Available online at www.sciencedirect.com
SciVerse ScienceDirect
ELSEVIER
Systems Engineering Procedia 2 (2011) 182 – 190
Procedia
Systems Engineering

Reversal Pattern Discovery in Financial Time Series Based on Fuzzy Candlestick Lines

Qiujun Lan[a]*, Dan Zhang[a1], Longling Xiong[b]

[a]*Hunan University, Hunan, Changsha 410082, PR China*

[b]Southwest University, Chongqing 400716, PR China

Abstract

This paper provides a new technical analysis method for finding reversal points of stock price. The proposed method applies the Fuzzy logic theory to the Japanese candlestick theory by converting the open, close, high and low prices into fuzzy candlestick chart. Using the model, we can define and interpret the "symptom sequence" before the reversal point, and then identify the reversal patterns of the candlestick chart. Transaction data of non- ST Shares in SSE A Share, SSE B Share, Shenzhen A share and Shenzhen B share are analyzed to validate the proposed model, respectively. Results demonstrate that reversal patterns are convincingly identified by the model, and basically the model could be applied to establish a stock price reversal early warning system in the financial market.

Keywords: Candlestick chart, reversal point, Fuzzy logic, pattern recognition

1. Introduction

To date, financial market, which is an essential part of social economy, has highly developed along with the rapid development of modern society and economy. As a complex system, financial market is not only an interactive system involving mass of participants, but also affected by many other factors, such as politics, cultures and economy. It usually consists of highly complex operation rules and significant temporal variations. However, financial market is not a random and unpredictable system. Much literature found that financial market explicitly shows characteristics of nonlinear systems [1-4]. This motivates a number of scholars to apply various methods and techniques of economic statistical analysis, including artificial neural network (ANN) [5,6], fuzzy neural [7,8], genetic algorithm (GA), support vector machine (SVM) [9], Naïve Bayes Models [10], and Fuzzy Time Series [11-15] for financial market investigations.

Generally speaking, these existing methods and techniques rely upon the historical information base trend prediction process. To a certain extent, relevant analysis tools are useful and applicable in financial markets. Unfortunately, these tools are generally difficult accessed for most investors due to following reasons.

* Corresponding author. Tel.: +86-731-88910146
E-mail address: zd1987xll@sina.com

2211-3819 © 2011 Published by Elsevier B.V. Open access under CC BY-NC-ND license.
doi:10.1016/j.sepro.2011.10.021

First, modern technical analysis mainly relies on mathematical models and statistical methods. The predictive models involve numerous and complex parameters, which are incomprehensible for many investors. Second, traditional time series method and its variations prefer to focus on close price and turning points rather than other informational variables such as open, high and low prices. It is hard to uncover patterns of stock price fluctuation accurately and completely.

Last, computer predicting systems based on traditional technical analysis devote to predict the future price and trend of the stock, whereas for investors what they really concern is the turning points of the stock price.

In summary, although being pervasive in negotiable securities market, existing technical analyses can hardly supply the demands of investors. For instance, Das et al. (1998) applied subsequence time series clustering to financial data, but failed to identify some special and meaningful patterns [16-18]. Chen (2005) improved the algorithm developed by Das et al., but barely contributed to normalize and symbolize the financial data [19]. Literature [20] symbolized financial data into one-dimension time series, and applied dynamic programming to analyze the time series. However, the identified patterns in the literature are limited by the dimensions of the symbolized time series. New methods and techniques are of great urgent in the financial field to combine multiple technologies, such as data mining, computing and artificial intelligence, and impose novel knowledge to establish a better intuitionistic financial technical analysis model for reversal point finding and stock price prediction.

In this article, we introduce and apply a workable and effective technical analysis framework– the Candlestick chart – in the stock market to help us represent and interpret financial time series. The proposed financial prediction model can be easily accepted and handled by investors. The model emphasizes on finding reversal point, defining the candlestick lines before the reversal point as "symptom sequence", symbolizing the accurate time series into fuzzy candlestick line chart based on fuzzy logic theory, and traversal patterns mining and validation in real financial markets. The reminder of this paper is organized as follows. Section 2 briefly introduces the candlestick chart theory and the fuzzy logic theory. Section 3 discusses the definition of symptom sequence. Section 4 emphasizes on the normalization and symbolization of symptom sequence. Section 5 introduces the experiment and analyzes the results. Section 6 draws a conclusion of the paper.

