ENGG4910CIJ1 Thesis I

The Automatic Trading System

Combining the Central Limit Theorem and Pattern Matching

BY

Wang Shuoyao
1155013381

A FINAL YEAR PROJECT REPORT
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF BACHELOR OF INFORMATION ENGINEERING
DEPARTMENT OF INFORMATION ENGINEERING
THE CHINESE UNIVERSITY OF HONG KONG

May, 2013

ABSTRACT

With the development of Technology and Internet, trading in financial instruments is becoming more and more depend on the use of computer-based trading systems.
These have, in the main, replaced the human-based trading systems which is the tradition way of financial trading. Although, there are still huge amount of people using the human-based trading system, they are also taking more and more reference on the computer tools. The price and turnover of the stocks are captured to electronic machines. Investment firms and their clients have also made increasing use of trading algorithms whereby orders are generated by computer algorithms responding to market data.
And our work is to build up our own trading algorithms by our own computer algorithms.
In this project, an automatic trading system will be developed. We plan to derive analytic tools that can predict the stock market by using central limit theorem and strong law of large numbers in probability theory. Also, we would like to analyze the relationship between volatility and options pricing based on Black-Scholes model so that we can make estimation on market fluctuations so as to minimize lose or maximize profit. With those analytic tools, we will be able to build a system to make decision and planning on investment.
As a result, we are using the central limit theory and pattern matching together to build up an automatic trading system. Basically, the ideas of our application are just two Chinese idioms, “樂極生悲否極泰來”and “循環往復前因後果 (四季循環)”.

ACKNOWLEDGEMENTS

I would like to take this unique opportunity to show my sincere thanks to Prof. Lee for his excellent guidance, academic help and considerate attention. Also I am really grateful to my partner Li Sixiang for his hard workings on our Final Year Project.

Contents

I. Introduction

Automatic trading system is a computer-based trading application which is set up to automatically buy and sell stock on an exchange. It's quite easy to get started with automatic trading. You just need to typically set up a strategy for trading stocks, options or foreign currency. If you are satisfied with our system's performance, you can continue to maintain the account. All over the world, there are so many automatic trading systems, some will have a little bit high revenue, some will be a little bit lower, some will be more suit for the Hong Kong market, other maybe not. But our system is combining the most two popular ways pattern matching and central limit theory which are seldom used together.
Schedule of our project:
1. To detect what market it is now
2. Strategy in a Trending Market
3. Strategy in a Trading Market

For first semester, we are focused on knowledge of stock market system, building up the trading rule, trading system and the system flow design. My partner Li Sixiang and I were actually working together for all the working days and we exchange our learning and working results frequently. In details, at the early beginning, we should have understood of the stock market for both us are form the engineering department having no knowledge about the stock. So both of us are reading the material of the technical analysis book, draw out the stock market rules, for example the market could be divided into trending market and trading market. And then we work together to build up the trading rule. And then we have a first version of our trading system only use a combination of indicator and some rule built by our self. And then we find we should do something more to be creative and have a better revenue, so we study and use the pattern matching and central limit theorem to modify our trading system.
For second semester, we are mainly focused on bi-direction trading, robustness and stability of our system. My partner Li Sixiang mainly focuses on determining the state of market and design of our own indicator Special MACD. I am mainly focus on pattern matching and design of our own pattern "V shape". After we combine all our skills and rules to do the bi-directional trading, the profit of our system is significantly improved from 30% to 50%. Also, our system can work well in different time windows, which means our system is robust.

II. Background

And for our project we should have a big rule:
We don't consider the finance part of market, that means we skip out all the information of that the company or the government has published some policy and so on, we regard that all the information are built into the price of the market.
And for this topic was raised up in the 20th Century, and a successful early automated trading system is Instinet. Instinet is best known as one of the first off-exchange trading alternatives, with its "green screen" terminals prevalent in the 1980s and 1990s, and, more recently, as the founder of Chi-X Europe and Chi-X Global.
And there are also many successful researches on the automatic trading system, such as the order flow analysis, empirical Results and Models, Arbitrage pricing and so on.

1. Hang Seng Index

Hang Seng Index (HSI) is one of the most important index in Hong Kong Stock Market, which is used to record and monitor daily changes of the largest companies of the Hong Kong Stock Market. It follows the freefloat-adjusted market capitalization weighted methodology, which takes into account long-term strategic holdings not readily available for trading in the market. [1]

The formula of calculating HSI is provided by Hang Seng Indexes Company Limited: [1]

$$\text{Current Index} = \frac{\text{Current Aggregate Freefloat-adjusted Market Capitalisation of Constituents}}{\text{Yesterday's Aggregate Freefloat-adjusted Market Capitalisation of Constituents}} \text{ X } \text{Yesterday's Closing Index}$$

$$= \frac{\sum(P_t \text{ x IS x FAF x CF})}{\sum(P_{t-1} \text{ x IS x FAF x CF})} \text{ X } \text{Yesterday's Closing Index}$$

P_t : Current Price at Day t

P_{t-1} : Closing Price at Day t-1

IS : Number of Issued Shares
(In case of H-share constituents, only H-share portio is taken into calculation)

FAF : Freefloat-adjusted Factor, which is between 0 and 1

CF : Cap Factor, which is between 0 and 1

There are 49 constituent stocks in HSI effective on Nov.1st, 2012, which are shown in Table.1.

In our testing process, we chose 20 constituents to emulate the HSI which are shown below:

Stock Code	Stock Name	Stock Code	Stock Name	Stock Code	Stock Name
0001.HK	Cheung Kong	0012.HK	Henderson Land	0101.HK	Hang Lung Prop
0002.HK	CLP Hldgs	0013.HK	Hutchison	0144.HK	China Mer Hldgs
0003.HK	HK & China Gas	0016.HK	SHK Prop	0267.HK	CITIC Pacific
0004.HK	Wharf (Hldgs)	0017.HK	New World Dev	0291.HK	China Resources
0005.HK	HSBC Hldgs	0019.HK	Swire Pacific 'A'	0293.HK	Cathay Pac Air
0006.HK	Power Assets	0023.HK	Bank of E Asia	0688.HK	China Overseas
0011.HK	Hang Seng Bank	0083.HK	Sino Land		

0004.HK	Wharf (Hldgs)	0293.HK	Cathay Pac Air	1199.HK	COSCO Pacific
0005.HK	HSBC Hldgs	0322.HK	Tingyi	1299.HK	AIA
0006.HK	Power Assets	0330.HK	Esprit Hldgs	1398.HK	ICBC
0011.HK	Hang Seng Bank	0386.HK	Sinopec Corp	1880.HK	Belle Int'l
0012.HK	Henderson Land	0388.HK	HKEx	1898.HK	China Coal
0013.HK	Hutchison	0494.HK	Li & Fung	1928.HK	Sands China
0016.HK	SHK Prop	0688.HK	China Overseas	2318.HK	Ping An
0017.HK	New World Dev	0700.HK	Tencent	2388.HK	BOC Hong Kong
0019.HK	Swire Pacific 'A'	0762.HK	China Unicom	2600.HK	CHALCO
0023.HK	Bank of E Asia	0836.HK	China Res Power	2628.HK	China Life
0066.HK	MTR Corporation	0857.HK	PetroChina	3328.HK	Bankcomm
0083.HK	Sino Land	0883.HK	CNOOC	3988.HK	Bank of China
0101.HK	Hang Lung Prop	0939.HK	CCB		
0144.HK	China Mer Hldgs	0941.HK	China Mobile		

Table.1

1. Technical Analysis & Technical Indicators

Technical analysis is a security analysis discipline used for predicting the trend of prices through the study of historical market data, mainly price and volume. Through technical analysis, there are many different kinds of technical indicators may be used:

Trend Indicators

Moving Average

Moving average (MA) is one of the basic and efficient methods to determine the trend of the stock market. If the price of a stock is above the moving average, the market is probably in an upping trend; otherwise, if the price of a stock is below the moving average, the market is probably in a downing trend. [2]

There are some properties of moving average:

(a) The larger the absolute value of slope is, the faster the market is going up/down.

(b) The smaller the absolute value of slope is, the slower the market is going up/down.

(c) If the MA is going nearly horizontally and the price is fluctuating around the MA, then the market is not a trending market.

(d) As to long-run MA, if it meets the bottom and turns up or meets the top and turn down, it means that the trend of the market is changing.

SMA

The simple moving average is simply the average of the price in a period.
Since the typical period of the stock market is 14 days, [2] the default period for SMA is also 14 days, which is also the default period for other indicators.

EMA

The exponential moving average is the MA that reduces the lag by applying different weights to different time. The more recent the price is, the more weight it has.
The weighting applied to the most recent price is dependent on the number of periods in the moving average. [3]

Calculation:

SMA = a period sum / the period;
Multiplier = (2 / (Time periods + 1));
EMA = [Close - EMA(previous day)] * multiplier + EMA(previous day).

Momentum Indicators

Momentum is one of the most important concepts in technical analysis since it can provide two important information of a stock market: the direction of the market and the range of changing. [2]

DMI & ADX

The Average Directional Index (ADX), Minus Directional Indicator (-DI) and Plus Directional Indicator (+DI) are a group of directional movement indicators. [3] They are useful indicators to determine the market is trading or trending.

Definition [2][3]:
"Directional movement is positive (plus) when the current high minus the prior high is greater than the prior low minus the current low. This so-called Plus Directional Movement (+DM) then equals the current high minus the prior high, provided it is positive. A negative value would simply be entered as zero.

Directional movement is negative (minus) when the prior low minus the current low is greater than the current high minus the prior high. This so-called Minus Directional Movement (-DM) equals the prior low minus the current low, provided it is positive. A negative value would simply be entered as zero.

True Range is the maximum of the range between current high and low price, the range between the current high and the previous close price and the range of between the current low and the previous close price."

Calculation [3]:

1. Calculate the True Range (TR), Plus Directional Movement (+DM) and Minus Directional Movement (-DM) for each period.

2. Use the Wilder's smoothing techniques, which are some kind of moving average, to smooth the periodic values.

3. Calculate Plus Directional Indicator (+DI).
   +DI=MA(+DM)/MA(TR)

4. Calculate Plus Directional Indicator (-DI).
   -DI=MA(-DM)/MA(TR)

5. The Directional Movement Index (DX) equals the absolute value of +DI less -DI divided by the sum of +DI and - DI.
   DX = abs(((+DI) - (-DI)) / ((+DI) + (-DI)))

6. After all these steps, it is time to calculate the Average Directional Index (ADX).
   ADX = SMA(DX).

RSI

Relative Strength Index (RSI) is a momentum indicator that can measure the speed and change of price movements. RSI oscillates between 0 and 100. The market is probably overbought when RSI is larger than 70 and oversold when RSI is less than 30. [3]
Calculation [4]:

RS=AU/AD
AU=the average of the price rise during a period
AD= the average of the price descend during a period
RSI=100-100/(1+RS)=AU/(AU+AD)*100.

Cyclic Indicators

The market is consisted of mainly two parts: the market trend and the market cyclicity. The cyclic indicators are developed to describe the cyclicity of a stock market. Stochastic oscillators are the most famous ones.

MACD

The Moving Average Convergence-Divergence (MACD) indicator is one of the simplest and most effective momentum oscillators. The MACD turns two trend-following indicators, moving averages, into a momentum oscillator by subtracting the longer moving average form the shorter moving average. Thus, the MACD offers messages of both trend following and momentum. [3]
Calculation [3]:

MACD Line: (12-day EMA - 26-day EMA)

Signal Line: 9-day EMA of MACD Line

MACD Histogram: MACD Line - Signal Line

2. Pattern Matching

Dow Theory:

The Dow Theory is not constructed to predict the stock market or indicate the investor; it's representing the whole trend of the stock market. And the most important idea of Dow Theory is that

1) The Manipulation can influence the weekly fluctuate and Secondary reaction, but the Primary trend will not be affected.
2) In the market, every professional economic and business person will combine all the polity, the speech by the famous person, and their own psychological factor together to do the strategic decision. So the market looks like difficult to understand.
3) Dow Theory is doing the objective analysis, so when you use it subjectively, you will get big lose.

And Dow Theory has 5 definitions:

1) Primary trend is long trend, and can last for several years. And the most difficult one to predict is the Secondary reaction, always last for three weeks to several months. And the amplitude is always larger than 20%.
2) There is no existing method to successful predict how long the Primary trend will last.
3) Primary Bear Markets are the long term decreasing market, and during the period, there are many rebounds.
4) Primary Bull Markets are the long term increasing market, the average increasing time is about 2 years.
5) Secondary Reactions are the decreasing period in the Primary Bull Markets or the increasing period in the Primary Bear Markets, which are always lasting for three weeks to several months, and the amplitude is about 33%~67% of the Primary Market.

So form the statistical data, you can see the market always behaves similarly and so we can do the pattern matching find the regular in the stock market.

Pattern matching:

Figure 2.3

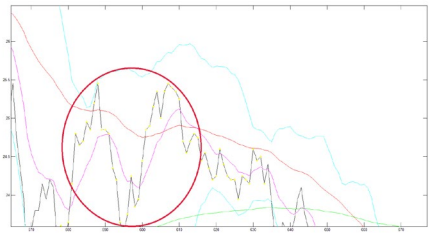


Figure 2.4

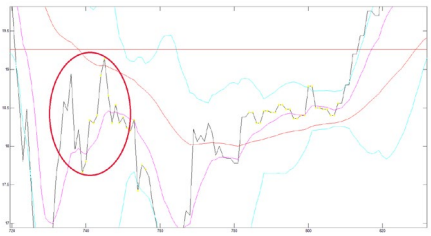


Figure 2.5

The Figure2.4 and Figure2.5 are capture from the Figure2.3, and you can see that they are behaving similar. And all the stock data can be regarded as the time series data. And we can regard several continues data as a query. So we are interested in similarity queries.
And a common way to definite the similarity is:
If the Euclidean distance between two time sequences is less than a user-defined threshold, we can call the two time sequence similar [15]. Other distance metrics, including the $\mathcal{L}_p$ Norms may also be used [16]. However, the time series are always being long. So the distance computation can be time consuming and CPU occupying.
A popular solution to this problem is translate the time sequences into the frequency domain using Fourier transform, and uses the first few coefficients to tell the similarity.[17] [18]
Later on, some improve has been achieved, such as The Haar Wavelet Transform [19] [20] and some similar techniques [16].
But there is still a question that how to do the similarity among a long series data of transform coefficients not been solved.
Shasha and Wang proposed an indexing method using triangular inequality [21]. But the space overhead is up to n^2. And Voronoi diagram work well for 2-dimensional spaces, but complicated to work in 3 or high dimensional [22].
Jagadish first suggested using minimum bounding rectangles to extract features from shapes and subsequently managing the resulting vectors using a spatial access method, like k-d-B-trees, grid files. And the multidimensional indexing methods are: R*-trees [23], the rest of the R-tree and k-d-Btree family [24], linear quadtrees [25], grid files [26].
However all the method above face two problems:
(a) Completeness of feature extraction
(b) Dimensionality “curse” (Don’t use the method grows exponentially with the dimensionality)
The wavelet transform can both represent the information in time domain and frequency domain and have a very acceptable error; meanwhile the Fourier transform can only represent the information in frequency domain. What’s more, the complexity of wavelet transform is O(n) and the complexity of Fourier transform is O(nlogn). So we replace the Fourier Transform with wavelet transform this semester.

III. Methodology

1. Flow Chart

At the first semester, we use the flow chart below to do trading.

