

STOCK INVESTMENT DECISION SUPPORT FOR HONG KONG MARKET USING RBFNN BASED CANDLESTICK MODELS

WING W. Y. NG, XUE-LING LIANG, PATRICK P. K.CHAN, DANIEL S. YEUNG

Machine Learning and Cybernetic Research Center, School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
E-MAIL: wingng@iee.org

Abstract:

Candlestick pattern, as an efficient method in technical analysis, is widely used in decision support of stock investment. From historical data, there is a no 100% guarantee for a stock price increasing after the appearance of a bullish candlestick pattern. The main aim of this paper is to enhance the prediction ability of Candlestick Patterns using a Multiple Classifier System (MCS) consisting of Radial Basis Function Neural Network (RBFNN) trained by a Localized Generalization Error Model (L-GEM). The RBFNN classifies particular candlestick pattern to be a real bullish candlestick pattern or not based on training with past data. The MCS fusing RBFNN for different patterns makes the final prediction of the stock price trend. In this paper, stock price data of 40 stocks in Hong Kong Hang Seng Component Index is collected to carry out the investment simulation experiment. Experimental result shows that the proposed method yields statistically significant profit when compared with a random investment strategy and candlestick investment without RBFNN.

Keywords:

Candlestick Pattern; Stock Investment; Radial Basis Function Neural Network; Localized Generalization Error Model

1. Introduction

Stock is a product of market economy. The circulation and trading of stocks promote the development of market economy. Stock analysis and prediction are vital to the success of stock investment. Fundamental analysis and technical analysis are two major branches in stock analysis and prediction. Fundamental analysis predict future stock price trend by analyzing the basic condition and determinant which make influences to stock price, such as current economic status, overall industry situation, and company's financial status, etc.[1]. In contrast, technical analysis [2, 3] is a method that analyzes future stock price trend from the aspect of stock market behavior. Among all technical analysis methods, candlestick chart is one of the

most widely adopted technical analysis technique in stock investment. However, the Efficient Market Hypothesis (EMH), proposed by Eugene Fama in [4], states that technical analysis on stock is invalid because all known factors affecting the stock price has been reflected by current stock price. In [5], it also tells us that making a very high return via fundamental analysis and technical analysis is very difficult.

In fact, more and more experts in investment field take the EMH as a conservative academic opinion. Technical analysis is still widely adopted in stock prediction. Harvey finds that the return in developing stock market is easier to be affected by local information than in developed stock market [6]. Authors in [7] show that the forecasting stock price based on historical data could gain high profit while the EMH still hold true. Sometimes, we may even break the EMH in New York Market [1].

In resent years, people did diversified researches on stock price prediction. Neural Network, with the virtue of self-organization and adaptability, has been widely used in this field. In [8], a three-layered feed forward neural network is utilized as a prediction model for stock market. A genetic network programming technique was adopted to enhance the stock trading model in [9]. An improving support vector machine (SVM) learning paradigm with a mixed kernel is used to predict stock market movement direction in terms of historical data series [10]. In this paper, we adopted Radial Basis Function Neural Network (RBFNN) trained by the Localized Generalization Error Model (L-GEM) to enhance candlestick patterns. A single candlestick pattern may not yield a good prediction performance, therefore we propose a Multiple Classifier System (MCS) with 10 RBFNNs to generate the final buy signal. Each RBFNN corresponds to classify a candlestick pattern.

The rest of the paper is organized as follows: we describe the proposed investment decision support system with MCS consisting of RBFNN trained by L-GEM in

Section 2. Section 3 introduces the 10 candlestick patterns adopted in this work and their data representations. Experimental results are presented and discussed in Section 4. Section 5 concludes this work.

2. Investment Decision Support System based MCS with RBFNN Trained by L-GEM

If a buy signal is generated from the MCS based on existing candlestick patterns of today and previous days, we will buy the stock in the next trading day. Since the candlestick technical analysis is useful to forecast short-term price movements and its effective period is usually less than ten days [11]. In this work, we will sell the stock within 3 days. This “3-day investment strategy” may not be the best trading method, but we use it as reference to compare different buy signal prediction methods. The stock will be sold if the closing price is higher than the buy-in price within three days. Otherwise, we sell it at the third day. The major research problem in this work is to predict the buy signal correctly with a high probability, i.e. high generalization capability. In the following sub-sections, we will introduce the method for buy-in signal prediction based on candlestick patterns and RBFNN trained by the L-GEM.

2.1. RBFNN trained by the L-GEM

Radial Basis Function Neural Network (RBFNN) is widely adopted in machine learning and financial prediction. RBFNN consists of the input layer, the hidden layer and the output layer. Its linear relationship between the hidden layer and the output layer makes it fast in training. Therefore, RBFNN is selected as the based classifier of the MCS. In our work, the number of input neurons depends on the number of features describing particular candlestick patterns and their data representation. Each RBFNN (base classifier) recognizes the real bullish candlestick from a given candlestick pattern. So we need only one output neuron to forecast the appearance of specified candlestick pattern. The number of hidden neurons determines the architecture of a RBFNN. So the selection of a suitable number of the hidden neurons becomes a key architectural component with respect to the performance of the RBFNN. Once the number of hidden neurons is determined, we can use standard learning algorithms (e.g. k-means clustering) for centers and widths and least square method for weights. Therefore, RBFNN is very suitable for large scale classification problems like stock market prediction.