2. Methodology

2.1. Japanese Candlestick theory

Candlestick chart is initially developed in the 16th century by Japanese rice trader of financial instruments. It is a combination of a line-chart and a bar-chart, in that each bar represents the range of price movement over a given time interval. It is most often used in technical analysis of equity and currency price patterns.

Candlestick is formed with the help of open, high, low and close price of the day (Figure 2.1). If the open price is above the close price then a filled candlestick is drawn. Normally, black colour is used for filling the candle. If the close price is above the open price, then a hollow candlestick (normally displayed in white with black border) is drawn. The filled or the hollow portion of the candle is known as body or real body which can be long, normal or short with proportion to the line above or below it. The lines (long or short) above and below the body or real body represent the high or low price range and these lines are known as shadows, tails or wicks. For the particular day, the highest price is declared by the top of the upper shadow and the lowest price is marked by the bottom of the lower tail. Note that the body may or may not have shadows, tails or wicks.

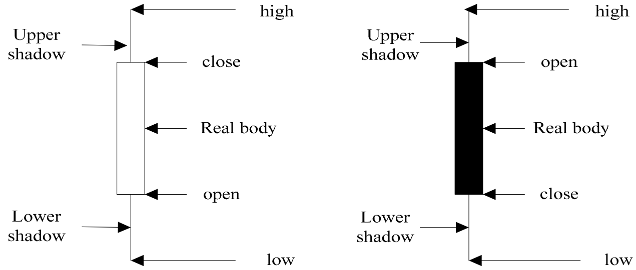

Figure 2.1 The basic candlestick

Candlestick charts are a visual aid for decision making in stock, commodity, and options trading. For example, when the bar is white and high relative to other time periods, it means buyers are very bullish. The opposite is true for a black bar. An experienced investor can easily understand the candlestick chart and make the investment decision. For instance, “shooting star” candlestick patterns usually indicate that the upward trend will be reversed. The candlestick chart theory perfectly transfers the numerical calculation of the financial data into meaningful pattern recognition. In other words, this theory provides a powerful forecasting method for those investors who prefer pattern recognition rather than real value calculation. However, the understanding and identification of candlestick rules are determined by the personal experiences of individual investor. Investors can obtain different information from the same candlestick chart. Due to its vagueness and impreciseness, effective usage of the candlestick patterns requires many years of investment experiences. To solve these problems, we apply fuzzy theory to transfer the time series data into symbolic representation and refer to the reversal pattern as fuzzy reversal pattern. Detail information about the fuzzy reversal pattern is included in Section 4.

2.2. Fuzzy logic theory

Fuzzy Set Theory was formalised by Professor Lofti Zadeh at the University of California in 1965. Fuzzy logic is derived from fuzzy set theory, and is the superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth-truth values between "completely true" and "completely false". As its name suggests, it is the logic underlying modes of reasoning which are approximate rather than exact. In contrast with traditional logic theory, where binary sets have two-valued logic: true or false, fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. The importance of fuzzy logic derives from the fact that most modes of human reasoning and especially common sense reasoning are approximate in nature. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Furthermore, when linguistic variables are used, these degrees may be managed by specific functions.

Definition of Fuzzy set：

A fuzzy set A, defined in the universal space X, is a function defined in X which assumes values in the range [0, 1].Fuzzy set A is written as a set of pairs {x, A(x)} as

$$A = \{\{x, A(x)\}\}, \text{ x in the set X}$$

Where x is an element of the universal space X, and A(x) is the value of the function A for this element. The value A(x) is the membership grade of the element x in a fuzzy set A.

3. Symptom Sequence Finding

The symptom sequence is determined by the reversal point. Reversal is a stock-market term, which refers to the stock price runs in a contrary direction of the original trend, divided into upward and downward. In this paper, the situation that the share price declines continuously in N days and rises continuously in M days later is defined as upward reversal , on the contrary, the case that share price rises continuously in N days and declines continuously in M days later is defined as downward reversal.

Before giving the definition of upward reversal, we have to make the conceptions of continuous rise and continuous decline clearly. Consecutive rise(or consecutive decline) does not require all candlestick lines in N days must be green(or red), actually, it's difficult to meet this condition(especially N>=5).In most cases, even under the trend of booms and busts, great part of the stock will be at an "one-or-two-days" adjustment or at range bound stages. So, taking into account the actual situation, we relax restriction on continuous rising and falling: in N days, the increase or decrease of stock (index), which falls within the prescribed limits, is defined as the consecutive rising or falling.

