

Investment Decision Making by Using Fuzzy Candlestick Pattern and Genetic Algorithm

Chiung-Hon Leon Lee, Yi-Ching Liaw, and Lindroos Hsu

Department of Computer Science and Information Engineering

Nanhua University

Dalin, Chia-Yi, 620, Taiwan

chlee@mail.nhu.edu.tw, ycliaw@mail.nhu.edu.tw, somewherebeing@gmail.com

Abstract—This paper proposed an approach to extract fuzzy candlestick patterns from financial time series and select a set of patterns for investment decision making. The candlestick chart in stock market is a widely used technical analysis model. The investor observes the candlestick chart and makes investment decisions by identifying patterns in the chart. We use fuzzy linguistic variables to model candlestick chart and extract patterns from the chart. A Genetic algorithm based approach is used to select a set of extracted pattern as the background knowledge in the system for investment decision making. The advantage of the proposed approach is the investment knowledge is comprehensible, editable, and visible. The user can set different range of historical financial time series to extract and select different set of patterns. The experimental results shows that the investment decisions based on selected fuzzy patterns have better investment performance than using original non-fuzzy patterns.

Keywords—component; fuzzy candlestick pattern; time series; genetic algorithm

I. INTRODUCTION

How to modeling and implementing domain knowledge is an important part when design and implement an intelligent decision supporting system. For example, if an intelligent investment expert system has rich background knowledge from human expert, the system can give more precise suggestions to the user [1].

There are a lot of fundamental and technical analysis concepts and techniques have been established in financial investment domain such as Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Fast and Slow Stochastic (KD-line), etc. These technical analysis tools are the basis when design some investment decision supporting systems [2].

However, the set of data used to calculate RSI, MACD, and KD indicators is the closing price in a time point. There is rich information which exists in the financial time series database, but most of the traditional approaches only scratch the surface of the wealth of knowledge buried in the data. Many financial time series prediction approaches only use daily closing price as raw data to construct the forecasting model.

Figure 1 shows three general ways to represent the stock trading price during a time period. Figure 1(a) indicates a

single closing price. Figure 1(b) represents the bar line which contains richer information than 1(a). The data required to produce a standard bar chart consists of the open, high, low, and close prices for the time period under study. A bar chart consists of vertical lines representing the high to low range in prices for that trading time period. The high price refers to the highest price that the issue traded during that trading time period. Likewise, the low price refers to the lowest price traded that trading time period. Figure 1(c) illustrates the candlestick line which is similar to the bar line but using a box to makes up the difference between the open and close price.

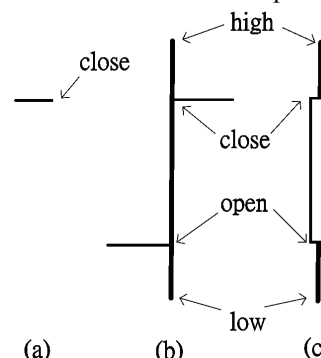


Figure 1. Different representation of financial data.

The candlestick chart in stock market is a widely used empirical model for investment decision making [3]. The investor observes the candlestick chart and makes investment decisions by identifying patterns in the chart. Figure 2 shows two examples of the daily candlestick chart in a stock market. Daily opening, closing, highest, and lowest prices are recorded in the candlestick lines for days d1, d2, d3, and d4. On d1, the stock opens in a higher price, but close at a lower price. On the next day d2, the opening price is lower than previous closing price, but closes at the highest price and is much higher than the lowest price. This situation might be interpreted by an experienced investor as the candlestick line on the day d1 reflecting a downtrend of the stock price, because there are many investors want to sell their stock and make the closing price much lower than the opening price. However, the downtrend might reverse itself on the day d2, because there are investors to buy the stock in the trading period that makes

the price close at the highest price and much higher than the lowest price. In other word, the candlestick lines at d1 and d2 might represent a bouncing back. At d3, the closing price is higher than the opening price, and at d4, the opening price is much higher than the previous closing price, but closes at lowest price. The lines at d3 and d4 might be interpreted by the investor that the trend is returned, because the uptrend is broken at d4. It is obvious that the interpretation between the lines at d1, d2 and d3, d4 is very different, but the closing prices of d1, d2 and d3, d4 are the same.

