



A formal approach to candlestick pattern classification in financial time series

Weilong Hu^a, Yain-Whar Si^{a,*}, Simon Fong^a, Raymond Yiu Keung Lau^b

^a Department of Computer and Information Science, University of Macau, China

^b Department of Information Systems, City University of Hong Kong, China

ARTICLE INFO

Article history:

Received 6 August 2018

Received in revised form 10 July 2019

Accepted 6 August 2019

Available online 19 August 2019

Keywords:

Financial time series

Candlestick chart patterns

Formal specifications

Pattern matching

Stock markets

ABSTRACT

Patterns with varying numbers of candlestick-shaped features are commonly used by analysts to predict future price trends in financial markets. Although general descriptions of candlestick patterns have been reported in literature, they are usually described in natural languages. Such descriptions are prone to ambiguity and misinterpretation by users. Hence, these descriptions written in natural language cannot be easily adopted for use in computational technical analysis. Since there is also no agreed-upon standard on describing the definitions of these patterns, inconsistencies can easily occur during the applications of these patterns. To alleviate these problems, we propose a comprehensive formal specifications of 103 known candlestick patterns. Our goal is to establish an unambiguous reference model which can be used in future pattern classification research without significant modifications. The formal specifications of these patterns are formulated in the form of the first-order logic, which is comprehensive, extensible, and reusable. To evaluate the proposed specifications, extensive experiments are performed for classifying candlestick patterns from synthetic and real datasets. The experimental results show that the proposed specifications can be used to effectively generate synthetic datasets for selecting best classifiers for candlestick pattern identification.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

A candle or a candlestick is a type of visualization for displaying four basic price indexes of a stock in a chronological order. Each candlestick represents an explicit price movement of one day, including opening, closing, high and low price. The notions of candlestick were developed in the 18th century by Munehisa Homma (1724–1803) [1], a Japanese rice merchant from Sakata, as a technical analysis tool to record and analyze the prices of rice. The use of candlesticks in stock market was popularized by Steve Nison in his book, “Japanese Candlestick Charting Techniques” [1] in 1990. Candlesticks contain more information and they are easier to understand compared to traditional bar charts. Therefore, candlesticks are often used to predict the future price trends in stock markets.

In [2], Thomas N. Bulkowski compiles the natural language descriptions of 103 known candlestick patterns from financial time series. In addition, a plethora of methods were also proposed in recent years to classify candlestick patterns from financial time series for predicting future price trend. The methods for the

classification of candlestick patterns in financial series stem from two main approaches: supervised and unsupervised learning. The former includes support vector machines (SVM) [3], neural networks [4], fuzzy reasoning [5–11] and rule-based (RB) methods [12–14]. In unsupervised learning, the existing approaches for the clustering of the candlestick patterns include genetic algorithm (GA) [6,15], hierarchical clustering [16], nearest neighbor clustering [17] and image based query [18]. Several hybrid methods for candlestick patterns classification were also proposed in recent years. These approaches include hybrid clustering algorithms based on particle swarm optimization with simulated annealing [19], adaptation of both traditional technical analysis and artificial intelligence [20], combination of neural network and fuzzy system [21], and approaches based on artificial neural networks (ANN), support vector machines (SVM) and GA [3,22,23].

Although candlestick patterns are commonly used for prediction of price trends, definitions of the patterns used in these applications often vary to a certain degree. Since there is no agreed-upon industry-wide standard for describing their properties, inconsistent pattern definitions could be inadvertently adopted by users. The lack of comprehensive pattern definitions can also lead to discrepancies when performance of different classification methods are compared. To alleviate these problems, in this paper, we propose a formal specification of 103 known candlestick patterns. To avoid any potential ambiguity in the definitions of these

* Correspondence to: Faculty of Science and Technology, University of Macau, E11-4022, Avenida da Universidade, Taipa, Macau, China.

E-mail addresses: 375962913@qq.com (W. Hu), fstasp@umac.mo (Y.-W. Si), ccfong@umac.mo (S. Fong), raylau@cityu.edu.hk (R.Y.K. Lau).

patterns, the formal specifications are formulated in first-order-logic representations. These specifications are also designed in such a way that they can be applied in software systems without significant modifications.

The formal specifications proposed in this paper can also be easily extended when a new pattern is to be defined. Therefore, the formal specifications compiled in this paper are unambiguous, comprehensive, extensible, and reusable. We hope that the specifications compiled in this study will become a reference model for future research in candlestick pattern classification. To evaluate the effectiveness of the proposed specifications, extensive experiments are performed on synthetic and real datasets. Synthetic datasets are generated from the rules which are translated from the proposed specifications. The experimental results show that the proposed specifications can be used to effectively generate synthetic datasets for selecting best classifiers for candlestick pattern identification.

This paper is structured as follows. In Section 2, we briefly review the related work on candlestick patterns classification. In Section 3, we establish the basic definitions of candlestick patterns. Based on these definitions, the formal specifications for 103 candlestick patterns are proposed in Section 4. In Section 5, we evaluate the effectiveness of the specifications with synthetic and real datasets from the stock markets. In Section 6, we conclude the paper with future work.

2. Related work

In recent years, Artificial Intelligence (AI) is increasingly used to develop new methods for financial time series classification and prediction. Advanced trading systems equipped with techniques such as fuzzy reasoning, artificial neural networks, genetic algorithms, and expert systems gain widespread popularity in finance for stock selection and forecasting [24].

In the context of supervised classification, one of the widely used methods is the rule based (RB) method. RB allows the most straightforward way to identify the candlestick patterns based on explicit rules. In [13], a one by four (1×4) vector about the ranking of the opening and closing prices of two candlesticks is used to systematically categorize the two-day candlestick patterns in three main European stock markets. The long-term profitability of candlestick trading strategies in six two-day reversal (i.e. bullish and bearish) patterns with the RB method is also investigated in [12]. Fuzzy logic reasoning is also used for the classification of candlestick patterns. In [5], the size, relation and color of candlesticks are described in fuzzy logic. In [7], a system based on fuzzy candlestick patterns is also introduced. The system can recognize the candlestick patterns with specific size, relation, color, and trend by comparing the distance between the patterns based on Hamming distance. In [9], a classification approach based on the “Hammer” pattern is used to predict the future trend by using the fuzzy inference mechanism. In [8], a candlestick pattern classification model for the frequent item sets (i.e. the common candlestick patterns) from symbolic stocks series is proposed. In the prototype system [10], fuzzy candlestick patterns are identified from a large amount of the financial trading data, and a knowledge base is used to store investment strategies. Also in [14], five groups of candlestick patterns (falling, rising, neutral, trend continuation and trend-reversal patterns) are classified by a expert system.

Clustering methods are also widely used for unsupervised classification of candlestick patterns. In [16], Karsten Martiny adopts an agglomerative hierarchical clustering method with the Euclidean distance measure for automatically discovering significant candlesticks patterns from a time series of price data. A new method called nearest neighbor clustering algorithm based on

the similarity match model of candlestick series is also proposed to test the predictive power of the patterns as well as to mine these patterns from the time series. In addition, image retrieval methods are also used to retrieve similar historical candlestick charts which are represented by the image features. In [18], content-based image retrieval (CBIR) technique is proposed to predict the stock price by extracting the wavelet texture and Canny edge in low-level images from candlestick charts.

Hybrid methods are also proposed for the classification of candlestick patterns. Genetic Algorithms (GA) are global search algorithms for finding optimal solutions through the mechanics of natural selection and natural genetics [15]. In [23], the authors described an autonomous trading system with a GA. An Echo State Network (ESN) is also employed in [23] to target the bear markets. A hybrid pattern clustering algorithm combining particle swarm optimization with simulated annealing is proposed in [19]. The proposed algorithm was applied to cluster candlestick patterns from Warsaw stock market. A nearest neighbor clustering algorithm based on the similarity match model is also used in [17] for data mining of candlestick patterns. A distance measure based on shape of the upper shadow, the lower shadow, and the body of the candlestick was used in [17] to calculate the similarity between the candlesticks. In [6], a genetic algorithm is used to select a set of fuzzy candlestick patterns for investment decision making. These patterns are extracted from the financial time series based on the model of fuzzy linguistic variables.

Recurrent neural networks are used in [25] to classify triangle patterns from financial time series. The proposed approach was used to extract sixteen triangle patterns from the time series data of Tokyo Stock Exchange. In [3], candlestick patterns are used to represent the behavior of the proposed features of a stock. In their approach, ANN and SVM as adopted as classifiers to learn and train with these patterns for predicting the stock price. In [20], a decision support system was proposed for the stock selection problem. The proposed system combines Japanese candlesticks philosophy with technical indicators, charts analysis, and Dow theory. The system also adopts several AI techniques and fuzzy decision rules to verify the results from the technical analysis. Financial time series data from Taiwan stock market was used to validate their proposed approach.

Several commercial software tools are available for the classification of candlestick patterns. For instance, Xpertrader [26] and Investor/RT [27] provide functions for developers or users to identify a set of default candlestick patterns. To the best of authors' knowledge, both software tools provide the functions for classification of approximately 40 candlestick patterns. In addition, libraries for identifying candlesticks patterns are provided in a multi-platform tool for market analysis called TA-Lib [28]. A MATLAB library [29] was also developed by Nate Jensen (N.J.) for identifying 90 candlestick patterns. In contrast to these systems, this paper compiles the formal specifications of 103 known candlestick patterns from the literature.

3. Candlestick patterns

In this section, we introduce the crucial definitions of candlestick patterns as well as all relevant terms and functions used in the specifications. For the purpose of consistency, we assume that the increase or decrease in maximum price in a single trading day is 10%. Depending on the market situation, this percentage can be adjusted by researchers or developers. Since the formal specifications of the candlestick patterns proposed in this work are transformed into rules, we will simply name the classification method used in this paper as *rule-based (RB) method*. Note that, the terms “candle” and “candlestick” are interchangeable in this paper.

3.1. Definitions for describing candlestick and candlestick patterns

A candlestick consists of four basic indicators, which are opening price, high price, low price and closing price. They are often denoted as open, high, low and close for simplicity. Two examples including a white and black basic candlesticks are depicted in Fig. 1(a). The chart, a single price bar, in the left of Fig. 1(a) is a white candle whose closing price is higher than the opening price. The candle is colored in white because the price closes above the opening price. The area between the opening and closing prices is drawn as a box, which is called “a body” or “a real body”. Different types of the body sorted by the height from the shortest to the longest one are called “Doji”, “short body”, “normal body”, “long body” and “extremely long body”. The opening price of a Doji is usually similar to the closing price. The vertical lines drawn above and below the real body are called the upper and lower shadows respectively. The upper shadow in the white candle shown in Fig. 1(a) is the line segment between the high and closing price. Likewise, the lower shadow is the line segment between the low and opening price, and vice versa in the black candle.

Candlesticks can also be represented in abbreviated forms to omit the opening and closing prices. For example, three kinds of abbreviated candlesticks are depicted in Figs. 1(b), 1(f) and 1(g). The candlestick in Fig. 1(b) is a simplified candle whose opening and closing prices are ignored. The candlestick in Fig. 1(c) is a kind of a specific candle called Doji. Candlesticks are often drawn in black and white (see Figs. 1(d) and 1(e)) to denote that both of them can be either a white candle or a black candle. However, in practice, only a white or a black body may be used. The candle shown in Fig. 1(d) will be used when the candle is not a Doji. The candle shown in Fig. 1(f) will be used when the candle is a Doji. An upward and downward price trends formed by candlesticks are depicted in Figs. 1(f) and 1(g). Fig. 1(h) illustrates the top and the bottom of a body. The height of a body is equal to the difference between the top and the bottom of a body. The height of a body can also be defined as the absolute value of the difference between the opening and closing price.

Four types of gap (a downside shadow gap, an upside shadow gap, a downside body gap, and an upside body gap) are depicted in Figs. 1(i)–1(l). In the upside shadow gap, the first candle's high price is lower than the second candle's low price. In the downside shadow gap, the first candle's low price is higher than the second candle's high price. Two candlesticks bodies' relative positions are only considered in body gaps. In the upside body gap, the top of the first candle's body is below the bottom of the second candle's body. Whereas in the downside body gap, the bottom of the first candle's body is higher than the top of the second candle's body. Note that a gap, regardless of a shadow gap or a body gap, is not a candlestick but only a visual void.

In some occasions, a body can be represented by a single horizontal line, i.e. using a Doji candle. However, a single candle cannot show the relationship among prices of multiple trading days. Therefore, multiple candle lines along with variations in the length of shadows and the shape of bodies are commonly used to describe the price movement. Candlestick pattern is a specific combination of candlesticks with similar or different shapes. An example of a complete candlestick pattern is depicted in Fig. 1(m). A complete candlestick pattern is composed of two parts: trend candlesticks and significant candlesticks (or a significant candlestick when there is only one candlestick in the significant candlestick part). Furthermore, trend candlesticks part is optional, whereas the significant candlesticks part is compulsory. Specifically, significant candlesticks part must contain at least one candlestick. The trend candlesticks form a preceding sequence of a candlestick pattern when the pattern embodies the part

of a price trend. To maintain the formation, there are normally eight candlesticks in the trend candlesticks, and the price of the trend can be derived from these eight candlesticks' closing prices. For the purpose of illustration, only three candlesticks are used to form the part of trend candlesticks in 1(m). The significant candlesticks shown in Fig. 1(m) contain five candlesticks with different colors, relative positions, and body shapes. The pattern shown in Fig. 1(m) belongs to the category of “five candlesticks pattern” because there are five candlesticks in the significant candlesticks part. In this paper, the maximum number of significant candlesticks considered for composing a pattern is 13.

3.2. Definitions and functions for describing rules of candlestick patterns

The definitions and functions used for describing the rules of classifying candlestick patterns are listed as follows:

Definition 1 (Candlestick). A candlestick $c = (date, op, hp, lp, cp)$ is a tuple consists of a date and four intra-day prices of a stock. A candlestick c is a basic element in the identification of the candlestick patterns. The $date, op, hp, lp$ and cp , represent the date, opening price, high price, low price and closing price at the “date” position in a time series. For the purpose of simplification, we use c_i to denote the i^{th} candlestick $(date_i, op_i, hp_i, lp_i, cp_i)$.

Definition 2 (Candlestick Time Series). A candlestick time series $T = \{c_1, c_2, \dots, c_n\}$ is a sequence of candlesticks of a stock and consists of n candlesticks from day 1 to day n . For the purpose of simplification, $T[c_i, c_j]$ is used to denote a subsequence of T from c_i to c_j .

Definition 3 (Significant Candlesticks). $S = \{c_1, c_2, \dots, c_n\}$ is a subsequence of T and includes n candlesticks.

Definition 4 (Trend Candlesticks). $TC = \{c_1, c_2, \dots, c_n\}$ is a subsequence of T and composed of n candlesticks. In this paper, we set $n = 8$ for defining a Trend Candlesticks. TC is always followed by S .

Definition 5 (Sub-trend Candlesticks). $STC = \{c_1, c_2, \dots, c_n\}$ is a subsequence of TC and includes n consecutive candlesticks. In this paper, we set $n = 5$ for a defining a Sub-trend Candlesticks. For the purpose of simplification, we use $STC[c_i, c_{i+5}]$ to denote the subsequence of TC from c_i to c_{i+5} .

Definition 6 (Candlestick Pattern). $CanP = \{TC, S\}$ is a subsequence of T that comprises of two parts: trend candlesticks and significant candlesticks.

Function 1. $len(T)$ returns the number of basic candlesticks of an input sequence. If $T = \{c_1, c_2, \dots, c_i\}$, then

$$len(T) = i. \quad (1)$$

Function 2. $op(c_i)$ returns the opening price of a candlestick c_i .

$$op(c_i) = op_i. \quad (2)$$

Function 3. $hp(c_i)$ returns the high price of a candlestick c_i .

$$hp(c_i) = hp_i. \quad (3)$$

Function 4. $lp(c_i)$ returns the low price of a candlestick c_i .

$$lp(c_i) = lp_i. \quad (4)$$

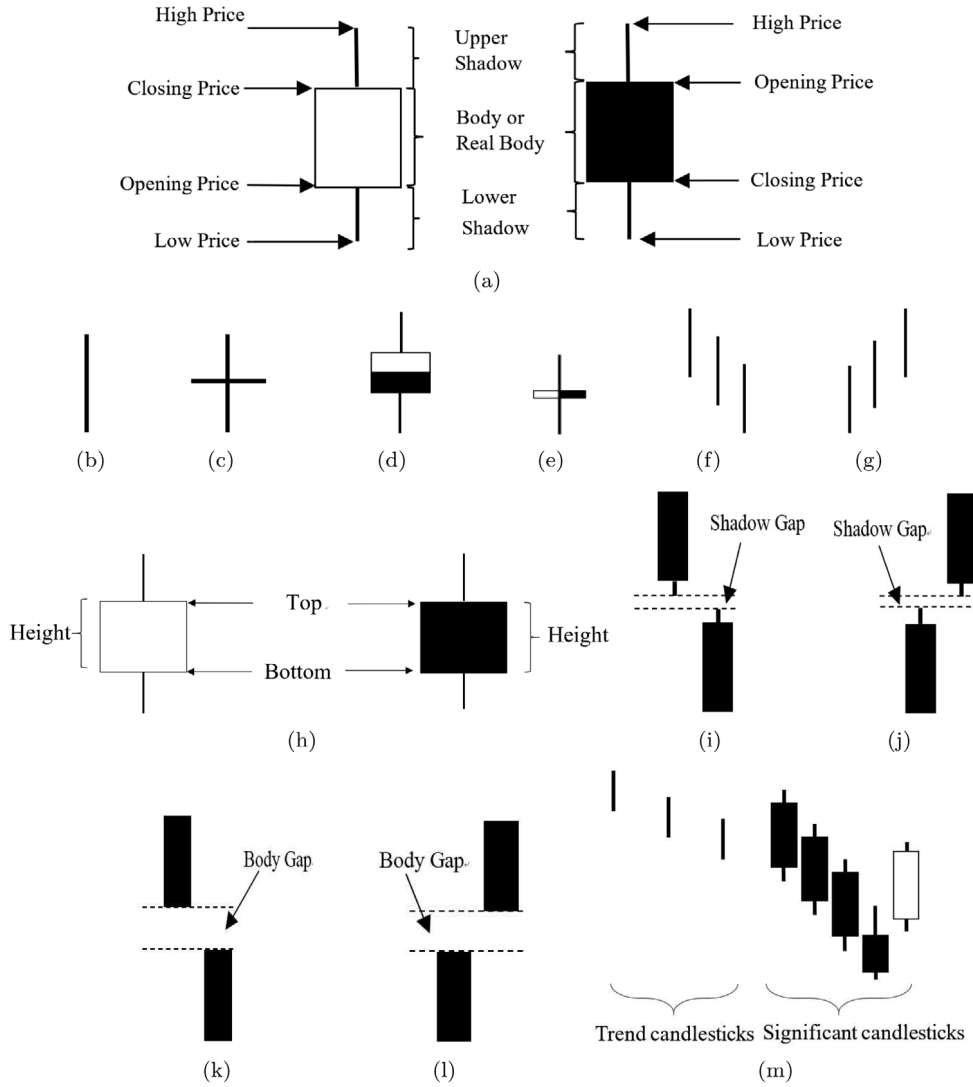


Fig. 1. (a) A white and a black candlesticks. (b) A candlestick in abbreviated form. (c) A Doji candlestick. (d) A white or black candlestick in horizontal direction. (e) A white or black candlestick in vertical direction. (f) A downward price trend. (g) An upward price trend. (h) The height, the top and the bottom from two types of candlesticks. (i) A downside shadow gap between two candlesticks. (j) An upside shadow gap between two candlesticks. (k) A downside body gap between two candlesticks. (l) An upside body gap between two candlesticks. (m) An example of a complete candlestick pattern.