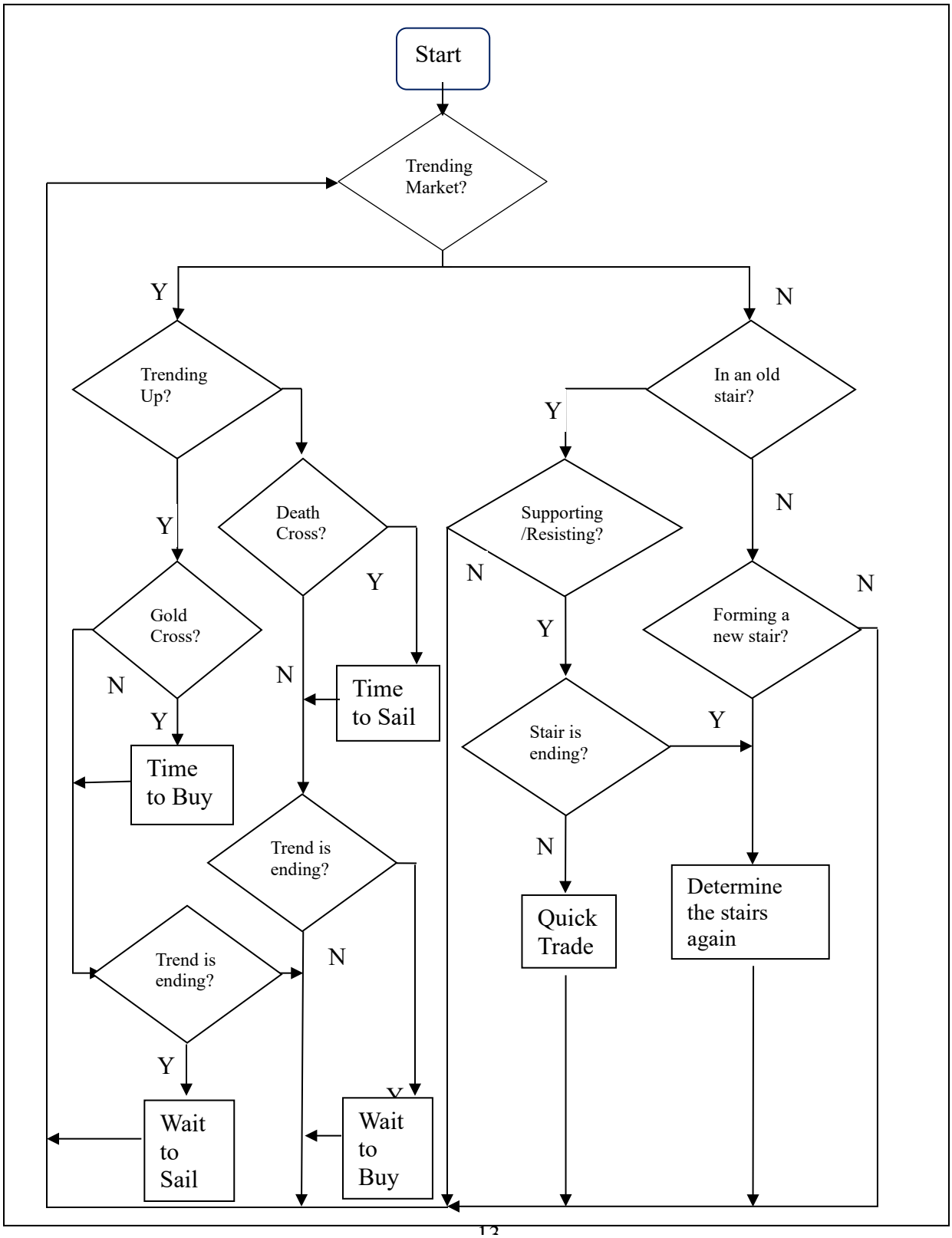


Figure 3.1.1

Flow Chart for Quick Trade

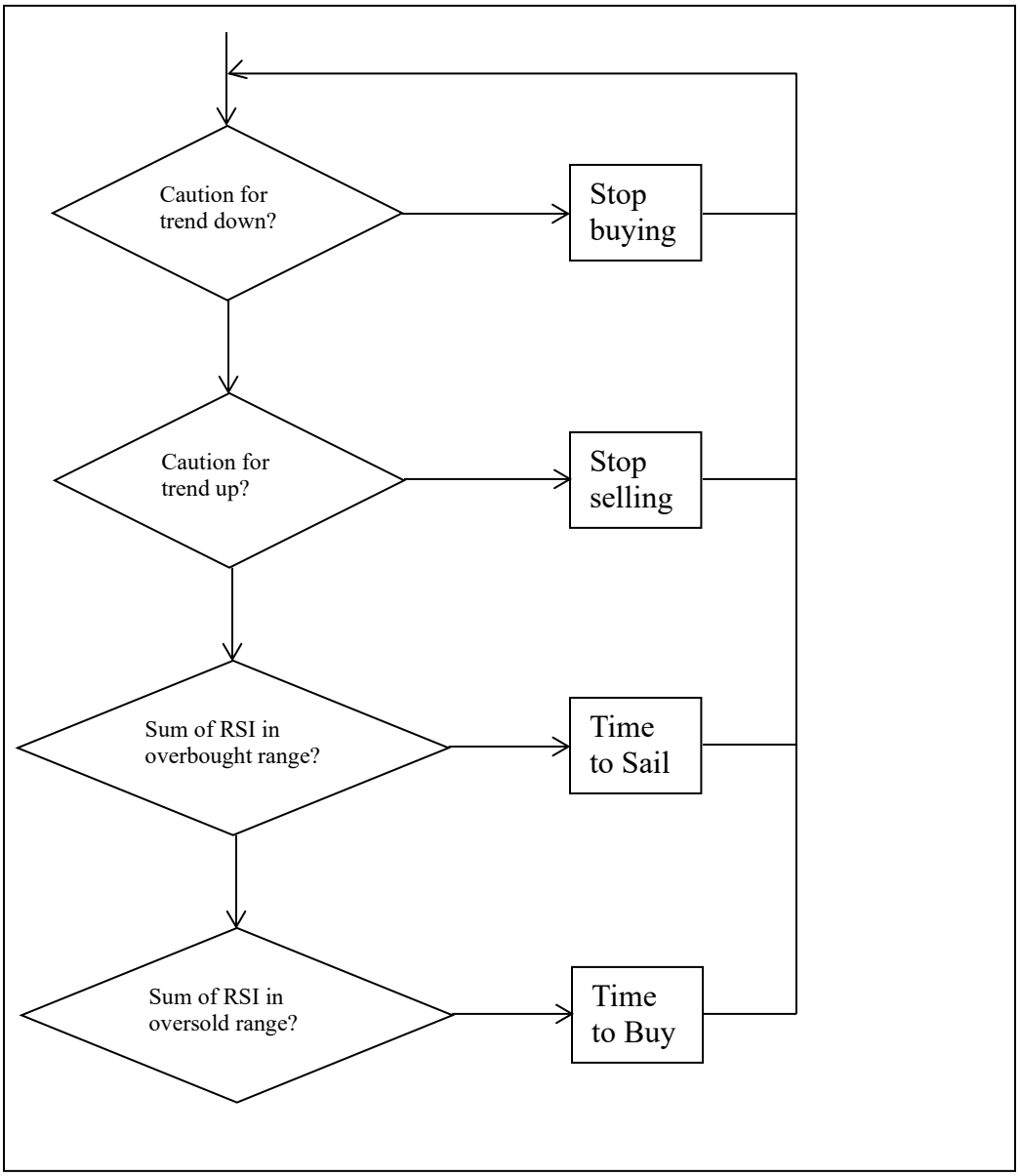


Figure 3.1.2

This flow chart works very well in time window 2007/11/01 to 2012/11/01. However, when we test the algorithm in a time period with market trending down for long time (such as 1999-11-01--2004-11-01 of HangSeng Index), we can't earn a lot money (As we will have a revenue of 8% per year).

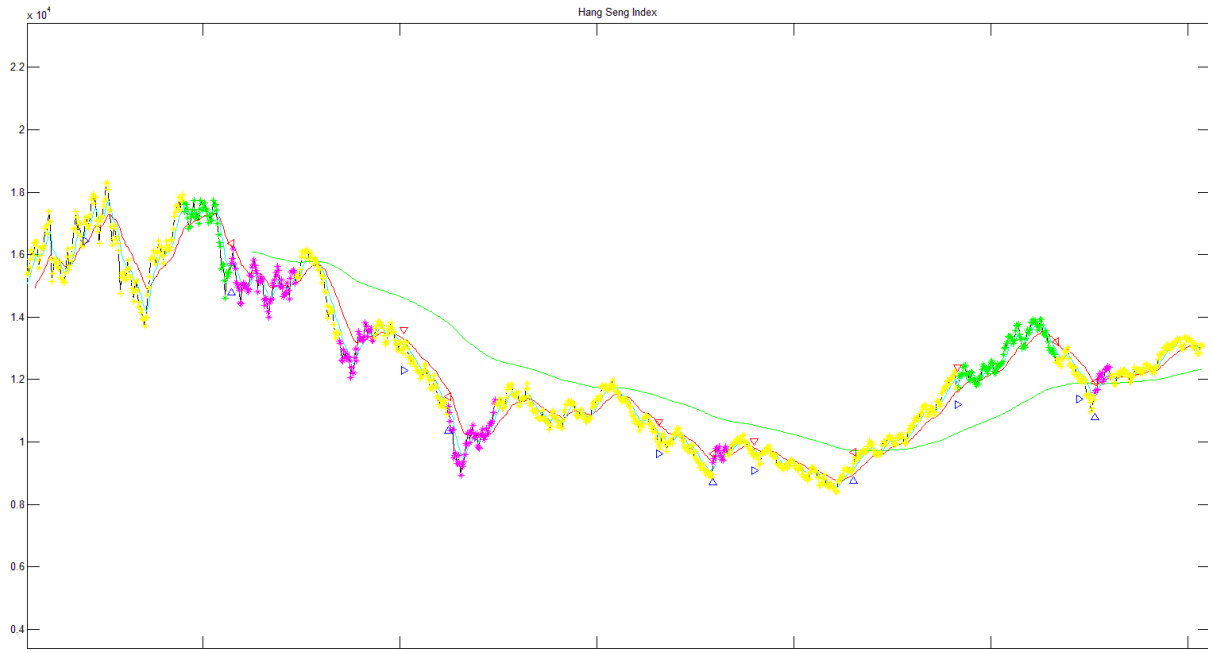
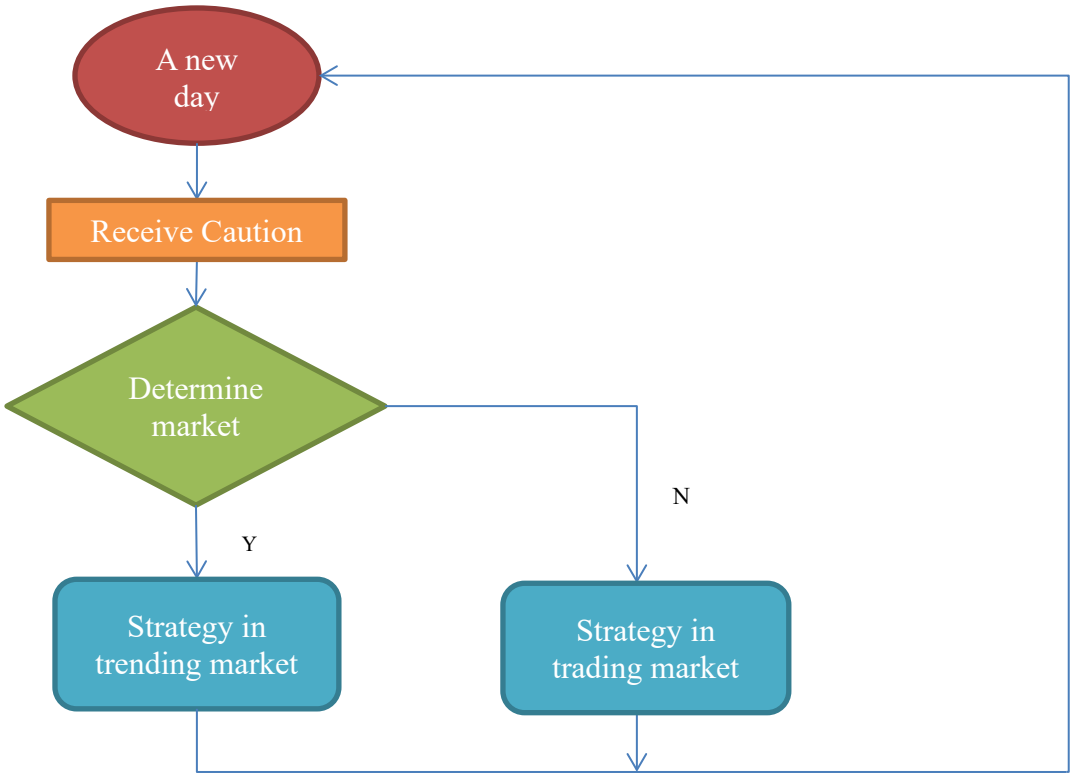
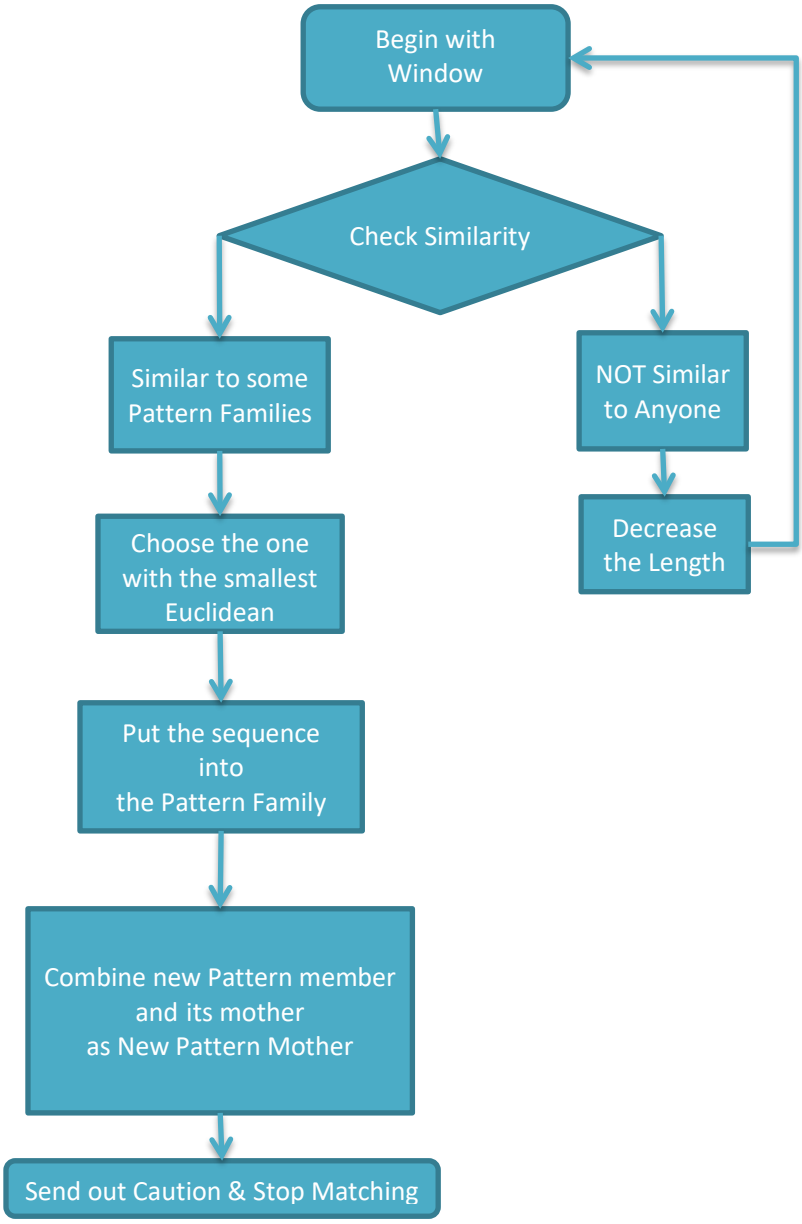


Figure 3.1.3

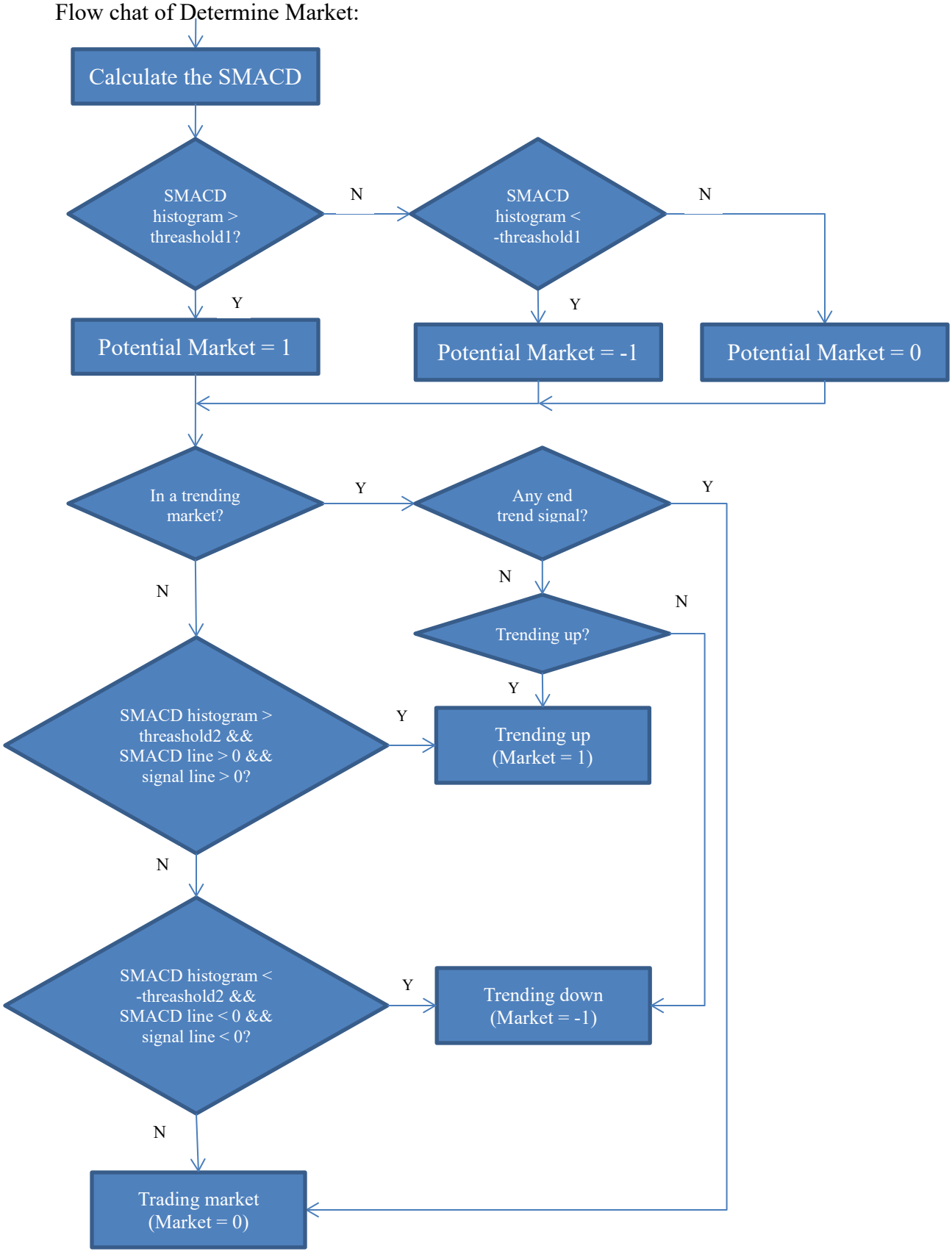
So we develop bi-direction trading rule to earn money at different time window.



