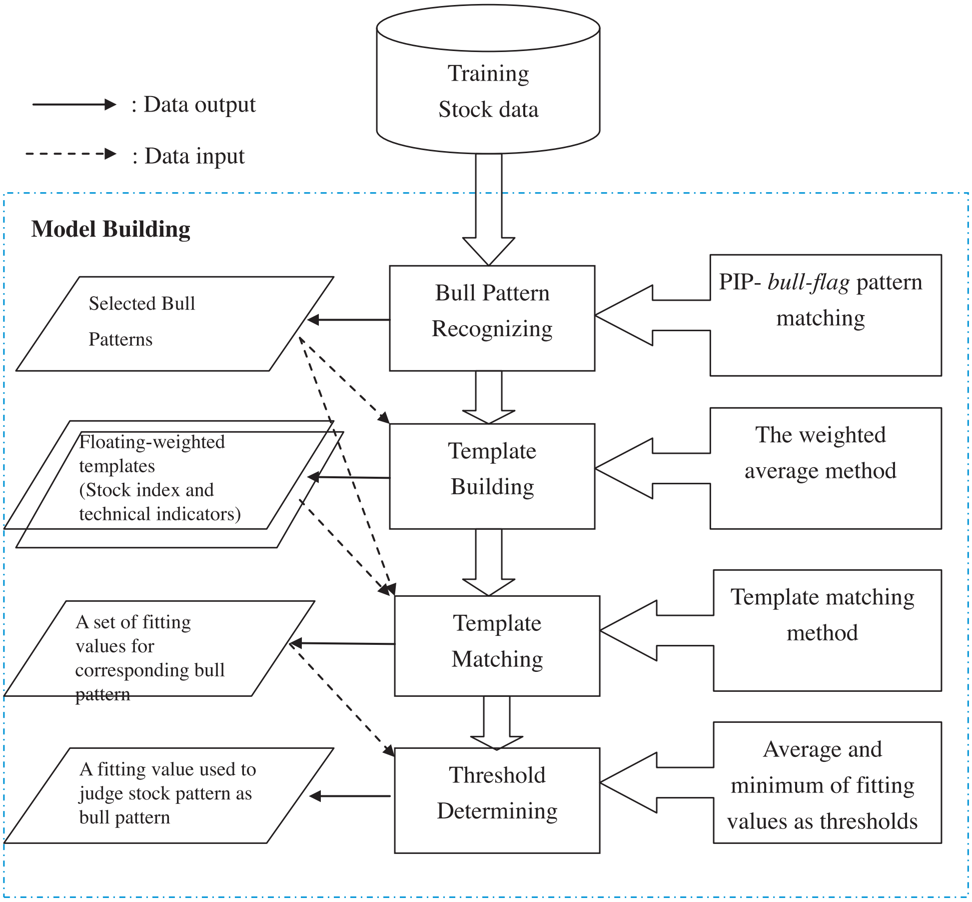

Fig. 6. The framework of the proposed model in training phase.

size of trading days for a stock pattern), *holding-period* (a certain period of trading days for investor to hold stock), and *distance-threshold* (a given value to define a "*bull-flag pattern*"). An example for a bull-flag pattern, based on the template matching method, is demonstrated in Table 1 (charting patterns) and Table 2 (technical indicators).

Step 2: Template building

This step is to build two types of templates (charting patterns and technical indicators) for representing the "*bull-flag template*". The weighted averages method is used to produce the weight for each cell of the template. The pseudo codes for the "*bull-flag template*" of charting pattern are illustrated as follows.

(a) $X = \{x_1, x_2, x_3, \ldots, x_k\}$ is a set of daily closing index containing a possible bull market pattern, where k is the fitting window size.
(b) Rank the index values in set X at a decrement.
(c) Calculate $I_{t,i}$ for trading day t ($1 \le t \le k$), and i is $1 \le i \le 10$.

$I_{t,i} = 1$ if $(i-1) \cdot k / 10 < Rank(x_t) \le i \cdot k/10$
$I_{t,i} = 0$ otherwise

(d) Repeat the procedure (c), until the last $I_{t,i}$ of possible bull flag pattern is calculated.
(e) Calculate $cp\_w_{t,i}$ in template girds for all possible bull flag patterns n, where n is the number of possible bull flag patterns.