Function 5. $cp(c_i)$ returns the closing price of a candlestick c_i .

$$cp(c_i) = cp_i. \quad (5)$$

Function 6. $hb(c_i)$ returns the height of the body of a candlestick c_i .

$$hb(c_i) = |cp(c_i) - op(c_i)|. \quad (6)$$

Function 7. $id(S)$ returns the sequence number of the first candlestick of S , where $S = \{c_i, c_{i+1}, \dots, c_{i+n}\}$.

$$id(S) = i. \quad (7)$$

Function 8. $tp_body(c_i)$ returns the top of the body of a candlestick c_i . The top is the maximum value between the opening price and the closing price.

$$tp_body(c_i) = \max(op(c_i), cp(c_i)) \quad (8)$$

Function 9. $bm_body(c_i)$ returns the bottom of the body of a candlestick c_i . The bottom is the minimum value between the opening

price and the closing price.

$$bm_body(c_i) = \min(op(c_i), cp(c_i)) \quad (9)$$

Function 10. $us(c_i)$ returns the height of the upper shadow of a candlestick c_i .

$$us(c_i) = hp(c_i) - tp_body(c_i). \quad (10)$$

Function 11. $ls(c_i)$ returns the height of the lower shadow of a candlestick c_i .

$$ls(c_i) = bm_body(c_i) - lp(c_i). \quad (11)$$

Function 12. $hs(c_i)$ returns the total height of the shadows of a candlestick c_i .

$$hs(c_i) = us(c_i) + ls(c_i). \quad (12)$$

Function 13. $ap(STC)$ returns the average closing price of STC .

$$ap(STC) = (cp_5 + cp_4 + cp_3 + cp_2 + cp_1)/5. \quad (13)$$

Function 14. $pt(TC)$ returns 1 when the price trend of TC is upward and return -1 when the price trend of TC is downward.

$$pt(TC) = \begin{cases} 1, & ap(STC[c_1, c_5]) < ap(STC[c_2, c_6]) \\ & < ap(STC[c_3, c_7]) < ap(STC[c_4, c_8]) \\ -1, & ap(STC[c_1, c_5]) > ap(STC[c_2, c_6]) \\ & > ap(STC[c_3, c_7]) > ap(STC[c_4, c_8]) \end{cases} \quad (14)$$

The short-term price trend is used to describe the trend of closing prices in the trend candlesticks where there are eight candlesticks. The short-term price trend could form either an increasing or a decreasing trend. We use 1 to denote that the trend of closing price is increasing or an upturn. We use -1 to denote that the trend of closing price is decreasing or a downturn. Although it is possible that the closing prices of all eight continuous candlesticks are the same, it is extremely unlikely and this situation has not been reported in literature [2]. Therefore, we do not consider such situation in this paper. Note that percentages used in the following functions can be adjusted based on the preference of the traders.

The following functions (15–27) are used for calculating the size of a candlestick body and the corresponding shadow. Variables x and y from function 15–27 can be any of the basic indicators of a candlestick (i.e. opening, closing, high and low price).

Function 15. $sli_greater(x, y)$ returns true when x is slightly greater than y in the range from 0.3% to 1%, where the interval is left-closed and right-open.

$$sli_greater(x, y) \equiv 0.3\% \leq [(x - y)/y] < 1\%. \quad (15)$$

Function 16. $mod_greater(x, y)$ returns true when x is moderately greater than y in the range from 1% to 2.5%, where the interval is left-closed and right-open.

$$mod_greater(x, y) \equiv 1\% \leq [(x - y)/y] < 2.5\%. \quad (16)$$

Function 17. $lar_greater(x, y)$ returns true when x is largely greater than y in the range from 2.5% to 5%, where the interval is left-closed and right-open.

$$lar_greater(x, y) \equiv 2.5\% \leq [(x - y)/y] < 5\%. \quad (17)$$

Function 18. $ext_greater(x, y)$ returns true when x is at least 5% greater than y .

$$ext_greater(x, y) \equiv (x - y)/y \geq 5\%. \quad (18)$$

Function 19. $sli_less(x, y)$ returns true when x is slightly less than y in the range from 0.3% to 1%, where the interval is left-closed and right-open.

$$lar_less(x, y) \equiv 0.3\% \leq [(y - x)/x] < 1\%. \quad (19)$$

Function 20. $mod_less(x, y)$ returns true when x is moderately less than y in the range from 1% to 2.5%, where the interval is left-closed and right-open.

$$lar_less(x, y) \equiv 1\% \leq [(y - x)/x] < 2.5\%. \quad (20)$$

Function 21. $lar_less(x, y)$ returns true when x is largely less than y in the range from 2.5% to 5%, where the interval is left-closed and right-open.

$$lar_less(x, y) \equiv 2.5\% \leq [(y - x)/x] < 5\%. \quad (21)$$

Function 22. $ext_less(x, y)$ returns true when x is at least 5% less than y .

$$ext_less(x, y) \equiv [(y - x)/x] \geq 5\%. \quad (22)$$

Function 23. $ext_near(x, y)$ returns true when the difference between x and y is less than or equal to 0.3%.

$$ext_near(x, y) \equiv [|x - y|/\max(x, y)] \leq 0.3\%. \quad (23)$$

Function 24. $mod_near(x, y)$ returns true when the difference between x and y is between 0.3% and 1%, where the interval is left-closed and right-open.

$$mod_near(x, y) \equiv 0.3\% \leq [|x - y|/\max(x, y)] < 1\%. \quad (24)$$

Function 25. $near(x, y)$ returns true when the difference between x and y is between 0 and 1%, where the interval is left-closed and right-open.

$$near(x, y) \equiv 0 \leq [|x - y|/\max(x, y)] < 1\%. \quad (25)$$

Function 26. $near_up(x, y)$ returns true when x is very close to y through the direction of the negative side of y and their difference is between 0.1% and 0.3%, where the interval is left-closed and right-open.

$$near_up(x, y) \equiv 0.1\% \leq [(y - x)/y] < 0.3\%. \quad (26)$$

Function 27. $near_down(x, y)$ returns true when x is very close to y through the direction of the positive side of y and their difference is between 0.1% and 0.3%, where the interval is left-closed and right-open.

$$near_down(x, y) \equiv 0.1\% \leq [(x - y)/x] < 0.3\%. \quad (27)$$

In Functions 26 and 27, the positive side of a price means the direction of approaching the price is from positive infinity to negative infinity. The negative side of a price means the direction of approaching the price is from negative infinity to positive infinity. For example, 3 approaches 3.1 from the negative side and 3.1 approaches to 3 from the positive side.

Function 28. $get_element(T, ele)$ returns the sequence of attribute ele of candlestick time series T . The parameter, ele , can be any basic attributes of a candle given in Fig. 1. For example:

$$get_element(T, hb) = \{|op(c_1) - cp(c_1)|, |op(c_2) - cp(c_2)|, \dots, |op(c_n) - cp(c_n)|\}. \quad (28)$$

where ele is the attribute hb .

Function 29. $avg(get_element(T, ele))$ returns the average of attribute ele of candlestick time series T . The parameter, ele can be any basic attribute of a candle given in Fig. 1. For example:

$$avg(get_element(T, hb)) = (|op(c_1) - cp(c_1)| + |op(c_2) - cp(c_2)| + \dots + |op(c_n) - cp(c_n)|)/n. \quad (29)$$

where ele is the attribute hb .

The following functions are used to describe the length of the body or the shadows of a candlestick.

Function 30. $Doji(c_i)$ returns true when opening price of a candlestick c_i is approximately equal to its closing price.

$$Doji(c_i) \equiv ext_near(op_i, cp_i). \quad (30)$$

Function 31. $small_body(c_i)$ returns true when the size of candlestick c_i is small.

$$small_body(c_i) \equiv sli_less(bm_body(c_i), tp_body(c_i)). \quad (31)$$

Function 32. *normal_body(c_i)* returns true when the size of candlestick c_i is normal.

$$\text{normal_body}(c_i) \equiv \text{mod_less}(\text{bm_body}(c_i), \text{tp_body}(c_i)). \quad (32)$$

Function 33. *long_body(c_i)* returns true when the size of candlestick c_i is long.

$$\text{long_body}(c_i) \equiv \text{lar_less}(\text{bm_body}(c_i), \text{tp_body}(c_i)). \quad (33)$$

Function 34. *el_body(c_i)* returns true when the size of candlestick c_i is exceedingly long.

$$\text{el_body}(c_i) \equiv \text{ext_less}(\text{bm_body}(c_i), \text{tp_body}(c_i)). \quad (34)$$

Function 35. *no_us(c_i)* returns true when the size of the upper shadow of the candlestick c_i is very small.

$$\text{no_us}(c_i) \equiv \text{ext_near}(\text{hp}(c_i), \text{tp_body}(c_i)). \quad (35)$$

Function 36. *small_us(c_i)* returns true when the size of the upper shadow of the candlestick c_i is small.

$$\text{small_us}(c_i) \equiv \text{sli_greater}(\text{hp}(c_i), \text{tp_body}(c_i)). \quad (36)$$

Function 37. *normal_us(c_i)* returns true when the size of the upper shadow of the candlestick c_i is normal.

$$\text{normal_us}(c_i) \equiv \text{mod_greater}(\text{hp}(c_i), \text{tp_body}(c_i)). \quad (37)$$

Function 38. *long_us(c_i)* returns true when the size of the upper shadow of the candlestick c_i is long.

$$\text{long_us}(c_i) \equiv \text{lar_greater}(\text{hp}(c_i), \text{tp_body}(c_i)). \quad (38)$$

Function 39. *el_us(c_i)* returns true when the size of the upper shadow of the candlestick c_i is exceedingly long.

$$\text{el_us}(c_i) \equiv \text{ext_greater}(\text{hp}(c_i), \text{tp_body}(c_i)). \quad (39)$$

Function 40. *no_ls(c_i)* returns true when the size of the lower shadow of the candlestick c_i is very small.

$$\text{no_ls}(c_i) \equiv \text{ext_near}(\text{lp}(c_i), \text{bm_body}(c_i)). \quad (40)$$

Function 41. *small_ls(c_i)* returns true when the size of the lower shadow of the candlestick c_i is small.

$$\text{small_ls}(c_i) \equiv \text{sli_less}(\text{lp}(c_i), \text{bm_body}(c_i)). \quad (41)$$

Function 42. *normal_ls(c_i)* returns true when the size of the lower shadow of the candlestick c_i is normal.

$$\text{normal_ls}(c_i) \equiv \text{mod_less}(\text{lp}(c_i), \text{bm_body}(c_i)). \quad (42)$$

Function 43. *long_ls(c_i)* returns true when the size of the lower shadow of the candlestick c_i is long.

$$\text{long_ls}(c_i) \equiv \text{lar_less}(\text{lp}(c_i), \text{bm_body}(c_i)). \quad (43)$$

Function 44. *el_ls(c_i)* returns true when the size of the lower shadow of the candlestick c_i is exceedingly long.

$$\text{el_ls}(c_i) \equiv \text{ext_less}(\text{lp}(c_i), \text{bm_body}(c_i)). \quad (44)$$

Function 45. *black_body(c_i)* returns true when the opening price of a candlestick c_i is greater than its closing price.

$$\text{black_body}(c_i) \equiv \text{op}(c_i) > \text{cp}(c_i). \quad (45)$$

Function 46. *white_body(c_i)* returns true when the opening price of a candlestick c_i is less than its closing price.

$$\text{white_body}(c_i) \equiv \text{op}(c_i) < \text{cp}(c_i). \quad (46)$$

Function 47. *small_black_body(c_i)* returns true when the candlestick c_i is a black candle with a small body.

$$\text{small_black_body}(c_i) \equiv \text{small_body}(c_i) \wedge \text{black_body}(c_i). \quad (47)$$

Function 48. *small_white_body(c_i)* returns true when the candlestick c_i is a white candle with a small body.

$$\text{small_white_body}(c_i) \equiv \text{small_body}(c_i) \wedge \text{white_body}(c_i). \quad (48)$$

Function 49. *normal_black_body(c_i)* returns true when the candlestick is a black candle with a normal-sized body.

$$\text{normal_black_body}(c_i) \equiv \text{normal_body}(c_i) \wedge \text{black_body}(c_i). \quad (49)$$

Function 50. *normal_white_body(c_i)* returns true when the candlestick is a white candle with a normal-sized body.

$$\text{normal_white_body}(c_i) \equiv \text{normal_body}(c_i) \wedge \text{white_body}(c_i). \quad (50)$$

Function 51. *long_black_body(c_i)* returns true when the candlestick is a black candle with a long body.

$$\text{long_black_body}(c_i) \equiv \text{long_body}(c_i) \wedge \text{black_body}(c_i). \quad (51)$$

Function 52. *long_white_body(c_i)* returns true when the candlestick is a white candle with a long body.

$$\text{long_white_body}(c_i) \equiv \text{long_body}(c_i) \wedge \text{white_body}(c_i). \quad (52)$$

Function 53. *el_black_body(c_i)* returns true when the candlestick is a black candle with an extremely long body.

$$\text{el_black_body}(c_i) \equiv \text{el_body}(c_i) \wedge \text{black_body}(c_i). \quad (53)$$

Function 54. *el_white_body(c_i)* returns true when the candlestick is a white candle with an extremely long body.

$$\text{el_white_body}(c_i) \equiv \text{el_body}(c_i) \wedge \text{white_body}(c_i). \quad (54)$$

Function 55. *down_shadow_gap(c_i, c_j)* returns true when there is a downside shadow gap between the candlesticks c_i and c_j , where $j > i$. The downside shadow gap means the low price of the c_i is greater than the high price of the c_j .

$$\text{down_shadow_gap}(c_i, c_j) \equiv \text{lp}(c_i) > \text{hp}(c_j). \quad (55)$$

Function 56. *up_shadow_gap(c_i, c_j)* returns true when there is an upside shadow gap between the candlesticks c_i and c_j , where $j > i$. The upside shadow gap means the high price of the c_i is less than the low price of the c_j .

$$\text{up_shadow_gap}(c_i, c_j) \equiv \text{hp}(c_i) < \text{lp}(c_j). \quad (56)$$

Function 57. *down_body_gap(c_i, c_j)* returns true when there is a downside body gap between the candlesticks c_i and c_j , where $j > i$. The downside body gap means the bottom of the body of c_i is greater than the top of the body of the c_j .

$$\text{down_body_gap}(c_i, c_j) \equiv \text{bm_body}(c_i) > \text{tp_body}(c_j). \quad (57)$$

Function 58. *up_body_gap(c_i, c_j)* returns true when there is an upside body gap between the candlesticks c_i and c_j , where $j > i$.



Fig. 2. (a) “Marubozu, Black” candle. (b) “Marubozu, White” candle.

The upside body gap means the top of the body of c_i is less than the bottom of the body of c_j .

$$up_body_gap(c_i, c_j) \equiv tp_body(c_i) < bm_body(c_j). \quad (58)$$

3.3. Categories of candlestick chart patterns

In Table 1, we categorize the 103 known candlestick patterns from [2] into five main groups. These categories are formed based on the number of candlesticks included in the significant candlesticks of the patterns. For example, One-Line Candle group contains patterns with a single candlestick in the significant candlestick part whereas Five-Line (or more) Candles group contains patterns with no less than five candles. Each group is then subdivided into multiple sub-groups. For example, One-line Candle group can be further subdivided into four sub-groups (No Shadow, Upper Shadow, Lower Shadow, and Shadows). The fourth column in Table 1 shows the ID of the patterns that belong to that sub-group as well as the total number of patterns in that sub-group (written in parentheses). The rightmost column indicates the ID of the pattern selected for our experiments that will be introduced in Section 5.

4. Formal specifications of candlestick chart patterns

In this section, we describe the formal specifications of each candlestick pattern in the form of first-order-logic. Note that the candlesticks used in these specifications are significant candlesticks unless stated otherwise.

4.1. One-line candles

One-line Candle patterns are divided into 4 subgroups: No Shadow, Upper Shadow, Lower Shadow and Shadows.

4.1.1. No shadow

In this group, the candle does not have a shadow. Each candlestick's opening, high, low and closing prices can be used to identify the patterns.

Pattern P1: “Marubozu, Black”

Description: The pattern shown in Fig. 2(a) contains a tall black candle without any shadow in the significant candlestick.

Rules:

$$(len(S) = 1) \wedge no_us(s_1) \wedge long_black_body(s_1) \wedge no_ls(s_1). \quad (59)$$

Pattern P2: “Marubozu, White”

Description: The pattern shown in Fig. 2(b) contains a tall white candle without any shadow in the significant candlestick.

Rule:

$$(len(S) = 1) \wedge no_us(s_1) \wedge long_white_body(s_1) \wedge no_ls(s_1). \quad (60)$$

4.1.2. Upper shadow

In this subgroup, there is no lower shadow in the candlestick of each candlestick pattern. To identify the patterns in this group, each candlestick's opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P3: “Belt Hold, Bullish”

Description: The pattern shown in Fig. 3(a) contains a price down-trend followed by a tall white candle in the significant candlestick, its price opens at the low price and closes near the high price.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = -1) \wedge tall_white_body(s_1) \wedge no_ls(s_1) \wedge mod_near(cp(s_1), hp(s_1)). \quad (61)$$

Pattern P4: “Marubozu, Closing Black”

Description: The pattern shown in Fig. 3(b) contains a tall black candle in the significant candlestick, with an upper shadow. However, it does not contain a lower shadow.

Rule:

$$(len(S) = 1) \wedge tall_black_body(s_1) \wedge !no_us(s_1) \wedge no_ls(s_1). \quad (62)$$

Pattern P5: “Marubozu, Opening White”

Description: The pattern shown in Fig. 3(c) has a tall white candle with an upper shadow. The pattern does not contain a lower shadow in the significant candlestick.

Rule:

$$(len(S) = 1) \wedge tall_white_body(s_1) \wedge !no_us(s_1) \wedge no_ls(s_1). \quad (63)$$

Pattern P6: “Shooting Star, One-Candle”

Description: The pattern shown in Fig. 3(d) begins with an up-trend. The candle in the significant candlestick has a tall upper shadow whose height is at least twice the body's height. The color of the body in the significant candlestick (for this pattern, there is only one candlestick in the significant candlestick) can be either white or black. The significant candlestick has a small body and the body should be at or near the candle's low. Specifically, the significant candlestick has either very small or no lower shadow.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = 1) \wedge tall_us(s_1) \wedge (us(s_1) > (2 * hb(s_1))) \wedge small_body(s_1) \wedge no_ls(s_1). \quad (64)$$

Pattern P7: “Doji, Gravestone”

Description: The pattern shown in Fig. 3(e) contains a Doji at the low price. The pattern also has a tall upper shadow in the significant candlestick.

Rule:

$$(len(S) = 1) \wedge Doji(s_1) \wedge no_ls(s_1) \wedge tall_us(s_1). \quad (65)$$

4.1.3. Lower shadow

In this subgroup, there is no upper shadow in the candlestick of each candlestick pattern. To identify patterns in this group, each candlestick's opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P8: “Belt Hold, Bearish”

Description: The pattern shown in Fig. 4(a) begins with an up-trend. The price of the candle in the significant candlestick opens at the high price and closes near the low price, creating a tall black candle.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = 1) \wedge no_us(s_1) \wedge small_ls(s_1) \wedge tall_black_body(s_1). \quad (66)$$

Table 1
Category of Candlestick patterns.