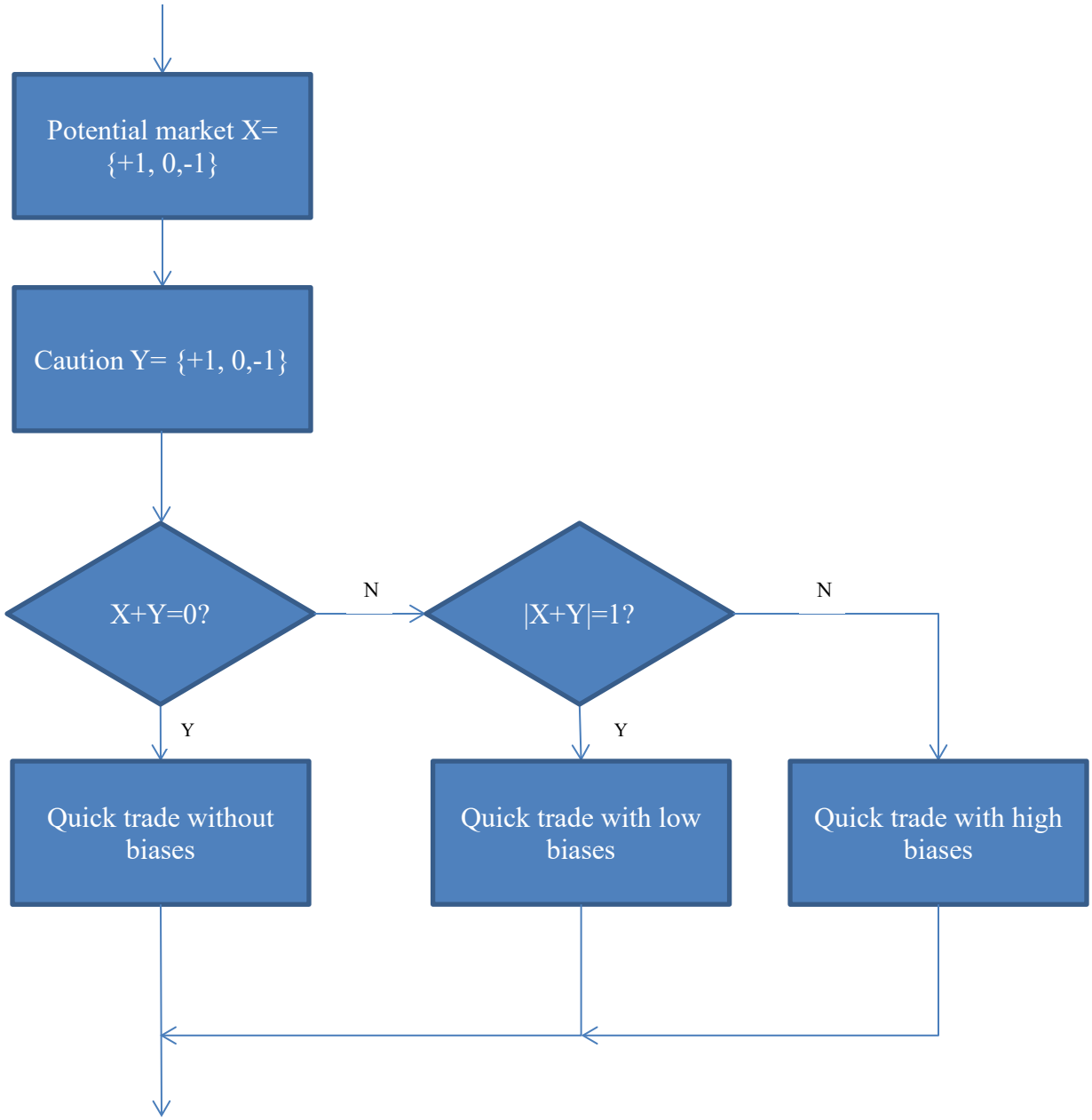
Flow chat of Receive Caution



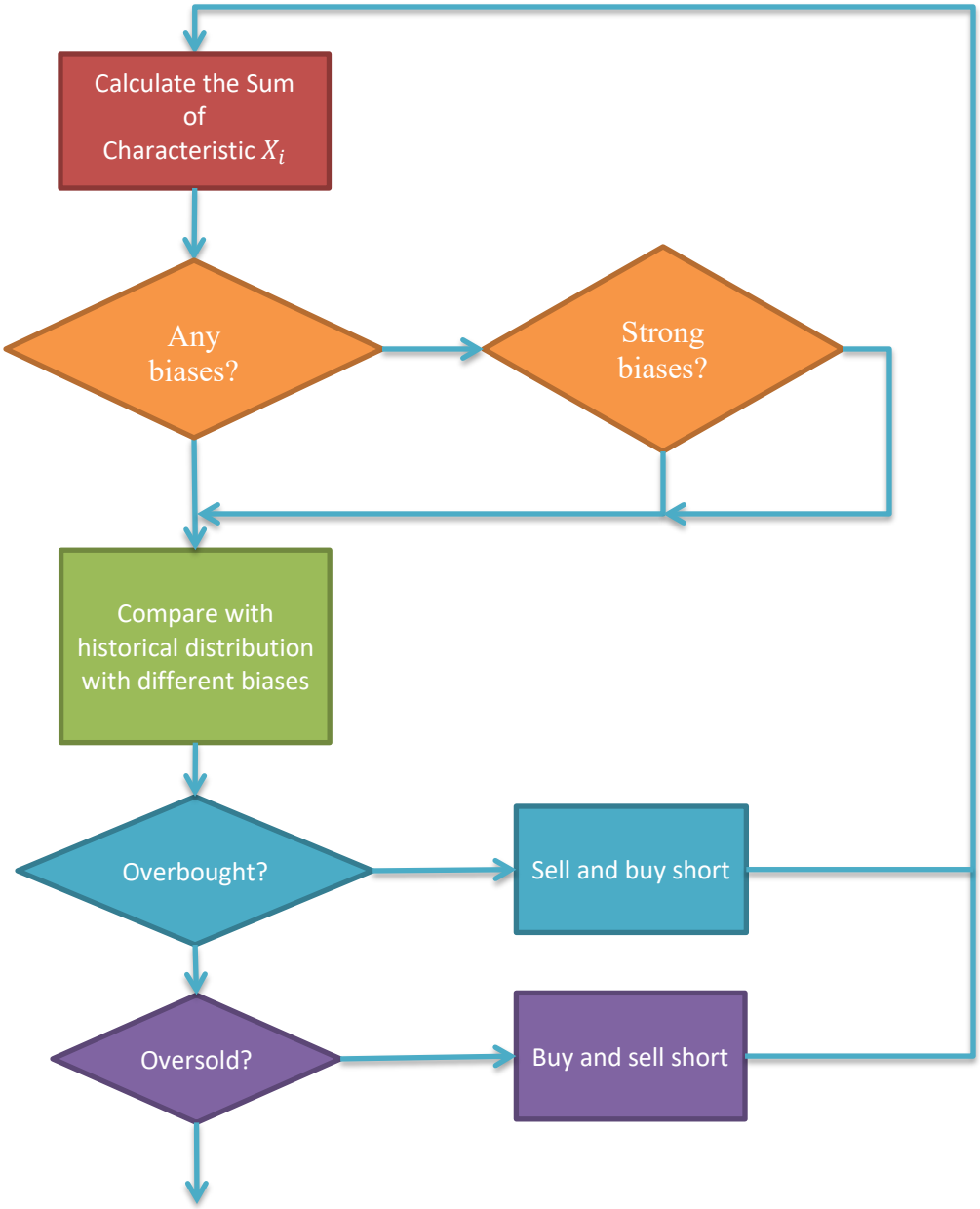
Flow chat of Determine Market:



Strategy in trading market:



Details of Quick trade:



2. Technical Analysis by Indicators
Definition of Special MACD

MACD & short-medium-term moving average
MACD describes the change of the difference between short-term moving average and long-term moving average. In addition, in technical analysis, it is a strong signal implies that the marketing is trending up when the short-term moving average is above the medium-term moving average and the medium-term is above the long-term moving average. Combing these two technical analysis skills, we defined our own indicator – Special MACD (SMACD).

Definition of SMACD
Instead of only calculate the difference between the short-term and long-term moving average, we calculate the difference between moving averages of different periods and calculate their weighted sum.

Calculation:

$$SMACD = \sum_{i=0}^{k-1} w_i\big(EMA(period_i) - EMA(period_{i+1})\big)$$
$$SMACD_{Signal} = EMA(SDFF, period_{signal})$$
$$SMACD_{Histogram} = SMACD - SMACD_{signal}$$
$$where\ w_i\ is\ the\ weight\ of\ the\ difference\ between\ EMA\ of\ period\ i\ and\ period\ i+1$$

Advantages of SMACD
Compared with MACD, SMACD describes the change of difference between short-term, medium-term and long-term moving average instead of only the difference between short-term and long-term moving average. Thus, SMACD contains more message of the market.
Compared with the observations of the position of the moving average with different periods, SMACD evaluates the market by the weight sum of the difference between moving averages with different periods.

Implementation
In our implementation, we choose five different periods for the moving average, 7 days, 14 days, 30 days, 50 days and 100 days, which are typical periods of moving average.
When we assign the weights to the difference between the moving averages with different periods, we assign larger weight to the difference between longer periodic moving averages. The reason is that the moving average with longer period can more powerfully describe the market in long term.

```
function [ SMACD ] = DetermineSMACD( cl, varargin  )
%DETERMINESMACD Summary of this function goes here
if isempty(varargin)%default periods
    period_s = 7;
    period_ms = 14;
    period_m = 30;
    period_ml = 50;
    period_l = 100;
    period_signal = 14;
else%optional input for periods
    period_s = varargin{1};
    period_ms = varargin{2};
    period_m = varargin{3};
    period_ml = varargin{4};
    period_l = varargin{5};
    period_signal = varargin{6};
end

ema_l=indicators(cl,'ema',period_l);
ema_ml=indicators(cl,'ema',period_ml);
ema_m=indicators(cl,'ema',period_m);
ema_ms=indicators(cl,'ema',period_ms);
ema_s=indicators(cl,'ema',period_s);

smacd1=0.1*(ema_s-ema_ms)+0.2*(ema_ms-ema_m)+0.3*(ema_m-ema_ml)+0.4*(ema_ml-
ema_l);
lengthof = length(smacd1);
nan_length = length(find(isnan(smacd1)));
ssignal1=smacd1;
ssignal1(nan_length+1:lengthof)=indicators(smacd1(nan_length+1:lengthof),'ema'
,period_signal);
smacd_h1=smacd1-ssignal1;

smacd2=0.1*(ema_s-ema_ms)+0.3*(ema_ms-ema_m)+0.6*(ema_m-ema_ml);
lengthof = length(smacd2);
nan_length = length(find(isnan(smacd2)));
ssignal2=smacd2;
ssignal2(nan_length+1:lengthof)=indicators(smacd2(nan_length+1:lengthof),'ema'
,period_signal);
smacd_h2=smacd2-ssignal2;

SMACD=[smacd1,ssignal1,smacd_h1,smacd2,ssignal2,smacd_h2];

end
```

Determine the Market is Trending or Trading

By ADX

When ADX is increasing, the market is probably trending; when ADX is decreasing, the market is probably trading.

By Bollinger Band

If the bandwidth of Bollinger Band is decreasing significantly, the market is probably changing from trading to trending. If the bandwidth of Bollinger Band is increasing significantly, the market is probably changing from trading to trending.

By Moving Average

When the short-run EMA is significantly above the long-run EMA, which means the market is trending up; when the short-run EMA is significantly below the long-run EMA, which means the market is trending down; when the short-run EMA is fluctuating around the long-run EMA, the trend is not clear, which means the market is trading.

By MACD

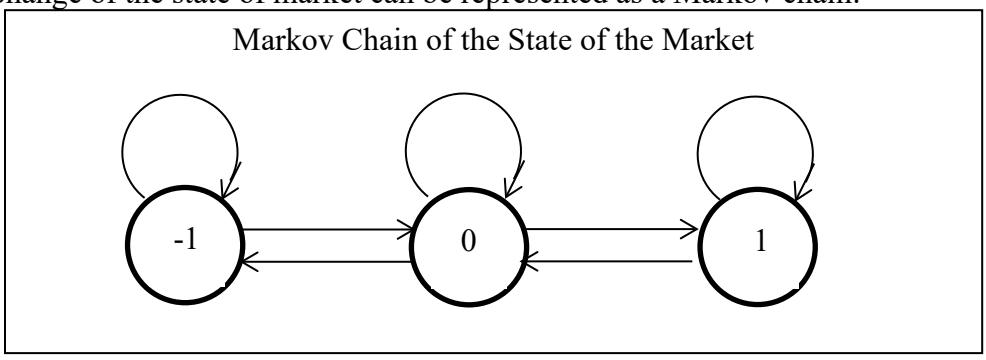
When MACD changes from negative to positive and is increasing quickly, the market is probably trending up; when MACD changes from positive to negative and is decreasing quickly, the market is probably trending down; when MACD fluctuates in a low level, the market is probably trading.

Definition

Let X(t) be a random variable which represents the state of the market at time t.

$$X(t) = \begin{cases} 1, & if\ the\ market\ is\ trending\ up \\ 0, & if\ the\ market\ is\ trading \\ -1, & if\ the\ market\ is\ trending\ down \end{cases}$$

Then the change of the state of market can be represented as a Markov chain:



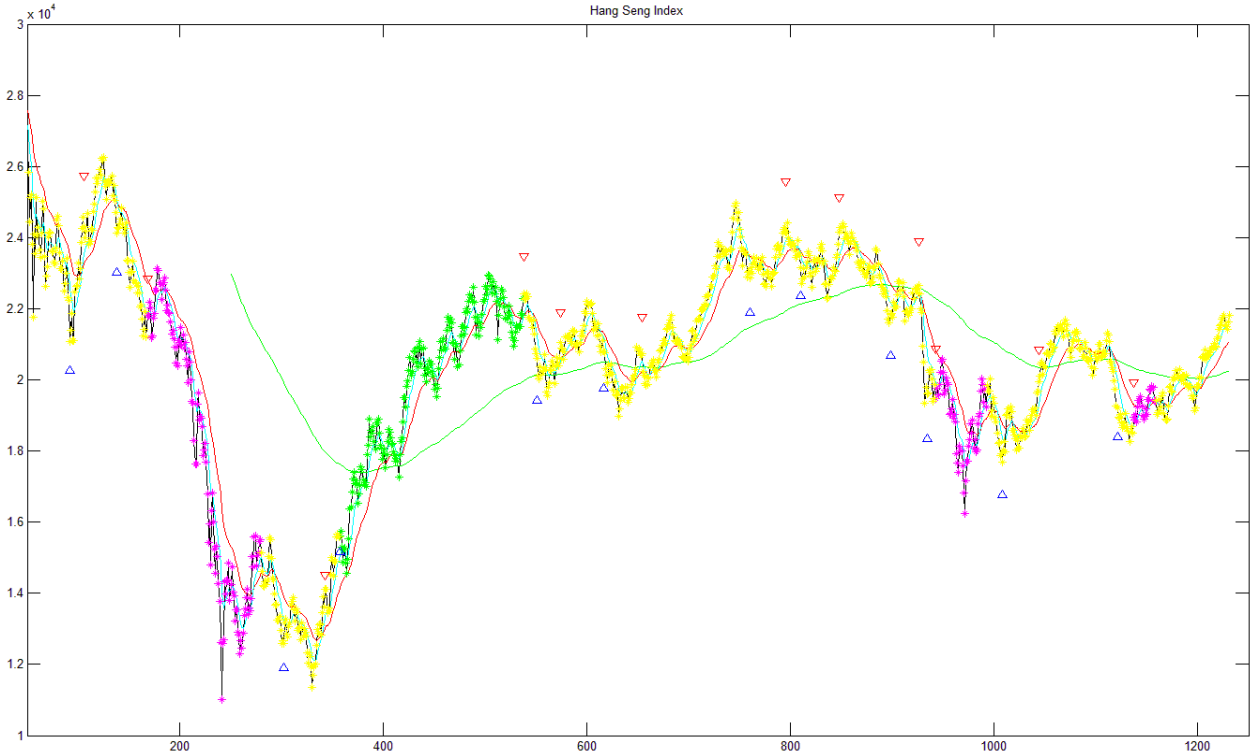
Implementation

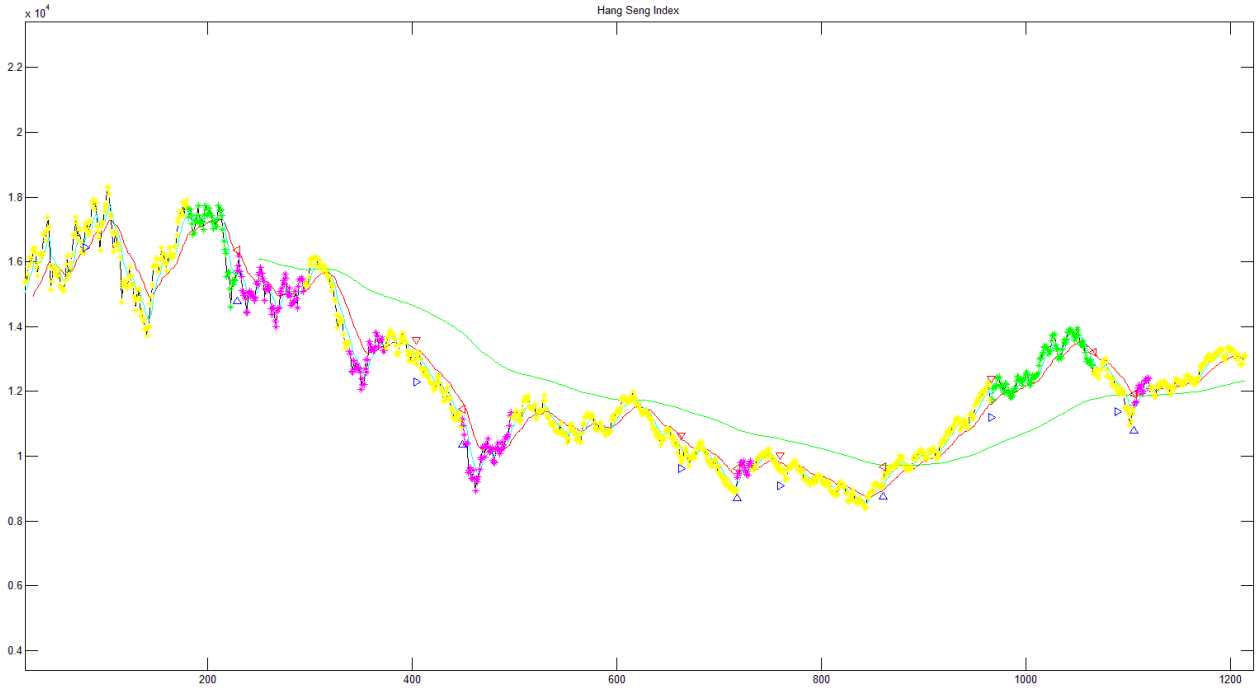
Last semester, we tried to determine the state of the market by the combination of ADX and the EMA. First, we tried to determine the market via ADX and EMA separately. Then we used a combination function to combine the two results of the market states.

However, the method only works well in the time window from 2007~2012, and the result in 1999~2004 was very bad.

So we develop another method to determine the states of the market by using SMACD.

We use SMACD histogram together with the SMACD line and the signal line to determine the state of the market.


Hang Seng Index
x 10^4
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
200
400
600
800
1000
1200


Hang Seng Index
x 10^4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
200
400
600
800
1000
1200

We set 3 thresholds to evaluate the level of the SMACD:
Threshold 1: the pre-initial threshold of a trending market, it implies the market is in which potential market.
Threshold 2: the initial threshold of a trending market. When the sign of SMACD histogram is the same as the sign of SMACD line and signal line, at the same time the SMACD histogram is higher (or lower) than the threshold 2 (or the negative threshold 2), it implies that the market is trending up (or down).
Threshold 3: the ending threshold of a trending market. When the SMACD histogram is beyond the thresholds 3, the signal of ending a trend will send.