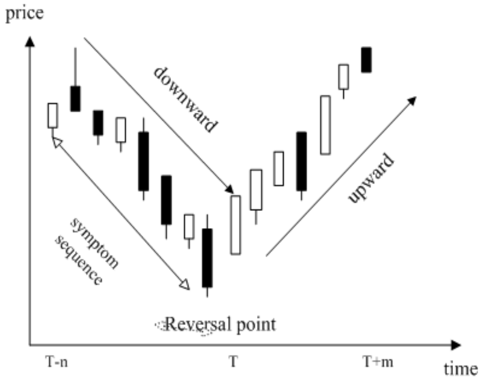

Figure 4.1

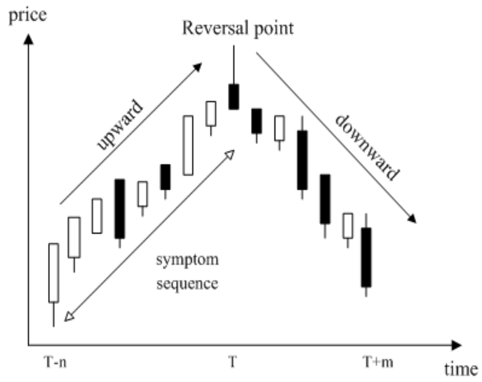

Figure 4.2

In the paper, we refer to that the stock price declines continuously in N days and rises continuously in M days later as (as shown in Figure 4.1), which should satisfy the two formulas below:

$$\frac{close_{T-n} - close_T}{close_{T-n}} \times 100 \geq a; \tag{1}$$

$$\frac{close_{T+m} - close_T}{close_T} \times 100 \geq b; \tag{2}$$

In the same way, we call that the stock price rises continuously in N days and declines continuously in M days later downward reversal(as shown in Figure 4.2), also satisfying the following two conditions:

$$\frac{close_T - close_{T-n}}{close_{T-n}} \times 100 \geq a; \tag{3}$$

$$\frac{close_T - close_{T+m}}{close_T} \times 100 \geq b; \tag{4}$$

$close_T$, $close_{T-n}$, $close_{T+m}$ are the closing prices at T,T-n, T+m; a and b are positive.

By Figure 4.1 and Figure 4.2, we can directly see that T is the reversal point in reverse pattern. In practical applications, there may be several continuous reverse points. That's to say, T, T+1, and T+2 are content with formula (1) and formula (2). In this situation, we define the first point (T) as the reverse point.

The candlestick lines between T-n and T compose a sequence, named symptom sequence of reverse pattern. In the article, we will find all the symptom sequences of each reversal point, and with the way of data mining, we can get the reversal patterns by calculating the fuzzy eigenvalues of the symptom sequence. Classification Algorithm is a important algorithm in data mining, which has been applied in a broader sense. The classical decision tree category methods (such as: ID3, C4.5) draw more and more attention because of its fast modelling. Alternating decision tree (ADTree) is a decision tree algorithm based on boosting．ADTree has a wide range of applications，because it can attain a more accurate hypothesis than other general decision tree algorithms and output confidence—rated predictions. So, in our article, we choose ADTree algorithm to dig the reversal patterns in fuzzy symptom sequences**.**

4. Fuzzy representation of symptom sequence

4.1. Fuzzy representation of candlestick line

In a candlestick chart, the lengths of the shadow and the body play an important role in the recognition of the candlestick pattern and determine the efficiency of the candlestick pattern. The description of a candlestick line is inaccurate or even vague. For example, the line is long, medium or short, which has no apparent value to define the length of the body and shadow line.

In our article, we use the method which has been proposed by Lee C-H and Liu A in TAIEX forecasting [11] to transfer the time series data into symbolic representation. Four fuzzy linguistic variables are used in fuzzy set to describe the length of shadow line and the body: equal, short, middle, and long. The concerned branch of a stock, in Shanghai and Shenzhen Securities Exchanges, can not be over the close price of previous day by 10%. So the universe of fuzzy set which describes the length of entity and line is [0, 0.2], we can get [0, 2] by normalization. In Figure 4, the footnote of x axis is the actual length of the box or the shadow, and the unit of x-axis is a percentage.