Group Name	Subgroup name	ID of patterns (The number of patterns in the subgroup)	Selected patterns for experiments
103 candlestick patterns	One-Line Candles	No Shadow	P1,-P2 (2)
		Upper Shadow	P3 to P7 (5)
		Lower Shadow	P8 to P14 (7)
		Shadows	P15 to P29 (15)
	Two-Line Candles	Gap A	P30 to P35 (6)
		Sink A	P36,-P37 (2)
		Sink B	P38 to P42 (5)
		Gap B	P43 to P50 (8)
		Engulfing A	P51 to P55 (5)
		Engulfing B	P56 to P59 (4)
		Tweezes	P60,-P61 (2)
	Three-Line Candles	One Star	P62 to P68 (7)
		Three Doji Star	P69,-P70 (2)
		With Gap	P71 to P78 (8)
		Monotonic	P79 to P89 (11)
		Fluctuant	P90 (1)
	Four-Line Candles	Four-Line Candles	P91 to P93 (3)
	Five-Line Candles	New Price Lines	P94 to P97 (4)
		Breakaway, Bearish	P98,-P99 (2)
		Three Method	P100 to P102 (3)
		Ladder Bottom	P103 (1)

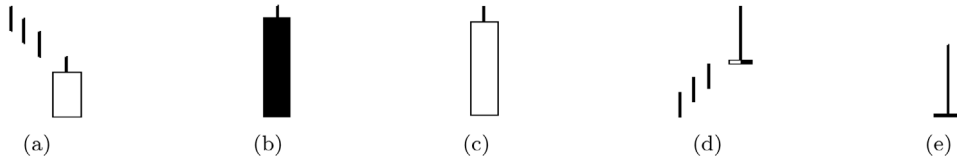


Fig. 3. (a) “Belt Hold, Bullish” candle. (b) “Marubozu, Closing Black” candle. (c) “Marubozu, Opening White” candle. (d) “Shooting Star, One-Candle” candle. (e) “Doji, Gravestone” candle.

Pattern P9: “Doji, Dragonfly”

Description: The pattern shown in Fig. 4(b) contains a Doji near the high price. The pattern also has a long lower shadow in the significant candlestick.

Rule:

$$(len(S) = 1) \wedge Doji(s_1) \wedge small_us(s_1) \wedge long_ls(s_1). \quad (67)$$

Pattern P10: “Hammer”

Description: The pattern shown in Fig. 4(c) begins with a down-trend. The significant candlestick has a small body. The candle also has a lower shadow and its height is between approximately two to three times the height of the body. The pattern also has small or no upper shadow. Body color can be either white or black.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = -1) \wedge small_body(s_1) \wedge !no_ls(s_1) \wedge (2 * hb(s_1) < ls(s_1) < 3 * hb(s_1)) \wedge (small_us(s_1) \vee no_us(s_1)). \quad (68)$$

Pattern P11: “Hanging Man”

Description: The pattern shown in Fig. 4(d) begins with an up-trend. In the significant candlestick, price opens at or near the high price. The candlestick forms a long lower shadow. The significant candlestick has a small body and it can be either white or black.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = 1) \wedge no_us(s_1) \wedge long_ls(s_1) \wedge small_body(s_1). \quad (69)$$

Pattern P12: “Marubozu, Opening Black”

Description: The pattern shown in Fig. 4(e) has a tall black candle in the significant candlestick, with a lower shadow. However, the pattern does not have an upper shadow.

Rule:

$$(len(S) = 1) \wedge tall_black_body(s_1) \wedge !no_ls(s_1) \wedge no_us(s_1). \quad (70)$$

Pattern P13: “Marubozu, Closing White”

Description: The pattern shown in Fig. 4(f) has a tall white body in the significant candlestick, with a closing price at the high price and a lower shadow.

Rule:

$$(len(S) = 1) \wedge tall_white_body(s_1) \wedge no_us(s_1) \wedge !no_ls(s_1). \quad (71)$$

Pattern P14: “Takuri Line”

Description: The pattern shown in Fig. 4(g) begins with a down-trend and the significant candlestick has a small body. The color of the body can be either white or black. The pattern may not have an upper shadow. If it has an upper shadow, it should be a very small one. The lower shadow of the pattern is at least three times the height of the body.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = -1) \wedge small_body(s_1) \wedge no_us(s_1) \wedge (ls(s_1) > (3 * hb(s_1))). \quad (72)$$

4.1.4. Shadows

In this subgroup, there are upper and lower shadows in the candlestick of each candlestick pattern. To identify patterns in this group, each candlestick's opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P15: "Candle, Black"

Description: The pattern shown in Fig. 5(a) has a normal-size black candle in the significant candlestick, with shadows that do not exceed the height of the body.

Rule:

$$(len(S) = 1) \wedge normal_black_body(s_1) \wedge !no_us(s_1) \wedge !no_ls(s_1) \wedge (us(s_1) < hb(s_1)) \wedge (ls(s_1) < hb(s_1)). \quad (73)$$

Pattern P16: "Candle, Short Black"

Description: The pattern shown in Fig. 5(b) is a small-size black candle in the significant candlestick, with shadows that do not exceed the height of the body.

Rule:

$$(len(S) = 1) \wedge small_black_body(s_1) \wedge !no_us(s_1) \wedge !no_ls(s_1) \wedge (us(s_1) < hb(s_1)) \wedge (ls(s_1) < hb(s_1)). \quad (74)$$

Pattern P17: "Candle, White"

Description: The pattern shown in Fig. 5(c) has a normal-size white candle in the significant candlestick, with shadows that do not exceed the height of body.

Rule:

$$(len(S) = 1) \wedge normal_white_body(s_1) \wedge !no_us(s_1) \wedge !no_ls(s_1) \wedge (us(s_1) < hb(s_1)) \wedge (ls(s_1) < hb(s_1)). \quad (75)$$

Pattern P18: "Candle, Short White"

Description: The pattern shown in Fig. 5(d) has a small-size white candle in the significant candlestick, with shadows that do not exceed the height of the body.

Rule:

$$(len(S) = 1) \wedge small_white_body(s_1) \wedge !no_us(s_1) \wedge !no_ls(s_1) \wedge (us(s_1) < hb(s_1)) \wedge (ls(s_1) < hb(s_1)). \quad (76)$$

Pattern P19: "Doji, Gapping Down"

Description: To identify this pattern, the closing price of the last day in the trend candlesticks is needed. The pattern shown in Fig. 5(e) begins with a downtrend. The candlestick from the significant candlesticks has a Doji and its upper shadow remains below the lower shadow of the eighth candles in the trend candlesticks, leaving a downside shadow gap on the chart.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = -1) \wedge ((\exists s_1 \in S \wedge \exists tc_8 \in TC) \Rightarrow Doji(s_1) \wedge down_shadow_gap(tc_8, s_1)). \quad (77)$$

Pattern P20: "Doji, Gapping Up"

Description: The pattern shown in Fig. 6(a) begins with an uptrend. The candlestick in the significant candlestick has a Doji and its lower shadow remains above the upper shadow of the eighth candlestick in the trend candlesticks, leaving an upside shadow gap on the chart.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = 1) \wedge \exists s_1 \in S \wedge \exists tc_8 \in TC \Rightarrow Doji(s_1) \wedge up_shadow_gap(tc_8, s_1). \quad (78)$$

Pattern P21: "Doji, Long-Legged"

Description: The pattern shown in Fig. 6(b) has a Doji, long upper and lower shadows in the significant candlestick.

Rule:

$$(len(S) = 1) \wedge Doji(s_1) \wedge long_us(cp(s_1)) \wedge long_ls(cp(s_1)). \quad (79)$$

Pattern P22: "Doji, Northern"

Description: The pattern shown in Fig. 6(c) begins with an uptrend and has a Doji in the significant candlestick.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = 1) \wedge Doji(s_1). \quad (80)$$

Pattern P23: "Doji, Southern"

Description: The pattern shown in Fig. 6(d) begins with a downtrend and has a Doji in the significant candlestick.

Rule:

$$(len(S) = 1) \wedge (pt(TC) = -1) \wedge Doji(s_1). \quad (81)$$

Pattern P24: "High Wave"

Description: The pattern shown in Fig. 6(e) has tall upper and lower shadows with a small body in the significant candlestick. Body can be either black or white. The main difference between this pattern and "Spinning Top, Black" (P28) or "Spinning Top, White" (P29) is the length of the shadows. "Spinning Tops" have smaller shadows and "High Wave" candles have unusually long shadows.

Rule:

$$(len(S) = 1) \wedge el_long_us(s_1) \wedge el_long_ls(s_1) \wedge small_body(s_1). \quad (82)$$

Pattern P25: "Long Black Day"

Description: The pattern shown in Fig. 7(a) has a tall black candle in the significant candlestick. The height of the body is at least three times taller than the average height of the body of preceding candlesticks from one or two prior weeks. The heights of the upper and lower shadows of the significant candlestick are shorter than the height of the body.

Rule:

$$(len(S) = 1) \wedge long_black_body(s_1) \wedge (\exists id \in id(S) \Rightarrow (hb(s_1) > (3 * avg(get_element(T[t_{id-7}, t_{id-1}], hb)))) \vee (hb(s_1) > (3 * avg(get_element(T[t_{id-14}, t_{id-1}], hb)))) \wedge (us(s_1) < hb(s_1)) \wedge (ls(s_1) < hb(s_1))). \quad (83)$$

Pattern P26: "Long White Day"

Description: The pattern shown in Fig. 7(b) is the same as the previous pattern "Long Black Day" except that it has a tall white candle in the significant candlestick.

Rule:

$$(len(S) = 1) \wedge long_white_body(s_1) \wedge (\exists id \in id(S) \Rightarrow (hb(s_1) > (3 * avg(get_element(T[t_{id-7}, t_{id-1}], hb)))) \vee (hb(s_1) > (3 * avg(get_element(T[t_{id-14}, t_{id-1}], hb)))) \wedge (us(s_1) < hb(s_1)) \wedge (ls(s_1) < hb(s_1))). \quad (84)$$

Pattern P27: "Rickshaw Man"

Description: The pattern shown in Fig. 7(c) has a Doji in the significant candlestick which body is near the middle of the candle. The significant candlestick has exceptionally long shadows.

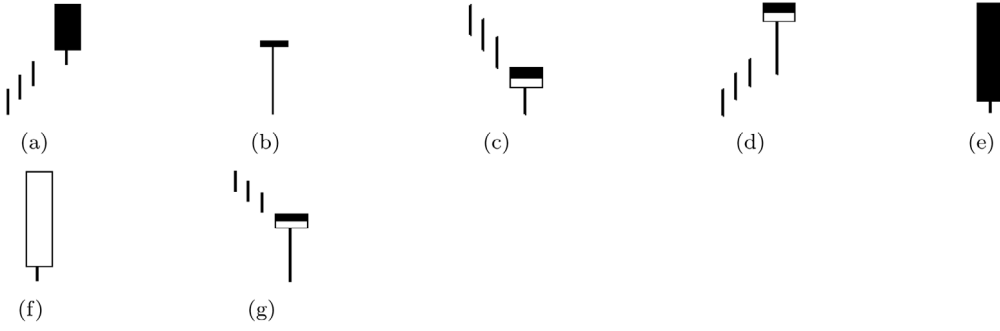


Fig. 4. (a) “Belt Hold, Bearish” candle. (b) “Doji, Dragonfly White” candle. (c) “Hammer” candle. (d) “Hanging Man” candle. (e) “Marubozu, Opening Black” candle. (f) “Marubozu, Closing White” candle. (g) “Takuri Line” candle.

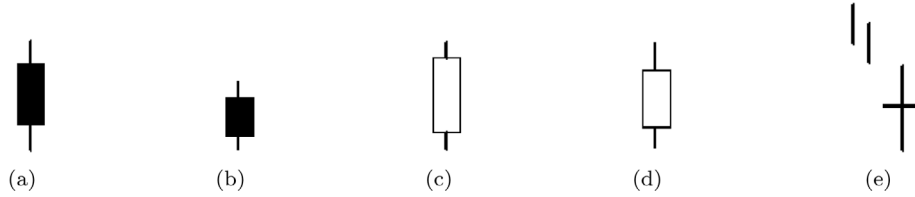


Fig. 5. (a) “Candle, Black” candle. (b) “Candle, Short Black” candle. (c) “Candle, White” candle. (d) “Candle, Short White” candle. (e) “Doji, Gapping Down” candle.

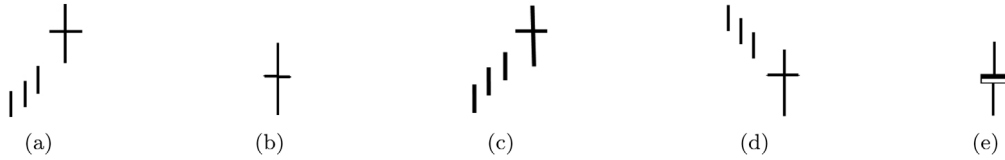


Fig. 6. (a) “Doji, Gapping Up” candle. (b) “Doji, Long-Legged” candle. (c) “Doji, Northern” candle. (d) “Doji, Southern” candle. (e) “High Wave” candle.

Rule:

$$(len(S) = 1) \wedge Doji(s_1) \wedge near(0.5 * (op(s_1) + cp(s_1)), 0.5 * (hp(s_1) + lp(s_1))) \wedge el_long_ls(s_1) \wedge el_long_us(s_1). \quad (85)$$

Pattern P28: “Spinning Top, Black”

Description: The pattern shown in Fig. 7(d) has a small black body and the heights of shadows of the significant candlestick are taller than the body.

Rule:

$$(len(S) = 1) \wedge small_black_body(s_1) \wedge (us(s_1) > hb(s_1)) \wedge (ls(s_1) > hb(s_1)) \wedge !no_ls(s_1) \wedge !no_us(s_1). \quad (86)$$

Pattern P29: “Spinning Top, White”

Description: The pattern shown in Fig. 7(e) has a small white body and the heights of shadows of the significant candlestick are taller than the body.

Rule:

$$(len(S) = 1) \wedge small_white_body(s_1) \wedge (us(s_1) > hb(s_1)) \wedge (ls(s_1) > hb(s_1)) \wedge !no_ls(s_1) \wedge !no_us(s_1). \quad (87)$$

4.2. Two-line candles

Two-Lines Candles are classified into 7 sub-groups by considering the trend candlesticks and relative positions of two significant candlesticks.

4.2.1. Gap upward

In this category, there is an upside body gap or shadow gap between the first day and the second day in each candlestick pattern. To identify the patterns in this group, each two consecutive

candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P30: “Doji Star, Bearish”

Description: The pattern shown in Fig. 8(a) begins with an uptrend. The first day in the significant candlesticks is a long white candle following with a uptrend body gap before the second day. The gap is formed when the bottom of second day's body is higher than the top of the first day's body. The second day in the significant candlesticks is a Doji without excessively long shadows. The sum of the shadows from the Doji candle is shorter than the body height of the prior day.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow long_white_body(s_1) \wedge up_body_gap(s_1, s_2) \wedge Doji(s_2) \wedge !el_us(s_2) \wedge !el_ls(s_2) \wedge (hs(s_2) < (hb(s_1)))). \quad (88)$$

Pattern P31: “Kicking, Bullish”

Description: This first day in the significant candlesticks shown in Fig. 8(b) is a “black Marubozu” candle which has a tall black candle with no shadows. There is a gap between the first and second significant candlesticks. The bottom of the second day's body is higher than the height of the first day's body. The second significant day is a “White Marubozu”.

Rule:

$$(len(S) = 2) \wedge (\exists s_1, s_2 \in S \Rightarrow long_black_body(s_1) \wedge no_us(s_1) \wedge no_ls(s_1) \wedge up_body_gap(s_1, s_2) \wedge long_white_body(s_2) \wedge no_us(s_2) \wedge no_ls(s_2)). \quad (89)$$

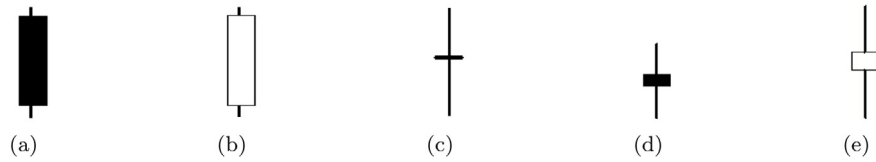


Fig. 7. (a) “Long Black Day” candle. (b) “Long White Day” candle. (c) “Rickshaw Man” candle. (d) “Spinning Top, Black” candle. (e) “Spinning Top, White” candle.

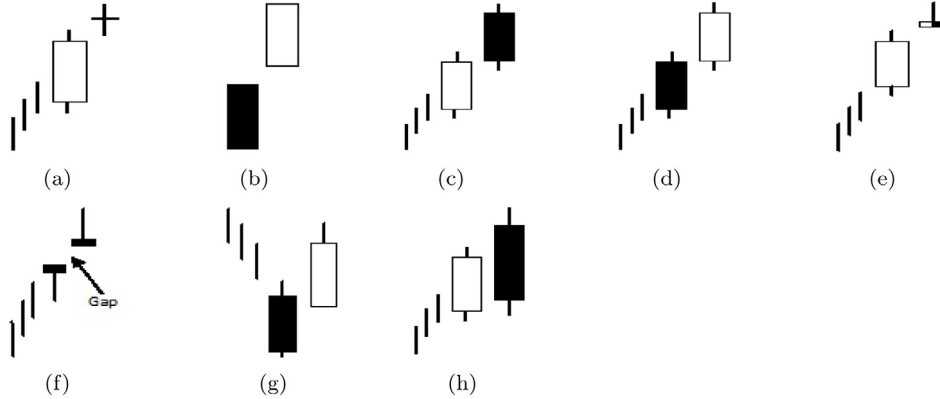


Fig. 8. (a) “Doji Star, Bearish” candle. (b) “Kicking, Bullish” candle. (c) “Meeting Lines, Bearish” candle. (d) “Separating Lines, Bullish” candle. (e) “Shooting Star, Two-Candle” candle. (f) “Window, Rising” candle. (g) “Above the Stomach” candle. (h) “Dark Cloud Cove” candle.

Pattern P32: “Meeting Lines, Bearish”

Description: The pattern shown in Fig. 8(c) begins with an up-trend. The first day in the significant candlesticks is a tall white candle and the second day is a tall black candle that closes at or near the prior day's closing price.

Rule:

$$(\text{len}(S) = 2) \wedge (\text{pt}(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow \text{long_white_body}(s_1) \wedge \text{long_black_body}(s_2) \wedge \text{near}(cp(s_1), cp(s_2))) \quad (90)$$

Pattern P33: “Separating Lines, Bullish”

Description: The pattern shown in Fig. 8(d) begins with an up-trend. The first day in the significant candlesticks is a tall black candle and the second day is a tall white candle that opens at or near the prior day's opening price.

Rule:

$$(\text{len}(S) = 2) \wedge (\text{pt}(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow \text{long_black_body}(s_1) \wedge \text{long_white_body}(s_2) \wedge \text{near}(op(s_1), op(s_2))) \quad (91)$$

Pattern P34: “Shooting Star, Two-Candle”

Description: The pattern shown in Fig. 8(e) begins with an up-trend. The first day in the significant candlesticks is a white candle. The second day has an upper shadow which is at least three times the height of the first day's body. The second day's small body is near the bottom of the second day's candle and the candle either has no lower shadow or has a very small one. The second day's body can be either black or white and it is above the prior day's body leaving an upside body gap.

Rule:

$$(\text{len}(S) = 2) \wedge (\text{pt}(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow \text{white_body}(s_1) \wedge \text{!no_us}(s_1) \wedge (\text{us}(s_2) > (3 * \text{hb}(s_1))) \wedge \text{small_body}(s_2) \wedge \text{no_ls}(s_2) \wedge \text{up_body_gap}(s_1, s_2)) \quad (92)$$

Pattern P35: “Window, Rising”

Description: The pattern shown in Fig. 8(f) begins with an up-trend. In the significant candlesticks, the low price of the second day is above the high price of the first day, which leaves an upside shadow gap between two days.