In order to set the 3 thresholds dynamically, we use the first 100-day data to train the system so that it could work well in different time windows.

Strategy in a Trending Market

In a trending market, the basic strategy is following the trend.
When the market is starting trending up, it is the time to buy and hold until the trend is end.
When the market is starting trending down, it is not the time to hold any stocks but it is the time to buy long and sell long until the trend is end.

Strategy in a Trading Market

In a trading market, we can use momentum indicators or stochastic oscillators to do quick trades. Since the indicators of the index cannot ensure the reliability of prediction, we choose to use the sum of random variables instead of simply using the indicators of the index.

Take RSI for an example, if RSI is larger than 75, the market is probably overbought; if RSI is less than 25, the market is probably oversold. We can construct a random variable x_i for each constituent stock in an index, which call the characteristic random variable of a stock.

$$x_i = \begin{cases} 1, & RSI(x_i) > 75 \\ -1, & RSI(x_i) < 25 \\ 0, & eslewhere \end{cases}$$

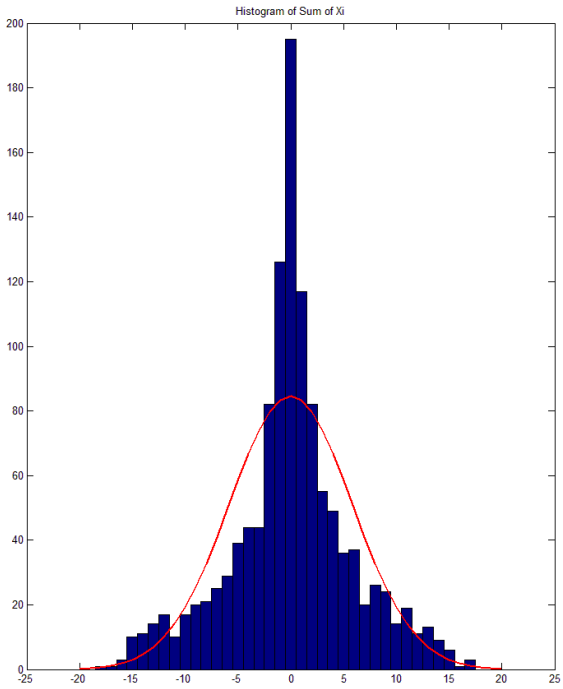
According to the Large Number Law and the Central Limit Theorem [7], the sum of x_i would be nearly normal distributed if the number of x_i is large enough, that is $\sum_i x_i \sim$ Normal (μ, σ).
The graph below (Figure.3.2.1) is shown the sum of characteristic random variable of 20 stocks over 1232 days.

$$y = \sum_{i=1}^{n=20} x_i$$

$$y \sim Normal(\mu, \sigma)$$

$$with\ mean\ \mu = -0.0975, stand\ deviation\ \sigma = 3.7586$$

Define the range of overbought/oversold by $\mu \pm k\sigma$, the coefficient k can be set variously according to the market. For example, when the market is potentially trending up, then the overbought level and the oversold level can be set higher; when the market is potentially trending down, then the overbought level and the oversold level can be set lower.



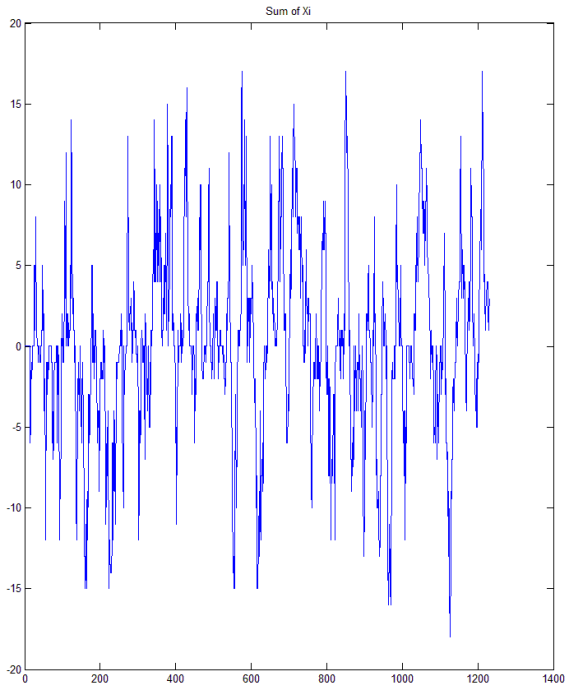


Figure 3.2.1

And these biases of coefficient k are set according to the potential market and the caution from the pattern matching.
If the signal from the potential market and the caution from the pattern matching are the same, then the biases will be set high, which means the system will try to buy or sell very aggressively. For example, if the market is probably starting to trend up, the system will try to enter the market although the sum of RSI is not very low.
If there is only the signal from the potential market or the caution from the pattern matching, then the biases will be set low, which means the system will a little aggressively try to buy or sell. For example, if the market is possibly starting to trend up, the system will try to enter the market if the sum of RSI is relatively low.
If the signal from the potential market and the caution from the pattern matching are different or there are no signal and caution, then there are no biases, which mean the system will try to buy or sell normally.

Drawback and Improvement

Although the Central Limit Theorem can mitigate the incident errors in analysis, technical indicators can only predict the market in a short period and always lagging.

According to Dow Theory, the market will repeat itself, so we can use pattern matching to predict the market in a longer period.

3. Pattern Matching

3.1 Aim

To accurately predict stock price changes through the temporal clustering of (immediately) prior price sequence. That is, given a stocks consecutive price information (day 0 – N), predict the price at the next day (day N+1).

3.2 Definition of pattern

A sequence of price ordered in chronological order.

3.3 Why transforms?

3.3.1 Why transforms (mainly focus on wavelet transform and Fourier transform)

There are several big advantages:

1) The transform method such as Fourier transform is widely used for data analysis in this and last century. So it's reliable and readily available.

2) And you can do the test for the past years stock prices, the transform will do a good job of concentrating the energy in the first few coefficients. Actually some research as tell that the stock price is belongs to Brownian Noise [28], and the Brownian noise has the spectrum like Figure 3.3.1. So if we are doing some transform on frequency domain, we will get a good result.

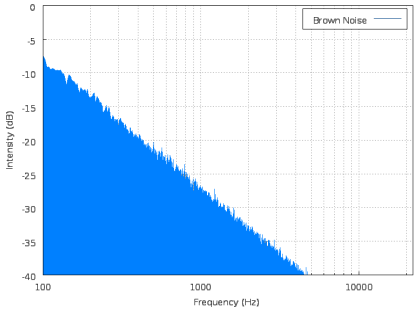


Figure 3.3.1

3) Convolution theorem says that "convolution in one domain (e.g., time domain) equals point-wise multiplication in the other domain (e.g., frequency domain)".[30]

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} * \mathcal{F}\{g\}$$

And it's will be a great equation when you do the things like moving-average.

For we can represent the process of moving-average as set the g=[1/n,1/n,1/n,…..,1/n,0,0,0……0] (set the first n elements as g=1/n, and others equal to 0), let the f equal to the data series. So when you do the transform, you can just calculate$\mathcal{F}\{f\} * \mathcal{F}\{g\}$.

4) When you do the normalization first, the first coefficients of transform or wavelet transform is zero, you can threw it out to reduce the dimension.

5) Compared to the data dependent transform, such as K-L transform, it doesn't need data reorganization.

3.3.2 Why Wavelet transforms:

1) The basic functions of wavelet transforms are non-zero only on a finite interval. What this means for an application is that wavelets are able to capture local (time dependent) properties of data, whereas Fourier transforms can only capture global properties.

2) The efficiency of the wavelet transform is superior even when compared with the Fast Fourier transform. The Fourier transform is O (n^2), where n is the length (number of attributes) of the data and Fast Fourier transform is O (nlogn). In general, the speed of wavelet transforms is linear in the length of the data.

3) The Fourier transform gives the set of frequency components, which exist in our signal. On the other hand, wavelet transforms give gradually refined representation of the signal of different

scales, which correspond to basic functions of different length. Hence, the wavelet transform is hierarchical and allows much finer tuning for a variety of applications.
4) Unlike the Fourier transform, wavelet transforms have an infinite set of possible basis functions. Thus, they provide access to information that can be obscured by other methods.

3.4 Definition

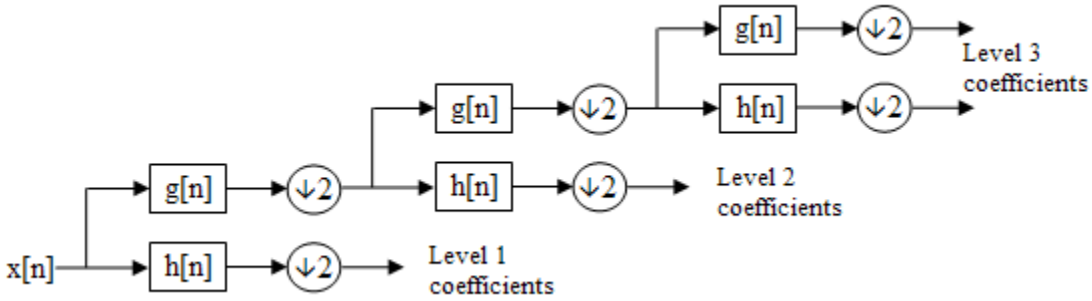
3.4.1 Definition of normalization:

Given a stock price sequence x_n , the corresponding normalized sequence is T days-modified-moving-average of the sequence y_n , where y_n=$\frac{x_n-\bar{x}}{stander\ deviation\ of\ x_n}$, T=$\frac{length\ of\ x_n}{3}$.
And what's more, when you take the normalization, you will find that the first coefficient of the transforms becomes zero becomes zero for every sequence.
The normalization idea comes from Davood Rafiei and Alberto Mendelzon.[31]
This is a modified version.

3.4.2 Definition of T days-modified-moving-average:

For a sequencex_n, $MV = \begin{cases} \frac{\sum_{i=0}^{k} x_i}{k} & for\ k < T \\ \frac{\sum_{i=k-n+1}^{k} x_i}{T} & for\ k \geq T \end{cases}$

3.4.3 Definition of V vector:



Given a sequence x[n], input into the left hand side of the system above, until we get a g[n] with length less or equal to 8. Then the g[n] becomes our V vector.

3.4.4 Definition of transform pattern similarity

Given sequence1 and sequence2, we can have their V vectors, V_1 and V_2. If the Euclidean distance smaller than the threshold1 which is set by the user, then we can call the sequence1 and sequence2 are transform pattern similar.

3.4.5 Definition of pattern similarity

Given sequence1 and sequence2, if the Euclidean distance between s1 and s2 is smaller than a threshold2 which is set by the user, then we can call the sequence1 and sequnce2 are pattern similarity, where ss1, ss2 is the normalized sequences of sequence1 and sequence2, and then

time wrap the one with small length to make them having the same length. The two sequences after the time wrapping are s1 and s2.

3.4.6 Definition of pattern family

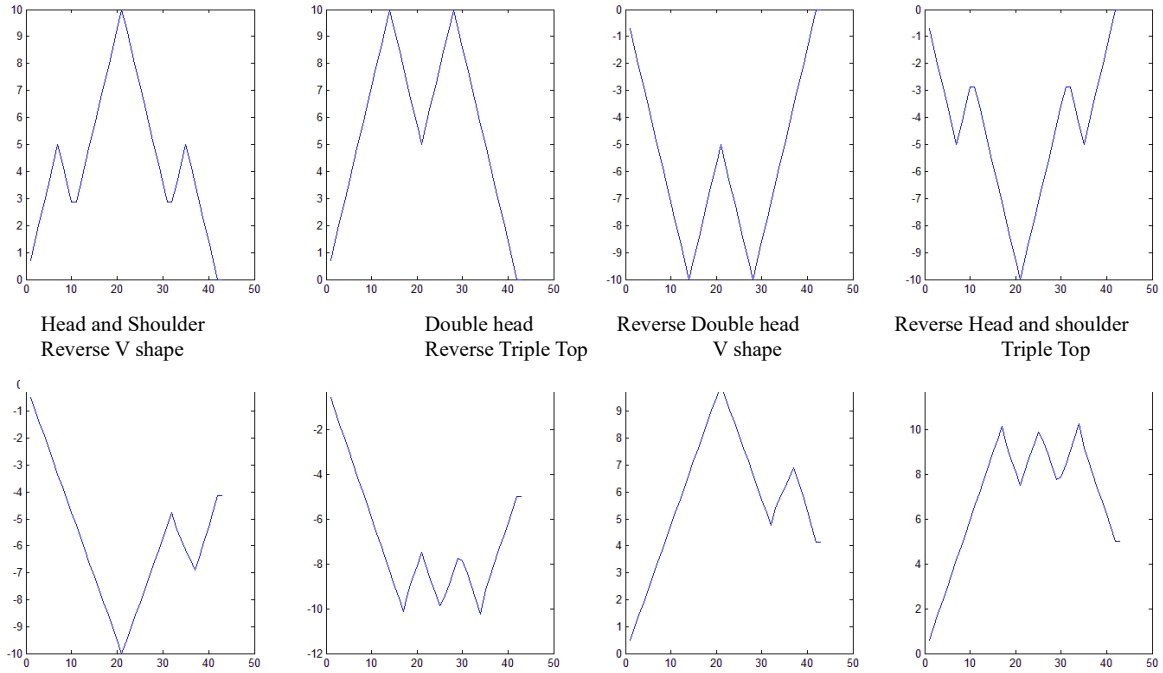
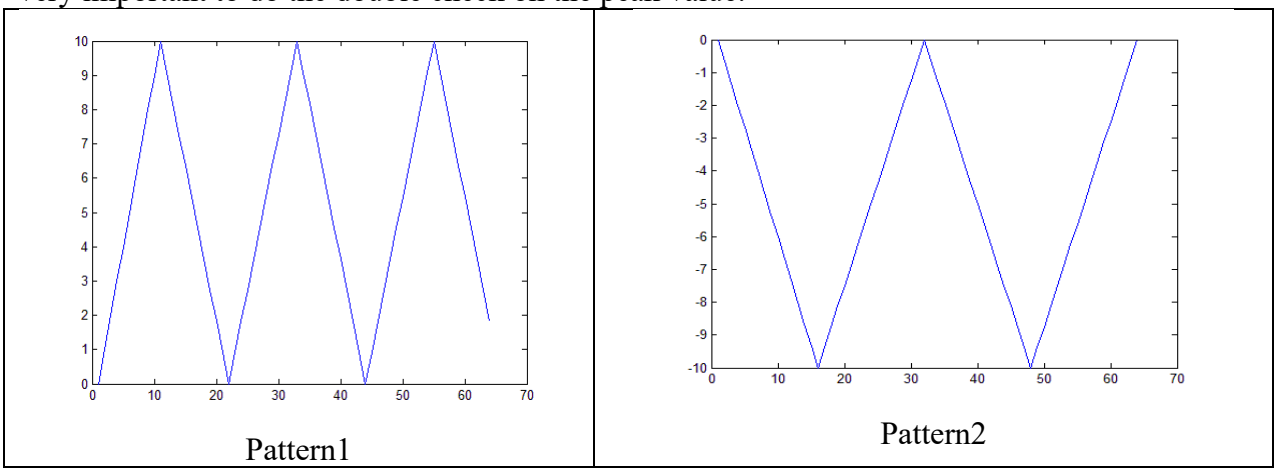
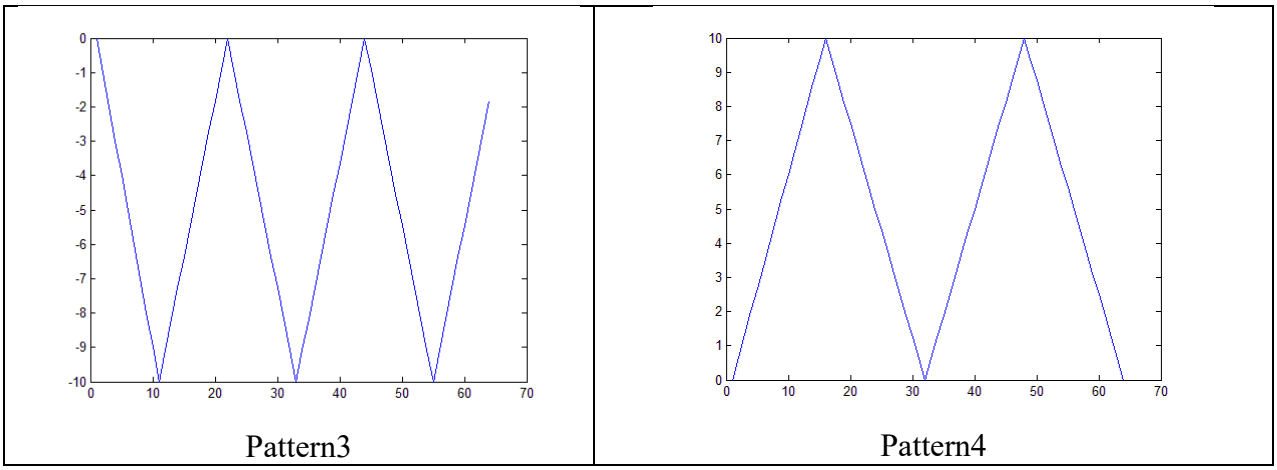


Figure 1

In the stock market, there are several famous patterns which represent important market signals. Standing the test of the history, we can regard them as "always true". We can select the most famous ones as the "pattern mothers".

However, if we directly do the matching on the pattern mothers and the real time series data, you will find the patterns are always appearing together. It's quite often to get the result that you figure out there is a double head at m day, but we will see a double bottom at m+5 day. So it's very important to do the double check on the peak value.





Pattern3 Pattern4

So we change the 'double head', 'double bottom', 'triple head', 'triple bottom', 'head and shoulder', 'inverse head and shoulder' into this four pattern. Actually the head and shoulder pattern and the triple head pattern only differ at the peak value, so we can use one pattern here with different requirement of peak value. That means our solution is matching to the patterns we defined and doing the flit according to our detail requirements.