$$cp\_w_{t,i} = \frac{\sum_{b=1}^{n} Bull\_flag\_pattern_b(I_{t,i})}{n}$$

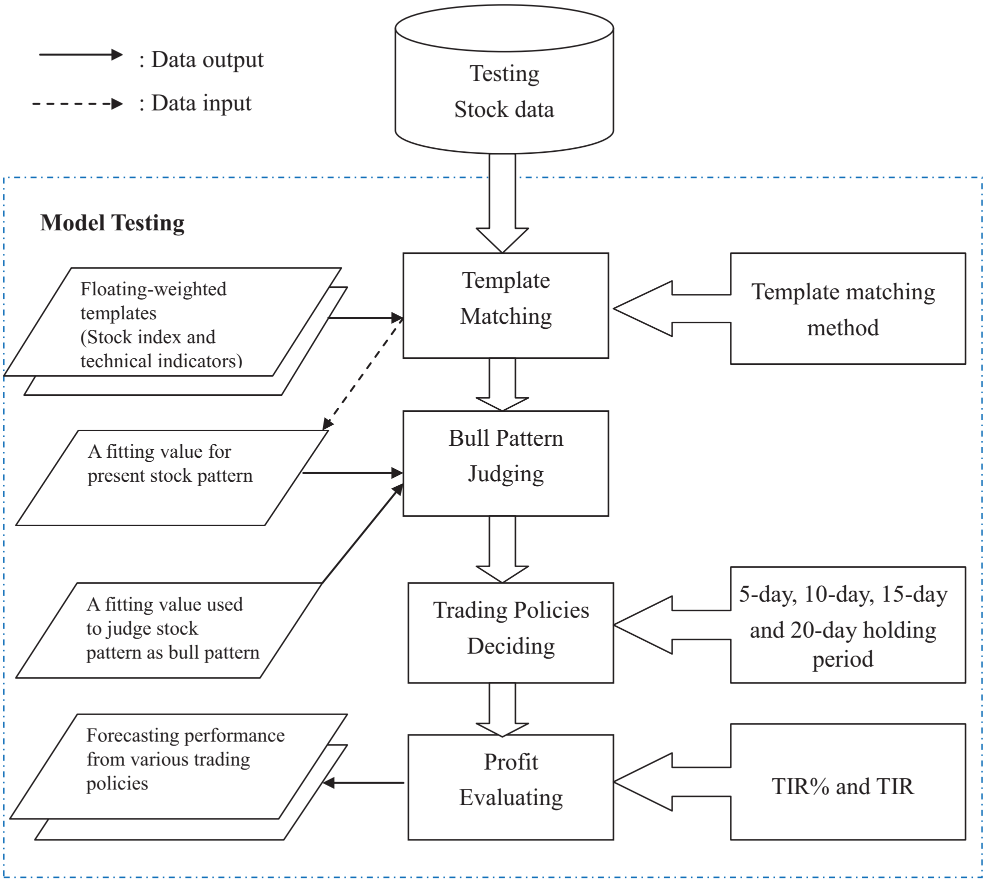

Fig. 7. The framework for the proposed model in testing phase.

For the template of technical indicator, the pseudo codes are listed in the followings.

(a) $X = \{x_{1,1}, x_{1,2}, ..., x_{1,9}, x_{2,1}, ..., x_{k,m}\}$ is a set of technical indicator linguistic values containing a possible bull market pattern, where k is the fitting window size, and m is the number of technical indicators.

(b) Calculate $I_{t,i,Lp}$ for trading day t ($1 \le t \le k$), i is the technical indicator number ($1 \le i \le 9$), and L_p is the linguistic value (where $p = 1, 2, 3$).

$I_{t,i,Lp} = 1$ if indicator(i)'s linguistic $x_{t,i} = L_p$
$I_{t,i,Lp} = 0$ otherwise

(c) Repeat the procedure (b), until the last $I_{t,i,Lp}$ of possible bull flag pattern is calculated.

(d) Calculate $ti\_w_{t,i,Lp}$ in template girds for all possible bull flag patterns n, where n is the number of possible bull flag patterns.

$$ti\_w_{t,i,Lp} = \frac{\sum_{b=1}^{n} Bull\_flag\_pattern_b(I_{t,i,Lp})}{n}$$

Tables 3 and 4 illustrate the examples of "*bull-flag template*" of charting pattern and technical indicator, respectively, based on a 20-day pattern fitting window from 1995/01/06 to 1995/02/06.

Step 3: Template matching

This step is to produce a fitting value for each "*bull-flag pattern*" selected from training data (**step 1**) and the value is used, in the next step, to define the thresholds for judging whether a stock pattern is the *bull-flag pattern*. The template

Table 1
An example for a bull-flag pattern based on template matching method (charting pattern).

i	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
	8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
t		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rank		1	2	3	5	4	7	9	12	10	11	6	8	15	13	19	20	18	17	16	14
Closing index		6,919.31	6,915.40	6,869.08	6,756.88	6,777.24	6,609.50	6,582.40	6,511.30	6,536.65	6,515.79	6,623.52	6,598.02	6,372.01	6,431.99	6,295.04	6,167.79	6,299.62	6,307.85	6,328.38	6,417.28
Date		1995/01/06	1995/01/07	1995/01/09	1995/01/10	1995/01/11	1995/01/12	1995/01/13	1995/01/14	1995/01/16	1995/01/17	1995/01/18	1995/01/19	1995/01/20	1995/01/21	1995/01/23	1995/01/24	1995/01/25	1995/01/26	1995/02/04	1995/02/06

matching technique [19] is applied to calculate the fitting value and the pseudo codes to produce a fitting value, based on the template of charting pattern and technical indicator, are illustrated in the following.

(a) $X = \{x_1, x_2, x_3, ..., x_k\}$ is a set of daily closing index containing a candidate stock pattern, where k is the fitting window size.
(b) Rank the index values in set X at a decrement.
(c) Calculate $J_{t,j}$ for trading day $t\,(1 \le t \le k)$, and j is $1 \le j \le 10$.

$J_{t,j} = 1$ if $(j-1)\cdot k/10 < Rank\,(x_t) \le j\cdot k/10$
$J_{t,j} = 0$ otherwise

(d) Calculate a fitting value (CP_Fit) for a stock pattern based on the template of charting pattern.

$$CP\_Fit = \sum_{t=1}^{k}\sum_{j=1}^{10} cp\_w_{tj} \times J_{tj}$$

(a) $X = \{x_{1,1}, x_{1,2}, ..., x_{1,9}, x_{2,1}, ..., x_{k,m}\}$ is a set of technical indicator linguistic values containing a candidate stock pattern, where k is the fitting window size, and m is the number of technical indicators.
(b) Calculate $J_{t,j,L}$ for trading day $t\,(1 \le t \le k)$, j is the technical indicator number $(1 \le j \le 9)$, and L_p is the linguistic value (where $p = 1,2,3$).