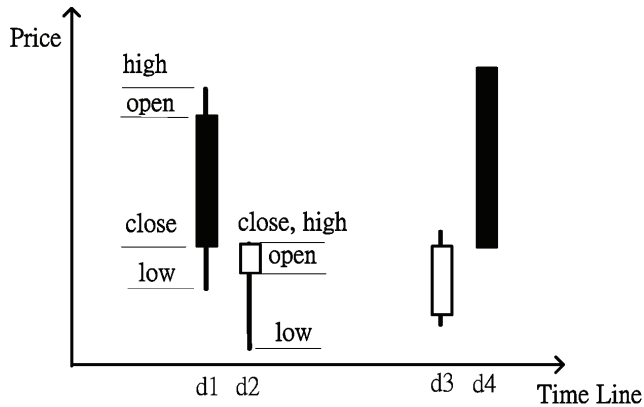


Figure 2. A candlestick chart.

Candlestick patterns which exist in the candlestick chart are empirical models of investment decision. The experienced investors can make their investment decision based on the observation of some specific candlestick patterns such as Bullish Engulfing, Hanging Man, Hammer, etc which are defined by the pioneer investors. However, to a human investor, to identify an effective pattern from a lot of imprecise and vague candlestick patterns needs years of investment experience and expertise. In addition, to retrieve the candlestick patterns from a large amount of financial trading data is time consuming.

In this paper, we propose a fuzzy candlestick pattern based financial decision supporting system to acquire the stock trading data in the Taiwan stock market (TAIEX) [4] from the Internet, provide an interface for the users to edit their own fuzzy candlestick patterns which are represented by fuzzy linguistic variables, extract fuzzy candlestick patterns from financial time series, and use a Genetic Algorithm (GA) based approach to select a set of extracted pattern as the background knowledge in the system for investment decision making. The advantage of the proposed approach is the investment knowledge is comprehensible, editable, and visible. The user can use different range of historical financial time series as basis for candlestick patterns construction.

This paper is organized as follows. In Section 2, the fuzzy-based candlestick pattern representation method is introduced. Section 3 describes the Genetic algorithm (GA) based fuzzy candlestick pattern extraction approach and the experimental results. Finally, Section 4 concludes this paper.

II. MODELING CANDLESTICK CHART

By the trading experience, the investor tries to identify the candlestick patterns to help themselves to make the investment decisions such as to buy, sell, or hold the stock. There are many existing defined candlestick patterns which are widely used by the investors. Figure 3(a) shows a hierarchy to represent the relationships among the candlestick pattern, candlestick line, and the prices. A visual candlestick pattern which called Bearish Engulfing is demonstrated in Figure 3(b).

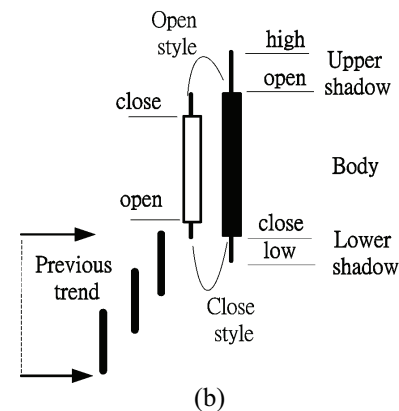
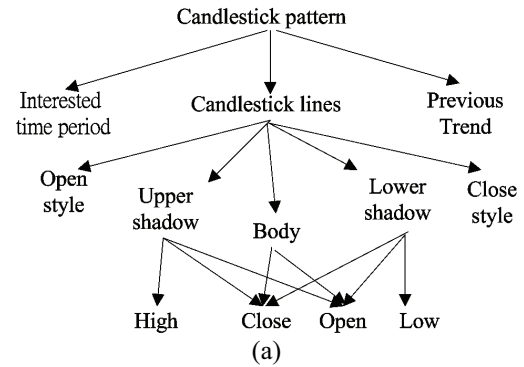


Figure 3. An example of the candlestick pattern.