Details requirements of peak value:

a) Double Head: i) The first peak must be lower or equal to the second peak
   ii) The beginning value should lower than the mean value of whole pattern.
   iii) The ending value should lower than the mean value of whole pattern.
   iv) The middle point shouldn't be the smallest value of the whole pattern. (The minimum value of the pattern should be located before ¼ or after 3/4)

b) Double Bottom: i) The first valley must be lower or equal to the second valley
   ii) The beginning value should higher than the mean value of whole pattern.
   iii) The ending value should higher than the mean value of whole pattern.
   iv) The middle point shouldn't be the largest value of the whole pattern. (The maximum value of the pattern should be located before ¼ or after 3/4)

c) Triple Head: i) The first peak must be no lower or equal to the second peak and the third.
   ii) The beginning value should lower than the mean value of whole pattern.
   iii) The ending value should lower than the mean value of whole pattern.
   iv) The two valley points shouldn't be the smallest value of the whole pattern. (The minimum value of the pattern should be located before ¼ or after 3/4)

d) Triple Bottom: i) The first valley must be no higher or equal to second valley and third.
   ii) The beginning value should higher than the mean value of whole pattern.
   iii) The ending value should higher than the mean value of whole pattern.
   iv) The two peak points shouldn't be the largest value of the whole pattern. (The maximum value of the pattern should be located before ¼ or after 3/4)

e) Triple Head: i) The first peak must be lower or equal to the second peak and the third.
   ii) The third peak should lower than the second peak.
   iii) The beginning and ending values should lower than the mean value of whole pattern.

iv) The two valley point shouldn't be the smallest value of the whole pattern. (The minimum value of the pattern should be located before ¼ or after 3/4)

f) Triple Bottom: i) The first valley must be higher or equal to the second valley and third.
ii) The third valley should higher than the second valley.
iii) The beginning and ending value should higher than the mean value of whole pattern.
iv) The two peak points shouldn't be the largest value of the whole pattern. (The maximum value of the pattern should be located before ¼ or after 3/4)

If a sequence is called pattern similar to one of the pattern mothers, then we can put this sequence into the pattern family generated by the specific pattern mother.
Put function:

Given a history data with length of n

According to dow theory, the The "medium swing", secondary reaction or intermediate reaction is always lasting from ten days to three months, and with a high probability that it's longer than three weeks

If the family K temp database is empty store the sec
Else {
If the Euclidean distance to the mother is smaller than the
sequence with the sequence now;
Else skip this sequence;
}

See the check similarity f unction and the put function next text window

choose the mother K with the smallest distance, put it into the family K temp database;
Else do nothing;
}
Merge the 8 temp databases with their corresponding global database;
}

3.5 Why this works?

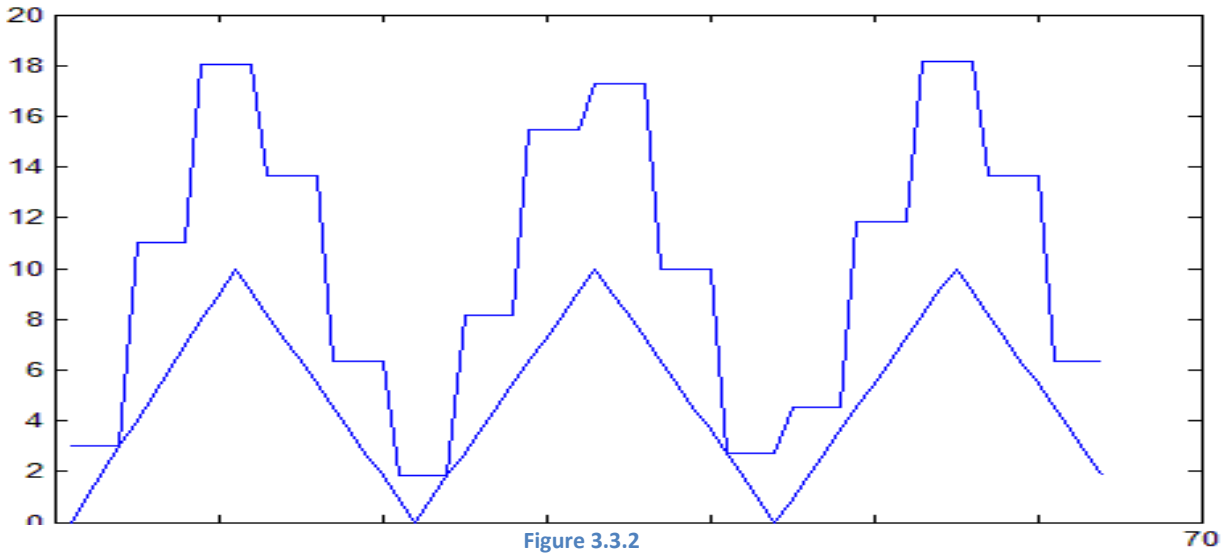


Figure 3.3.2

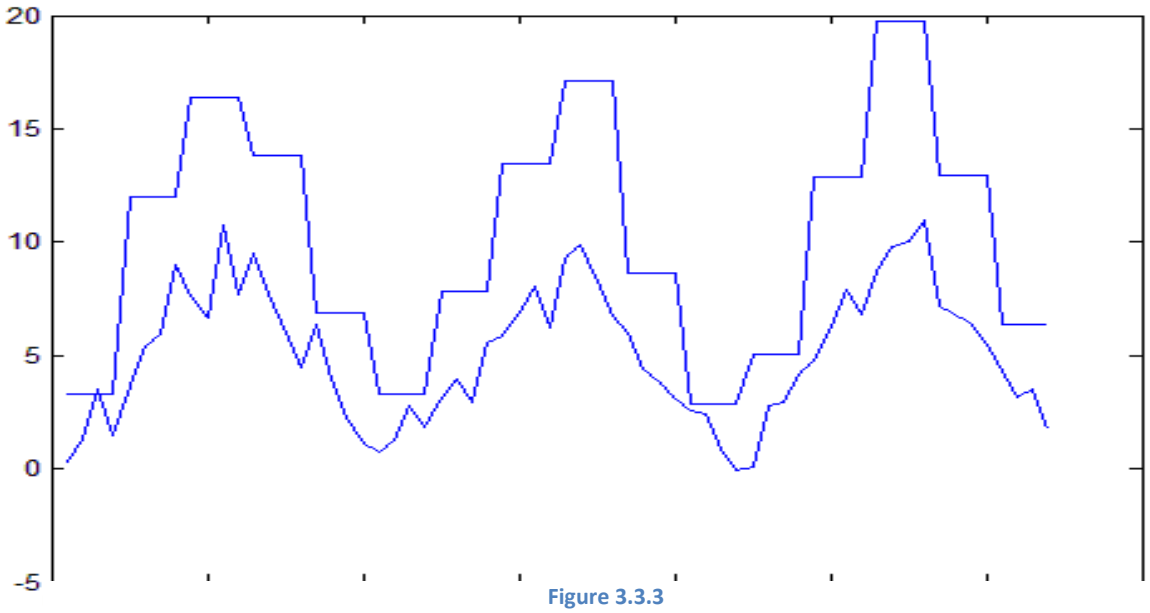


Figure 3.3.3

Let's see the two graph above, the first one is our ideal testing signal and its wavelet transform, and the second one is the combination of the ideal testing signal and Additive white Gaussian noise and its wavelet transform. It's easy to figure out that as long as the trending of the original signal is almost the same; the difference between two wavelet transform signals is quite small.

3.6 A way to match in the relative less complexity:

You can see in the Figure3.3.5 below, in the long time history data sequence, there are many sequences behave monotonically increasing. Imaging you are at the point the arrow pointed, and you are using the last 30 days price to find the similar sequences in the history. For all the data we marked with red circle, there are all broad sense monotonically increasing with different slope and duration. However, with the time wrapping and the zoom in and out of the window slide, the slope and duration don't make any sense (You can see the detail in the matching algorithm). That means, as long as you do the atching for the monotonically increasing sequence, you will get a lot of similar sequence. Unfortunately, you can find that the behavior of the stock price seems quite random after every monotonically increasing sequence. And you can see this from the figure3.3.4.

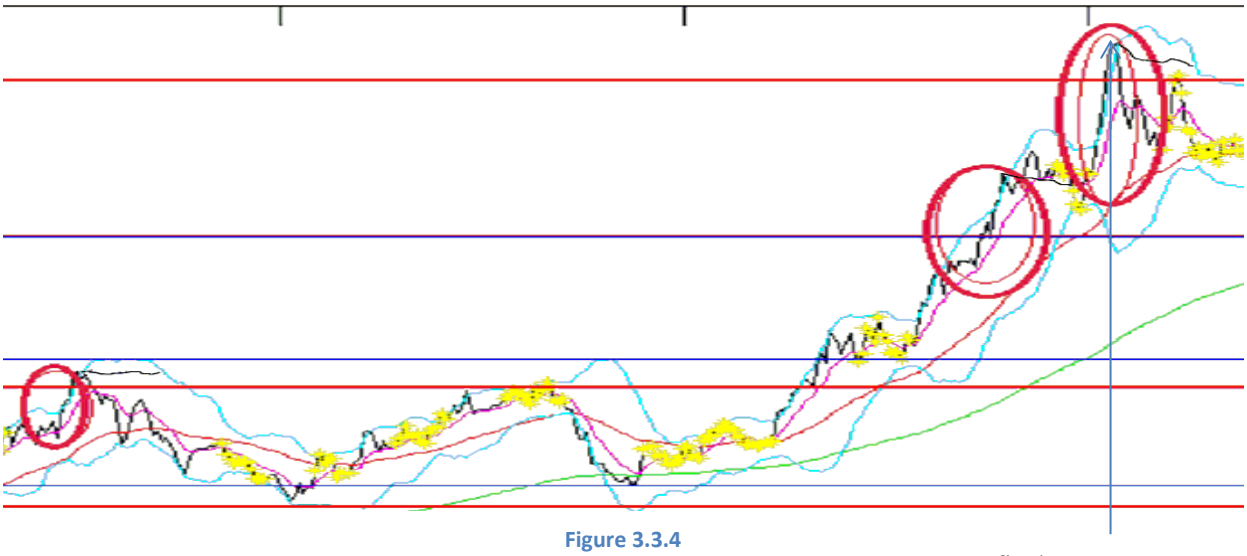


Figure 3.3.4

Now you find a monotone increasing sequence, you want to find the similar data in the history

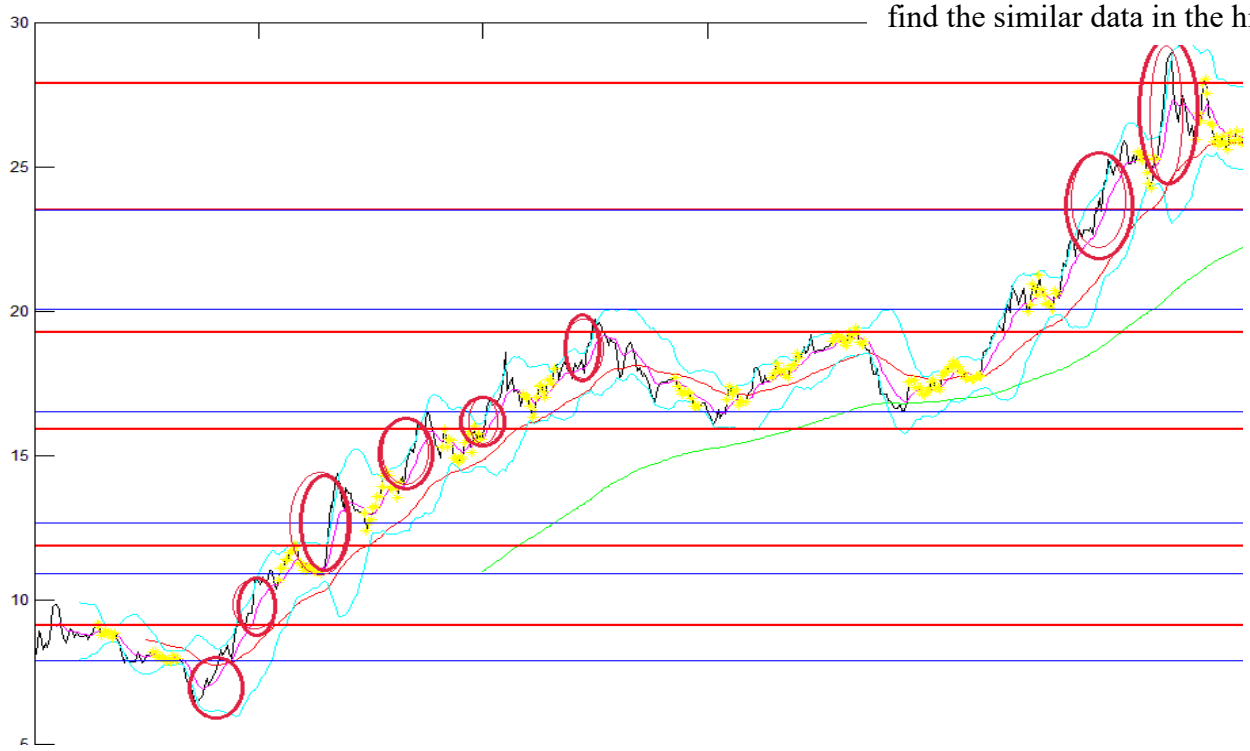


Figure 2.3.5

The result of the matching is similar to a horizontal line. So we decide to skip the broad sense monotonically increasing sequences when doing the pattern matching, for (1) the matching result is not quite well, and (2) to the most important thing, we are doing the pattern matching to find the market signal, and according to the Dow theory, you can never tell how long the trending market will be until it changes to trading market.

And as you can see from the figure6, as long as skip all the broad sense monotonically increasing part, you can greatly decrease the time for matching at least half (obviously larger than half).
Change the time complexity from real time to pre-programming:
As nowadays the stock market is trading online with computers, so when you take the preemptive opportunities for 1 microsecond, so the real time complexity is significantly important than the pre-programming (We can do it in the evening).
So we suggest that do the pattern matching for history first and select out the pattern families as the pattern stencils into the data, when it comes to real time matching, we only matching the last 30(90) sequence with the pattern stencils in the database.

3.7 Algorithm

3.7.1 Check the similarity function:

```
 Do the normalization to the sequencei-j+1 to j;
Get the V vector of the normalized sequence;
Calculate the Euclidean distance1 to the normalized mother patterns' V vectors which are
        Calculated before and stored in the database;
If the Euclidean distance1 is no larger than threshold1 for some mother pattern K
        {
        Calculate the Euclidean distance2 for the sequencei-j+1 to j and the mother pattern K
                                        //Both sequence are normalized
        If the Euclidean distance2 is no larger than threshold2
                {
                Return they are similar;
                }
                Else
                {Return they are not similar ;}
        }
Else{Return they are not similar;}
```

3.7.2 Algorithm of our real time pattern matching:

```
Construct the database of our pattern family;
                                                    //which is explain in the definition of pattern family
Then comes the new price at day W;
Start the matching as the window length equal to 90;
Check the similarity with index length window length/ (database sequence) to the pattern families one by one
{
If some sequence in the database is similar to the window sequence
        {
        Choose the one with the smallest Euclidean distance of real time data to be the matched pattern;
        }
Else
        {
        There is no matched pattern;
        }
}
If there is a matched pattern
{
        Send out the market signal of the pattern family, and the history data of the matched sequence as a reference;
}
Else
{
Do nothing;
}
```

3.7.3 Market signal of the pattern family

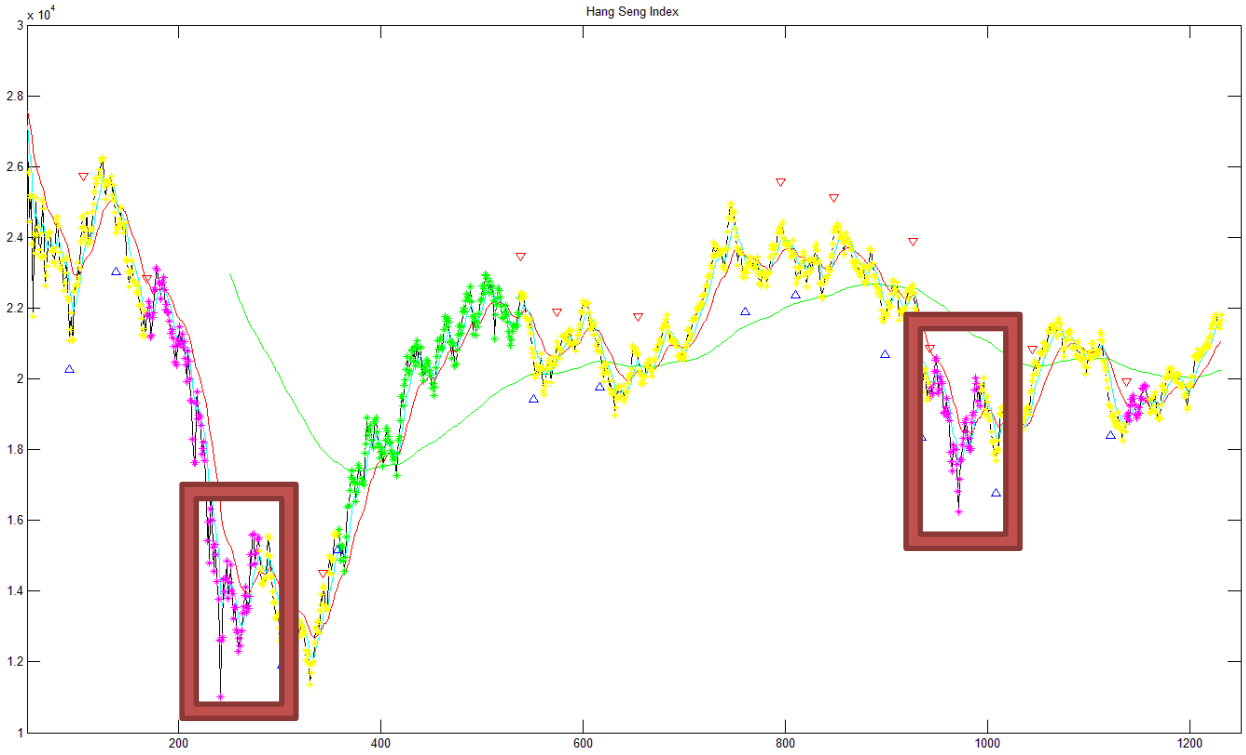
It's a caution to inform the market. It behaves like the lunar calendar. For example, when vernal equinox comes, it informs the farmer to prepare to sow seeds. And the farmer will look at the temperature and hydrometeor to determine the exact time to sow seeds. And the market signal is like the vernal equinox.