$J_{t,j,Lp} = 1$ if indicator(j)'s linguistic $x_{t,j} = L_p$
$J_{t,j,Lp} = 0$ otherwise

(c) Calculate a fitting value (TI_Fit) for a stock pattern based on the template of technical indicator

$$TI\_Fit = \sum_{t=1}^{k}\sum_{j=1}^{10} ti\_w_{tj,Lp} \times J_{tj,Lp}$$

Table 2
An example for a bull-flag pattern based on template matching method (technical indicator).

i, L_p	MA_L1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	MA_L2	1	0	0	0	1	0	1	0	1	1	0	0	0	1	0	1	0	1	1	0
	MA_L3	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	0	0	1
	RSI_L1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	RSI_L2	1	0	0	0	1	0	1	0	1	1	0	0	0	1	0	1	0	1	1	0
	RSI_L3	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	0	0	1
	STOD_L1	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0
	STOD_L2	1	0	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0	0	1	0
	STOD_L3	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1
	OBV_L1	0	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0	0	1	0	0
	OBV_L2	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
	OBV_L3	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1
	ROC_L1	0	0	1	0	1	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0
	ROC_L2	1	0	0	1	0	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0
	ROC_L3	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1
	VR_L1	0	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0	0	1	0	0
	VR_L2	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
	VR_L3	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1
	PSY_L1	0	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0	0	1	0	0
	PSY_L2	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
	PSY_L3	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1
	AR_L1	0	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0	0	1	0	0
	AR_L2	1	1	0	1	0	0	1	0	0	1	1	0	1	0	0	1	0	0	1	1
	AR_L3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
	DIS_L1	0	0	1	0	1	1	1	0	0	0	0	1	0	1	1	1	0	0	0	0
	DIS_L2	1	1	0	1	0	0	0	0	1	1	1	0	1	0	0	0	0	1	1	1
	DIS_L3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
	t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Technical indicator linguistic value	MA	L2	L3	L1	L3	L2	L3	L2	L3	L2	L2	L3	L1	L3	L2	L3	L2	L3	L2	L2	L3
	RSI	L2	L3	L1	L3	L2	L3	L2	L3	L2	L2	L3	L1	L3	L2	L3	L2	L3	L2	L2	L3
	STOD	L2	L3	L1	L2	L1	L2	L2	L3	L1	L2	L3	L1	L2	L1	L2	L2	L3	L1	L2	L3
	OBV	L2	L3	L1	L2	L1	L1	L2	L3	L1	L2	L3	L1	L2	L1	L1	L2	L3	L1	L2	L3
	ROC	L2	L3	L1	L2	L1	L1	L2	L3	L2	L2	L3	L1	L2	L1	L1	L2	L3	L2	L2	L3
	VR	L2	L3	L1	L2	L1	L1	L2	L3	L1	L2	L3	L1	L2	L1	L1	L2	L3	L1	L2	L3
	PSY	L2	L3	L1	L2	L1	L1	L2	L3	L1	L2	L3	L1	L2	L1	L1	L2	L3	L1	L2	L3
	AR	L2	L2	L1	L2	L1	L1	L2	L3	L1	L2	L2	L1	L2	L1	L1	L2	L3	L1	L2	L2
	DIS	L2	L2	L1	L2	L1	L1	L1	L3	L2	L2	L2	L1	L2	L1	L1	L1	L3	L2	L2	L2
	Date	1995/01/06	1995/01/07	1995/01/09	1995/01/10	1995/01/11	1995/01/12	1995/01/13	1995/01/14	1995/01/16	1995/01/17	1995/01/18	1995/01/19	1995/01/20	1995/01/21	1995/01/23	1995/01/24	1995/01/25	1995/01/26	1995/02/04	1995/02/06

Table 3
The template grids of weights for charting pattern analysis.

Relative high value of stock index																				
1	0.22	0.24	0.1	0.08	0.06	0.08	0.06	0.06	0.08	0.05	0.06	0.08	0.02	0.03	0.02	0	0.08	0.14	0.24	0.3
2	0.06	0.1	0.21	0.21	0.11	0.08	0.06	0.08	0.13	0.1	0.06	0.05	0.1	0.08	0.06	0	0.03	0.19	0.14	0.16
3	0.08	0.08	0.1	0.1	0.21	0.06	0.1	0.16	0.08	0.06	0.03	0.16	0.13	0.16	0.1	0.03	0.13	0.06	0.11	0.08
4	0.14	0.1	0.06	0.03	0.03	0.19	0.17	0.16	0.03	0.11	0.13	0.11	0.13	0.05	0.08	0.11	0.11	0.08	0.11	0.06
5	0.06	0.03	0.08	0.1	0.1	0.11	0.14	0.13	0.22	0.16	0.17	0.13	0.1	0.14	0.08	0.1	0.03	0.03	0.06	0.03
6	0.02	0.1	0.06	0.03	0.06	0.06	0.14	0.05	0.21	0.24	0.19	0.14	0.14	0.11	0.08	0.06	0.06	0.05	0.06	0.13
7	0.06	0.05	0.05	0.05	0.1	0.13	0.08	0.19	0.08	0.06	0.16	0.06	0.16	0.13	0.17	0.02	0.06	0.24	0.06	0.1
8	0.08	0.03	0.06	0.14	0.14	0.14	0.13	0.05	0.08	0.05	0.06	0.14	0.14	0.13	0.05	0.13	0.19	0.06	0.11	0.08
9	0.05	0.1	0.14	0.24	0.08	0.08	0.06	0.05	0.03	0.13	0.08	0.08	0.03	0.11	0.19	0.13	0.21	0.13	0.08	0.02
10	0.22	0.19	0.14	0.03	0.11	0.06	0.05	0.08	0.06	0.05	0.05	0.05	0.06	0.06	0.17	0.43	0.1	0.02	0.02	0.05
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Stock index time-series.