The advantage of candlestick chart to investors is that the candlestick chart is visual, and a reversal or continuation candlestick pattern can be easily identified by an experienced investor. However, identifying the candlestick pattern from a large amount of trading data is time consuming, and there are no crisp and standard definitions to the candlestick patterns. The imprecise and vague definitions of the candlestick patterns make the automated searching, mining, and processing the candlestick patterns with computer software difficult. In this paper, we adapted the approach proposed in [5] to solve this problem.

A. Modeling Candlestick Patterns

A candlestick chart is composed by a series of continual candlestick lines. The lines consist of the open, high, low, and close prices in a time period under study. The time period can be of any duration, such as five minute, fifty minute, one hour, one day, one week, or one month. The first trading price in the interested trading time period is called open price; the last

trading price is called close price; the highest price is called high price, and the lowest price is called low price. These four prices compose a candlestick line. A candlestick pattern is composed by several serial candlestick lines.

The box that makes up the difference between the open and close is called the real body of a candlestick line. The height of the body is the range between a trading day's open price and the day's close price. When the body is black, it means that the closing price was lower than the opening price. When the closing price is higher than the opening, the body is white.

Four fuzzy linguistic variables EQUAL, SHORT, MIDDLE, and LONG are defined to indicate the fuzzy sets of the shadows and body length. Figure 4 shows the fuzzy membership function $\mu(x)$ of the linguistic variables.

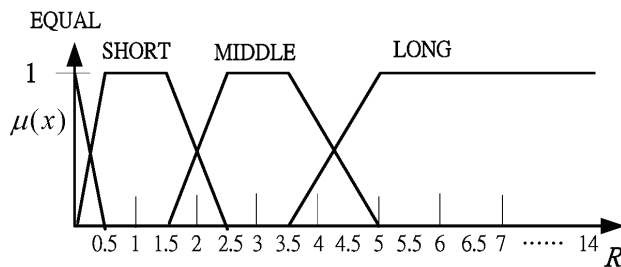


Figure 4. The membership functions of the length of the body and shadows.

The time period used in this paper is one day and the range of body and shadow length are set to 0 to 14 percent of the fluctuation of stock price, because the varying percentage of the stock prices are limited to 14 percent in TAIEX. Although we limit the fluctuation of body and shadow length in 14 percent in this paper, it is up to the system designer to change the range of the lengths to fit the needs of other applications.

In Figure 4, the unit of X axis is the percentage of price change in a stock and it indicates the lengths of body or shadows. The crisp input value of the membership function can be calculated by the following equations.

$$L_{upper} = [high - \max(open, close)] / open$$

$$L_{lower} = [\min(open, close) - low] / open$$

$$L_{body} = [\max(open, close) - \min(open, close)] / open$$

(1)

The character “L” of the equation indicates the length of the upper shadow, lower shadow, and body. The terms of open, close, high, and low are the prices in an interested time period. The function of **max** is used to calculate the greater value between the open price and the close price while and the function **min** is for the smaller value between them.

A right linear membership function is used to model the EQUAL fuzzy set and is defined by the following formula. The parameters (a, b) are equal to (0, 0.5) in this paper.

$$right_linear(x : a, b) = \begin{cases} 1 & x < a \\ (b - x) / (b - a) & a \leq x \leq b \\ 0 & x > b \end{cases} \quad (2)$$

The LONG fuzzy set is defined by the following left linear membership function. The parameters (a, b) are equal to (3.5, 5).

$$left_linear(x : a, b) = \begin{cases} 0 & x < a \\ (x - a) / (b - a) & a \leq x \leq b \\ 1 & x > b \end{cases} \quad (3)$$

The membership function of SHORT and MIDDLE is a trapezoid function and the following formula is used.

$$trapezoid(x : a, b, c, d) = \begin{cases} 0 & x < a \\ (x - a) / (b - a) & a \leq x < b \\ 1 & b \leq x < c \\ (d - x) / (d - c) & c \leq x < d \\ 0 & x \geq d \end{cases} \quad (4)$$

Four parameters (a, b, c, d) of this function to describe the linguistic variables SHORT and MIDDLE are (0, 0.5, 1.5, 2.5) and (1, 2.5, 3.5, 5).