Rule:

$$(\text{len}(S) = 2) \wedge (\text{pt}(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow \text{upside_shadow_gap}(s_1, s_2)) \quad (93)$$

4.2.2. Sink upward

In this subgroup, the bottom of the body of the second day is within the body of the first day and the top of the second day's body is higher than the top of the body of the first day. To identify patterns in this group, each two consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P36: “Above the Stomach”

Description: The pattern shown in Fig. 8(g) begins with a down-trend. The first day in the significant candlesticks is a black candle. The second day is a white candle opening and closing at or being above the midpoint of the previous day's body.

Rule:

$$(\text{len}(S) = 2) \wedge (\text{pt}(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow \text{black_body}(s_1) \wedge \text{white_body}(s_2) \wedge (\text{op}(s_2) \geq (0.5 * (\text{cp}(s_1) + \text{op}(s_1)))) \wedge (\text{cp}(s_2) \geq (0.5 * (\text{cp}(s_1) + \text{op}(s_1)))))) \quad (94)$$

Pattern P37: “Dark Cloud Cover”

Description: The pattern shown in Fig. 8(h) begins with an up-trend. The first day in the significant candlesticks is a tall white candle. The second day is a black candle whose opening price is above the previous day high price and whose closing price is below the midpoint of the previous day's body.

Rule:

$$(\text{len}(S) = 2) \wedge (\text{pt}(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow \text{long_white_body}(s_1) \wedge \text{black_body}(s_2) \wedge (\text{op}(s_2) > \text{hp}(s_1)) \wedge (\text{cp}(s_2) \geq (0.5 * (\text{cp}(s_1) + \text{op}(s_1)))))) \quad (95)$$

4.2.3. Sink downward

In this sub-group, the top of the body of the second day is within the body of the first day. The bottom of the body of the second day is lower than or equal to the bottom of the body of the first day. To identify patterns in this group, each two consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P38: "Below the Stomach"

Description: The pattern shown in Fig. 9(a) begins with an up-trend. The first day in the significant candlesticks is a tall white candle and the second day can be either a white or black candle. However, the second day must open below the middle of the first day's body and close at or below the first day's midpoint.

Rule:

$$\begin{aligned} & (len(S) = 2) \wedge (pt(TC) = 1) \wedge \\ & (\exists s_1, s_2 \in S \Rightarrow long_white_body(s_1) \wedge \\ & ((0.5 * (cp(s_1) + op(s_1))) \geq op(s_2)) \wedge \\ & (cp(s_2) \leq (0.5 * (cp(s_1) + op(s_1)))). \end{aligned} \quad (96)$$

Pattern P39: "Matching Low"

Description: The pattern shown in Fig. 9(b) begins with a down-trend. The first day in the significant candlesticks is a black candle with a long body and the second day is a black candle with a close that matches the prior close.

Rule:

$$\begin{aligned} & (len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow \\ & long_black_body(s_1) \wedge black_body(s_2) \wedge ext_near(cp(s_1), cp(s_2))). \end{aligned} \quad (97)$$

Pattern P40: "Piercing Pattern"

Description: The pattern shown in Fig. 9(c) begins with a down-trend. The first day in the significant candlesticks is a black candle and the second day is a white candle that opens below the prior candle's low price and closes between the midpoint and the opening price of the first day's black body.

Rule:

$$\begin{aligned} & (len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow \\ & black_body(s_1) \wedge white_body(s_2) \wedge \\ & (op(s_2) < lp(s_1)) \wedge \\ & ((0.5 * (cp(s_1) + op(s_1))) < cp(s_2) < op(s_1))). \end{aligned} \quad (98)$$

Pattern P41: "Two Black Gapping Candles"

Description: To identify this pattern, the closing price of the last day in the trend candlesticks is needed. The pattern shown in Fig. 9(d) begins with a downtrend. The first day in the significant candlesticks is a black candle whose high price is lower than the prior day's low price (i.e. the last day's low price in the trend candlesticks). In the significant, the second day is a black candle and it has a lower high price compared to the first day's high price.

Rule:

$$\begin{aligned} & (len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists tc_8 \in TC \wedge \\ & \exists s_1, s_2 \in S \Rightarrow \\ & black_body(s_1) \wedge down_shadow_gap(tc_8, s_1) \wedge \\ & black_body(s_2) \wedge (hp(s_1) > hp(s_2))). \end{aligned} \quad (99)$$

Pattern P42: "Thrusting"

Description: The pattern shown in Fig. 9(e) begins with a down-trend. The first day in the significant candlesticks is a black candle and the second day is a white candle that opens below the prior low and closes near but below the midpoint of the prior body.

Rule:

$$\begin{aligned} & (len(S) = 2) \wedge (pt(TC) = -1) \\ & \wedge (\exists s_1, s_2 \in S \Rightarrow black_body(s_1) \wedge \\ & white_body(s_2) \wedge (op(s_2) < lp(s_1)) \wedge \\ & near_up(cp(s_2), 0.5 * (op(s_2) + cp(s_2)))). \end{aligned} \quad (100)$$

4.2.4. Gap downward

In all the patterns from this sub-group, there is a downside body gap or shadow gap between the first day and the second day in each candlestick pattern. To identify these patterns, each two consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P43: "Doji Star, Bullish"

Description: The pattern shown in Fig. 10(a) begins with a down-trend. The first day in the significant candlesticks is a tall black candle and the closing price of the first day is above the body of the second day. The second day is a Doji. The opening and closing prices of the Doji candle remain below the prior day's close, even though the shadows may overlap. However, those patterns with extremely long shadows in the second day are not considered as the "Doji Star, Bullish" patterns.

Rule:

$$\begin{aligned} & (len(S) = 2) \wedge (pt(TC) = -1) \\ & \wedge (\exists s_1, s_2 \in S \Rightarrow long_black_body(s_1) \wedge \\ & down_body_gap(s_1, s_2) \wedge Doji(s_2) \wedge !el_us(s_2) \wedge !el_ls(s_2)). \end{aligned} \quad (101)$$

Pattern P44: "Hammer, Inverted"

Description: The pattern shown in Fig. 10(b) begins with a down-trend. The first day in the significant candlesticks is a tall black candle. The closing price and the low price of the first day are similar. The second day is a small-bodied candle with a tall upper shadow and has little or no lower shadow. The second day can be either a black or white candlestick. The body of the second day cannot be a Doji. Otherwise, this candlestick is considered as a "Doji, Gravestone" (P7). The opening price of the second day is below the prior day's closing price.

Rule:

$$\begin{aligned} & (len(S) = 2) \wedge (pt(TC) = -1) \wedge \\ & (\exists s_1, s_2 \in S \Rightarrow long_black_body(s_1) \wedge \\ & small_ls(s_1) \wedge small_body(s_2) \wedge long_us(s_2) \wedge \\ & no_ls(s_2) \wedge down_body_gap(s_1, s_2)). \end{aligned} \quad (102)$$

Pattern P45: "In Neck"

Description: The pattern shown in Fig. 10(c) begins with a down-trend. The first day in the significant candlesticks is a long black candle. The second day is a white candle with an opening price below the low price of the first day and with a closing price that is near the bottom of the body of the first day. The second day closes into the body of the first day's body but not by much.

Rule:

$$\begin{aligned} & (len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow \\ & long_black_body(s_1) \wedge \\ & white_body(s_2) \wedge (op(s_2) < lp(s_1)) \wedge \\ & near_up(cp(s_2), cp(s_1))). \end{aligned} \quad (103)$$

Pattern P46: "Kicking, Bearish"

Description: This pattern is shown in Fig. 10(d). The first day in the significant candlesticks is a "White Marubozu" candle and it contains a long white candle with no shadows. Then the second day is a "Black Marubozu" candle and it contains a tall black candle with no shadows. The second day's high price should be

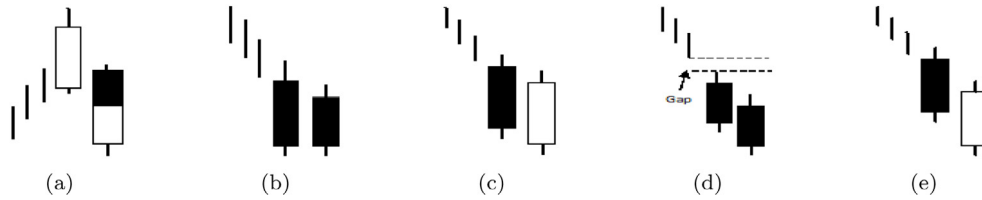


Fig. 9. (a) “Below the Stomach” candle. (b) “Matching Low” candle. (c) “Piercing Pattern” candle. (d) “Two Black Gapping” candle. (e) “Thrusting” candle.

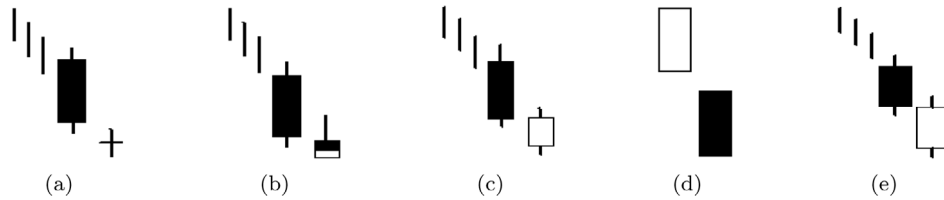


Fig. 10. (a) “Doji Star, Bullish” candle. (b) “Hammer, Inverted” candle. (c) “In Neck” candle. (d) “Kicking, Bearish” candle. (e) “Meeting Lines, Bullish” candle.

below the first day's low price leading to a downside shadow gap between two days.

Rule:

$$(len(S) = 2) \wedge (\exists s_1, s_2 \in S \Rightarrow no_us(s_1) \wedge no_ls(s_1) \wedge long_white_body(s_1) \wedge no_us(s_2) \wedge no_ls(s_2) \wedge long_black_body(s_2) \wedge down_shadow_gap(s_1, s_2)). \quad (104)$$

Pattern P47: “Meeting Lines, Bullish”

Description: The pattern shown in Fig. 10(e) begins with a downtrend. The first day in the significant candlesticks is a tall black candle. The second day is a tall white candle with a closing price at or near the prior day's closing price.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow long_black_body(s_1) \wedge long_white_body(s_2) \wedge ext_near(cp(s_1), cp(s_2))). \quad (105)$$

Pattern P48: “On Neck”

Description: The pattern shown in Fig. 11(a) begins with a downtrend. The first day in the significant candlesticks is a tall black candle. In the second day, the high price is lower than the first day's low price. The candle in the second day is a white candle with a close that matches the prior low.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow long_black_body(s_1) \wedge down_shadow_gap(s_1, s_2) \wedge white_candle(s_2) \wedge ext_near(lp(s_1), cp(s_2))). \quad (106)$$

Pattern P49: “Separating Lines, Bearish”

Description: The pattern shown in Fig. 11(b) begins with a downtrend. The first day in the significant candlesticks is a tall white candle. The second day is a tall black candle that opens at or near the same price as the opening price of the prior candle.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow long_white_body(s_1) \wedge long_black_body(s_2) \wedge near(op(s_1), op(s_2))). \quad (107)$$

Pattern P50: “Window, Falling”

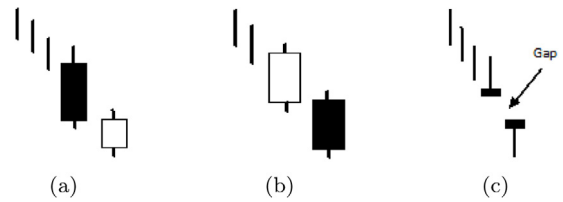


Fig. 11. (a) “On Neck” candle. (b) “Separating Lines, Bearish” candle. (c) “Window, Falling” candle.

Description: The pattern shown in Fig. 11(c) begins with a downtrend. The high price of the second day in the significant candlesticks is below the low price of the first day, leaving a downside shadow gap on the chart.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow down_hadow_gap(s_1, s_2)). \quad (108)$$

4.2.5. Engulfing right

In this group, the body of the first candle overlaps the second candle's body or the first candle overlaps the second candle. Such overlapping means that there are two areas in the body of candles that experience the same price interval or the second day's high-low prices are within the first day's high-low prices. To identify the patterns from this group, each two consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P51: “Harami, Bearish”

Description: This pattern is shown in Fig. 12(a). The first day in the significant candlesticks is a tall white candle. The second day is a small black candle. The open and close in second day is within the body of the first day regardless of the shadows. Either the tops or the bottoms of two bodies can be equivalent but not both, which can result two different situations. The first situation is that the top of the first day's body is equal to the top of the second day's body, while the bottom of the first day's body is not equal to the bottom of the second day's body. The second situation is that the top of the first day's body is not equal to the top of the second day's body, while the bottom of the first day's body is equal to the bottom of the second day's body.

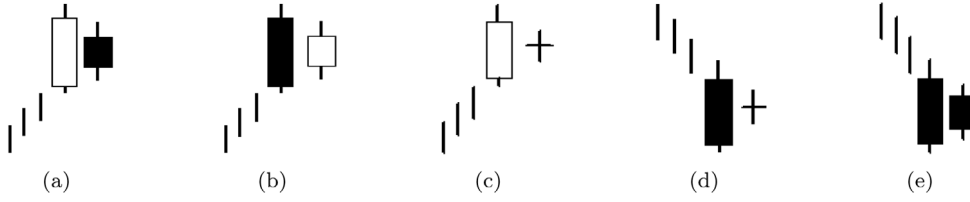


Fig. 12. (a) “Harami, Bearish” candle. (b) “Harami, Bullish” candle. (c) “Harami Cross, Bearish” candle. (d) “Harami Cross, Bullish” candle. (e) “Homing Pigeon” candle.

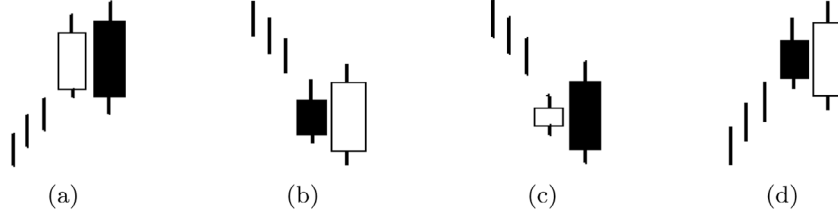


Fig. 13. (a) “Engulfing, Bearish” candle. (b) “Engulfing, Bullish” candle. (c) “Last Engulfing Bottom” candle. (d) “Last Engulfing Top” candle.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow long_white_body(s_1) \wedge small_black_body(s_2) \wedge ((cp(s_1) \geq op(s_2)) \wedge (cp(s_2) > op(s_1))) \vee ((cp(s_1) > op(s_2)) \wedge (cp(s_2) \geq op(s_1))))). \quad (109)$$

Pattern P52: “Harami, Bullish”

Description: This pattern is shown in Fig. 12(b). The pattern begins with an uptrend. The first day in the significant candlesticks is a tall black candle and the second day is a small-bodied white candle. The range of the body of the second day is in the range of the body of the first day. Either the tops or the bottoms of two bodies can be equivalent but not both, which can result two different situations. The first situation is that the top of the first day's body is equal to the top of the second day's body, while the bottom of the first day's body is not equal to the bottom of the second day's body. The second situation is that the top of the first day's body is not equal to the top of the second day's body, while the bottom of the first day's body is equal to the bottom of the second day's body.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow long_black_body(s_1) \wedge small_white_body(s_2) \wedge ((op(s_1) \geq cp(s_2)) \wedge (op(s_2) > cp(s_1))) \vee ((op(s_1) > cp(s_2)) \wedge (op(s_2) \geq cp(s_1)))). \quad (110)$$

Pattern P53: “Harami Cross, Bearish”

Description: The pattern shown in Fig. 12(c) begins with an uptrend. The first day in the significant candlesticks is a tall white candle. The second day is a Doji with a trading range inside the price range of the prior day.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow long_white_body(s_1) \wedge Doji(s_2) \wedge (lp(s_1) < lp(s_2)) \wedge (hp(s_2) < hp(s_1))). \quad (111)$$

Pattern P54: “Harami Cross, Bullish”

Description: The pattern shown in Figure Fig. 12(d) begins with a downtrend. The first day in the significant candlesticks is a tall

black candle and the second day is a Doji with a price range that is inside the range of the prior day.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow long_black_body(s_1) \wedge Doji(s_2) \wedge (lp(s_1) < lp(s_2)) \wedge (hp(s_2) < hp(s_1))). \quad (112)$$

Pattern P55: “Homing Pigeon”

Description: The pattern shown in Fig. 12(e) begins with a downtrend. The first day in the significant candlesticks is a tall black body. The second day is a small black body whose range is inside the range of the body of the first day.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow long_black_body(s_1) \wedge small_black_body(s_2) \wedge (cp(s_1) < cp(s_2)) \wedge (op(s_2) < op(s_1))). \quad (113)$$

4.2.6. Engulfing left

There are four patterns in this sub-group. In these patterns, the body of second candle overlaps the first candle's body or the second candle overlaps the first candle. To identify patterns in this group, each two consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P56: “Engulfing, Bearish”

Description: The pattern shown in Fig. 13(a) begins with an uptrend. The first day in the significant candlesticks is a white candle. The second day is a black candle and its body overlaps the white candle's body.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow white_body(s_1) \wedge black_body(s_2) \wedge (op(s_1) > cp(s_2)) \wedge (op(s_2) > cp(s_1))). \quad (114)$$

Pattern P57: “Engulfing, Bullish”

Description: The pattern shown in Fig. 13(b) begins with a downtrend. The first day in the significant candlesticks is a black candle. The second day is a white candle that opens below the prior body and closes above that body. The range of the price of the body in the second day does not need to engulf the shadows in the first day.

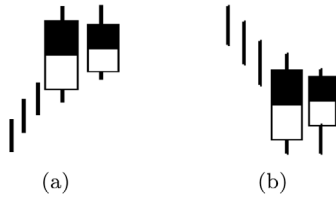


Fig. 14. (a) "Tweezers Top" candle. (b) "Tweezers Bottom" candle.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow black_body(s_1) \wedge white_body(s_2) \wedge (cp(s_1) > op(s_2)) \wedge (cp(s_2) > op(s_1))). \quad (115)$$

Pattern P58: "Last Engulfing Bottom"

Description: The pattern shown in Fig. 13(c) begins with a downtrend. The first day in the significant candlesticks is a white candle. The second day is a black candle that opens above the prior body and closes below the prior body. The range of the price of the body in the second day does not need to engulf the shadows in the first day.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow white_body(s_1) \wedge black_body(s_2) \wedge (op(s_1) > cp(s_2)) \wedge (op(s_2) > cp(s_1))). \quad (116)$$

Pattern P59: "Last Engulfing Top"

Description: The pattern shown in Fig. 13(d) begins with an uptrend. The first day in the significant candlesticks is a black candle. The second day is a white candle and the body of which overlaps the prior black candle's body.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow black_body(s_1) \wedge white_body(s_2) \wedge (cp(s_1) > op(s_2)) \wedge (cp(s_2) > op(s_1))). \quad (117)$$

4.2.7. Tweezers

The two candles in each pattern from this group have the same high or low price. To identify patterns in this group, each two consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P60: "Tweezers Top"

Description: The pattern shown in Fig. 14(a) begins with an uptrend. Two adjacent candlesticks in significant candlesticks share the same high price and they can be in any color.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2 \in S \Rightarrow ext_near(hp(s_2), hp(s_1))). \quad (118)$$

Pattern P61: "Tweezers Bottom"

Description: The pattern shown in Fig. 14(b) begins with a downtrend. Two adjacent candlesticks in significant candlesticks share the same low price and they can be in any color.