Table 4
The example for the "*bull-flag template*" of technical indicator.

Relative technical indicator linguistic value																				
MA_L1	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.38	0.38	0.37
MA_L2	0.32	0.32	0.32	0.3	0.3	0.3	0.3	0.3	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.3	0.32	0.3	0.3	0.32
MA_L3	0.3	0.3	0.3	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.33	0.32	0.32	0.32	0.32
RSI_L1	0.25	0.22	0.27	0.27	0.29	0.32	0.27	0.32	0.3	0.32	0.3	0.29	0.29	0.32	0.3	0.33	0.32	0.22	0.21	0.22
RSI_L2	0.4	0.44	0.43	0.41	0.38	0.37	0.38	0.35	0.35	0.33	0.37	0.37	0.37	0.32	0.35	0.4	0.32	0.37	0.41	0.43
RSI_L3	0.35	0.33	0.3	0.32	0.33	0.32	0.35	0.33	0.35	0.35	0.33	0.35	0.35	0.37	0.35	0.27	0.37	0.41	0.38	0.35
STOD_L1	0.37	0.38	0.38	0.32	0.32	0.35	0.32	0.35	0.41	0.37	0.35	0.33	0.38	0.43	0.52	0.67	0.22	0.06	0.03	0.08
STOD_L2	0.22	0.17	0.24	0.27	0.33	0.3	0.3	0.27	0.14	0.25	0.27	0.27	0.25	0.19	0.22	0.25	0.48	0.3	0.19	0.17
STOD_L3	0.41	0.44	0.38	0.41	0.35	0.35	0.38	0.38	0.44	0.38	0.38	0.4	0.37	0.38	0.25	0.08	0.3	0.63	0.78	0.75
OBV_L1	0.13	0.16	0.08	0.08	0.06	0.1	0.1	0.1	0.14	0.19	0.16	0.13	0.16	0.13	0.19	0.17	0.14	0.08	0.08	0.1
OBV_L2	0.51	0.46	0.56	0.57	0.57	0.56	0.52	0.56	0.48	0.49	0.56	0.54	0.59	0.67	0.6	0.62	0.57	0.56	0.49	0.46
OBV_L3	0.37	0.38	0.37	0.35	0.37	0.35	0.38	0.35	0.38	0.32	0.29	0.33	0.25	0.21	0.21	0.21	0.29	0.37	0.43	0.44
ROC_L1	0.19	0.25	0.25	0.21	0.24	0.22	0.27	0.21	0.27	0.27	0.19	0.24	0.24	0.29	0.27	0.35	0.17	0.03	0.02	0
ROC_L2	0.49	0.38	0.38	0.37	0.41	0.44	0.44	0.49	0.37	0.4	0.49	0.48	0.44	0.43	0.49	0.52	0.62	0.65	0.57	0.33
ROC_L3	0.32	0.37	0.37	0.43	0.35	0.33	0.29	0.3	0.37	0.33	0.32	0.29	0.32	0.29	0.24	0.13	0.21	0.32	0.41	0.67
VR_L1	0.21	0.24	0.29	0.19	0.17	0.13	0.21	0.22	0.25	0.29	0.24	0.27	0.35	0.21	0.21	0.22	0.16	0.08	0.1	0.08
VR_L2	0.35	0.33	0.32	0.44	0.44	0.49	0.33	0.43	0.3	0.4	0.44	0.38	0.37	0.54	0.59	0.57	0.46	0.51	0.35	0.41
VR_L3	0.44	0.43	0.4	0.37	0.38	0.38	0.46	0.35	0.44	0.32	0.32	0.35	0.29	0.25	0.21	0.21	0.38	0.41	0.56	0.51
PSY_L1	0.25	0.27	0.24	0.19	0.22	0.16	0.17	0.16	0.19	0.25	0.22	0.21	0.24	0.22	0.27	0.29	0.19	0.17	0.1	0.11
PSY_L2	0.38	0.4	0.41	0.43	0.37	0.44	0.41	0.44	0.37	0.33	0.32	0.37	0.33	0.33	0.29	0.37	0.44	0.38	0.52	0.43
PSY_L3	0.37	0.33	0.35	0.38	0.41	0.4	0.41	0.4	0.44	0.41	0.46	0.43	0.43	0.44	0.44	0.35	0.37	0.44	0.38	0.46
AR_L1	0.3	0.29	0.32	0.32	0.32	0.32	0.35	0.32	0.3	0.33	0.3	0.33	0.35	0.33	0.35	0.37	0.29	0.29	0.29	0.27
AR_L2	0.37	0.43	0.38	0.33	0.38	0.37	0.35	0.38	0.38	0.33	0.35	0.37	0.37	0.37	0.41	0.41	0.41	0.41	0.41	0.43
AR_L3	0.33	0.29	0.3	0.35	0.3	0.32	0.3	0.3	0.32	0.33	0.35	0.3	0.29	0.3	0.24	0.22	0.3	0.3	0.3	0.3
DIS_L1	0.24	0.22	0.24	0.19	0.21	0.22	0.24	0.21	0.27	0.24	0.22	0.21	0.25	0.27	0.25	0.37	0.1	0	0	0.02
DIS_L2	0.46	0.33	0.38	0.52	0.48	0.48	0.44	0.44	0.33	0.44	0.43	0.48	0.41	0.38	0.56	0.56	0.65	0.43	0.33	0.22
DIS_L3	0.3	0.44	0.38	0.29	0.32	0.3	0.32	0.35	0.4	0.32	0.35	0.32	0.33	0.35	0.19	0.08	0.25	0.57	0.67	0.76
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Stock index time-series.