The body color is also an import feature of a candlestick line and can be simply defined by three terms BLACK, WHITE, and CROSS. The situation where open price equals close price has specific meaning in the candlestick pattern, so a “CROSS” term is defined to describe this situation. In this case, the height of the body is 0, and the shape is represented with a horizontal bar. The definition of body color is defined as follows.

If open - close > 0 then the body color is BLACK.

If open - close < 0 then the body color is WHITE. (5)

If open - close = 0 then the body color is CROSS.

Figure 5 shows the membership function of the linguistic variables of the open style and close style. The candlestick line in the bottom of Figure 5 is the candlestick line of previous trading time. The unit of X axis is the trading prices of previous day and the unit of Y axis is the possibility values of the membership function.

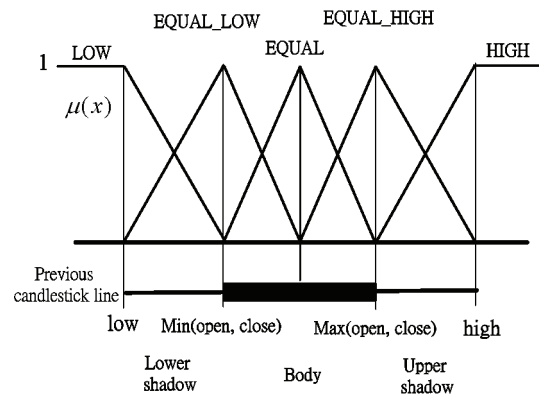


Figure 5. The membership function of the open and close style.

Five linguistic variables are defined to represent the open style relationships: OPEN LOW, OPEN EQUAL_LOW, OPEN EQUAL, OPEN EQUAL_HIGH, and OPEN HIGH, and five linguistic variables are defined to represent the close style relationships: CLOSE LOW, CLOSE EQUAL_LOW, CLOSE EQUAL, CLOSE EQUAL_HIGH, and CLOSE HIGH.

The function used to represent OPEN LOW and CLOSE LOW is the right linear function in (2), and (3) is also used to represent the fuzzy sets of OPEN HIGH and CLOSE HIGH. The other fuzzy sets are described by the triangle function in (6).

$$\text{triangle}(x : a, b, c) = \begin{cases} 0 & x < a \\ (x - a)/(b - a) & a \leq x \leq b \\ (c - x)/(c - b) & b < x \leq c \\ 0 & x > c \end{cases} \quad (6)$$

The parameters for the linguistic variables of open and close style are determined by the prices of previous candlestick line. For example, if the open price in a interested time period is equal to the price of $\min(\text{open}, \text{close})$, then the open style is OPEN EQUAL_LOW and if the close price is equal to the price $\max(\text{open}, \text{close})$, then the close style is CLOSE EQUAL_HIGH.

B. Modeling the Investment Knowledge

The description of a fuzzy candlestick pattern stored in the system consists of two parts: the pattern description part and the pattern information part. The description part is composed by the name of the candlestick pattern, a previous trend of the pattern, candle lines which composed the pattern, and the time period information. The information part records the candlestick pattern related information such as the pattern identification rules and the meaning of the pattern, etc. The information part is optional and is defined by the investors if the investor is interested in the pattern. If a candlestick pattern is described with the description information, it becomes more comprehensible to the other investors.

Table 1 shows a fuzzy candlestick pattern example which demonstrates a possible way to represent the Bearish Engulfing candlestick pattern and other candlestick patterns can be defined in the same way. The previous trend defined here is a crisp rule such as “down 15% in recent 10 days” to represent a downtrend or “up 15% in recent 10 days” to represent an uptrend. It should be noted that, for implementation convenience, we define the candlestick line which appear at the day t line 0, and the line appear in previous trading day $t-1$ line 1.

Assume that the pattern appears at day t , the close price of day t is denoted $\text{close}(t)$, the highest close price in recent k days is denoted $\text{highest}(k)$, and the lowest close price in recent k days is denoted $\text{lowest}(k)$. The rules to determine the downtrend and uptrend can be defined as follows.