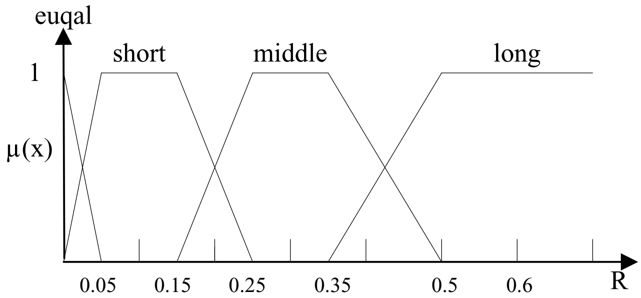

Figure 4.1 the membership function of the length of the body and the shadow

The input values of membership functions can be calculated by the following equation:

$$\begin{aligned}
\text{Lupper} &= \left(\left[\text{high} - \max\left(\text{open}, \text{close}\right)\right] / \text{open}\right) \times 10 \\
\text{Llower} &= \left(\left[\min\left(\text{open}, \text{close}\right) - \text{low}\right] / \text{open}\right) \times 10 \\
\text{Lbody} &= \left(\left[\max\left(\text{open}, \text{close}\right) - \min\left(\text{open}, \text{close}\right)\right] / \text{open}\right) \times 10
\end{aligned} \tag{5}$$

The character “L” indicates the length of the upper shadow, lower shadow and body. Some subscripts such as open, close, high, low represent the price in some period of time that you are interested. The long fuzzy set is defined by the following left linear membership function, and the parameters (a, b) are equal to (0.35, 0.5).

$$left\_linear\left(x : a, b\right) = \begin{cases} 0 & x < a \\ \left(x-a\right)/\left(b-a\right) & a \le x < b \\ 1 & x > b \end{cases} \tag{6}$$

The membership function of short and middle is a trapezoid function, the following is the expression:

$$trapezoid(x : a, b, c, d) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \le x < b \\ 1 & b \le x < c \\ (d-x)/(d-c) & c \le x < d \\ 0 & x \ge d \end{cases} \tag{7}$$

The four parameters (a, b, c, d) used to describe the linguistic variables short and middle in the function are（0，0.05，0.15，0.25），（0.15，0.25，0.35，0.5）. A right linear membership function is used to model the equal fuzzy set, described by the following formula. In this paper the parameters (a, b) are equal to (0, 0.05).

$$right\_linear(x:a,b)=\begin{cases}1 & x<a\\ (b-x)/(b-a) & a\le x<b\\ 0 & x>b\end{cases} \tag{8}$$

After establish the fuzzy set, we symbolic the candlestick chart with subordinate function maximum method. And we define the length of the body or shadow, of which the membership function value is maximum, to the fuzzy set.

The colour of the body is also an important indicator of candlestick lines. Three parameters can be simply defined: green, red, and cross. When the open price is equal to the close price, cross is used to define this kind of situation, for there is a special meaning in the pattern of candlestick. And the body length is 0. The shape is a horizontal line. The definition of the body colour can be expressed by the following formula:

$$\begin{array}{ll}\text{If open-close} > 0 & \text{“green”}\\ \text{If open-close} < 0 & \text{“red”}\\ \text{If open-close} = 0 & \text{“cross”}\end{array} \tag{9}$$

4.2. Fuzzy representation of the candlestick lines relationship

Only defining the length of the body, shadow to model the candlestick line is not enough, the model of which can only determine the shape of a candlestick line. The relationship between two adjacent candlestick lines should be taken into consideration to determine a correct position for the candlestick as well.

Once two characteristics are defined, they can be used to establish the relational model between the candlestick lines: the open style and the close style. The open style and the close style models are built by the related position between the open price and close price of two adjacent candlestick lines.

Five linguistic variables represent the styles of opening: open_low, open_equal_low, open_equal, open_equal_high and open_high. 5 linguistic variables represent the styles of closing: close_low，close_equal_low，close_equal，close_equal_high and close_high. The function used to represent the open_low and close_low is the right-linear function in(10). At the same manner, the left-linear function in (8) is used to represent the open_high and close_high. Other fuzzy sets can be described by a triangle function:

$$\text{triangle}（x:a,b,c）=\begin{cases}0 & x<a;\\ (x-a)/(b-a) & a\le x\le b;\\ (c-x)/(c-b) & b<x\le c;\\ 0 & x>c;\end{cases} \tag{10}$$

The parameters of linguistic variables for open style and close style are determined by the prices of previous candlestick line. For example, if the open price is equal to the lower one between previous candlestick line's open price and close price, the open style is open_equal_low. And if the close price is the same with the higher one between previous candlestick line's open price and close price, the close style is close_equal_high.