Rule:

$$(len(S) = 2) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2 \in S \Rightarrow ext_near(lp(s_2), lp(s_1))). \quad (119)$$

4.3. Three-line candles

Patterns in Three-Line Candles category have varying attributes. Specifically, they can be classified by the number of the star, Doji star, the existence of a gap and the movement of the trend price of the significant candlesticks.

4.3.1. One star

Each pattern in this group has only one star-shape candle (including a Doji candle). To identify such kind of patterns, each three consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend in each candlestick pattern should be calculated.

Pattern P62: "Abandoned Baby, Bearish"

Description: The pattern shown in Fig. 15(a) begins with an uptrend. The first day in the significant candlesticks is a white candle and it can be either short or tall. The second day is a Doji whose lower shadow is above the prior and following days' highs so that it leaves an upside shadow gap between the first and second day and a downside shadow gap between the second and third day. The third day is a black candle which can be either short or tall.

Rules:

$$(len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow (small_white_body(s_1) \vee normal_white_body(s_1) \vee long_white_body(s_1)) \wedge Doji(s_2) \wedge up_shadow_gap(s_1, s_2) \wedge down_shadow_gap(s_2, s_3) \wedge ((small_black_body(s_3) \vee normal_black_body(s_3) \vee long_black_body(s_3)))). \quad (120)$$

Pattern P63: "Abandoned Baby, Bullish"

Description: The pattern shown in Fig. 15(b) begins with a downtrend. The first day in the significant candlesticks is a black candle. The second day is a Doji that is below the prior and following days' lows so that it leaves a downside shadow gap between the first and second day and an upside shadow gap between the second and third day. The third day is a white candle.

Rule:

$$(len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow black_body(s_1) \wedge Doji(s_2) \wedge down_shadow_gap(s_1, s_2) \wedge up_shadow_gap(s_2, s_3) \wedge white_body(s_3)). \quad (121)$$

Pattern P64: "Doji Star, Collapsing"

Description: The pattern shown in Fig. 15(c) begins with an uptrend. The first day in the significant candlesticks is a white candle. The second day is a Doji, which is below the yesterday's low and forms a downside shadow gap between the first day and the second day. The third day is a black day, which is below the Doji's low and forms a downside shadow gap between the second day and the third.

Rule:

$$(len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow white_body(s_1) \wedge Doji(s_2) \wedge down_shadow_gap(s_1, s_2) \wedge down_shadow_gap(s_2, s_3) \wedge black_body(s_3)). \quad (122)$$

Pattern P65: "Evening Doji Star"

Description: The pattern shown in Fig. 15(d) begins with an uptrend. The first day in the significant candlesticks is a tall white

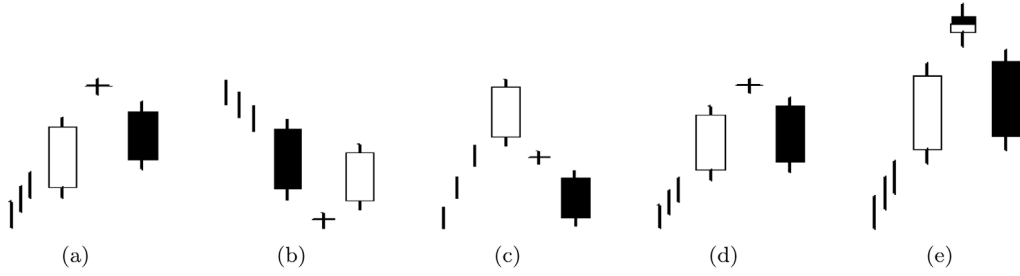


Fig. 15. (a) “Abandoned Baby, Bearish” candle. (b) “Abandoned Baby, Bullish” candle. (c) “Doji Star, Collapsing” candle. (d) “Evening Doji Star” candle. (e) “Evening Star” candle.

candle. The second day is a Doji, which is above the bodies of the surrounding candles. Two body gaps are formed and they are located between the first and the second candles as well as between the second candles and the third candles, respectively. The shadows are less important for identifying the patterns. The third day is a tall black candle that closes at or below the midpoint (well into the body) of the white candle.

Rule:

$$(len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow long_white_body(s_1) \wedge Doji(s_2) \wedge up_body_gap(s_1, s_2) \wedge down_body_gap(s_2, s_3) \wedge long_black_body(s_3) \wedge op(s_1) < (cp(s_3) \leq (0.5 * (op(s_1) + cp(s_1)))))) \quad (123)$$

Pattern P66: “Evening Star”

Description: The pattern shown in Fig. 15(e) begins with an uptrend. The first day in the significant candlesticks is a tall white candle. The second day is a candle with a small body that is above the bodies of the surrounding candles so that two body gaps are formed. These two body gaps located between the first and the second candles as well as between the second candles and the third candles. The candle can be either black or white. The third day is a tall black candle that is below the prior candle and closes at least half down the body of the white candle.

Rule:

$$(len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow long_white_body(s_1) \wedge small_body(s_2) \wedge up_body_gap(s_1, s_2) \wedge down_body_gap(s_2, s_3) \wedge long_black_body(s_3) \wedge (cp(s_3) > op(s_1)) \wedge (op(s_1) < cp(s_3) \leq (0.5 * (op(s_1) + cp(s_1)))))) \quad (124)$$

Pattern P67: “Morning Doji Star”

Description: The pattern shown in Fig. 16(a) begins with a downtrend. The first day in the significant candlesticks is a tall black candle. The second day is a Doji and its body is below the prior body so that a body gap is formed between the first day and the second day. The third day is a tall white candle and its body remains above the Doji’s body so that a body gap is formed between the second day and the third day.

Rule:

$$(len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow long_black_body(s_1) \wedge Doji(s_2) \wedge down_body_gap(s_1, s_2) \wedge long_white_body(s_3) \wedge up_body_gap(s_2, s_3)) \quad (125)$$

Pattern P68: “Morning Star”

Description: The pattern shown in Fig. 16(b) begins with a downtrend. The first day in the significant candlesticks is a tall black

candle. The second day is a candle with a small body that is lower than the prior body so that a body gap is formed between the first day and the second day. The color of the second candle can be either black or white. The third day is a tall white candle that is above the body of the second day so that a body gap is formed between the second and the third day. The third candle closes above the middle of the first day.

Rule:

$$(len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow long_black_body(s_1) \wedge small_body(s_2) \wedge down_body_gap(s_1, s_2) \wedge long_white_body(s_3) \wedge up_body_gap(s_2, s_3) \wedge (cp(s_3) > (0.5 * (op(s_1) + cp(s_1)))))) \quad (126)$$

4.3.2. Three Doji stars

Each pattern in this sub-group has three Doji candlesticks. To identify these patterns, each three consecutive candlesticks’ opening and closing price are needed and the short-term price trend should be calculated.

Pattern P69: “Tri-Star, Bearish”

Description: The pattern shown in Fig. 16(c) begins with an uptrend. The candles for the first three days in the significant candlesticks are three Doji candles. The middle candle has a body that is above the bodies of other two Doji.

Rule:

$$(len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow Doji(s_1) \wedge Doji(s_2) \wedge Doji(s_3) \wedge up_body_gap(s_1, s_2) \wedge down_body_gap(s_2, s_3)) \quad (127)$$

Pattern P70: “Tri-Star, Bullish”

Description: The pattern shown in Fig. 16(d) begins with a downtrend. The candles for the first three days in the significant candlesticks are Doji candles. The middle candle has a body that is below the bodies of other two candles.

Rule:

$$(len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow Doji(s_1) \wedge Doji(s_2) \wedge Doji(s_3) \wedge down_body_gap(s_1, s_2) \wedge up_body_gap(s_2, s_3)) \quad (128)$$

4.3.3. With gap

Each pattern in this sub-group has a gap. To identify patterns in this group, each three consecutive candlesticks’ opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P71: “Side-by-Side White Lines, Bearish”

Description: The pattern shown in Fig. 17(a) begins with a downtrend. The first day in the significant candlesticks is a black candle. The second and third day are both white candles. Their

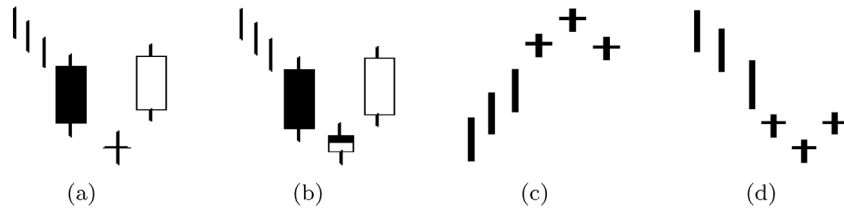


Fig. 16. (a) “Morning Doji Star” candle. (b) “Morning Star” candle. (c) “Tri-Star, Bearish” candle. (d) “Tri-Star, Bullish” candle.

bodies and opening prices are also similar. The closing prices of both candles remain below the body of the black candle.

Rule:

$$\begin{aligned} & (len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\ & black_body(s_1) \wedge white_body(s_2) \wedge white_body(s_3) \wedge \\ & near(cp(s_2), cp(s_3)) \wedge near(op(s_2), op(s_3)) \wedge \\ & down_body_gap(s_1, s_2) \wedge down_body_gap(s_1, s_3)). \end{aligned} \quad (129)$$

Pattern P72: “Side-by-Side White Lines, Bullish”

Description: The pattern shown in Fig. 17(b) begins with an uptrend. The first day in the significant candlesticks is a white candle. The second and third day are both white candles. They also have similar bodies and opening prices. The bodies of both last two candles are over the body of the first white candle, leaving a body gap between the first and the second and the third candlestick.

Rule:

$$\begin{aligned} & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\ & white_body(s_1) \wedge white_body(s_2) \wedge white_body(s_3) \wedge \\ & near(op(s_2), op(s_3)) \wedge near(cp(s_2), cp(s_3)) \wedge up_body_gap(s_1, s_2) \wedge \\ & up_body_gap(s_1, s_3)). \end{aligned} \quad (130)$$

Pattern P73: “Two Crows”

Description: The pattern shown in Fig. 17(c) begins with an uptrend. The first day in the significant candlesticks is a tall white candle. Then the second day is a black candle with a body that is over the prior body, which forms a body gap between the first and the second candlestick. The third day is a black candle that opens within the prior body and closes within the white candle's body (i.e. the first day).

Rule:

$$\begin{aligned} & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\ & long_white_body(s_1) \wedge black_body(s_2) \wedge \\ & up_body_gap(s_1, s_2) \wedge black_body(s_3) \wedge \\ & (cp(s_2) < op(s_3) < op(s_2)) \wedge (op(s_1) < cp(s_3) < cp(s_1))). \end{aligned} \quad (131)$$

Pattern P74: “Upside Gap Three Methods”

Description: The pattern shown in Fig. 17(d) begins with an uptrend. The first and second day in the significant candlesticks are both tall white candles with an upside shadow gap between them. The last day is a black candle which fills the gap.

Rule:

$$\begin{aligned} & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\ & long_white_body(s_1) \wedge long_white_body(s_2) \wedge \\ & up_body_gap(s_1, s_2) \wedge black_body(s_3) \wedge (cp(s_3) < cp(s_1)) \wedge \\ & (op(s_2) < op(s_3))). \end{aligned} \quad (132)$$

Pattern P75: “Downside Gap Three Methods”

Description: The pattern shown in Fig. 17(e) begins with a downtrend. The first day in the significant candlesticks is a long black-bodied candle. The second day is a long candle with a black

body. There is a downside shadow gap between current day and previous day. The third day is a white candle and it opens within the body of the second day and closes within the body of the first candle, thus closing the gap between the two black candles.

Rule:

$$\begin{aligned} & (len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\ & long_black_body(s_1) \wedge long_black_body(s_2) \wedge \\ & down_shadow_gap(s_1, s_2) \wedge white_body(s_3) \wedge \\ & (op(s_2) > op(s_3) > cp(s_2)) \wedge (op(s_1) > cp(s_3) > cp(s_1))). \end{aligned} \quad (133)$$

Pattern P76: “Downside Tasuki Gap”

Description: This pattern shown in Fig. 18(a) begins with a downtrend. The first day in the significant candlesticks is a black candle. The second day is also a black candle. A downside body gap is between the first and the second candles. The third day is a white candle which opens in the body of the prior candle and closes within the gap. In other words, the closing price of the third day must be below the close of the first day and above the open of the second day.

Rule:

$$\begin{aligned} & (len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\ & black_body(s_1) \wedge black_body(s_2) \wedge \\ & down_body_gap(s_1, s_2) \wedge white_body(s_3) \wedge \\ & (op(s_2) > op(s_3) > cp(s_2)) \wedge (op(s_2) < cp(s_3) < cp(s_1))). \end{aligned} \quad (134)$$

Pattern P77: “Upside Tasuki Gap”

Description: The pattern shown in Fig. 18(b) begins with an uptrend. The first day in the significant candlesticks is a white candle. The second day is also a white candle. The low price of the second day is higher than the high price of the first day, leaving a shadow gap between these two candles. The third day is a black candle which opens in the body of the prior candle and closes within the gap. The close of the last day is above the high price of the first candle but below the low of the second candle.

Rule:

$$\begin{aligned} & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\ & white_body(s_1) \wedge white_body(s_2) \wedge \\ & up_shadow_gap(s_1, s_2) \wedge black_body(s_3) \wedge (op(s_2) < \\ & op(s_3) < cp(s_2)) \wedge (cp(s_3) > hp(s_1)) \wedge (cp(s_3) < lp(s_2))). \end{aligned} \quad (135)$$

Pattern P78: “Upside Gap Two Crows”

Description: The pattern shown in Fig. 18(c) begins with an uptrend. The first day in the significant candlesticks is a tall white candle. The second day is a black candle with a body that is above the prior candle's body, leaving an upside body gap between these two candlesticks. The third day is a black candle whose body engulfs the body of the prior day. The close of the third day remains above the close of the first day.

Rule:

$$\begin{aligned} & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\ & long_white_body(s_1) \wedge black_body(s_2) \wedge \\ & up_body_gap(s_1, s_2) \wedge black_body(s_3) \wedge (op(s_3) > op(s_2)) \wedge \\ & (cp(s_2) > cp(s_3)) \wedge (cp(s_3) > cp(s_1))). \end{aligned} \quad (136)$$

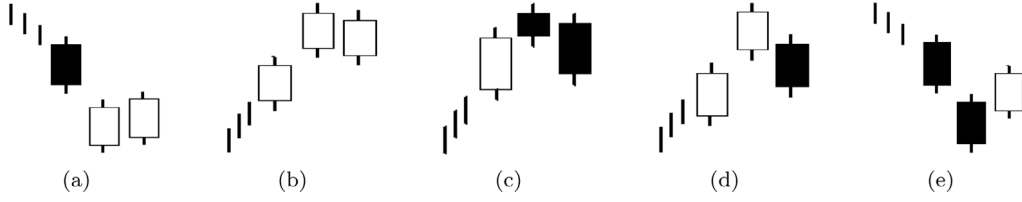


Fig. 17. (a) “Side-by-Side White Lines, Bearish” candle. (b) “Side-by-Side White Lines, Bullish” candle. (c) “Two Crows” candle. (d) “Upside Gap Three Methods” candle. (e) “Downside Gap Three Methods” candle.

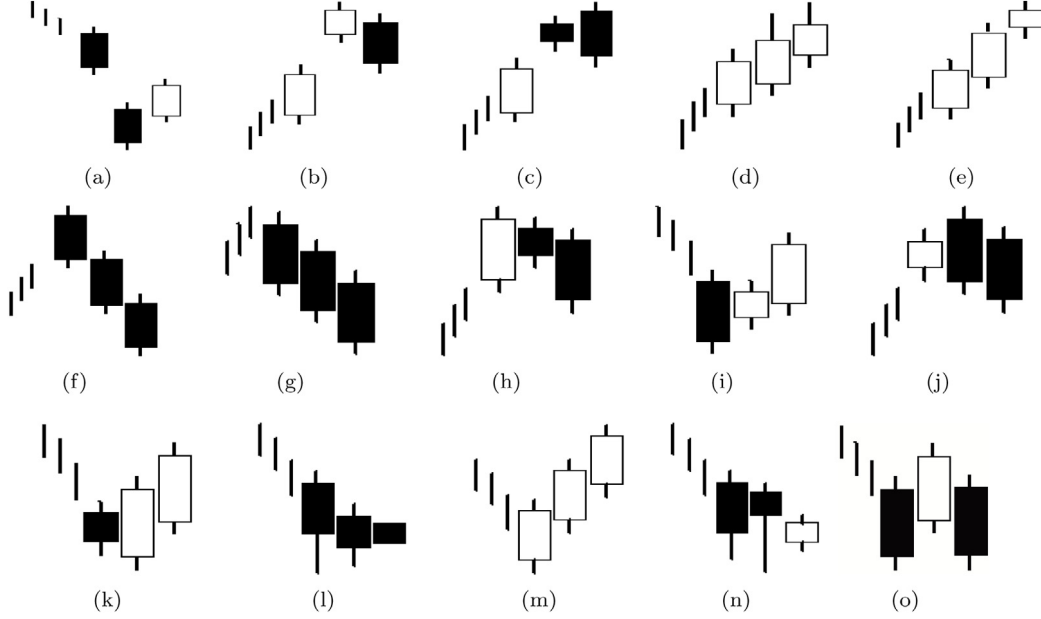


Fig. 18. (a) “Downside Tasuki Gap” candle. (b) “Upside Tasuki Gap” candle. (c) “Upside Gap Two Crows” candle. (d) “Advance Block” candle. (e) “Deliberation” candle. (f) “Identical Three Crows” candle. (g) “Three Black Crows” candle. (h) “Three Inside Down” candle. (i) “Three Inside Up” candle. (j) “Three Outside Down” candle. (k) “Three Outside Up” candle. (l) “Three Stars in the South” candle. (m) “Three White Soldiers” candle. (n) “Unique Three-River Bottom” candle. (o) “Stick Sandwich” candle.