If $\frac{\text{highest}(k) - \text{close}(t)}{\text{close}(t)} > p$ then the stock price is in a downtrend.

If $\frac{\text{close}(t) - \text{lowest}(k)}{\text{close}(t)} > q$ then the stock price is in an uptrend,

where p and q are the thresholds to determine the trend of stock price is in uptrend or downtrend, and the unit of p and q is percentage of the fluctuation of stock prices.

Table 1

An example of the fuzzy candlestick pattern.

Pattern description part	Pattern information part
Pattern name: Bearish Engulfing	Confirmation suggest: Suggest
Previous trend: Uptrend	Confirmation information:
Candle lines:	The open price after the pattern should not be higher than the open price of candle line 0.
Candle line 0:	
Open style: OPEN HIGH	
Close style: CLOSE LOW	
Upper shadow: null	
Body: ABOVE MIDDLE	
Body color: BLACK	
Lower shadow: null	
Candle line 1:	
Open style: ABOVE OPEN EQUAL_LOW	
Close style: CLOSE HIGH	
Upper shadow: null	
Body: ABOVE SHORT	
Body color: WHITE	
Lower shadow: null	
Interested time period: DAY	
	Recognition rule:
	1. A definite downtrend must be underway.
	2. The second day's body must completely engulf the prior day's body.
	3. The first day's color should reflect the trend: black for a downtrend and white ...
	Pattern explanation:
	The first day of the Engulfing pattern has a small body and the second day has a long real body. Because the second day's move.

C. Pattern recognition

Since the patterns have been defined by the investor, the defined patterns can be easily transferred into fuzzy rules. For example, the Bearish Engulfing pattern can be transferred as following fuzzy rule.

IF trend = UP_TREND
AND line0.open_style = OPEN HIGH
AND line0.close_style = CLOSE LOW
AND line0.body = ABOVE MIDDLE
AND line0.body_color = BLACK
AND line1.open_style = ABOVE OPEN EQUAL_LOW
AND line1.close_style = CLOSE HIGH
AND line1.body = ABOVE SHORT
AND line1.body_color = WHITE
THEN the pattern = BEARISH ENGULFING.

A pattern recognition rule consists of the crisp part and the fuzzy part. The crisp part includes the previous trend of the pattern and the body color. The others of the rule are the fuzzy part such as the body and shadow length and the open and close style. From observation, well arranged identification rule will reduce the pattern recognition processing time.

Comparing with the processing time of the fuzzy part, the crisp part takes less processing time. For example, the body

color includes three possibilities: BLACK, WHITE, and CROSS. For judging the value of the body color, the pattern recognition module only needs to compare the value of open price and close price. The pattern identifying time can be reduced if the judgment of the crisp part is placed before the process of the fuzzy part.

The concept of Hamming distance [6] is used to measure the similarity among fuzzy candlestick patterns. Assume that A and B represent two different fuzzy candlestick patterns which described by n linguistic variables, the similarity is defined as:

$$S(A, B) = \frac{1}{n} \sum_{i=1}^n |\mu_A(x_i) - \mu_B(x_i)|, \quad 0 < S(A, B) \leq 1. \quad (7)$$

The user can set a threshold T to determine a pattern is recognized or not. If $S(A, B) \geq T$ then the pattern is identified.

III. INVESTMENT KNOWLEDGE EXTRACTING

When using the candlestick chart, the investment decision is based on a set of candlestick patterns rather than only use a single pattern. In other word, how to select a set of patterns becomes an important problem.

Using fuzzy candlestick pattern to make investment decisions has several steps.

1. Transfer a range of financial time series into candlestick chart for pattern extraction.
2. Extracting original fuzzy candlestick patterns by using predefined heuristic rules.
3. Selecting patterns from extracted patterns by GA.
4. Using the selected patterns as the investment knowledge for decision making.

Step 1 follows the method introduced in section 2 to transfer financial time series into candlestick chart. Three heuristic rules are defined for pattern extraction in step 2.