Step 4: Threshold determining

In this step, average and minimum thresholds are applied to define the criteria to judge whether a stock pattern is the *bull-flag pattern* or not. They are generated with the fitting values of the "*bull-flag patterns*" selected from training stock data. For the template of charting pattern, the thresholds are defined by Eqs. (3) and (4), where n is the number of possible bull market patterns.

$$CP\_Threshold_{avg} = \frac{\sum_{b=1}^{n} Bull\_market\_pattern_b(CP\_Fit)}{n} \tag{3}$$

$$CP\_Threshold_{\min} = \min_{b=1\ to\ n} (Bull\_market\_pattern_b(CP\_Fit)) \tag{4}$$

For the template of technical indicator, the thresholds are defined by Eqs. (5) and (6), where n is the number of possible bull market patterns.

$$TI\_Threshold_{avg} = \frac{\sum_{b=1}^{n} Bull\_market\_pattern_b(TI\_Fit)}{n} \tag{5}$$

$$TI\_Threshold_{\min} = \min_{b=1\ to\ n} Bull\_market\_pattern_b(TI\_Fit) \tag{6}$$

From Eq. (3) to Eq. (6), we can obtain the "thresholds" based on the templates of charting pattern and technical indicator. With these thresholds, a trading rule is provided as follows: "*If the two fitting values* (*charting pattern and technical indicator) for a stock pattern are both "greater than" or "equal to" the thresholds, the stock pattern should be defined as a "bull-flag pattern" and we should buy the stock at the next day to make profit in the future.*"

Phase II: Model testing

Step 5: Template matching

This step is to produce two fitting values for each stock pattern selected from testing data. One is from the "*bull-flag template*" of charting pattern and the other is from the template of technical indicator.

Step 6: Bull pattern judging

This step is to judge whether the present pattern is the "*bull-flag pattern*" or not, and the trading (judging) rule is listed in **step 4**.

Step 7: Trading policies deciding

Two experimental factors, holding-period (a certain period of trading days for investor to hold the investing stock) and "stop loss" are employed in the trading rule (**step 4**) to examine investing profit and avoid a heavy loss when stock market trend is going in the other direction for a long time. In the proposed model, the holding-period is set as four vales: 5-days, 10-days, 15-days and 20-days and the "stop loss" is set as 10% of buying price for the stock.

Step 8: Profit evaluating

Two performance indicators, total index return percentage (TIR%) and total index return (TIR), are employed to evaluate model performance because both are adopted in past research models [5,31]. TIR% is defined by Eq. (7) which measures a gain rate of price index for total transactions, and TIR is defined by Eq. (8) which measures an absolute gain of price index for total transactions. The major difference between the performance indicators is that TIR is more convenient to produce actual profit return for investor when transaction cost is determined.

$$\text{Total index return}(\%) = \sum_{n=1}^{m} \frac{Closing\_index_s - Closing\_index_b}{Closing\_index_b} \tag{7}$$

$$\text{Total index return} = \sum_{n=1}^{m} Closing\_index_s - Closing\_index_b \tag{8}$$

4. Model verification

In this section, we conduct an experiment with 3 factors (pattern fitting window, PIP-distance for stock pattern, and holding-period) for evaluating the proposed model, and employ 4 advanced models as benchmarks: Wang and Chan's model [31], the RST algorithms [23,27], the genetic algorithms [13], and Cheng et al.'s model [5].

4.1. Performance evaluation and model comparison

The experimental datasets of NASDAQ contains four testing-periods: (1)1989/01/07 to 1992/10/27; (2) 1992/10/28 to 1996/08/19; (3)1996/08/12 to 200/05/26 and (4) 2000/05/27 to 2004/40/30. The TAIEX contains two testing-periods: (1) 1990/08/15 to 1997/02/17 and (2) 1997/02/18 to 2004/03/24. The training–testing ratio is 7:1 (year) for the proposed model.

Because two "*bull-flag*" templates, charting pattern and technical indicator, are employed in forecasting process, the proposed model can be classified as three models based on the usage of templates: (1) model A employs the template of technical indicator to generate a total fit value, (2) model B employs the template of charting pattern to generate a total fit value, and (3) model C employs the two templates simultaneously to a total fit value.

Initially, Wang and Chan' [31] model is used as benchmark in forecasting performance. Table 5 shows the performance data for the proposed and the comparison models. It is clear that the proposed model (models A, B, C) bears much better TIR% than Wang and Chan's for forecasting the NASDAQ and the TAIEX; model B performs better than the others in the NASDAQ; and models A and C give the best TIR% for the TAIEX.