5. Empirical studies and analysis

Take upward reversal for instance, the chapter will verify the validity of the method with experimental date from Shanghai A-share, B-share, Shenzhen A-share and B-share. The data for experimental research is from CSMAR (www.gtarsc.com) and calculated by MATLAB7.0, including 100 non-ST stocks taken at random from the Shanghai

A-share and Shenzhen A-share (from 1 January, 2000 to December 30, 2010), respectively. Considering the small amount of the non-ST shares, we will take all the non-ST shares from Shanghai B-share and Shenzhen B-share from 1992 to December 30, 2010. Experimental procedures look like the following illustration:

Extract the symptom sequence → The fuzzification of the symptom sequence → Mining Classification for the symptom sequence

Figure 5.1 Experimental procedures

In the practical operating，we illustrate with shanghai non-ST shares. The data is composed of the open price, high price, low price and close price of each stock in every trading day. First of all, based on the definition of reversal point in chapter 3, we find out all reversal points satisfying formula (1) and formula (2) by close price, a total of 2187, and mark it as a positive example with yes. And the point, satisfying formula (1) but not meeting formula (2), is marked as a counter example with no, a total of 2120. Of course, there may be several consecutive T (counter example) satisfying the situation as positive example, we still only record the first occurrence.

Identifying the turning points and their counter examples, we should extract the close price, open price, high price and low price of the stock in the last 5 days (reversal point included). Take point T for example, we need extract the trading data at T, T-1, T-2, T-3 and T-4. Then, we will translate the data to the K-chart entities, the fuzzy eigenvalues of the length of the lines, colour of the entities, open style and close style, according to the following table:

Table 5.1: the fuzzy eigenvalues of the symptom sequence

lin1 Lupper	lin1 Llower	…	lin5 Body_Color	lin5 Open_Style	lin5 Close_Style	class
short	short	…	green	open_equal_low	close_low	yes
equal	equal	…	green	open_equal	close_low	no
short	short	…	red	open_equal_high	close_high	no
short	short	…	red	open_equal_low	close_equal_high	no
…	…					…
short	short	…	red	open_equal	close_equal	yes
equal	short	…	green	open_equal_high	close_equal_high	yes
short	short	…	red	open_low	close_low	yes
short	short	…	green	open_equal_low	close_low	no
equal	equal	…	green	open_equal	close_low	yes
short	equal	…	green	open_equal_low	close_equal_high	yes
short	short	…	green	open_equal_low	close_equal	yes

In the next, we will divide the fuzzy eigenvalues of the symptom sequence into two sets in the ratio of 2:1. The former is training set and the latter is testing set. Then we train the training set with ADTree to achieve the optimal decision tree. At last, we predict the trend using the optimal decision tree.

The effect of the prediction can be evaluated by statistical indicators. Recall rate and precision rate are the most important concepts and evaluation in data mining. Both of them are available through Confusion Matrix. In the paper, the elements of Confusion Matrix are divided into four categories, as shown in Figure 5.1.

	Classified to be true	Classified to be not
Real	A	B
Not real	C	D

Figure 5.1

In Figure 5.1, A means the true reversal point is classified as true; B indicates that the real turning point is not convicted to be real; C refers to the point is not a reversal point but sentenced to be a real one; D implies that the point is not a turning point and adjudged not. Recall rate reflect the ratio of the accurate prediction in total reversal points. Precision rate display the ratio of the accurate prediction in all the marked reversal points. Their formulas are as follows:

$$\begin{cases} recall = \dfrac{A}{A+B}; \\ precision = \dfrac{A}{A+C}; \end{cases} \tag{11}$$

The experimental result of four share markets is shown in Table 5.2:

Table 5.2: the result of prediction

Stock market	Reversal point	Recall rate	Precision rate
Shanghai A	yes	75.6%	84.2%
	no	86.0%	78.1%
Shanghai B	yes	70.7%	73.9%
	no	74.8%	71.6%
Shenzhen A	yes	71.7%	82.8%
	no	85.4%	75.0%
Shenzhen B	yes	87.1%	76.1%
	no	70.2%	83.4%

From the experimental results, we can conclude that the recall rate and precision rate of the prediction of the reversal point in the four markets are far greater than 50%, reaching an acceptable level. So we can take it as that the model plays an early warning role of the appearance of reversal point.