1. Bullish pattern recognition rule: if the previous trend is downtrend and the up_index of following N days up $P\%$ then the pattern is a bullish pattern.
2. Bearish pattern recognition rule: if the previous trend is uptrend and the down_index of following N days down $P\%$ then the pattern is a bearish pattern.
3. Other patterns: don't care.

The up and down indexes are used to indicate the maximum and minimum variation of the stock price after a specific pattern has happened. The calculation of up and down indexes is similar to previous trend of the candlestick pattern. Assume that the pattern appears on day t , the close price of day t is denoted $close(t)$, the highest close price in recent k days is denoted as $highest(k)$, and the lowest close price in recent k days is denoted as $lowest(k)$. The up and down index are defined in (8).

$$up_index = \frac{highest(k) - close(t)}{close(t)} \times 100$$

$$down_index = \frac{close(t) - lowest(k)}{close(t)} \times 100 \quad (8)$$

The $total_up_down$ index is used to indicate the possibility investment result when using a specified pattern in an assigned time period, and the $mean_fluctuation$ index is used to reflect the mean fluctuation of the stock prices when the pattern appears. To a reversal pattern, these two parameters should as large as possible. The total up times and total down times are also calculated. If the up index is higher than the down index, the total up times are increased by 1, and if the up index is lower than the down index, the total down times are increased by 1. The total up and down index and mean fluctuation are calculated as follows.

$$total_up_down = \sum (up_index - down_index)$$

$$mean_fluctuation = \frac{1}{T} \sum_{i=1}^T (up_index + down_index) \quad (9)$$

$$T = total_up_times + total_down_times$$

Figure 6 shows a reversal pattern called "Hanging Man" testing results. The selected stock number is 2409 (AUO), and the validation range is 1000 trading days before 2004-12-10. The maximum and minimum close prices from trading day t to $t - 5$ are selected to calculate the up and down index. In the selected range of trading days, the pattern appeared 10 times. In other words, the frequency of this pattern is 0.01. The $total_up_down$ index is equal to about 20 percent, and $mean_fluctuation$ index is equal to about 11 percent.

Stock No.	Pattern Name	Day	Up Index	Down Index
2409	Bearish Hanging Man	day:2003-10-22	up:0.0	down:3.8076181
2409	Bearish Hanging Man	day:2002-11-04	up:7.4999967	down:2.5000014
2409	Bearish Hanging Man	day:2002-03-06	up:0.0	down:14.876034
2409	Bearish Hanging Man	day:2001-11-14	up:21.105532	down:0.0
2409	Bearish Hanging Man	day:2001-11-07	up:8.152174	down:1.0869503
2409	Bearish Hanging Man	day:2001-10-12	up:7.5268836	down:8.960573
2409	Bearish Hanging Man	day:2001-04-19	up:6.7885175	down:1.8276783
2409	Bearish Hanging Man	day:2001-03-23	up:12.462001	down:0.0
2409	Bearish Hanging Man	day:2001-03-07	up:0.0	down:9.354837
2409	Bearish Hanging Man	day:2001-01-04	up:3.8135576	down:4.2372885
total up times:5			total down times:5	mean pattern times:0.01
total up down:20.697681			average up down:2.0697682	mean fluctuation:11.399965

Figure 6. A pattern validation results.

In step 3, we use the extracted patterns as gene to consist a chromosome for GA. The relationship of pattern rules and chromosomes is shown in Figure 7.

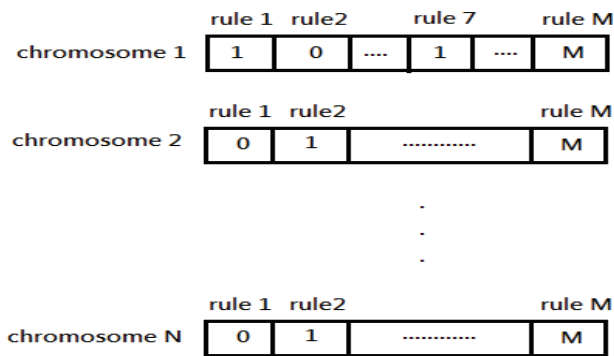


Figure 7: The chromosome used in GA.