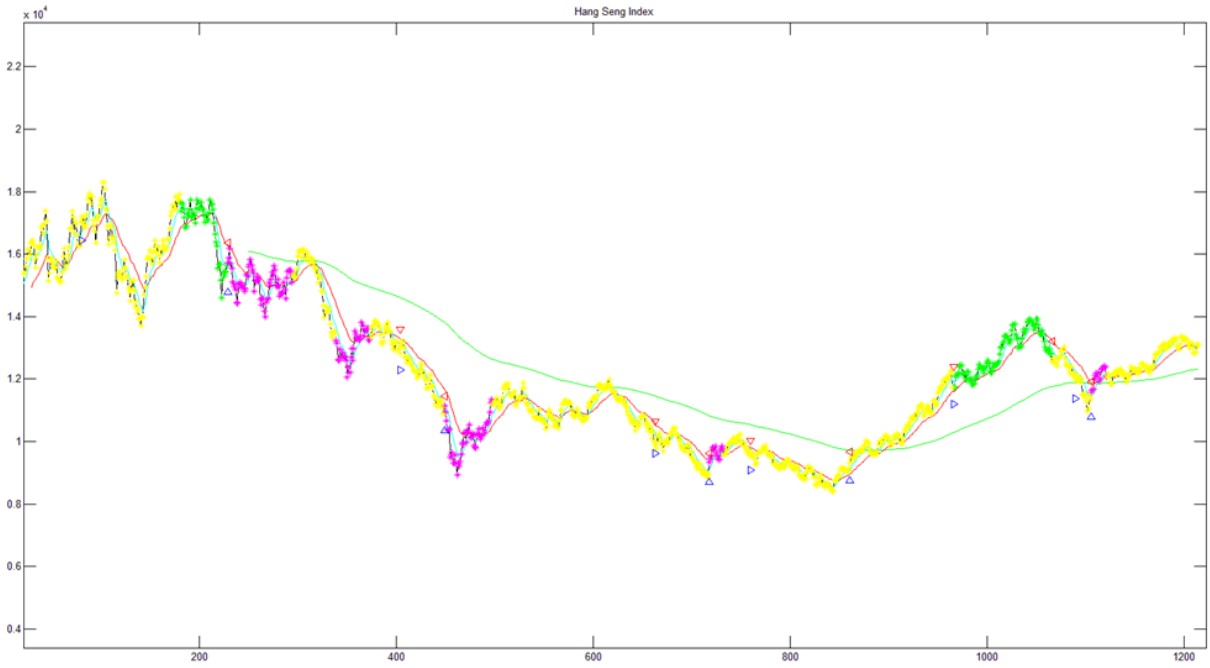
Header and Shoulder(Triple Top)	Strong signal that the market may trend down
Reverse Header and Shoulder(Reverse Triple Top)	Strong signal that the market may trend up
V shape	Heavy Strong signal that the market may trend up
Reverse V shape	Heavy Strong signal that the market may trend down
Double Header	the market may trend down
Reverse Double Header	the market may trend up

IV. Testing Result

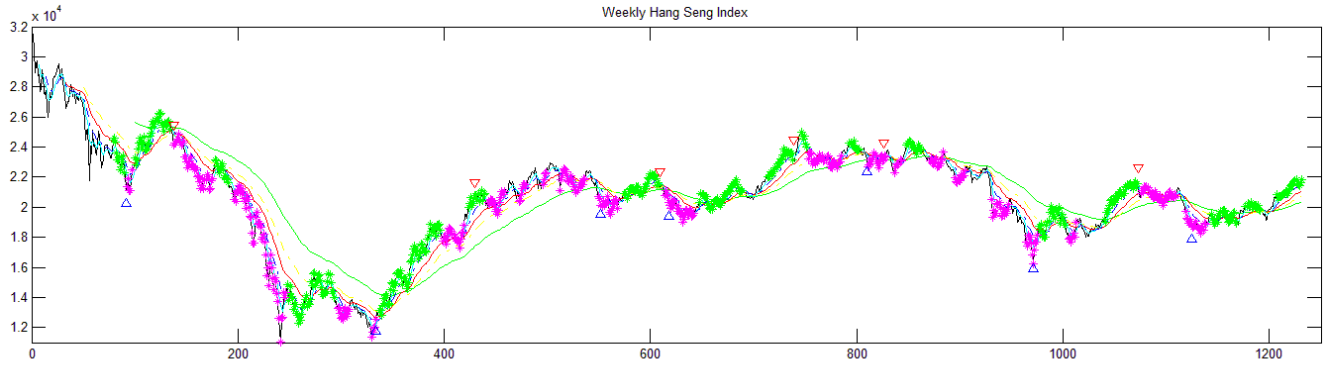
We get the revenue of 24% per year in the time window 2007/11/01~2012/11/01 in the first semester. However, there are still some serious problems.

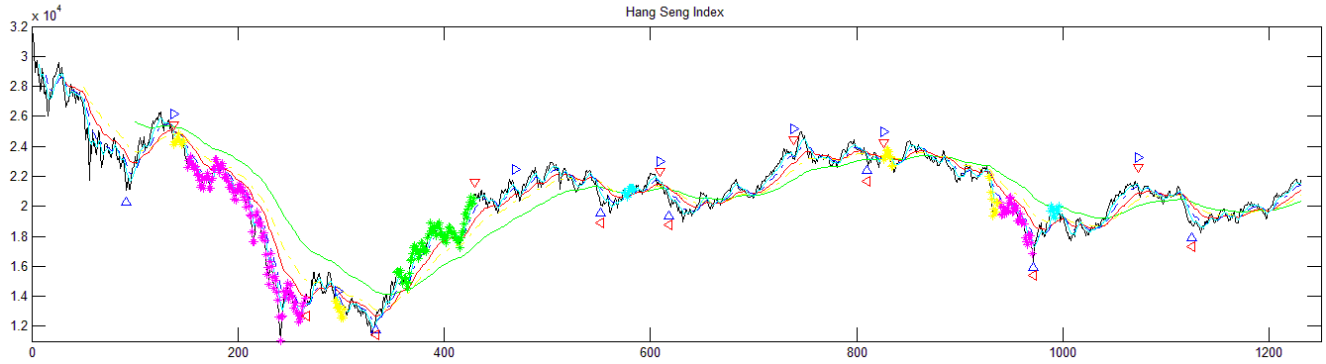
1) The judging of market have a quite large delay.(Signed with □)
2) In the long term trending down market, we can't earn any money.
3) And in some time window like 1999/11/01~2004/11/01 which don't have definite trend, the market judging can be problematical.(only have 8% revenue per year)



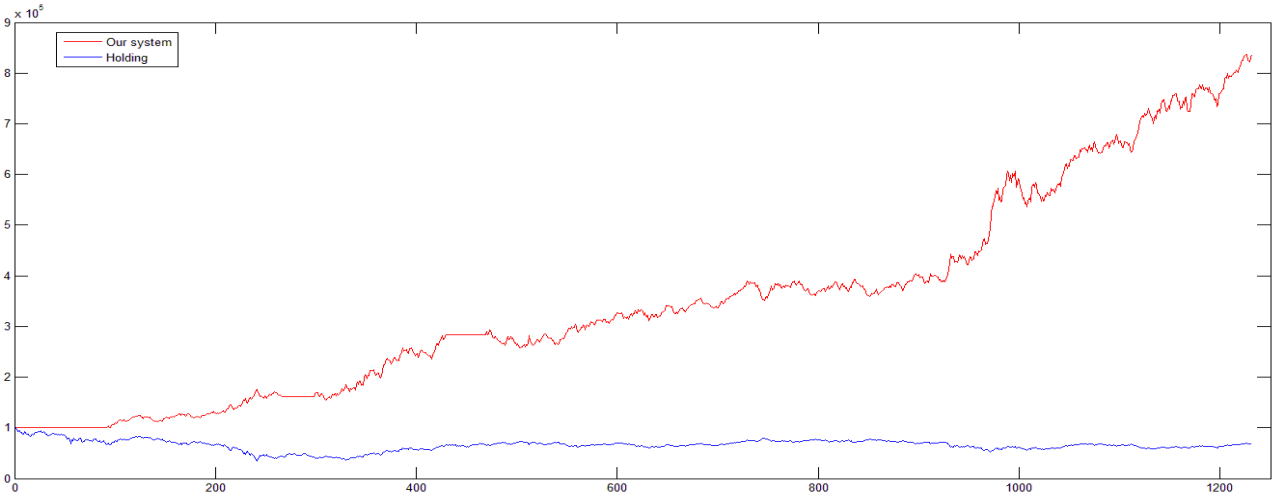


After the modification clarified in the Methodology, we get the revenue of 51% per year in the time window 2007/11/01~2012/11/01.

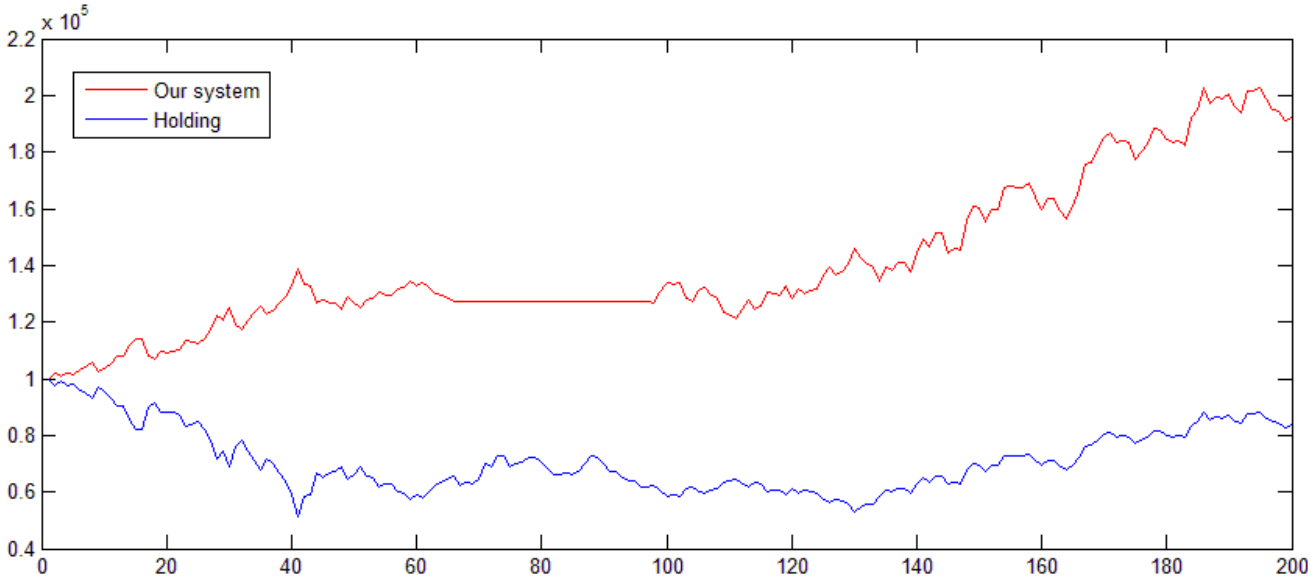




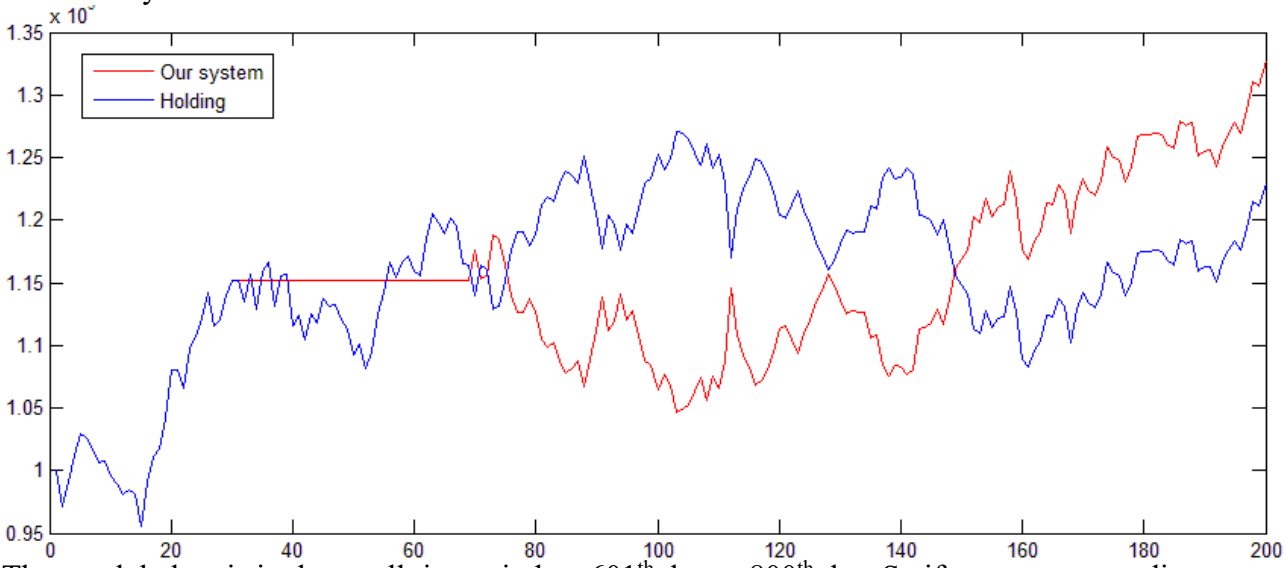
Here the comparison between our system and just holding the market in Hang Seng Index from 2007/11/01~2012/11/01:
For we will need 100 days to collect the data to find the situation of the market, we ride the fence on the market until 101th day. Finally, we can earn 830% money than the start point. (This means 51% per year) Meanwhile, if you just hold on the market, actually you will lose money for the whole trend is trending down.



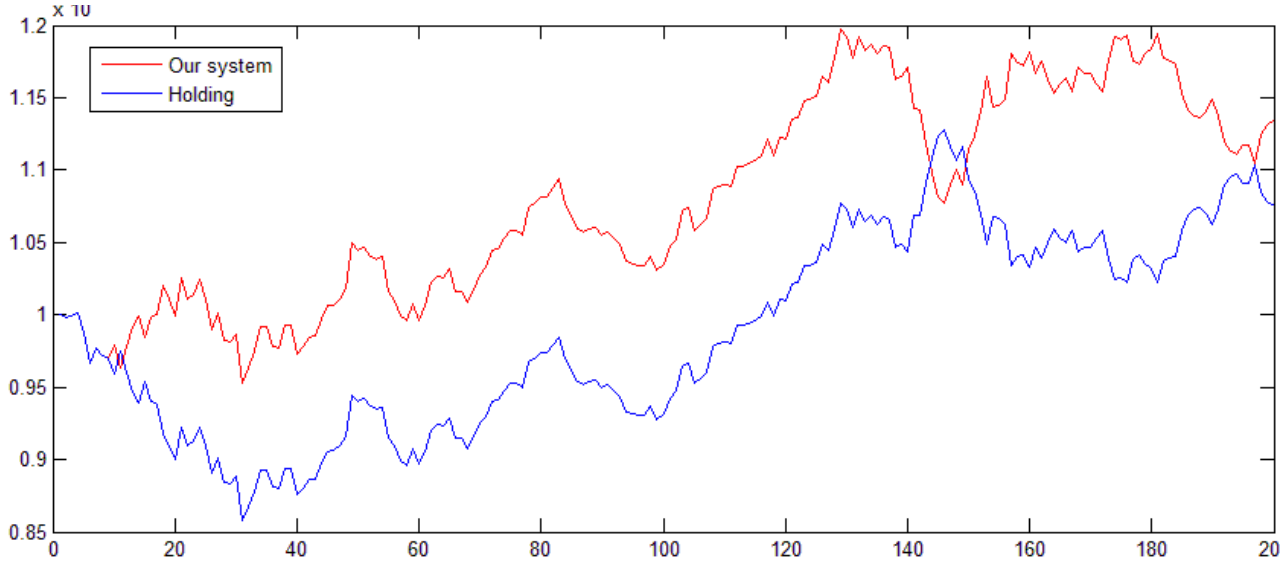
The graph below is in the small time window 201th day to 400th day. So if we start our trading system at 201th day. You can see at the 65th day, we stand out the market, for we find it's too noisy and the last decision we made seems not very exact. We can earn90% money in this 200 days' time window.



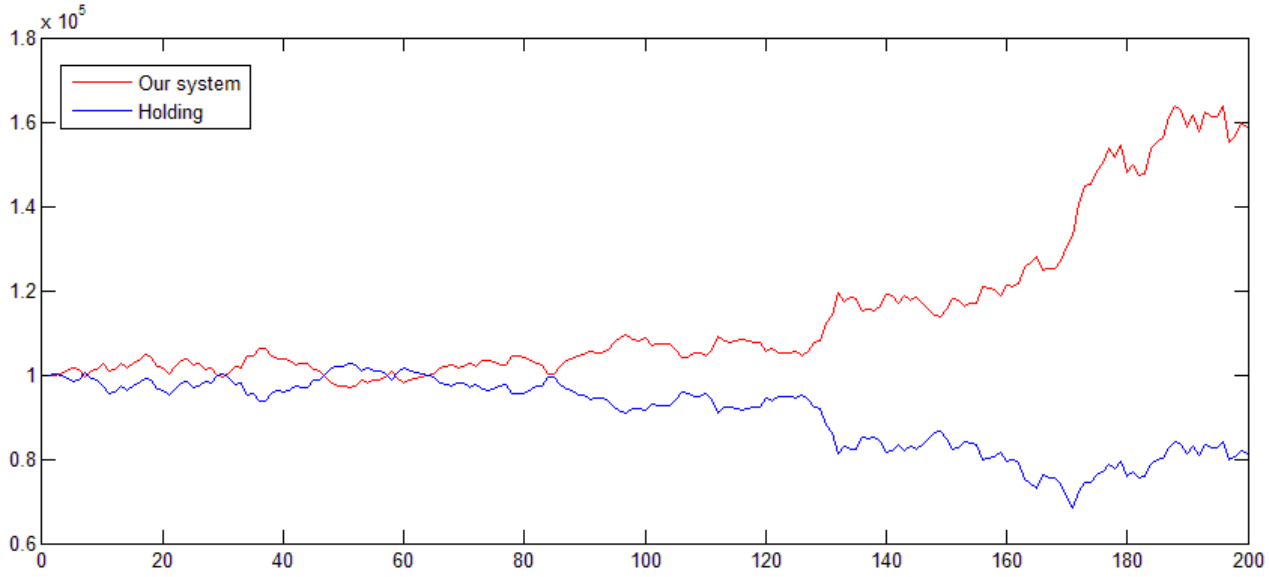
The graph below is in the small time window 401th day to 600th day. So if we start our trading system at 401th day. You can see at the beginning, we follow the trend to trade and see something strange at 30th day. So we stand out the market at 30th. We can earn 33% money in this 200 days' time window.