4.3.4. Monotonic

Either the opening prices or the closing prices in the significant candlesticks of each pattern in this sub-group is monotonically increasing or decreasing. To identify patterns in this group, each three consecutive candlesticks’ opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P79: “Advance Block”

Description: The pattern shown in Fig. 18(d) begins with an uptrend. The color of all three candles in the significant candlesticks is white. The last two day’s opening prices must be within the previous body. Shadows on the second and the third days should be taller than ones on the first day and their own bodies.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad white_body(s_1) \wedge white_body(s_2) \wedge \\
 & \quad white_body(s_3) \wedge (op(s_1) < op(s_2) < cp(s_1)) \wedge \\
 & \quad (op(s_2) < op(s_3) < cp(s_2)) \wedge \\
 & \quad (hs(s_3) > hb(s_3)) \wedge (hs(s_2) > hb(s_2)) \wedge \\
 & \quad (hs(s_3) > hs(s_1)) \wedge (hs(s_2) > hs(s_1))).
 \end{aligned} \quad (137)$$

Pattern P80: “Deliberation”

Description: The pattern shown in Fig. 18(e) begins with an uptrend. The first and second days in the significant candlesticks are white candles with long bodies. The third day is a small white candle that opens near the second day’s close. Moreover, for the

last two candlesticks, each candle opens and closes higher than its previous one’s opening price and closing price.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad long_white_body(s_1) \wedge long_white_body(s_2) \wedge \\
 & \quad small_white_body(s_3) \wedge near(op(s_3), cp(s_2)) \wedge \\
 & \quad (op(s_1) < op(s_2)) \wedge (op(s_2) < op(s_3)) \wedge \\
 & \quad (cp(s_1) < cp(s_2)) \wedge (cp(s_2) < cp(s_3))).
 \end{aligned} \quad (138)$$

Pattern P81: “Identical Three Crows”

Description: The pattern shown in Fig. 18(f) begins with an uptrend. All three days in the significant candlesticks are tall black candles and the last two candles must respectively open at or near their prior closing prices.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad long_black_body(s_1) \wedge long_black_body(s_2) \wedge \\
 & \quad long_black_body(s_3) \wedge near(cp(s_2), op(s_3)) \wedge near(cp(s_1), op(s_2))).
 \end{aligned} \quad (139)$$

Pattern P82: “Three Black Crows”

Description: The pattern shown in Fig. 18(g) begins with an uptrend. All three days in the significant candlesticks are tall black candles. For the last two candles, each one close at a new low compared to its prior one. The last two candles open within

the body of the previous candle. Moreover, all candles should close at or near their lows.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad long_black_body(s_1) \wedge long_black_body(s_2) \wedge \\
 & \quad long_black_body(s_3) \wedge (cp(s_1) > cp(s_2)) \wedge \\
 & \quad (cp(s_2) > cp(s_3)) \wedge (cp(s_1) < op(s_2) < op(s_1)) \wedge \\
 & \quad (cp(s_2) < op(s_3) < op(s_2)) \wedge near(cp(s_1), lp(s_1)) \wedge \\
 & \quad near(cp(s_2), lp(s_2)) \wedge near(cp(s_3), lp(s_3))).
 \end{aligned} \quad (140)$$

Pattern P83: “Three Inside Down”

Description: The pattern shown in Fig. 18(h) begins with an uptrend. The first day in the significant candlesticks is a tall white candle. The second day is a small black candle and its opening and closing price are within the body of the first day regardless of their shadows. Either the tops or the bottoms of the bodies of first two candles can be similar, but not both, which leads to two situations. The first situation is that the top of the first day's body is equal to the top of the second day's body, while the bottom of the first day's body is not equal to the bottom of the second day's body. The second situation is that the top of the first day's body is not equal to the top of the second day's body, while the bottom of the first day's body is equal to the bottom of the second day's body. The third day candle is a black candle and it must close lower compared to the first and second day's closes.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad long_white_body(s_1) \wedge small_black_body(s_2) \wedge \\
 & \quad cp(s_1) \geq op(s_2) \wedge cp(s_2) \geq op(s_1) \vee \\
 & \quad (ext_near(cp(s_1), op(s_2)) \wedge !ext_near(cp(s_2), op(s_1))) \vee \\
 & \quad (!ext_near(cp(s_1), op(s_2)) \wedge ext_near(cp(s_2), op(s_1))) \wedge \\
 & \quad black_body(s_3) \wedge (cp(s_1) > cp(s_3)) \wedge (cp(s_2) > cp(s_3))).
 \end{aligned} \quad (141)$$

Pattern P84: “Three Inside Up”

Description: The pattern shown in Fig. 18(i) begins with a downtrend. This pattern is similar with the P83 except the color and positions are contrary.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad long_black_body(s_1) \wedge small_white_body(s_2) \wedge \\
 & \quad cp(s_1) \leq op(s_2) \wedge cp(s_2) \leq op(s_1) \vee \\
 & \quad (ext_near(cp(s_1), op(s_2)) \wedge !ext_near(cp(s_2), op(s_1))) \vee \\
 & \quad (!ext_near(cp(s_1), op(s_2)) \wedge ext_near(cp(s_2), op(s_1))) \wedge \\
 & \quad white_body(s_3) \wedge (cp(s_1) < cp(s_3)) \wedge (cp(s_2) < cp(s_3))).
 \end{aligned} \quad (142)$$

Pattern P85: “Three Outside Down”

Description: The pattern shown in Fig. 18(j) begins with an uptrend. The first day in the significant candlesticks is a white candle. The second day is a black candle and opens higher and closes lower than the prior candle's body. The third day is a black candle and its close is lower than other two days' bodies. Shadows in this candlestick pattern are ignored.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad white_body(s_1) \wedge black_body(s_2) \wedge \\
 & \quad (cp(s_1) < op(s_2)) \wedge (cp(s_2) < op(s_1)) \wedge \\
 & \quad (cp(s_3) < cp(s_2)) \wedge black_body(s_3)).
 \end{aligned} \quad (143)$$

Pattern P86: “Three Outside Up”

Description: The pattern shown in Fig. 18(k) begins with a downtrend. The first day in the significant candlesticks is a black candle. The second day is a white candle and opens below and closes above the prior body. The prices of the second day does not need to engulf the shadows of the first day. The third day is a white candle in which price closes higher than the closes of other two days.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad black_body(s_1) \wedge white_body(s_2) \wedge (op(s_2) < cp(s_1)) \wedge \\
 & \quad (cp(s_2) > op(s_1) \wedge white_body(s_3)) \wedge (cp(s_2) < cp(s_3))).
 \end{aligned} \quad (144)$$

Pattern P87: “Three Stars in the South”

Description: The pattern shown in Fig. 18(l) begins with a downtrend. The first day in the significant candlesticks is a tall black candle with a long lower shadow. The second day is similar to the first day but the height of the body of the second is smaller than the height of the body of the first day. The second day possesses a low that is above the previous day's low. The third day is a “Black Marubozu” type candle which fits inside of the high–low trading range of the prior day.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad long_black_body(s_1) \wedge long_ls(s_1) \wedge \\
 & \quad black_body(s_2) \wedge (lp(s_1) < lp(s_2)) \wedge \\
 & \quad (hb(s_1) > hb(s_2)) \wedge no_us(s_3) \wedge long_black_body(s_3) \wedge no_ls(s_3) \wedge \\
 & \quad (hp(s_2) > hp(s_3)) \wedge (lp(s_2) < lp(s_3))).
 \end{aligned} \quad (145)$$

Pattern P88: “Three White Soldiers”

Description: The pattern shown in Fig. 18(m) begins with a downtrend. All three days in the significant candlesticks are tall white candles. In the last two candles, the high price of each candle is higher than their previous one's high price. Furthermore, for these two days, price opens within the previous body. For each day in the significant candlesticks, the closing price should be near the high price.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad long_white_body(s_1) \wedge long_white_body(s_2) \wedge \\
 & \quad long_white_body(s_3) \wedge (hp(s_1) < hp(s_2)) \wedge (hp(s_2) < hp(s_3)) \wedge \\
 & \quad (op(s_1) < op(s_2) < cp(s_1)) \wedge (op(s_2) < op(s_3) < cp(s_2)) \wedge \\
 & \quad near(cp(s_1), hp(s_1)) \wedge near(cp(s_2), hp(s_2)) \wedge near(cp(s_3), hp(s_3))).
 \end{aligned} \quad (146)$$

Pattern P89: “Unique Three-River Bottom”

Description: The pattern shown in Fig. 18(n) begins with a downtrend. The first day in the significant candlesticks is a black candle with a tall body. The second day is a black body that is inside the prior body, but the low price drawing a long lower shadow is below the prior day's low. The third day is a white candle with a short body and it is below the body of the prior day.

Rule:

$$\begin{aligned}
 & (len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow \\
 & \quad long_black_body(s_1) \wedge black_body(s_2) \wedge (op(s_1) > op(s_2)) \wedge \\
 & \quad (cp(s_1) < cp(s_2)) \wedge \\
 & \quad long_ls(s_2) \wedge (lp(s_1) > lp(s_2)) \wedge small_white_body(s_3) \wedge \\
 & \quad down_body_gap(s_2, s_3)).
 \end{aligned} \quad (147)$$

4.3.5. Fluctuation

As the name suggests, the movement of the price trend in the significant candlesticks in this sub-group fluctuates. To identify the pattern in this group, each three consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P90: "Stick Sandwich"

Description: The pattern shown in Fig. 18(o) begins with a down-trend. The first day in the significant candlesticks is a black candle. The second day is a white candle that trades above the close of the first day, which indicates that the high of the second day is above the close of the first day. The third day is a black candle that closes at or near the close of the first day.

Rule:

$$(len(S) = 3) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3 \in S \Rightarrow black_body(s_1) \wedge white_body(s_2) \wedge (cp(s_1) < hp(s_2)) \wedge black_body(s_3) \wedge near(cp(s_3), cp(s_1))) \quad (148)$$

4.4. Four-line candles

There is no sub-group in this category and all patterns in this group include four candlesticks. To identify the patterns in this group, each four consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P91: "Concealing Baby Swallow"

Description: The pattern shown in Fig. 19(a) begins with a down-trend. The first and second day in the significant candlesticks both are "Black Marubozu" candles. The third day is a black candle with a tall upper shadow. The body of the third candle is lower than the body of the preceding candles. The third day trades into the body of the prior day, which indicates that the high of the third day is within the body of the second day. The fourth day is also a black candle that completely engulfs the prior day, including the shadows.

Rule:

$$(len(S) = 4) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3, s_4 \in S \Rightarrow long_black_body(s_1) \wedge long_black_body(s_2) \wedge no_us(s_1) \wedge no_us(s_2) \wedge no_ls(s_1) \wedge no_ls(s_2) \wedge black_body(s_3) \wedge long_us(s_3) \wedge down_body_gap(s_1, s_3) \wedge down_body_gap(s_2, s_3) \wedge (op(s_2) > hp(s_3) > cp(s_2)) \wedge black_body(s_4) \wedge (hp(s_4) > hp(s_3)) \wedge (lp(s_3) > lp(s_4))) \quad (149)$$

Pattern P92: "Three-Line Strike, Bearish"

Description: The pattern shown in Fig. 19(b) begins with a down-trend. The first to third days in the significant candlesticks are three black candles that form lower closes compared to their preceding candlestick. That is, the first day's closing price is higher than the second day's closing price. The second day's closing price is higher than the third day's closing price. The last day is a white candle that opens below the prior close and closes above the first day's open.

Rule:

$$(len(S) = 4) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3, s_4 \in S \Rightarrow black_body(s_1) \wedge black_body(s_2) \wedge black_body(s_3) \wedge (cp(s_1) > cp(s_2)) \wedge (cp(s_2) > cp(s_3)) \wedge white_body(s_4) \wedge (op(s_4) < cp(s_3)) \wedge (cp(s_4) > op(s_1))) \quad (150)$$

Pattern P93: "Three-Line Strike, Bullish"

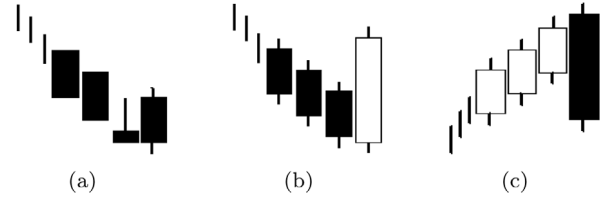


Fig. 19. (a) "Concealing Baby Swallow" candle. (b) "Three-Line Strike, Bearish" candle. (c) "Three-Line Strike, Bullish" candle.

Description: The pattern shown in 19(c) begins with an uptrend. The first to third days in the significant candlesticks are three white candles. The first day's closing price is lower than the second day's closing price. The second day's closing price is lower than the third day's closing price. The last day is a black candle that opens higher than the prior close and closes below the first day's opening price.

Rule:

$$(len(S) = 4) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3, s_4 \in S \Rightarrow white_body(s_1) \wedge white_body(s_2) \wedge white_body(s_3) \wedge (cp(s_1) < cp(s_2)) \wedge (cp(s_2) < cp(s_3)) \wedge black_body(s_4) \wedge (op(s_4) > cp(s_3)) \wedge (cp(s_4) < op(s_1))) \quad (151)$$

4.5. Five-line (or more) candles

There are four sub-groups of patterns in this category.

4.5.1. New price lines

For every pattern in this group, the high price of candlesticks in the significant candlesticks except the first candlestick is greater than the former day's high price. To identify these patterns, each eight consecutive candlesticks' high price are needed.

Pattern P94: "8 New Price Lines"

Description: This pattern is shown in Fig. 20(a). The candles for eight consecutive days in the significant candlesticks can be either black or white and each day except the second to last day in the significant candlesticks must make a higher high price compared to their own prior candle.

Rule:

$$(len(S) = 8) \wedge (\forall s_1, s_2, \dots, s_8 \in S, i \in \{1, 2, \dots, 7\} \Rightarrow (hp(s_i) < hp(s_{i+1}))) \quad (152)$$

Pattern P95: "10 New Price Lines"

Description: This pattern is shown in Fig. 20(b). The properties of this pattern is similar to pattern P94 except that there are 10 significant candlesticks in the pattern.

Rule:

$$(len(S) = 10) \wedge (\forall s_1, s_2, \dots, s_{10} \in S, i \in \{1, 2, \dots, 9\} \Rightarrow (hp(s_i) < hp(s_{i+1}))) \quad (153)$$

Patterns P96: "12 New Price Lines"

Description: This pattern is shown in Fig. 20(c). The properties of this pattern is similar to pattern P94 except that there are 12 significant candlesticks in the pattern.

Rule:

$$(len(S) = 12) \wedge (\forall s_1, s_2, \dots, s_{12} \in S, i \in \{1, 2, \dots, 11\} \Rightarrow (hp(s_i) < hp(s_{i+1}))) \quad (154)$$

Pattern P97: "13 New Price Lines"

Description: This pattern is shown in Fig. 20(d). The properties of this pattern is similar to pattern P94 except that there are 13

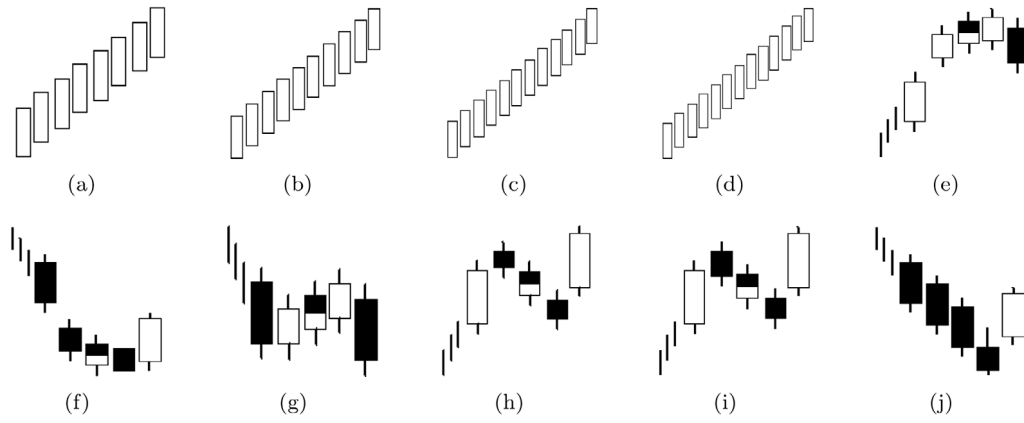


Fig. 20. (a) “8 New Price Lines” candle. (b) “10 New Price Lines” candle. (c) “12 New Price Lines” candle. (d) “13 New Price Lines” candle. (e) “Breakaway, Bearish” candle. (f) “Breakaway, Bullish” candle. (g) “Falling Three Methods” candle. (h) “Mat Hold” candle. (i) “Rising Three Methods” candle. (j) “Ladder Bottom” candle.

significant candlesticks in the pattern.

Rule:

$$(len(S) = 13) \wedge (\forall s_1, s_2, \dots, s_{13} \in S, i \in \{1, 2, \dots, 12\} \Rightarrow (hp(s_i) < hp(s_{i+1}))) \quad (155)$$

4.5.2. Breakaway

In this sub-group, there is a breakaway in each pattern. To identify these patterns, each five consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend in each pattern should be calculated.

Pattern P98: “Breakaway, Bearish”

Description: The pattern shown in Fig. 20(e) begins with an uptrend. The first day in the significant candlesticks is a tall white candle. The second day is a white candle followed with a upside body gap between the body of the first two candles, but the shadows can overlap. The third candle day can be in any color and has a higher close. The fourth day is a white candle with a close that is higher than the closes of previous candles. The fifth day is a tall black candle with a close within the gap between the first two candle bodies.

Rule:

$$(len(S) = 5) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3, s_4, s_5 \in S \Rightarrow \begin{aligned} &long_white_body(s_1) \wedge white_body(s_2) \wedge \\ &up_body_gap(s_1, s_2) \wedge (cp(s_3) > cp(s_2)) \wedge \\ &white_body(s_4) \wedge (cp(s_4) > cp(s_3)) \wedge \\ &long_black_body(s_5) \wedge (cp(s_1) < cp(s_5) < op(s_2))) \end{aligned} \quad (156)$$

Pattern P99: “Breakaway, Bullish”

Description: The pattern shown in Fig. 20(f) begins with a downtrend. The first day in the significant candlesticks is a tall black candle. The second day is a black candle followed with a downside a body gap between the body of the first two candles, but the shadows can overlap. The third day candle can be in any color and has a lower close. The fourth day is a black candle with a lower close. The fifth day is a tall white candle. The fifth day has a close within the body gap between the first two candle bodies, which means the close of the fifth day is below the closing price of the first day but above the opening price of the second day.

Rule:

$$(len(S) = 5) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3, s_4, s_5 \in S \Rightarrow \begin{aligned} &long_black_body(s_1) \wedge black_body(s_2) \wedge \\ &down_body_gap(s_1, s_2) \wedge (cp(s_3) < cp(s_2)) \wedge \\ &black_body(s_4) \wedge (cp(s_4) < cp(s_3)) \wedge \\ &long_white_body(s_5) \wedge (op(s_2) < cp(s_5) < cp(s_1))) \end{aligned} \quad (157)$$

4.5.3. Three methods

For each pattern in this group, the price trend of three middle candles is monotonic. To identify such kind of patterns, each five consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend in each pattern should be calculated.

Pattern P100: “Falling Three Methods”

Description: The pattern shown in Fig. 20(g) begins with a downtrend. The first day in the significant candlesticks is a tall black candle. The second to fourth days are short candles and the middle candle can be either black or white. The rest of the candles are white. The price trend of three middle candles is upward but the close remain within the high–low range of the first day. The fifth day is a tall black candle with a close below the first day's close.

Rule:

$$(len(S) = 5) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3, s_4, s_5 \in S \Rightarrow \begin{aligned} &long_black_body(s_1) \wedge \\ &small_white_body(s_2) \wedge small_body(s_3) \wedge small_white_body(s_4) \wedge \\ &(cp(s_2) < cp(s_3)) \wedge (cp(s_3) < cp(s_4)) \wedge (hp(s_1) > cp(s_4)) \wedge \\ &(lp(s_1) < cp(s_2)) \wedge long_black_body(s_5) \wedge (cp(s_5) < cp(s_1))) \end{aligned} \quad (158)$$

Pattern P101: “Mat Hold”

Description: The pattern shown in 20(h) begins with an uptrend. The first day in the significant candlesticks is a long white candle. The second day is a small black candle and its body is above the first day and shadows can be ignored. The candle in third day can be in any color. However, the fourth one is a black candle. Both candles in the third and the fourth day have small bodies with the closing price easing lower, but the bodies remain above the low of the first day. The fifth day is a white candle with a close above the highs of the prior four candles.