6. Conclusion

In this paper, we proposed a financial time series reversal pattern model based on fuzzy candlestick lines. This model is a visual financial technical analysis model which can warn the reversal point from historical statistics. With the way of data mining, we can get the statistical properties of reversal patterns by calculating the fuzzy eigenvalues of the symptom sequence. Experiments show that this idea is feasible, and the predict result is fine. This model can guide the investors to invest in a more accurate way and to get a more satisfactory return. In future work, we will further improve the model, using more fuzzy variables to represent candlestick lines, for example, the relative position of body and shadows, in order to get better prediction results.

7. Copyright

Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities of China under Grant No. 531107040018.

References

1.Tsaih Ray, Hsu Yenshan, Lai Charles C. Forecasting S&P 500 stock index futures with a hybrid AI system[J].Decision Support Systems.2 (1998) 161-174.

2.Abu-Mostafa Yaser S. Atiya Amir F. Introduction to financial forecasting [J].Applied Intelligence.3 (1996) 205-213.

3.WEN F.H.,YANG X.G. Skewness of Return Distribution and Coeffcient of Risk Premium[J].Journal of Systems Science and Complexity.22 (2009) 360–371.

4.WEN F.H., LIU Z.F. A Copula-based Correlation Measure and Its Application in Chinese Stock Market[J].International Journal of Information Technology & Decision Making.4 (2009) 1-15.

5.E.W.Saad,D.V.Prokhorov, D.C. II Wunsch. Comparative Study of Stock Trend Prediction Using Time Delay, Recurrent and Probabilistic Neural Networks.IEEE Trans. Neural Networks..96 (1998) 1456-1470.

6.A.P.N.Refenes,W.T. Holt. Forecasting Volatility with Neural Regression: A Contribution to Model Adequacy.IEEE Trans.Neural Networks. 4 (2001) 850-864.

7.Chang Pei-Chann,Liu Chen-Hao. A TSK type fuzzy rule based system for stock price prediction[J].Expert Systems with Applications.1 (2008) 135-144.

8. Wang Y F. Predicting stock price using fuzzy grey prediction system[J].Expert Systems with Applications.1(2002) 33-38.

9.Cao L.J,F.E.H.Tay.Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting.IEEE Trans. Neural Networks.6 (2003) 1506-1518.

10.V.Dharand Chou D. A Comparison of Nonlinear Methods for Predicting Earnings Surprises and Returns.IEEE Trans. Neural Networks.4 (2001) 907-921.

11.Lee C-H L,Liu A. Pattern discovery of Fuzzy Time Series for financial prediction[J].IEEE Transactions on Knowledge and Data Engineering. 5 (2006).

12.Chen T L, Cheng C H, Teoh H J. Fuzzy time series based on Fibonacci sequence for stock price forecasting[J].Physica A.38 (2007) 377-390.

13.Singh S R. A computational method 0f forecasting based on fuzzy times series[J].Mathematic and Computers in Simulation.79 (2008) 539-554.

14.Li S T, Cheng Y C,Lin S Y. A FCM-based deterministic forecasting model for fuzzy time series[J].Computers and Mathematics with Applications.56 (2008) 3052-3063.

15.Chen S.M., Hsu C.C. A New Method to Forecast Enrollments Using Fuzzy Time Series.Applied Science and Eng.3 (2004) 234-244.

16.Cotofrei P, Stoffel K. Classification Ruls+Time=Temporal Rules[C]// proceedings of the 2002 Int'1 Conference on Computational Science.4 (2002) 572-581.

17.Das G, Lin K, Mannila H,et al. Rule discovery from time series[C]// Proceedings of the 4th Int'1 Conference on knowledge Discovery and Data Mining.8 (1998) 16-22.

18.Fu T Chung F L, et a1. Pattern discovery for stock time series using Self-Organizing Maps[C]//Workshop Notes of KDD2001 Workshop on Temporal Data Mining,.8 (2001):27-37.

19.Chen J R. Making subsequence time series clustering meaningful[C]// Proceedings of the Fifth IEEE International Conference on Data Mining USA:Washington.(.2005) 114-121.

20.Yamano T,Sato K,Kaizoji T,et a1. Symbolic analysis of indicator time series by quantitative sequence alignment[J].Computational Statistics and Data Analysis.53 (2008) 486-495.