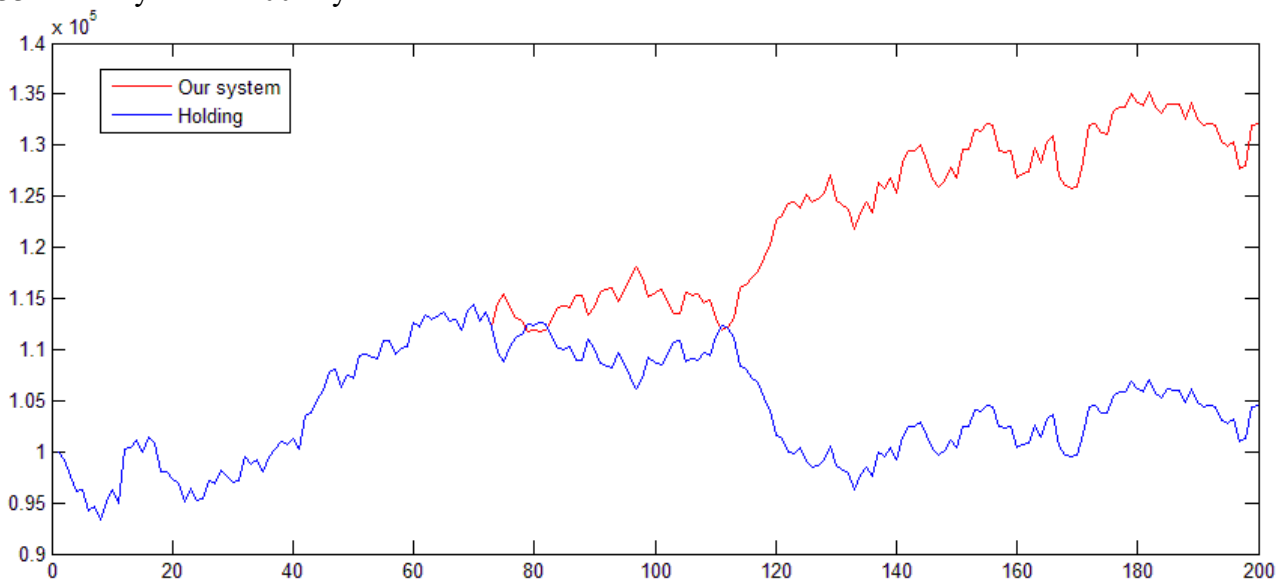
The graph below is in the small time window 601th day to 800th day. So if we start our trading system at 601th day. We can earn 13% money in this 200 days' time window.



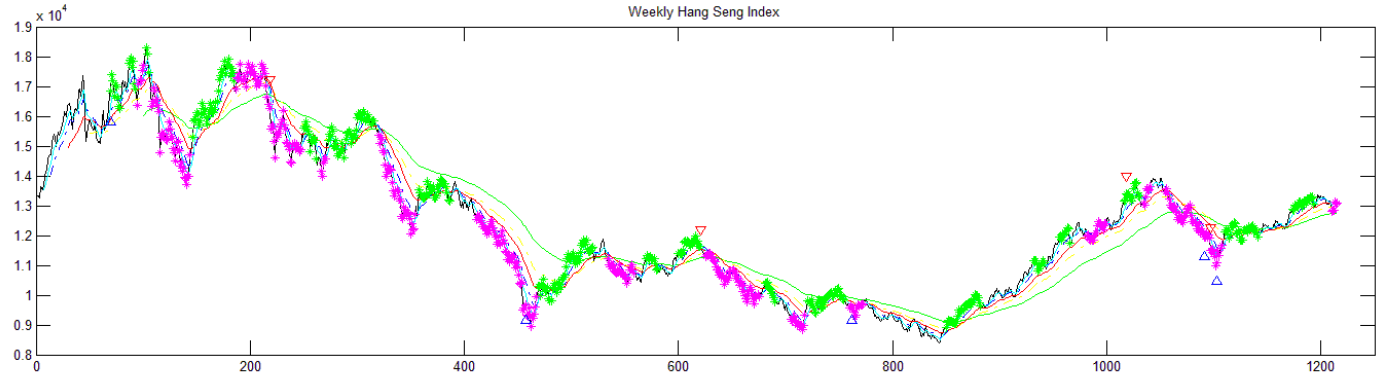
The graph below is in the small time window 801th day to 1000th day. So if we start our trading system at 801th day. In this time window, the price is steady at the beginning and decrease rapidly. And the key point is to sketch the time for buying short. We can earn 60% money in this 200 days' time window.

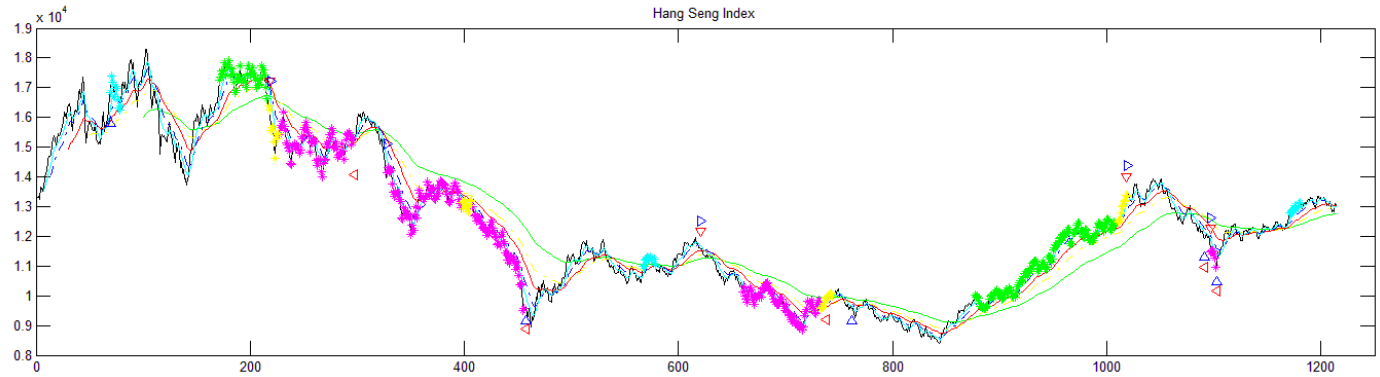


The graph below is in the small time window 1001th day to 1200th day. So if we start our trading system at 1001th day. You can see at the beginning, we follow the trend to trade. We can earn 33% money in this 200 days' time window.

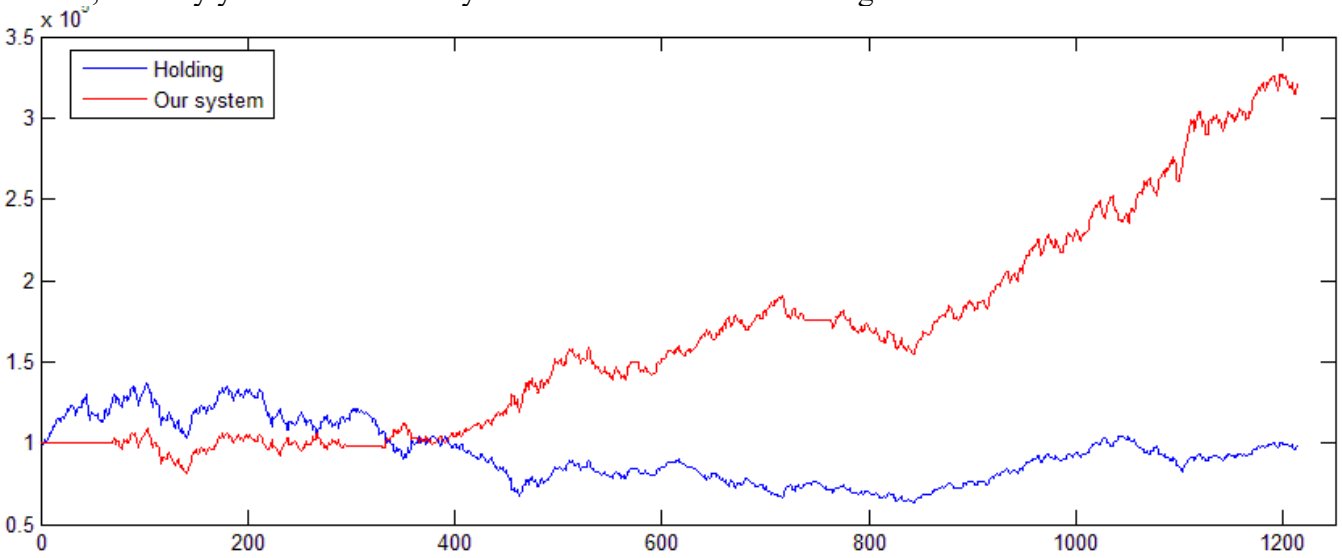


On the others hand, the other main contribution of the new algorithm is that it can earn money in different time window, no matter how the market goes on, let see another time window of Hang Seng Index.(1999/11/01~2004/11/01).

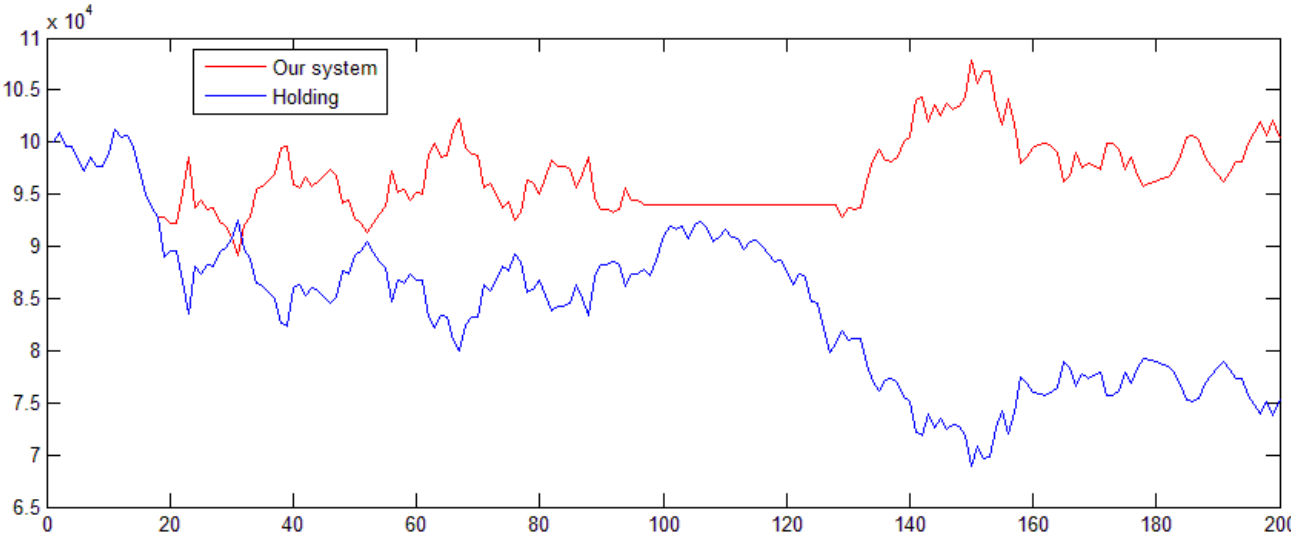




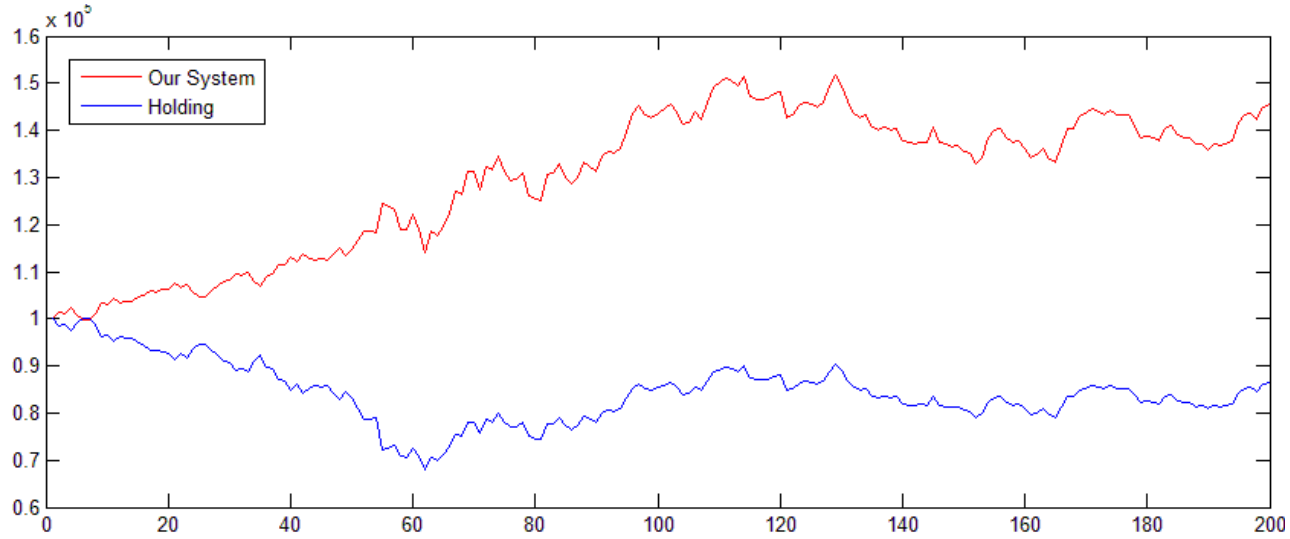
Here the comparison between our system and just holding the market in Hang Seng Index from 1999/11/01-2004/11/01:
For we will need 100 days to collect the data to find the situation of the market, we ride the fence on the market until 101th day. In this time window, the market is increasing at the early beginning, so we actually start at an inferior position. However finally, we can earn 350% money than the start point, which means around 30% profit per year. Meanwhile, if you just hold on the market, actually you will lose money for the whole trend is trending down.



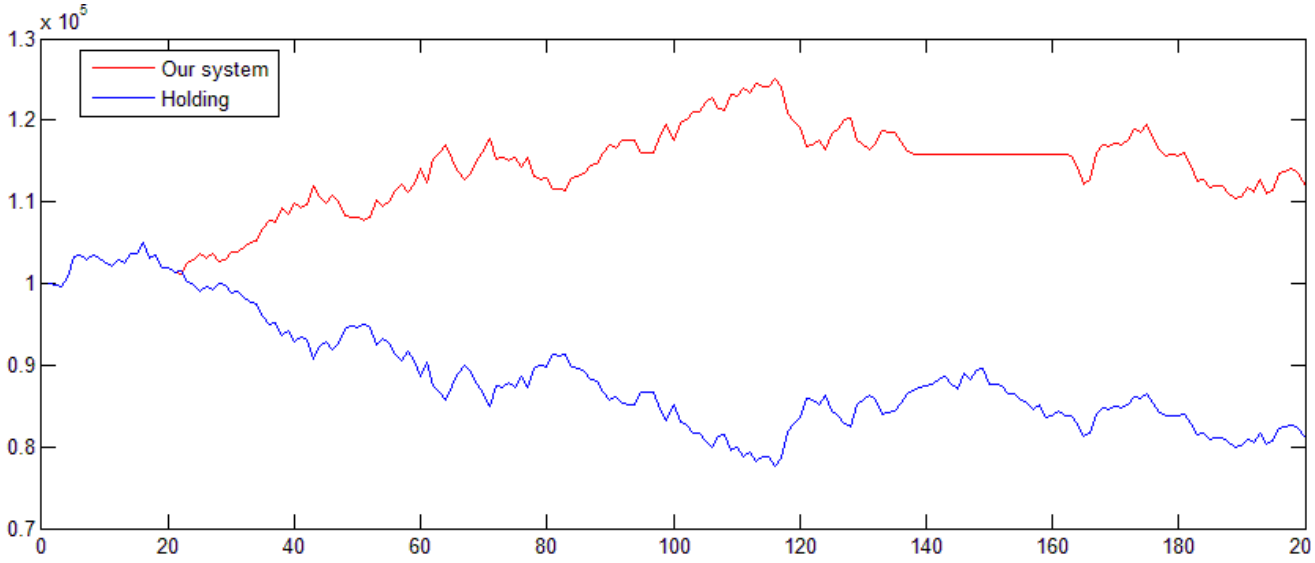
The graph below is in the small time window 201th day to 400th day. So if we start our trading system at 201th day. Although you can see from the price at the last page, the noise of the period is quite big, and the price doesn't behavior normally and don't have a trend. But we will not lose money at this period, for we can correct the wrong decision. You can see at the 95th, we stand out the market, for we find it's too noisy and the last decision we made seems not very exact.