As it is mentioned above, the selection of hidden neurons is a key point. We select the architecture (i.e. number of hidden neurons) of RBFNN yielding the least

Localized Generalization Error Model (L-GEM). Compare with other training algorithms for RBFNN (e.g. sequential learning and ad-hoc choice), the L-GEM enhances the generalization capability of RBFNN more effectively [12]. It has been applied to image classification [13] and bankruptcy prediction [14], which achieves good results.

The L-GEM provides an upper bound for the mean square error of a RBFNN for unseen samples within a Q-union of training samples. With probability 1- η , the localized generalization error(R^*_{SM}) of a RBFNN is:

$$R_{SM}(Q) \leq \left(\sqrt{R_{emp}} + \sqrt{E_s((\Delta y)^2)} + 1 \right)^2 + \varepsilon = R^*_{SM}(Q) \quad (1)$$

Where R_{emp} , $E_s((\Delta y)^2)$ denote the training mean square error and the stochastic sensitivity measure, respectively. For RBFNN using Gaussian activation function, the stochastic sensitivity can be expressed as follow:

$$E_s((\Delta y)^2) \approx \sum_{j=1}^M \varphi_j \left[\frac{\sum_{i=1}^n \sigma_{\Delta x_i}^2 (\sigma_{x_i}^2 + (\mu_{x_i} - u_{ji})^2 + 0.2\sigma_{\Delta x_i}^2)}{v_j^4} \right] \quad (2)$$

$$\text{where } \varphi_j = (w_j)^2 \exp\left(\left(Var(s_j)/2v_j^4\right) - \left(E(s_j)/v_j^2\right)\right),$$

$$s_j = \|x - u_j\|^2, \quad E(s_j) = \sum_{i=1}^n (\sigma_{x_i}^2 + (\mu_{x_i} - u_{ji})^2),$$

$$Var(s_j) = \sum_{i=1}^n \left(\frac{E_D[(x_i - \mu_{x_i})^4] - (\sigma_{x_i}^2)^2}{4E_D[(x_i - \mu_{x_i})^3] (\mu_{x_i} - u_{ji}) + 4\sigma_{x_i}^2 (\mu_{x_i} - u_{ji})^2} \right)$$

and n , w_j , v_j , u_{ji} , μ_{x_i} denote the number of features, the connection weight between the j^{th} hidden neuron and the output neuron, the width of the j^{th} hidden neuron, the i^{th} input feature of the j^{th} center of the hidden neuron, the expected value of the i^{th} input feature, respectively.

Then, we perform some equivalent transformation to simplify Equation (2) and the stochastic sensitivity could be rewritten as:

$$E_s((\Delta y)^2) \approx \frac{1}{3} Q^2 \sum_{j=1}^M v_j + \frac{0.2}{9} Q^4 n \sum_{j=1}^M \varsigma_j \quad (3)$$

$$\text{Where } v_j = \varphi_j \left(\sum_{i=1}^n (\sigma_{x_i}^2 + (\mu_{x_i} - u_{ji})^2) \right) / v_j^4, \quad \varsigma_j = \varphi_j / v_j^4.$$

In the case that the input features are scaled into [0,1], the L-GEM models the generalization capability of the RBFNN with unseen samples having less than 1% difference from the training samples when the Q value equals to 0.01.

2.2. Buy Signal Prediction by MCS

In [15], only one candlestick pattern is applied to the prediction of stock price trend. In fact, we can hardly assure that a particular candlestick pattern occurring frequently will have strong indication of up-trend, and vice versa. On the other hand, the effects of different candlestick patterns differ from stock to stock. Therefore, we select ten candlestick patterns to compose the proposed MCS for buy signal prediction. Ten RBFNNs are trained and each of them corresponds to a specified candlestick pattern to recognize the occurrence of the real bullish candlestick. By training a MCS to fuse the outputs from the 10 base RBFNNs by a weighted sum, we finally make an investment decision. The MCS combines the real or false bullish signal predicted by each base classifier. The architecture of the MCS is described in Figure 1.

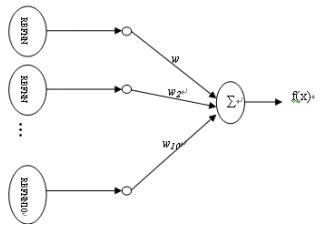


Figure 1. The architecture of the MCS

3. Candlestick patterns and Data representation

The prediction results from the MCS with RBFNN trained by L-GEM help us to make investment decisions. If a buy signal is generated from MCS based on existing candlestick patterns, we will buy the stock in the next trading day. A “3-day investment strategy” is employed in this work for all methods in comparison. This may not be the best investment strategy, but provide a fair comparison basis to our experiments.