To examine the forecasting performance further, we use the RST algorithms, the GAs algorithms and Cheng et al.'s [5] model as comparison in the second experiment. The experimental datasets contains 5 testing-periods, the last two months of the year from 2001 to 2005. The testing periods for the proposed model are the same with Cheng et al.'s model. To make the model comparison in fair condition, only the performance data of model A is provided because it employs the same forecasting factors (the nine technical indicators: MA, RSI, STOD, OBV, ROC, VR, PSY, AR, and DIS) with the three algorithms. Table 6 lists all performance data with TIR and it also has revealed that model A outperforms the listing advanced algorithms.

Table 5
Performance comparison with the model advanced by Wang and Chan [31].

Model		Wang and Chan's			Proposed model A			Proposed model B		Proposed model C	
Dataset	Testing period	TIR (%)	Average IR (%)	*N* (trades)	PIP-distance	TIR (%)	*N* (trades)	TIR (%)	*N* (trades)	TIR (%)	*N* (trades)
NASDAQ	1989/01/17	291	1.94	150	5	132.29	232	*429.12	241	−7.04	52
	|				10	−31.30	150	253.10	170	−25.35	40
	1992/10/27				15	−57.13	80	273.43	116	−22.14	24
	1992/10/28	201.96	1.32	153	5	0.00	0	*338.94	278	0.00	0
	|				10	0.00	0	217.04	224	0.00	0
	1996/08/09				15	0.00	0	163.17	149	0.00	0
	1996/08/12	444.27	2.51	177	5	99.94	32	*1084.17	446	46.65	23
	|				10	94.16	34	1056.40	403	40.87	25
	2000/05/26				15	93.51	41	937.34	356	51.22	28
	2000/05/27	24.96	0.39	64	5	*168.24	70	−35.06	242	4.31	40
	|				10	−15.15	115	−62.81	240	−166.18	62
	2004/03/20				15	−227.94	169	−103.50	242	−336.86	86
TAIEX	1990/08/15	286.88	1.63	176	100	*1291.20	319	1156.96	395	1159.79	216
	|				200	694.97	319	1190.92	443	787.83	210
	1997/02/17				300	742.23	257	1216.23	430	592.98	162
	1997/02/18	314.5	1.7	185	100	479.40	635	818.71	667	*940.98	269
	|				200	340.95	633	495.73	604	752.13	239
	2004/03/24				300	123.94	498	166.57	441	352.94	120

* Denotes the best among comparison models.

Table 6
Performance comparison with the models advanced by RST, GAs and Cheng et al. [5].

TAIEX Testing period	Comparison model				Proposed model A (pattern fitting window = 60; PIP-distance = 200; minimum threshold)			
	Buy-and-hold	Rough set theory	Genetic algorithms	Cheng et al.'s	5-day holding	10-day holding	15-day holding	20-day holding
2001(Nov.–Dec.)	1612.16	923.9	1272.05	1683.51	4923.25	11203.01	16524.59	*20303.3
2002(Nov.–Dec.)	−144.24	304.28	353.56	421.82	−454.07	689.46	2085.70	*3402.29
2003(Nov.–Dec.)	−163.62	70.78	247	336.25	100.08	1045.00	3026.30	*5369.40
2004(Nov.–Dec.)	414.04	196.06	481.65	780.26	*1034.71	985.25	724.19	322.84
2005(Nov.–Dec.)	740.59	105.85	163.8	210.83	1982.12	4733.39	6355.68	*7517.35
Average	491.79	320.17	503.61	686.53	1517.22	3731.22	5743.29	*7383.04

Performance indicator: TIR.
* The best among comparison models.

4.2. Findings and discussions

In this section, we give a summarization for experimental results. Form the comparison results with the listing 4 advanced algorithms, 6 findings are discovered in the followings.

(1) The proposed model performs differently for forecasting the TAIEX and NASDAQ because of their stock patterns are dissimilar. From Table 5, we can see that the PIP-distance values with better TIR (%) for the NASDAQ (5, 10 and 15) are much smaller than TAIEX (100, 200 and 300). This finding is easy to be understood because the spread of TAIEX is greater than NASDAQ and, therefore, the bull-flag patterns of TAIEX should be much larger than NASDAQ.
(2) Small stock pattern is better than lager pattern for forecasting stock market. From Table 5, it is shown that the TIR (%) for a lower PIP-distance value is better than a higher value in the TAIEX and NASDAQ. It tells that the bull-flag patterns are usually in the pattern with smaller height.
(3) Charting pattern analysis is more effectively than technical indictor for forecasting the NASDAQ. From Table 5, it is clear that model B performs better than model A in 3 testing datasets of the NASDAQ. However, in the TAIEX, model A outperforms model B in the first testing dataset. This phenomenon implies that charting pattern analysis can discover more bull-flag patterns in the NASDAQ and the signal of the bull-flag pattern for technical indictor in the NASDAQ is not very clear. However, in the TAIEX, the signal of the bull-flag pattern is easy to be found by using technical indictor analysis.
(4) The hybrid model (model C) using the two templates simultaneously to generate the total fit value can reduce the transaction cost and maintain a high TIR (%). From Table 5, it is found that the amount of trades (N) for model C is much less than model A and B for the TAIEX. For the first testing period, the amount of trades for model C is 216, model A is 319, and model B is 395. For the second testing period, model C is 269, model A is 635, and model B is

667. Besides, in the first testing period, the TIR (%) for model C is 1159.79 that is a little less than the best TIR (%), 1291.20. And, in the second testing period, model C bears the best TIR (%), 940.98, with the least amount of trades among three models. This evidence indicates that model C brings more profit when forecasting the TAIEX.