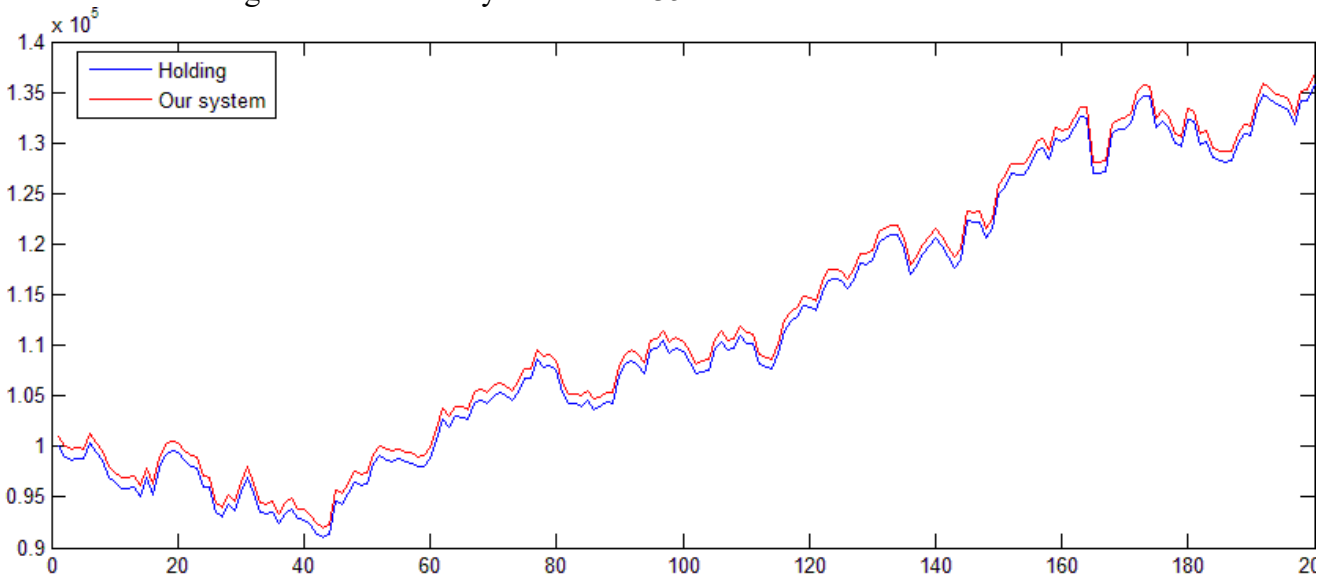
The graph below is in the small time window 401th day to 600th day. We earn about 50% in this time window.



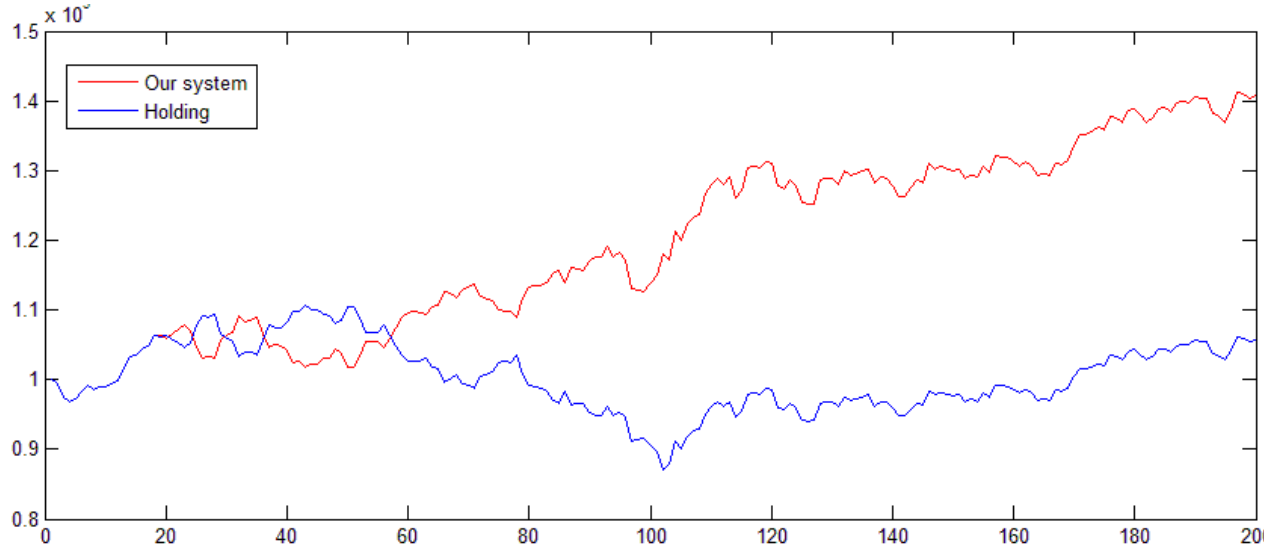
The graph below is in the small time window 601th day to 800th day. We earn about 12% in this time window.



The graph below is in the small time window 801th day to 1000th day. For this time window, the market is in a long term trending up, so we just follow the trend, so the curve is exactly same as the curve of holding the market. And you can earn 35% in this time window.



The graph below is in the small time window 1001th day to 1200th day. We can earn 41% in this time window.



V. Conclusion and future direction

In our project, we designed a stable auto-trading system which can make 30%~ 50% profits per year in different time window and with high probability we can declare that we will never lose money at a medium period. We mainly use two methods to do trading, those are central limit theorem and the pattern matching. We not only combined these two methods but also designed our own indicator (SMACD) and our modified pattern (V shape). Furthermore, we do bi-directional trading to maximize our profit.

And we can use the system to build Android Application and Website to susgest users to do the correct choice of trading.

VI. Reference

1. Hang Seng Indexes http://sc.hangseng.com/gb/www.hsi.com.hk/HSI-Net/HSI-Net
2. 【港 黃柏中】技术分析原理_金融分析指标及买卖系统大全
3. StochCharts.com – ChartSchool http://stockcharts.com/school/doku.php?id=chart_school
4. Investopedia http://www.investopedia.com/
5. Applied Data Mining statistical methods for business and industry, Paolo Giudici
6. Artificial Intelligence through Simulated Evolution, L. J. Fogel, A. J. Owens, and M. J. Walsh
7. Electronics, Circuits and Systems, 2004. ICECS 2004. Proceedings of the 2004 11th IEEE International Conference on
8. Topic(s): Components, Circuits, Devices & Systems
9. Digital Object Identifier: 10.1109/ICECS.2004.1399598
10. Publication Year: 2004
11. A New Evolution System for Evolving Artificial Neural Networks, Xin Yao, Yong Liu
12. Artificial Intelligence A Modern Approach, by Stuart Russell and Peter Norvig.
13. Applied optimal control: optimization, estimation, and control, by Arthur Earl Bryson, Yu-Chi Ho
14. Neural Networks for Pattern Recognition, by Christopher M. Bishop
15. 12.Bates, R. G. ; Dempster, M. A. H. ; Romahi, Y. S. Evolutionary reinforcement learning in FX order book and order flow analysis. Proceedings of the 2003 IEEE International Conference on Computational Intelligence for Financial Engineering, Hong Kong, March 2003, 355–362.
16. Bouchard, J. ; Mezard, M. ; Potters, M. Statistical Properties of Stock OrderBooks: empirical Results andModels. Quantitative Finance, vol. 2, no. 4, Aug. 2002.
17. Chiu, Kai-Chun ; Xu, Lei. Arbitrage pricing theory-based Gaussian temporal factor analysis for adaptive portfolio management. Decision Support Systems, Volume 37, Issue 4, September 2004, Pages 485-500.
18. R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity search in sequence databases. In *Proc. of the FODO Conf.*,volume 730. Springer, 1993.
19. K. Chakrabarti and S. Mehrotra. The Hybrid Tree: An index structure for high dimensional feature spaces. In Proc. of the ICDE Conf., pages 440–447, Sydney, Austrialia, 1999.
20. Rakesh Agrawal, Christos Faloutsos, and Arun Swami. Efficient similarity search in sequence databases. In Foundations Of Data Organizations and algorithms (FODO) conference, October 1993.
21. C.Faloutsos, M.Ranganathan, and Y.Manolopoulos. Fast subsequence matching in time-series databases. In Intl.Conf. on Management of Data –SIGMOD 94, pages 419-429, Minieapolis, May 1994.
22. C. Faloutsos, M. Ranganathan, and Y. Manolopoulos. Fast subsequence matching in time-series databases. In Proc.of the ACM SIGMOD, pages 419–429, Minneapolis, Minnesota,USA, May 1994.
23. Yi-Leh Wu, Divyakant Agrawal, Ame EI Abbadi. A Comparison of DFT and DWT Based Similarity Search in Time-Series Databases. Department of Computer Science University of California, Santa Barbara.

24. Tsong-Li Wang,Dennis Shasha. Query processing for distance metrics. Proceedings of the sixteenth international conference on Very large databases Pages 602-613
25. Franz Aurenhammer (1991). Voronoi Diagrams – A Survey of a Fundamental Geometric Data Structure. ACM Computing Surveys, 23(3):345–405, 1991
26. N.Beckmann, H.-P.Kriegel, R.Schneider, and B.Seeger, "The R*-tree:an efficient and robust access method for points and rectangles", ACM SIGMOD, pages 322-331, May 1990.
27. A.Guttman, "R-trees: a dynamic index structure for spatial searching", Proc. ACM SIGMOD, pages 47-57, June 1984.
28. H.Samet, The Design and Analysis of Spatial Data Structures, Addison-Wesley, 1989
29. J.Nievergelt, H.Hinterberger, and K.C.Sevcik, "The grid file:an adaptable, symmetric multikey file structure", ACM TODS, 9(1):38-71, March 1984.
30. Norbert Beckmann, Hans-Peter begel,Ralf Schneider, Bernhard Seeger. The R*-tree: An Efficient and Robust Access Method for Points and Rectangles+ 1990 ACM 089791365 5/!90/0@35/0322
31. R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity searchin sequence databases. In Proc. of the FODO Conf.,volume 730. Springer, 1993.
32. Parseval des Chênes, Marc-Antoine "Mémoire sur les séries et sur l'intégration complète d'une équation aux différences partielles linéaire du second ordre, à coefficients constants" presented before the Académie des Sciences (Paris) on 5 April 1799. This article was published in *Mémoires présentés à l'Institut des Sciences, Lettres et Arts, par divers savans, et lus dans ses assemblées. Sciences, mathématiques et physiques. (Savans étrangers.)*, vol. 1, pages 638-648 (1806).
33. Katznelson, Yitzhak (1976), An introduction to Harmonic Analysis, Dover, ISBN 0-486-63331-4
34. Davood Rafiei and Alberto Mendelzon, Similarity_Based Queries for Time Series Data, 2002, IEEE.
35. Jensen, J.R, Remote sensing of the environment an Earth resource perspective, 2nd edition, Prentice Hall. ISBN 0-13-188950-8, 2007.

VII. Appendix A. Logbook

Term 1:

Pre meeting:
12th, September

1) Read the papers given by professor again and drew the outline
   Conclusion: The papers use some indicators such as RSI to predict the future trending of the data, which is the most widely used indicator in the analysis.
2) To do:
   Find more papers by other university and social analyzer s.

Deadline: 20th, September

Pre meeting 2:
20th, September

1) To do:
   Draw Outline.

Deadline: 25th, September

First meeting:
27th, September
Content:

1) Divided work to do.
   We should have a feeling of the finance indicators and analysis.
   There is a book named: 技术分析原理_金融分析指标及买卖系统大全.
   Steven: Chapter2: Market Trending indictors (Moving Average, Bollinger's Band)
   Chapter4: Market cycle indictors (MACD STC SMI Stoch RSI)
   Allay: Chapter3: Market Trading Indictors (RSI RMI TSI)

Deadline:11th, October

Second meeting:
13th, October
Content:

What we should do now?
Allay: Select the indicators from so many indicators and Study how can we judging a market is what kind of market, trending or trading.
Steven: Select the indicators from so many indicators and write the outline of the analysis part.

Deadline: 19th, October

Third Meeting:
20th, October

To do:
Doing the preliminary schedule into Matlab do the test:
Allay: The ADS and Stoch RSI
Steven: Bollinger's bound

Deadline: 28th, October

Forth meeting:
To do:
Steven: Translate the algorithm of Bollinger into codes which is containing in the main function not only the indicator function.
Allay: Translate the algorithm of ADS and Stoch RSI into codes which is contain in the main function not only the indicator function.

Deadline: 6th, November

Fifth meeting:
6th, November
To do list:
Both of us to find a way to detect the header and shoulder

Sixth meeting:
13th, November
To do list:
Find an exactly to find the resistance and support in the stock.

Seventh meeting:
22th, November
To do list:
We should make the whole process into a completed one.
Steven: Pattern matching part using data mining
Allay: Central limit part to modify the trading rule

Eighth meeting
25th Nov

1. Outcome: the trading rule and first version of pattern matching.
2. To do list:

   Steven: Write the pattern matching code

   Allay: The neutral network to combine the system signal and the indicators.

Ninth Meeting
30th Nov

1. Outcome: The first version code and an outline of the neutral network

2. To do list:

   Steven: Write the thesis for his part
   Allay: Write the thesis for his part

Tenth Meeting
4th Dec

1. The part of each one's thesis

2. To do list:

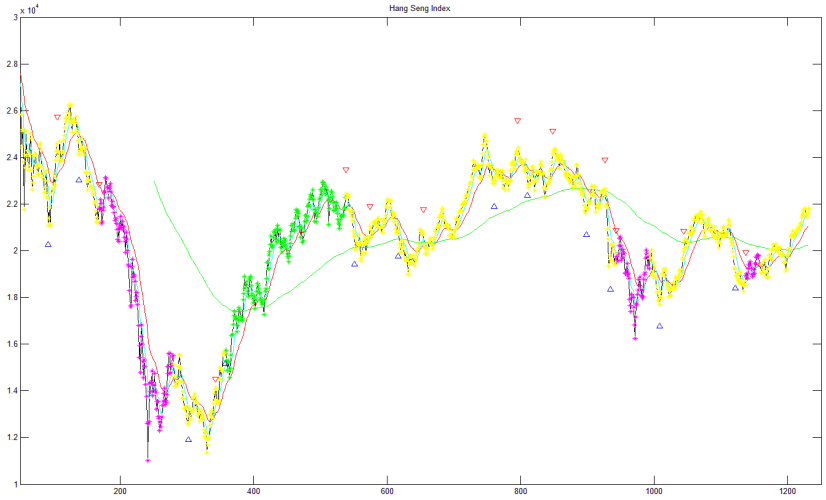
   Merge into one.

Term2:

Pre meeting:
12th, January

1) Review the process of project

   Conclusion:
   Last semester, we designed an auto-trading system to trade the Hang Seng Index (HSI) during the Nov. 2007 to Nov. 2012. We did the one-direction trading and achieved around 20% profit per year.



2) Problem found:
   a) We did not test our trading system in other time window.
   b) We did not trade in two directions to maximize the profit.

3) To do:
   a) Test our trading system in other time window.
   b) Trade in two directions to maximize the profit.

Deadline: 30th, January

Pre meeting 2:
30th, January

1) Review:
   Allay: Collect the data from 1998 to 2003, and test the system
   Steven: Try to trade in two directions.

2) Problem found:
   a) Our system does not work well in other time window, which means the system is not robust.
   b) Fourier Transform does not work well in bi-direction trading.

3) To do:
   a) Find method to improve our algorithm to determine the market is trending or not.
   b) Find other method to do the bi-direction trading.

Deadline: 5th, February

First meeting:
7th, February
Content:

1) Report the problems in our trading system

   a) Our system does not work well in other time window, which means the system is not robust.
   b) Fourier Transform does not work well in bi-direction trading.

2) Share our idea to improve the our system

   a) Tune the parameters in our system to make it work well in different time window.

   b) Use Wavelet Transform instead of Fourier Transform to do the bi-direction trading.

Deadline:21st, February

Pre meeting 3:
20th, February

1) Review:
   Allay: Try to tune the parameters in our system to make it work well in different time window.
   Steven: Try to trade in two directions by using Wavelet Transform to do pattern matching.

2) Problem found:
   a) No matter how the parameters are tuned, the algorithm of determining the market is trending did not work well.
   b) Need to rewrite the code to use the Wavelet Transform to do pattern matching.

3) To do:
   a) Design a new algorithm to determine the market is trending or not.
   b) Implement the Wavelet Transform to do the bi-direction trading.

Deadline: 4th, March

Third Meeting:
4th, March
Content:

1) Report:
   Allay: Try to use Gold Cross and Dead Cross to determine the market.
   Steven: Implement the Wavelet Transform to do the bi-direction trading.
   By trading in bi-direction, the profit was improved from around 20% per year to around 40% per year.

2) Professor's Advice:
   There are 3 layers of data in the stock market – weekly, daily, and hourly.
   The higher layer would master the lower layer, which means the daily data could not break the rule of the weekly data.

3) To do:
   a) Use the weekly data to determine the trend of the market in long run.
   b) Modify the structure of the code (separate the main function into different function), which will benefit and convenient for parameter changing test.
   c) Optimize the logic of the code which makes the computer occupation lower.

Deadline: 19th, March (After the last midterm)

Forth Meeting:
20th, March
Content:

1) Report and problem:
   Allay: Use the Gold Cross and Dead Cross to determine the market. Find the weekly data is much slower than the dairy data, which will make our decision slower and loss some opportunities to make money.
   Steven: Modify the structure and do the optimization.

2) Discussion and searching the solution:
There is a model named ARIMA, which is used to predict the price of water & electricity fees. So we can use the ARIMA to predict the weekly data to release the delay.

3) To do:
   a) Use the ARIMA to predict the weekly data.
   b) Use the new market signal to do the trading.

Deadline: 27th, March

Fifth Meeting:
28th, March
Content:

1) Report:
   Allay: Implement the prediction of the weekly data with the function in Matlab.

   Steven: Use the new market signal to do the trading. Find the revenue is not very high.

2) Problem:

a) The ARIMA can predict the trending of the price but can't predict the changing of the price. We can't predict
b) The new market signal is not precise and has some error.

3) Discussion:

a) According to the Dow Theory, the long run trending market has some law of statistic, such as the time window is usually 8 to 12 month.

b) The problem in the new market signal is that in the trading market, the cross will appear frequently.

4) To do:

a) Add the rule of time window into the market signal

b) To move the cross in the trading market.

Deadline: 8th April(After the Easter Day)

Sixth Meeting:
9th, April
Content:

1) Report:
   Allay: Add the rule of time window into the judging of market signal.
   Steven: Implement using matrix of the indicators to help judging the market signal.

2) Problem:
   a) The error is erased, but the delay is still very serious.
   b) Find the matrix need large standard data for bull markets; we need to collect the data again.

3) To do:
   a) Decrease the influence from the delay.
   b) Collect the data for the bull market in the history.

Deadline: 18th April (After the Easter Day)

Seventh Meeting:
19th, April
Content:
1) Report:
   Allay: Do some modification and optimization.
   Steven: Finish Implement using matrix of the indicators to help judging the market signal.
2) Problem:
a) The delay in second term trending will case loss of money.
b) The matrix pattern only works for the long run.

3) To do:
a) Do pattern matching to help judging the market.

Deadline: 27th April

Eighth Meeting:
29th, April

1) Report:
   Design our own pattern "V shape" and do the testing.

2) Result:
a) The profit from 2007 to 2012 was improved from 40% to 50% per year.
b) The profit from 1999 to 2004 was improved from 20% to 30% per year.

3) To do:
a) Write the thesis and design the poster.

Deadline: 20th May