If the pattern extracted in step 2 is 576 the M will be 576. In chromosome, bit 1 indicate the rule is used for investment decision making, bit 0 indicate this rule do not be used. In the beginning of GA, we set bits in chromosome randomly.

The fitness function is define as:

$$f(x_n) = \frac{x + m}{\sum_{i=1}^N (x_i + m)} \quad (10)$$

Where x is the earned or lost point when using patterns indicated in chromosome and m is the stop value if earned point reached this value. We use Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) from 2001 to 2005 for patterns selection. The experimental results by running GA are shown in Figure 8.

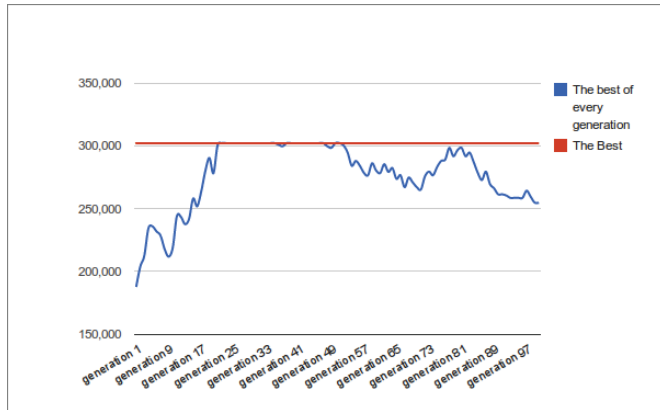


Figure 8: A comparison of different GA generation.

Finally we compared the performance by using different pattern sets. First results use extracted fuzzy candlestick pattern as background knowledge, second use user defined patterns, and third use GA to get better pattern set. The experimental results in Table 2 show fuzzy candlestick patterns with GA have best performance.

Table 2: a comparison among using different pattern set.

	Time series for pattern extraction	Evaluation Time series	Results
Extracted patterns	2001~2005	2006~2009	5027
Predefined patterns	No	2006~2009	6114
Selected patterns	2001~2005	2006~2009	8452

IV. CONCLUSION

In this paper, we proposed a knowledge based method to represent the financial time series. The variance of the stock price is represented in fuzzy candlestick patterns which make the imprecise and vague investment knowledge comprehensible, computable, visual, and editable. We also propose a GA based approach for patterns selection. The system can give investment suggestion to the investor by using these patterns as background knowledge.

The fuzzy candlestick patterns carry rich information and also can be used to increase the efficiency of the data mining, machine learning, and pattern recognition models. Pattern construction and recognition approach is introduced and implemented in a system prototype to illustrate the usage of the fuzzy candlestick patterns. Moreover, investors can save and share their investment experience. By reusing and modifying the stored candlestick pattern information, the investor can also increase the efficiency of their investing strategies. We believe that the fuzzy candlestick pattern representation method can still be used in other areas in which the time series has some properties like the stock price.

ACKNOWLEDGEMENT

This research was partially supported by the National Science Council in Taiwan through Grant NSC-99-2221-E-343-006.

REFERENCES

- [1] J. C. P. Shieh, *Contemporary Investments—Analysis And Management*, Taipei, Taiwan, Best-Wise, pp. 317-431, 1998.
- [2] R.S.T. Lee, "iJADE stock advisor: an intelligent agent based stock prediction system using hybrid RBF recurrent network," *IEEE Trans. on Systems, Man and Cybernetics, Part A*, vol. 34, no. 3, pp. 421-428, 2004.
- [3] G.L. Morris, *Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures*, 2nd ed., McGraw-Hill Trade, 1995.
- [4] http://www.twse.com.tw/en/products/indices/tsec/tsec_index.php
- [5] C.H.L. Lee, A. Liu, and Wen-Sung Chen, "Pattern Discovery of Fuzzy Time Series for Financial Prediction," *IEEE Trans. on Knowledge and Data Engineering*, vol. 18, no. 5, pp. 613-625, 2006.
- [6] Klir, G.J. & Yuan, B. (1995). *Fuzzy sets and fuzzy logic theory and application*. Upper Saddle River, NJ: Prentice Hall.