(5) A longer holding period can make higher stock return. From Table 6, the performance data has clearly shown that model A gives excellent forecasting performance (the average TIR of model A is 2–10 times of Cheng et al.'s model) and four best TIRs are produced by using 20-day holding period. From the evidence, it is strongly suggested that investor should invest the TAIEX for a longer holding period, such as 20 days at least. This phenomenon was also found in the research conducted by Wang and Chan [31].

(6) Overall, model B is a better profitable model among the testing models and especially for forecasting the TAIEX. From Table 5, it is clear that model B performs well for forecasting the NASDAQ and the TAIEX except the testing period of the NASDAQ (from 2000/05/27 to 2004/03/20). In practical applications for stock market forecasting, model A is a traditional method refined by the framework of the proposed model. Model B is a "new" method based on technical indicators provided by the framework of the proposed model. Model C is a "hybrid" model of model A and model B. From the experimental results, model B is the most profitable one among three models but the trading number is the largest and model C bears the smallest trading number with acceptable profit return for the TAIEX. Therefore, model B is suggested when the investor has a "positive" or "bullish" view for the present stock market and model C is employed when the investor has a "negative" or "bearish" view.

5. Conclusion

In this paper, we proposed an intelligent pattern recognition model to ascertain the bull-flag patterns contained in historical stock patterns, and extract two types of bull-flag templates, the chart pattern and the technical indicator, as judging criteria for the bull-flag pattern.

The contributions for the proposed model are: (1) address a new definition for the bull-flag pattern (see Fig. 4) beside the subjective the bull-flag pattern defined by researchers [8,18,31]; (2) provide a new weighted method, based on the occurrence of observations, to effectively fuse information of stock pattern and indicators; (3) provide an effective granulating method, the CPDA method [5], to discretize technical indicators to reduce computing complexity; (4) provide a new concept which integrates the advantages of the past two pattern-matching methods, PIP identification-matching and template-matching, in the recognition process beside the past models which devoted to find the bull-flag pattern only with the chart pattern [8,18,31]. Therefore, it provided effective rules (bull-flag templates) to judge a stock pattern more accurately.

Based on the trading rule for the bull-flag pattern, the proposed model gave an unprecedented stock index return (TIR% and TIR) in forecasting the NASDAQ and TAIEX. In the experiments for model verification, the performance data proved superior to the proposed models by comparing them with the three advanced stock market forecasting algorithms, including RST (LEM2), GAs and their hybrid model, issued by Cheng et al. [5]. The experimental results also demonstrate a robust stock return, when used to forecast the TAIEX. Although the proposed model (model B) had a negative TIR% for forecasting the fourth testing period of the NASDAQ, the average TIR% was still at a high, positive level (454.29%), and was much better than Wang and Chan's model (240.55%). Additionally, the proposed model, with two templates (model C), can give a high level of stock index return (TIR) with a low number of trades.

The experimental data indicate that this model can provide a highly accurate forecast for the bull-flag pattern, thereby allowing stock investors to accrue huge profits by employing it. Further, this model also allows investors to examine the stock information of selected bull-flag patterns, including basic and technical indicators. This information can help stock analysts or senior stock investors to check stock patterns more carefully.

In the future, in order to strengthen the proposed model, two suggestions are recommended, as follows: (1) use other stock databases, such as the S&P 500 and the NYSE, to evaluate the proposed model; and (2) apply other advanced pattern-searching methods to the proposed model, such as wavelet theory, to extract the bull-flag pattern to produce the bull-flag template and evaluate stock return.

References

[1] S. Asadi, E. Hadavandi, F. Mehmanpazir, M.M. Nakhostin, Hybridization of evolutionary Levenberg-Marquardt neural networks and data pre-processing for stock market prediction, Knowledge Based Syst. 35 (2012) 245–258.
[2] L. Blume, D. Easley, M. O'Hara, Market statistics and technical analysis: The role of volume, J. Finance 49 (1) (1994) 153–181.
[3] W. Brock, J. Lakonishok, B. Lebaron, Simple technical trading rules and the stochastic properties of stock returns, J. Finance 47 (5) (1992) 1731–1764.
[4] C.H. Chen, Neural networks for financial market prediction, in: Paper presented at IEEE International Conference on Neural Networks, Orlando, Florida, 1994.
[5] C.H. Cheng, T.L. Chen, L.Y. Wei, A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting, Inf. Sci. 180 (9) (2010) 1610–1629.
[6] V. Cho, MISMIS – a comprehensive decision support system for stock market investment, Knowledge Based Syst. 23 (2010) 626–633.
[7] F.L. Chung, T.C. Fu, R. Luk, V. Ng, Flexible time series pattern matching based on perceptually important points, in: Paper presented at International Joint Conference on Artificial Intelligence Workshop on Learning from Temporal and Spatial Data, Seattle, Washington, 2001.
[8] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, John Wiley & Sons, New York, 1973.
[9] R.D. Edwards, J. Magee, W.H.C. Bassetti, Technical Analysis of Stock Trends, ninth ed., CRC Press, Florida, 2007.
[10] T.C. Fu, A review on time series data mining, Eng. Appl. Artif. Intell. 24 (2011) 164–181.
[11] T.C. Fu, F.L. Chung, R. Luk, C.M. Ng, Stock time series pattern matching: Template-based vs. rule-based approaches, Eng. Appl. Artif. Intell. 20 (2007) 347–364.