Rule:

$$(len(S) = 5) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3, s_4, s_5 \in S \Rightarrow \begin{aligned} &long_white_body(s_1) \wedge small_black_body(s_2) \wedge \\ &up_body_gap(s_1, s_2) \wedge small_body(s_3) \wedge small_black_body(s_4) \wedge \\ &(cp(s_2) > cp(s_3)) \wedge (cp(s_3) > cp(s_4)) \wedge (lp(s_1) < cp(s_4)) \wedge \\ &white_body(s_5) \wedge (cp(s_5) > \max(hp(s_1), hp(s_2), hp(s_3), hp(s_4)))) \end{aligned} \quad (159)$$

Pattern P102: “Rising Three Methods”

Description: The pattern shown in 20(i) begins with an uptrend. The first day in the significant candlesticks is a tall white candle. The second to fourth days are short candles. As for the second to fourth days, middle candle can be either black or white and the rest of the candles are black. The price trend of the middle three candles is downward but remains within the high–low range of the first day. The fifth day is a tall white candle with a close above the highs of the prior four candles.

Rule:

$$\begin{aligned} & (len(S) = 5) \wedge (pt(TC) = 1) \wedge (\exists s_1, s_2, s_3, s_4, s_5 \in S \Rightarrow \\ & long_white_body(s_1) \wedge small_black_body(s_2) \wedge small_body(s_3) \wedge \\ & small_black_body(s_4) \wedge (cp(s_3) < cp(s_2)) \wedge \\ & (cp(s_4) < cp(s_3)) \wedge (lp(s_1) < cp(s_4)) \wedge \\ & (hp(s_1) > cp(s_2)) \wedge long_white_body(s_5) \wedge \\ & (cp(s_5) > \max(hp(s_1), hp(s_2), hp(s_3), hp(s_4)))) \end{aligned} \quad (160)$$

4.5.4. Ladder bottom

There is only one pattern in this group. To identify this pattern, each five consecutive candlesticks' opening, high, low and closing prices are needed and the short-term price trend should be calculated.

Pattern P103: "Ladder Bottom"

Description: The pattern shown in 20(j) begins with a downtrend. The first to third days in the significant candlesticks are tall black candles and each candle opens and closes lower than the prior one. The fourth day is a black candle with an upper shadow. The fifth day is a white candle whose body is above the body of the prior day, which forms a body gap between the fourth and the fifth candle.

Rule:

$$\begin{aligned} & (len(S) = 5) \wedge (pt(TC) = -1) \wedge (\exists s_1, s_2, s_3, s_4, s_5 \in S \Rightarrow \\ & long_black_body(s_1) \wedge long_black_body(s_2) \wedge \\ & long_black_body(s_3) \wedge \\ & (op(s_2) < op(s_1)) \wedge (op(s_3) < op(s_2)) \wedge \\ & (cp(s_2) < cp(s_1)) \wedge (cp(s_3) < cp(s_2)) \wedge black_body(s_4) \wedge \\ & !no_us(s_4) \wedge white_body(s_5) \wedge up_body_gap(s_4, s_5)) \end{aligned} \quad (161)$$

5. Experiments

The main purpose of the experiments is to validate the fact that the formal specifications proposed in this paper can be used to create synthetic data for training and selecting best classifiers for identifying candlestick patterns and to validate the fact that specifications are effective to a certain degree. All the experiments were conducted in a computer with an Intel Core (TM) i7-6700HQ CPU with a 2.60 GHz basic frequency and 8 GB memory. From each subgroups, we select at least 1 sample pattern for the experiments. In total, we have selected 30 representative patterns for evaluation. The experiments are conducted on both real and synthetic data sets.

For the synthetic datasets, we first transform the formal specifications proposed in the previous section into rules. Next, we design standard templates for 30 selected patterns. Finally, by using an algorithm, we generate 1000 subsequences for each candlestick pattern comprising 500 positive and 500 negative samples based on the templates. The pseudo code for the generation of the synthetic data is described in Appendix A. The algorithm add noises to the standard templates. All in all, we have generated 30,000 samples for 30 selected patterns.

For the real datasets, historical price data of Apple Inc. (APPL), Microsoft Corporation (MSC) and Amazon.com, Inc. (AMZN), from

9 March 2007 to 8 March 2017, were used for the experiments. Historical prices were downloaded from Yahoo Finance.¹

In the preliminary experiments, we have tested 38 functions provided by Weka Data Mining Software² for classification. The synthetic data sets are used to train the classifiers by ten folds cross validation for different functions. Based on the results, we selected five top performing functions: *Bagging*, *Random Committee*, *Random SubSpace*, *PART*, *Random Forest* and two popular functions: *Artificial Neural Network (ANN)* and *Support Vector Machine (SVM)* for further comparison with the rule-based approach proposed in this paper.

In [30], Arratia defined the size of the body of a candlestick, i.e. the size of the body of a candlestick is a measure relative to the population the candlestick belongs to. In a similar way, in the proposed approach, we normalize the basic attributes of all the candlesticks which are extracted for pattern matching. These candlesticks are extracted using a sliding window. Since the patterns selected for the experiments belong to different subgroups, we use sliding-window of different sizes. The minimum width is one day and the maximum width is thirteen days. Therefore, subsequences extracted from the datasets may have different number of candlesticks. Each day in the dataset represents one candlestick and each candlestick comprises of five basic attributes: date, opening price, closing price, high price, and low price.

The size of the body of a candlestick and its adjacent candlesticks can affect the accuracy of the identification of the candlestick patterns. Therefore, we should also consider the neighboring candlesticks in pattern matching. In our approach, we propose a measure to take into account the population (neighborhood) of the individual candlestick. Basic attributes of every candlestick (opening price, high price, closing price, and low price) are normalized in Eqs. (162) and (163) based on the minimum and maximum values determined from all the candlesticks extracted with a sliding window. After this step, each attribute of a candlestick will be normalized.

$$\alpha'_i = (\alpha_i - \min(\Theta)) / (\max(\Theta) - \min(\Theta)) \quad (162)$$

where α is one of the basic attributes of a candlestick and $\alpha \in \{op, hp, cp, lp\}$. Θ is a set of prices containing all the related prices of the candlestick pattern based on the attribute α , from the first day to the last day in the corresponding window size. The output α'_i is the normalized value about the corresponding price at i th day. For example, when α is op , i is 1 and $\Theta \in \{op_1, op_2, \dots, op_n\}$, the normalized i th day's opening price can be calculated as follows:

$$\begin{aligned} op'_1 &= (op_1 - \min(op_1, op_2, \dots, op_n)) / (\max(op_1, op_2, \dots, op_n) \\ &\quad - \min(op_1, op_2, \dots, op_n)) \end{aligned} \quad (163)$$

5.1. Experiment results

5.1.1. Results for the synthetic datasets

The percentages of correctly classified instances from the synthetic datasets are shown in Table 2. From these results, we can observe that the *Random Forest* is the best classifier of classifying the candlestick patterns with up to 95.30% accuracy and *SVM* is the worst classifier with only 73.49% accuracy. Note that the RB method was not evaluated for the synthetic datasets since the formal specifications proposed in this paper were used to create the normal templates and synthetic datasets.

In Fig. 21, we depict the experiment results from synthetic data sets with a bar chart. The average number of the true positive and the true negative results are depicted for each classifier.

¹ <https://finance.yahoo.com/>.

² <https://www.cs.waikato.ac.nz/ml/weka/>.

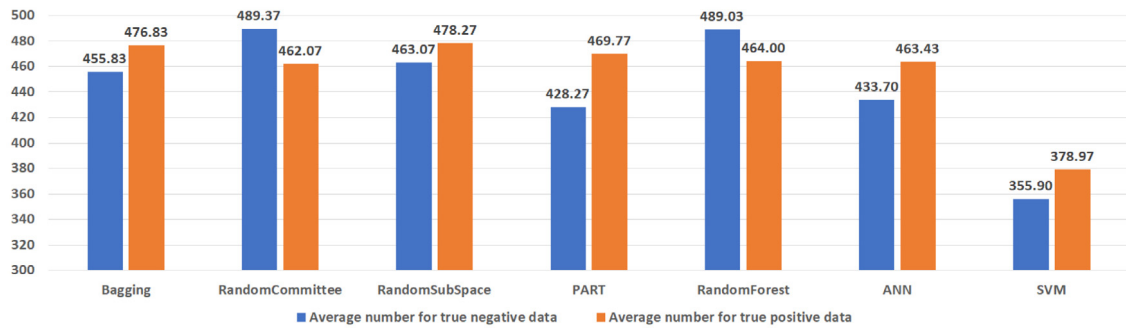


Fig. 21. Results of classification by different classifiers.

Table 2

The performance of the seven classifiers for the synthetic datasets.

Pattern ID	Pattern name	Correctly classified instances (%)						
		Classifier						
		Bagging	Random committee	Random subspace	PART	Random forest	ANN	SVM
P1	Marubozu, Black	97.7	96.5	95.2	97.6	97.1	97.4	55
P6	Shooting Star, One-Candle	97.2	98.4	97.4	96.6	98.3	92.9	77.2
P14	Takuri Line	97	98.7	97.3	96.1	98.4	93	65.8
P19	Doji, Gapping Down	97.1	97.8	97.4	95.5	98.5	94.3	68.6
P35	Window, Rising	91.9	95.9	93.9	89.9	95.5	91.3	66.2
P36	Above the Stomach	94.1	96	95.3	89.3	95.7	88.1	68.3
P41	Two Black Gapping Candles	93.1	95	95.1	88.3	95.3	88.1	65.5
P50	Window, Falling	93.2	96.2	93	91.6	95.5	92	84.1
P53	Harami Cross, Bearish	93.1	95.5	94.2	90.7	95.9	88.9	71.4
P56	Engulfing, Bearish	93.8	95.9	95.5	89.7	96.1	91.1	78.3
P60	Tweezers Top	92.1	95.3	92.3	90.9	95.4	92.3	78.4
P62	Abandoned Baby, Bearish	93.8	95.2	93.4	88.3	95.4	87.9	63.6
P67	Morning Doji Star	91.7	94	93.1	86.6	94.7	85.8	77.6
P69	Tri-Star, Bearish	92	94.2	92.4	88.7	94.1	89.8	76.1
P71	Side-by-Side White Lines, Bearish	93.1	95.5	93.3	92.6	95	91.6	74.3
P73	Two Crows	93.4	95.6	94.5	91.8	95.5	89.2	84.3
P74	Upside Gap Three Methods	92.5	94.1	94.1	88.6	95.3	88.7	82.5
P77	Upside Tasuki Gap	93.6	94.3	94.2	89.7	94.6	88.3	83.2
P79	Advance Block	91.3	93.6	93.5	86	94.3	88	81.8
P81	Identical Three Crows	89.8	92.4	91.7	81.9	93.2	86.5	63.8
P84	Three Inside Up	91	93.9	93.3	85.7	93.9	87.5	79.3
P86	Three Outside Up	94.8	95.3	95.3	85.8	95.3	91.1	87.1
P87	Three Stars in the South	89.8	93.5	92.4	83.9	93.9	86.2	81.8
P90	Stick Sandwich	93.1	95.3	93.9	89.9	95.3	89.4	73.2
P91	Concealing Baby Swallow	91.5	93.5	92.3	86.3	93.6	85.7	81
P92	Three-Line Strike, Bearish	92.4	93.6	92.2	90.7	93.7	89	75.8
P94	8 New Price Lines	98.4	95.1	93.3	87.5	94.8	90.4	66.8
P98	Breakaway, Bearish	91.9	92.2	92.7	86.8	92.1	89.7	67.1
P100	Falling Three Methods	91.7	92.9	93.4	88.6	93.5	89.4	65.4
P103	Ladder Bottom	91.9	98.9	98.4	98.5	99.2	87.8	61.1
Average performance		93.27	95.14	94.13	89.80	95.30	89.71	73.49

These results are obtained from the synthetic datasets containing 500 positive instances and 500 negative instances of each pattern.

From Fig. 21, we can observe that *RandomCommittee* and *SVM* achieve the best and worst results in classifying negative patterns. We can also observe that *RandomSubSpace* and *SVM* achieve the best and worst results in classifying positive patterns. Among the 7 classifiers tested in the experiments, *RandomCommittee* and *RandomForest* should be favored when maximizing the probability of identifying the true negative patterns. Correspondingly, *RandomSubSpace* and *Bagging* should be favored when maximizing the probability of identifying the true positive patterns.

In addition, we depict an example classification from synthetic dataset for further illustration in Fig. 22. A standard template, false positive and false negative results of pattern P100 “Falling Three Methods” are shown in Fig. 22. From the experiment results, we found that the classifier *RandomForest* cannot correctly classify the pattern because it cannot accurately recognize the positions of the candlesticks from the given subsequence. Specifically, in the standard pattern, the third day’s closing price is lower than the fourth day’s closing price. However, in the false positive

case (in Fig. 22(b)), the third day’s closing price is higher than the fourth day’s closing price and in the false negative case (in Fig. 22(c)), the third day’s closing price is lower than the fourth day’s closing price.

5.1.2. Results for real datasets

In order to validate the effectiveness of the specifications proposed in this paper, experiments for real datasets are conducted in two phases. In the first phase, we compare the results of 7 classifiers which are trained with the synthetic datasets with the MATLAB library developed by Nate Jensen (N.J.) [29]. The results are shown in Tables 3–5. Note that the MATLAB library contains rules for 90 patterns only and some of the pattern names used in the library are slightly different from ours. Among the 30 patterns tested in the previous experiments for synthetic datasets, we found 20 corresponding patterns in the MATLAB library. Accordingly, we choose these 20 patterns for comparison in the experiments for real datasets. The last column in Tables 3–5 represents the pattern found by the Nate Jensen’s approach. “N/A” in the last column indicates that the pattern is not available in [29].

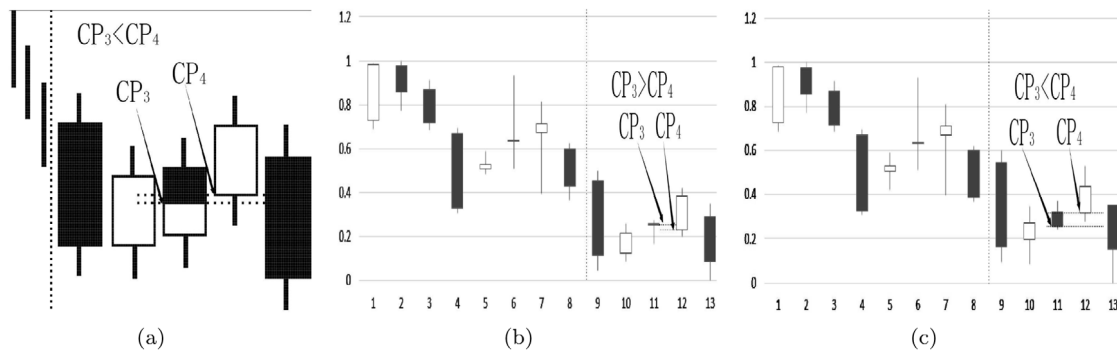


Fig. 22. Classification examples of *RandomForest* for pattern P100 “Falling Three Methods” from the synthetic datasets. (a) normal template of P100, (b) false positive result, and (c) false negative result.

Table 3

The number of patterns identified by seven classifiers and N.J. approach [29] in the stock APPL.

Pattern No.	Pattern name	No. of patterns found							
		Classifier							
		Bagging	Random committee	Random subspace	PART	Random forest	ANN	SVM	N.J.
P1	Marubozu, Black	40	38	249	40	39	56	114	36
P6	Shooting Star, One-Candle	206	108	129	277	91	15	14	N/A
P14	Takuri Line	429	487	586	197	333	72	175	69
P19	Doji, Gapping Down	290	76	94	825	32	6	6	N/A
P35	Window, Rising	40	751	838	165	1258	263	484	N/A
P36	Above the Stomach	34	0	65	1	0	9	30	N/A
P41	Two Black Gapping Candles	12	86	22	7	23	11	28	N/A
P50	Window, Falling	285	434	147	417	459	752	437	N/A
P53	Harami Cross, Bearish	23	435	38	25	8	2	2	0
P56	Engulfing, Bearish	1	1	6	5	1	0	58	N/A
P60	Tweezers Top	1005	159	74	205	109	1246	46	N/A
P62	Abandoned Baby, Bearish	18	75	6	2	12	1	3	0
P67	Morning Doji Star	528	7	296	3	19	193	0	0
P69	Tri-Star, Bearish	39	8	26	104	3	12	32	1
P71	Side-by-Side White Lines, Bearish	137	170	218	95	240	64	157	11
P73	Two Crows	856	1021	1031	0	457	26	313	0
P74	Upside Gap Three Methods	40	982	6	4	604	2	0	0
P77	Upside Tasuki Gap	1107	205	1109	58	437	1	20	0
P79	Advance Block	393	222	245	136	96	23	80	N/A
P81	Identical Three Crows	365	213	32	1	0	2	86	0
P84	Three Inside Up	16	12	114	8	0	5	185	16
P86	Three Outside Up	0	0	0	1	0	1	10	6
P87	Three Stars in the South	15	37	19	1	13	1	28	0
P90	Stick Sandwich	91	518	1126	4	372	9	167	0
P91	Concealing Baby Swallow	0	4	0	27	0	1	3	0
P92	Three-Line Strike, Bearish	94	114	86	12	154	50	130	0
P94	8 New Price Lines	40	14	10	53	9	152	196	N/A
P98	Breakaway, Bearish	8	0	1	21	0	568	1	0
P100	Falling Three Methods	1	0	0	0	0	14	195	0
P103	Ladder Bottom	0	6	0	6	0	286	121	1

In the second phase, we translate the specifications proposed in this paper into rules to build a rule-based pattern classification method (RB) and compare directly with the results obtained from using MATLAB library developed by Nate Jensen [29]. The experiment results are listed in Table 6. We have also analyzed the MATLAB library and found three significant differences in the rules defined for pattern classifications.

1. The first significant difference is that the rules in [29] do not consider the price trend. According to the specifications proposed in this paper, the price trend can be calculated by using trend candlesticks (see Fig. 1(m)) based on Eq. (14) from Section 3.
2. The second significant difference is that, in our approach, we first normalize the candles within the sliding window. Next, the type of the length of an upper or lower shadow is determined by only considering certain attributes of the current candlestick. In Nate Jensen's approach, two cases are considered to identify the type of the length of an upper

or lower shadow. In the case of long or very long upper or lower shadow, the type of the lengths of the shadows is determined by only considering certain attributes of the current candlestick. In the case of small or very small upper or lower shadow, the type of the lengths of the shadows are determined by comparing with the average length of the shadows of its ten past candles.

3. The last significant difference is similar to the second one. In our approach, we first normalize the candles within the sliding window. Next, we determine the type of the length of a body by only considering certain attributes of the current candlestick. In Nate Jensen's approach, the type of the length of the body is determined by comparing with the average length of its ten past candlesticks' bodies.

To further illustrate the differences between two approaches, a specific case of pattern P1. *Marubozu, Black* found by RB method and Nate Jensen's approach in APPL is shown in Fig. 23. Fig. 23(a) is the template of P1 used in the experiment. Patterns found by

Table 4

The number of patterns identified by seven classifiers and N.J. approach [29] in stock MSC.

Pattern No.	Classifier	No. of patterns found							
		Bagging	Random committee	Random subspace	PART	Random forest	ANN	SVM	N.J.
P1	Marubozu, Black	25	25	199	27	25	35	80	21
P6	Shooting Star, One-Candle	174	66	119	284	68	22	17	N/A
P14	Takuri Line	421	447	603	176	291	72	175	79
P19	Doji, Gapping Down	219	49	83	664	24	6	6	N/A
P35	Window, Rising	45	820	743	105	1255	188	407	N/A
P36	Above the Stomach	56	0	81	3	0	8	9	N/A
P41	Two Black Gapping Candles	15	59	30	10	8	4	11	N/A
P50	Window, Falling	251	358	131	368	362	679	352	N/A
P53	Harami Cross, Bearish	38	363	54	28	9	1	2	33
P56	Engulfing, Bearish	0	0	1	1	2	1	58	N/A
P60	Tweezers Top	951	123	58	166	83	1336	51	N/A
P62	Abandoned Baby, Bearish	24	70	16	7	10	0	2	0
P67	Morning Doji Star	495	8	283	10	14	0	157	0
P69	Tri-Star, Bearish	36	12	21	77	2	7	28	2
P71	Side-by-Side White Lines, Bearish	142	136	202	124	182	29	87	14
P73	Two Crows	953	1125	1069	2	516	21	215	0
P74	Upside Gap Three Methods	46	995	5	1	623	0	1	0
P77	Upside Tasuki Gap	1154	209	1142	79	430	2	18	0
P79	Advance Block	346	240	229	147	87	10	50	N/A
P81	Identical Three Crows	319	180	24	0	1	2	56	0
P84	Three Inside Up	29	16	122	9	2	1	157	7
P86	Three Outside Up	0	0	0	0	0	0	10	5
P87	Three Stars in the South	17	33	16	0	12	0	20	0
P90	Stick Sandwich	70	397	975	10	267	8	88	0
P91	Concealing Baby Swallow	0	2	0	25	0	0	7	0
P92	Three-Line Strike, Bearish	49	80	79	9	104	39	77	1
P94	8 New Price Lines	22	6	1	28	2	102	128	N/A
P98	Breakaway, Bearish	8	0	3	37	0	539	3	0
P100	Falling Three Methods	0	0	0	2	0	16	167	0
P103	Ladder Bottom	0	5	0	3	0	279	109	0

Table 5

The number of patterns identified by seven classifiers and N.J. approach [29] in stock AMZN.

Pattern No.	Classifier	No. of patterns found							
		Bagging	Random committee	Random subspace	PART	Random forest	ANN	SVM	N.J.
P1	Marubozu, Black	29	25	248	29	29	30	59	16
P6	Shooting Star, One-Candle	183	98	118	276	98	22	14	N/A
P14	Takuri Line	436	489	609	194	305	77	187	52
P19	Doji, Gapping Down	228	34	79	690	24	3	7	N/A
P35	Window, Rising	47	799	751	142	1239	156	325	N/A
P36	Above the Stomach	45	0	78	2	0	5	18	N/A
P41	Two Black Gapping Candles	6	65	22	6	8	9	32	N/A
P50	Window, Falling	235	351	101	333	331	661	363	N/A
P53	Harami Cross, Bearish	28	379	39	29	5	0	1	25
P56	Engulfing, Bearish	0	1	0	1	0	1	61	N/A
P60	Tweezers Top	954	158	57	188	102	1323	47	N/A
P62	Abandoned Baby, Bearish	17	50	9	0	9	1	2	0
P67	Morning Doji Star	492	5	281	5	17	159	1	0
P69	Tri-Star, Bearish	40	9	29	80	3	8	21	0
P71	Side-by-Side White Lines, Bearish	113	110	171	93	141	33	117	3
P73	Two Crows	873	1111	1010	2	487	22	281	0
P74	Upside Gap Three Methods	32	951	6	0	609	2	0	0
P77	Upside Tasuki Gap	1154	209	1142	79	430	2	18	2
P79	Advance Block	355	190	229	135	89	11	65	N/A
P81	Identical Three Crows	313	190	32	0	0	1	66	0
P84	Three Inside Up	23	4	117	7	1	5	130	12
P86	Three Outside Up	1	0	0	0	0	0	5	4
P87	Three Stars in the South	9	18	10	0	6	0	16	0
P90	Stick Sandwich	50	415	1028	2	249	8	106	0
P91	Concealing Baby Swallow	1	3	0	14	0	0	5	0
P92	Three-Line Strike, Bearish	50	60	53	6	81	35	113	1
P94	8 New Price Lines	22	3	3	3	3	56	73	N/A
P98	Breakaway, Bearish	5	0	2	25	0	565	3	0
P100	Falling Three Methods	2	0	0	2	0	15	188	0
P103	Ladder Bottom	1	12	0	3	0	275	93	0

Table 6
Comparison between the number of patterns found by the RB method and N.J. approach in real datasets.

No.	Pattern name in [2]	No. of patterns found by RB method			No.	Pattern name in N.J. [29]	No. of patterns found by N.J method		
		AAPL	MSC	AMZN			AAPL	MSC	AMZN
P1	Marubozu, Black	20	13	21	39 ^a	Marubozu	36	21	16
P14	Takuri Line	23	19	10	2 ^c	Inverted Takuri/Umbrella	69	79	52
P53	Harami Cross, Bearish	1	14	15	18 ^a	Harami Cross	0	33	25
P62	Abandoned Baby, Bearish	1	0	0	1 ^a	Abandoned Baby	0	0	0
P67	Morning Doji Star	0	0	0	4 ^b	Morning Doji Star	0	0	0
P69	Tri-Star, Bearish	1	0	0	22 ^a	Tri Star	1	2	0
P71	Side-by-Side White Lines, Bearish	0	0	0	33 ^a	Side-by-Side White Lines	11	14	3
P73	Two Crows	0	0	0	9 ^a	Upside Gap Two Crows	0	0	0
P74	Upside Gap Three Methods	0	0	0	33 ^b	Upside Gap Three Methods	0	0	0
P77	Upside Tasuki Gap	0	0	0	34 ^b	Upside Tasuki Gap	0	0	2
P81	Identical Three Crows	0	0	0	19 ^a	Identical Three Crows	0	0	0
P84	Three Inside Up	0	0	2	7 ^b	Three Inside Up	16	7	12
P86	Three Outside Up	2	12	9	8 ^b	Three Outside Up	6	5	4
P87	Three Stars in the South	0	0	0	23 ^b	Three Stars in the South	0	0	0
P90	Stick Sandwich	2	0	2	22 ^b	Stick Sandwich	0	0	0
P91	Concealing Baby Swallow	0	0	0	2 ^b	Concealing Baby Swallow	0	0	0
P92	Three-Line Strike, Bearish	0	0	0	36 ^b	Three Line Strike	0	1	1
P98	Breakaway, Bearish	0	0	0	11 ^a	Breakaway	0	0	0
P100	Falling Three Methods	0	0	0	28 ^a	Falling Three Methods	0	0	0
P103	Ladder Bottom	0	0	0	18 ^b	Ladder Bottom	1	0	0

Patterns are categorized into three groups [29].

^aThe pattern belongs to bear group.

^bThe pattern belongs to the bull group.

^cThe pattern belongs to neutral group.

RB method and Nate Jensen's approach are depicted in Figs. 23(b) and 23(c). In Fig. 23(b), the candle on 18-Feb-16 is identified as the pattern and in Fig. 23(c), the candle on 18-Oct-16 is identified as the pattern. The figure shows that the identified patterns are basically consistent with the template except that RB method only considers the length of shadow in the candlestick that is being identified, while the method proposed by N.J. considers the lengths of the shadows of the candlestick that is being identified and its past 10 candlesticks.

6. Conclusion

In this paper, we describe the formal specifications of 103 known candlestick patterns from financial time series. These candlestick patterns are classified into sub-groups based on their characteristics. In the experiments, we selected 30 representative patterns from each subgroup for evaluation. Next, the formal specifications of these patterns are formulated in the form of first-order-logic. These specifications are later translated into rules which can be used to generate synthetic data. A rule-based classification method (RB) was also implemented based on the translated rules.

The synthetic datasets generated are then used for training and testing 7 selected classifiers namely (a) *Bagging*, (b) *Random-Committee*, (c) *RandomSubSpace*, (d) *PART*, (e) *RandomForest*, (f) *ANN*, and (g) *SVM*. Furthermore, we compare the performance of both 7 selected classifiers and RB method with the MATLAB library developed by [29] on real datasets. The scientific and technical contributions made in this article are as follows:

1. In this paper, we endeavor to compile the complete formal specifications of 103 known candlestick patterns from the financial trading literature.
2. Since the formal specifications are formulated in first-order-logic, they are unambiguous and can be easily extended when a new pattern is to be described.
3. The formal specifications can be easily translated into rules which can be used to generate synthetic datasets for selecting best classifiers.

4. The work described in this paper is intended as a reference model for further research in candlestick pattern matching and stock trend prediction. For example, classifiers tested in this paper for identification of patterns can be combined with trend following or other methods from technical analysis for forecasting market trends.

Although 30 patterns are selected in the experiments for real datasets to compare the RB approach and the MATLAB Library developed by [29], further analysis is necessary in evaluating the rules defined in [29] and the formal specifications proposed in this paper. Specifically, additional experiments should be conducted in designing trading strategies which are based on the candlestick patterns identified by various classifiers and approaches. The remaining patterns should also be tested to evaluate the effectiveness of the formal specifications in generating the synthetic datasets. As for the future work, we are planning to develop a real-time (online) decision support system for candlestick pattern classification based on the formal specifications proposed in this paper.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.asoc.2019.105700>.

Acknowledgments

This work was funded by the Research Committee of the University of Macau, China [grant numbers MYRG2016-00148-FST, MYRG2017-00029-FST].

Appendix A. Algorithm for synthetic data generation

See Algorithm 1.

Algorithm 1 The synthetic data generator

```

1: procedure SYNTHETICGENERATOR(basDa, patternID,  $\mathcal{R}$ )
Require: basDa: Sets of basic data saved in CSV files for 30 selected candlestick patterns. patternID: The ID of pattern needed to be
generated.  $\mathcal{R}$ : Sets of rules of all candlestick patterns.
Ensure: CSVFiles: 30 CSV type of files about the synthetic data. Each one of them contains 500 positive or 500 negative records relating
to the corresponding candlestick pattern.
2: Let maxNumberOfCandle be an integer number of candles that the candlestick pattern contains with the patternID;
3: Let date be the date of a candle;
4: Let sizeBDs be the size of the basDa;
5: Let numberOfCandle be an integer number indicating the number of candlesticks that are needed to be changed. The range of the
value is (0, maxNumberOfCandle];
6: Let numberOfAttribute be an integer number indicating the number of attributes that needs to be changed. The range of the value
is (0,  $4 * \text{numberOfCandle}$ ];
7: Let range be the range of change of generated data;
8: Let maxRange be the max range of change of generated number;
9: Let offset be the deviation about how many the generated price is away from the original price;
10: Let whichAttribute[ ] be a group of integer numbers indicating which attributes need to be changed;
11: Let sytData[ ] be the sets of records of synthetic data saved in a CSV type of file and each set (or record) in sytData[ ] includes
numbers of groups of opening, high, low and closing prices according to the numberOfCandle. Each generated price is decided
together by the numberOfCandle, numberOfAttribute, range and whichAttribute;
12: Let sytDataP[ ] be the sets of new positive data that are fulfilling the rules of specific patterns;
13: Let sytDataN[ ] be the sets of new negative data that are not fulfilling the rules of specific patterns;
14: Let ceil(p) be a function. After processing, the parameter within the brackets will be plus by one and the decimal part of the
parameter will be set to zero as long as its decimal part is not equaling to zero.;
15: Let random() be a function. A random number will be generated and its value is between [0,1]
16: numberOfCandle  $\leftarrow \text{ceil}(\text{maxNumberOfCandle} \times \text{random}());$ 
17: numberOfAttribute  $\leftarrow \text{ceil}(4 * \text{numberOfCandle} \times \text{random}());$ 
18: repeat
19:   whichAttribute[ ]  $\leftarrow (\text{numberOfAttributes of}) 4 * \text{numberOfCandle} * \text{random}().$   $\triangleright$  There are numberOfAttribute of elements in
whichAttribute[ ]
20: until There is no same elements in whichAttribute[ ]
21:   range  $\leftarrow \text{ceil}(\text{maxRange} \times \text{random}() + \text{offset}).$ 
22:   for i  $\leftarrow 1$  to 500 do
23:     for j  $\leftarrow 1$  to sizeBDs do
24:       repeat
25:         interimData[ ]  $\leftarrow \text{whichAttribute}[ ]$  in basDa  $\times$  range
26:         if interimData[ ] fulfills  $\mathcal{R}$  regarding to the patternID then
27:           sytDataP[ ]  $\leftarrow \text{interimData}[ ]$ 
28:         else
29:           sytDataN[ ]  $\leftarrow \text{interimData}[ ]$ 
30:         end if
31:       until Each candle in interimData[ ] fulfills the basic requirements of a normal candle.  $\triangleright$  A normal candle is that the high
price is the largest one and the low price is the smallest one among four basic elements: high, opening, closing and low price.
32:     end for
33:   end for
34: CSVFiles  $\leftarrow \text{sytDataP}[ ]$  and sytDataN[ ].
35: return CSVFiles.
36: end procedure

```

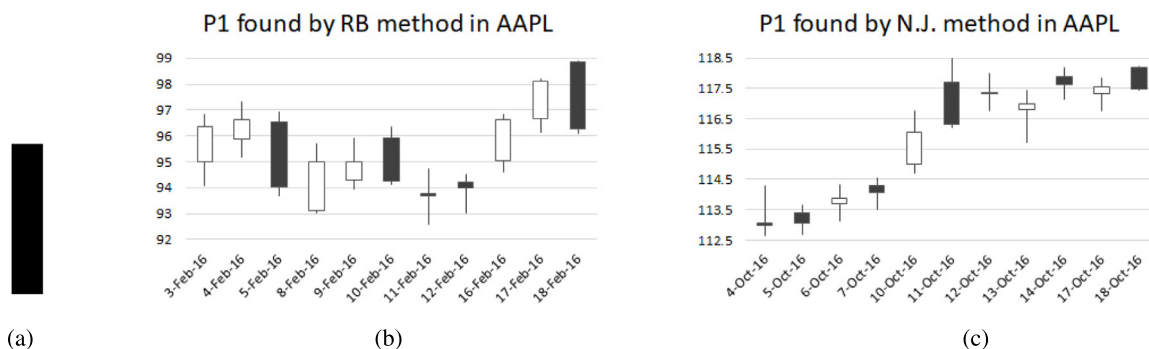


Fig. 23. Comparison of pattern *P1*. *Marubozu*, *Black* identified by N.J. and RB method in AAPL. (a) Normal template of pattern *P1* used in the experiment, (b) Pattern found by RB method, and (c) Pattern found by N.J. method.

References

- [1] N. Steve, *Japanese Candlestick Charting Techniques*, New York Institute of Finance, 1991.
- [2] T.N. Bulkowski, *Encyclopedia of Candlestick Charts*, 2008.
- [3] L.C. Tiong, D.C. Ngo, Y. Lee, Stock price prediction model using candlestick pattern feature, *Int. J. Interact. Digital Media* 1 (3) (2013).
- [4] W. Xiao, W.W.Y. Ng, M. Firth, D.S. Yeung, G.Y. Cai, J.C. Li, B.B. Sun, L-GEM based MCS aided candlestick pattern investment strategy in the ShenZhen stock market, in: *International Conference on Machine Learning and Cybernetics*, 2009, pp. 243–248.
- [5] S. Etschberger, H. Fock, C. Klein, B. Zwergel, The classification of candlestick charts: laying the foundation for further empirical research, *IET Comput. Digital Techniq.* 2 (1) (2006) 6–11.
- [6] C.H.L. Lee, Y.C. Liaw, L. Hsu, Investment decision making by using fuzzy candlestick pattern and genetic algorithm, in: *2011 IEEE International Conference on Fuzzy Systems*, 2011, pp. 2696–2701.
- [7] C.-H.L. Lee, A. Liu, Applying fuzzy candlestick pattern ontology to investment knowledge management, *J. Internet Technol.* 9 (4) (2008) 307–315.
- [8] M.L. Vásquez, F.A. González Osorio, D.F. Hernández Losada, Mining candlesticks patterns on stock series: A fuzzy logic approach, in: *Advanced Data Mining and Applications*, Springer Berlin Heidelberg, 2009, pp. 661–670.
- [9] P. Roy, D. Ramesh Kumar, D. Sanjay Sharma, Fuzzy candlestick based stock market trading system using hammer pattern, *Int. J. Res. Sci.* 1 (2014) 6–10.
- [10] C.-H.L. Lee, A. Liu, W.-S. Chen, Pattern discovery of fuzzy time series for financial prediction, *IEEE Trans. Knowl. Data Eng.* 18 (5) (2006) 613–625.
- [11] A.J. Mayne, Fuzzy sets, uncertainty, and information, *J. Oper. Res. Soc.* 41 (9) (1990) 884–886.
- [12] T.-H. Lu, Y.-M. Shiu, T.-C. Liu, Profitable candlestick trading strategies—The evidence from a new perspective, *Rev. Financ. Econ.* 21 (2) (2012) 63–68.
- [13] T.H. Lu, J. Chen, Candlestick charting in European stock markets, *Finsia J. Appl. Finance* (2) (2013) 20–25.
- [14] K. Lee, G. Jo, Expert system for predicting stock market timing using a candlestick chart, *Expert Syst. Appl.* 16 (4) (1999) 357–364.
- [15] D.E. Goldberg, *Genetic Algorithm in Search Optimization and Machine Learning*, vol. 8(7), Addison Wesley, 1989, pp. 2104–2116.
- [16] K. Martiny, Unsupervised discovery of significant candlestick patterns for forecasting security price movements, in: *KDIR 2012 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, *Int. Conf. Knowl. Discov. Inform. Retrieval*. (2012) 145–150.
- [17] T. Lv, Y. Hao, Y. Hao, C. Shen, K-line patterns' predictive power analysis using the methods of similarity match and clustering, *Math. Probl. Eng.* 2017 (30) (2017) 1–11.
- [18] Z.Y. Quan, Stock prediction by searching similar candlestick charts, in: *IEEE International Conference on Data Engineering Workshops*, 2013, pp. 322–325.
- [19] L.J. Chmielewski, M. Janowicz, A. Orłowski, Clustering algorithm based on molecular dynamics with nose-hoover thermostat. Application to Japanese candlesticks, in: *Artificial Intelligence and Soft Computing*, Springer International Publishing, 2015, pp. 330–340.
- [20] S.-c.T. Chou, H.-j. Hsu, C.-c. Yang, F. Lai, A stock selection DSS combining AI and technical analysis, *Ann. Oper. Res.* 75 (1997) 335–353.
- [21] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, 1992, pp. 956–957.
- [22] W.W.Y. Ng, X.L. Liang, P.P.K. Chan, D.S. Yeung, Stock investment decision support for Hong Kong market using RBFNN based candlestick models, in: *2011 International Conference on Machine Learning and Cybernetics*, vol. 2, 2011, pp. 538–543.
- [23] X. Lin, Z. Yang, Y. Song, Intelligent stock trading system based on improved technical analysis and Echo State Network, *Expert Syst. Appl.* 38 (9) (2011) 11347–11354.
- [24] G.J. Deboeck, *Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic financial Markets*, John Wiley & Sons, Inc., 1994, pp. 54–61.
- [25] K. Kamijo, T. Tanigawa, Stock price pattern recognition—a recurrent neural network approach, in: *1990 IJCNN International Joint Conference on Neural Networks*, vol. 1, 1990, pp. 215–221.
- [26] ChartNexus XPertTrader, http://www.chartnexus.com/software/xpt_content.php, (Accessed 15 April 2018).
- [27] Investor/RT, <https://www.linnsoft.com/techind/candlestick-pattern-recognition-cpr>, (Accessed 15 April 2018).
- [28] TALib: Technical Analysis Library, https://www.ta-lib.org/hdr_dev.html, (Accessed 03 March 2019).
- [29] Candlesticks version 1.0.0.0 (7.79 KB) by Nate Jensen, <https://www2.mathworks.cn/matlabcentral/fileexchange/33782-candlesticks>, (Accessed 03 March 2019).
- [30] A. Arratia, *Computational Finance: An Introductory Course with R*, Springer, 2014.