[12] B.D. Garland, SV mixture models with application to S&P 500 index returns, J. Financ. Econ. 85 (3) (2007) 822–856.
[13] J.H. Holland, Adaptation in Nature and Artificial Systems, University of Michigan, 1975.
[14] K.J. Kim, I. Han, The extraction of trading rules from stock market data using rough sets, Expert Syst. 18 (4) (2001) 194–202.
[15] M.J. Kim, S.H. Min, I. Han, An evolutionary approach to the combination of multiple classifiers to predict a stock price index, Expert Syst. Appl. 31 (2) (2006) 241–247.
[16] W. Leigh, N. Modani, R. Hightower, A computational implementation of stock charting: abrupt volume increase as signal for movement in New York stock exchange composite index, Decis. Support Syst. 37 (4) (2004) 515–530.
[17] W. Leigh, N. Modani, R. Purvis, T. Roberts, Stock market trading rule discovery using technical charting heuristics, Expert Syst. Appl. 23 (2) (2002) 155–159.
[18] W. Leigh, R. Purvis, J.M. Ragusa, Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural networks, and genetic algorithm: a case study in romantic decision support, Decis. Support Syst. 32 (4) (2002) 361–377.
[19] A. Lo, H. Mamaysky, J. Wang, Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation, J. Finance 55 (4) (2000) 1705–1765.
[20] R. Majhi, G. Panda, Prediction of S&P500 and DJIA stock indices using particle swarm optimization technique, in: Paper presented at 2008 IEEE Congress on Evolutionary Computation, Hong Kong, China, 2008.
[21] C. Neely, P. Weller, R. Dittmar, Is technical analysis in the foreign exchange market profitable? A genetic programming approach, J. Financ. Quant. Anal. 32 (4) (1997) 405–426.
[22] S.N. Neftci, Naive trading rules in financial markets and Wiener-Kolmogorov prediction theory: a study of “Technical analysis”, J. Bus. 64 (4) (1991) 549–571.
[23] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.
[24] Z. Pawlak, Rough sets and intelligent data analysis, Inf. Sci. 147 (2002) 1–12.
[25] Z. Pawlak, A. Skowron, Rough sets: some extensions, Inf. Sci. 177 (2007) 28–40.
[26] Z. Pawlak, A. Skowron, Rudiments of rough sets, Inf. Sci. 177 (2007) 3–27.
[27] B. Predki, R. Slowinski, J. Stefanowski, R. Susmaga, S. Wilk, ROSE-software implementation of the rough set theory, Lecture Notes in Artificial Intelligence, 1424th ed.,, Springer-Verlag, Berlin, 1998, pp. 605–608.
[28] W. Shen, X. Guo, Wu C., D. Wu, Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm, Knowledge Based Syst. 24 (2011) 378–385.
[29] C.H. Su, T.L. Chen, C.H. Cheng, Y.C. Chen, Forecasting the stock market with linguistic rules generated from the minimize entropy principle and the cumulative probability distribution approaches, Entropy 12 (2011) 2397–2471.
[30] H.J. Teoh, C.H. Cheng, Chu H.H., J.S. Chen, Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets, Data Knowl. Eng. 67 (2008) 103–117.
[31] J.L. Wang, S.H. Chan, Stock market trading rule discovery using pattern recognition and technical analysis, Expert Syst. Appl. 33 (2) (2007) 304–315.
[32] W. Wuang, Y. Nakamori, S. Wang, Forecasting stock market movement direction with support vector machine, Comput. Oper. Res. 32 (10) (2005) 2513–2522.
[33] C.R. Roberto, G. Francisco, M. Karolina, Stock market trading rule based on pattern recognition and technical analysis: Forecasting the DJIA index with intraday data, Expert Syst. Appl. 42 (14) (2015) 5963–5975.
[34] S.H. Cheng, S.M. Chen, W.S. Jian, Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures, Inf. Sci. 327 (2016) 272–287.
[35] M.Y. Chen, B.T. Chen, A hybrid fuzzy time series model based on granular computing for stock price forecasting, Inf. Sci. 294 (2015) 227–241.
[36] T. Xionga, Y. Bao, Z. Hu, R. Chiong, Forecasting interval time series using a fully complex-valued RBF neural network with DPSO and PSO algorithms, Inf. Sci. 305 (2015) 77–92.
[37] L.A. Laboissierea, R.A.S. Fernandes, G.G. Lage, Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks, appl. soft comput. 35 (2015) 66–74.
[38] H. Li, C.-J. Li, X.-J. Wu, J. Sun, Statistics-based wrapper for feature selection: An implementation on financial distress identification with support vector machine, appl. soft comput. 19 (2014) 57–67.
[39] R.K. Nayak, D. Mishra, A.K. Rath, A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices, appl. soft comput. 35 (2015) 670–680.
[40] M. Pulido, P. Melin, O. Castillo, Particle swarm optimization of ensemble neural networks with fuzzy aggregation for time series prediction of the Mexican Stock Exchange, Inf. Sci. 280 (2014) 188–204.