3.1. Candlestick Pattern

Candlestick patterns could consist of a single candlestick or multiple candlesticks. In this paper, we use ten of the most widely used candlestick patterns from [3] that indicate an uptrend of stock price in the future stock trading strongly. Six out of the ten patterns are bullish single lines while the other four are bullish reversal patterns. The bullish single lines could be applied to any condition, but the bullish reversal patterns should be used in downtrends only (three declining parallel lines in Table 1).

Pattern 1: Long White Candle	Pattern 2: White Marubozu
Pattern 3: Closing White Marubozu	Pattern 4: Opening White Marubozu
Pattern 5: White and Black Paper Umbrella Either one)	Pattern 6: Doji
Pattern 7: Jump empty gap	Pattern 8: Bullish Engulfing
Pattern 9: Piercing Line	Pattern 10: Bullish Harami

Table 1. The ten Candlestick Patterns

3.2. The data representation of candlestick pattern

A candlestick chart provides trend and current information about a stock. However, a neural network can not recognize the candlestick charts directly. We need to extract features from the candlestick charts.

A candlestick of the i^{th} day is represented by the day's open price (open_i), close price (close_i), high price and low price, where the high price (high_i) and low price (low_i) denote the highest price and the lowest price of the i^{th} day, respectively. A candlestick is defined by its body color, body length, upper shadow and lower shadow. The body color (YY_i) is determined by the difference between open price and close price. If the close price is higher than the open price, the body will be a white body otherwise a black one. The body length (body_i) represents the gap between the

close price and open price. The upper shadow (us_i) of a white (black) candlestick indicates gap between the high price and close (open) price. While the lower shadow (ls_i) of a white (black) candlestick indicates gap between the low price and open (close) price. Patterns 1 to 6 consist of only a single candlestick and could be described by Equations (4) to (7).

$$us_i = high_i - \max(open_i, close_i) \quad (4)$$

$$ls_i = \min(open_{i-1}, close_{i-1}) - low_i \quad (5)$$

$$body_i = |close_i - open_i| \quad (6)$$

$$YY_i = \begin{cases} 1 & \text{if } close_i - open_i \geq 0 \\ 0 & \text{else} \end{cases} \quad (7)$$

Patterns 7 to 10 consist of two candlesticks. Two descriptors (i.e. $gapu_i$ and $gapl_i$) are added to indicate the gap between upper ends of two consecutive candlesticks and the gap between lower ends of two consecutive candlesticks. Besides, a downtrend signal ($flag_i$) is the prerequisite of the occurrence of bullish reversal patterns.

$$gapu_i = \max(close_{i+1}, open_{i+1}) - \max(close_i, open_i) \quad (8)$$

$$gapl_i = \min(close_{i+1}, open_{i+1}) - \min(close_i, open_i) \quad (9)$$

$$flag_i = \begin{cases} 1 & \text{if } close_{i-3} > close_{i-2} > close_{i-1} \\ 0 & \text{else} \end{cases} \quad (10)$$

In our work, we adopt a 20-day normalization method to make a good representation for the candlestick patterns.

$$us'_i = \frac{us_i - \overline{us}_{i-19,i}}{\sigma_{us_{i-19,i}}} \quad (11)$$

$$ls'_i = \frac{ls_i - \overline{ls}_{i-19,i}}{\sigma_{ls_{i-19,i}}} \quad (12)$$

$$body'_i = \frac{body_i - \overline{body}_{i-19,i}}{\sigma_{body_{i-19,i}}} \quad (13)$$

$$gapu'_i = \frac{gapu_i - \overline{gapu}_{i-19,i}}{\sigma_{gapu_{i-19,i}}} \quad (14)$$

$$gapl'_i = \frac{gapl_i - \overline{gapl}_{i-19,i}}{\sigma_{gapl_{i-19,i}}} \quad (15)$$

Therefore, the ten patterns adopted in this paper can be transformed into numerical values as follow:

Pattern 1:	Pattern 2:
1. $YY_i = 1$	1. $YY_i = 1$
2. $0.7 \leq \frac{body'_i}{body'_i + us'_i + ls'_i} < 1$	2. $\frac{body'_i}{body'_i + us'_i + ls'_i} = 1$

Pattern 3: 1. $YY_i = 1$ 2. $0.7 \leq \frac{body'_i}{body'_i + us'_i + ls'_i} < 1$ 3. $us'_i = 0$	Pattern 4: 1. $YY_i = 1$ 2. $0.7 \leq \frac{body'_i}{body'_i + us'_i + ls'_i} < 1$ 3. $ls'_i = 0$
Pattern 5: 1. $body'_i + us'_i + ls'_i \neq 0$ 2. $us'_i = 0$ 3. $0 \leq \frac{body'_i}{body'_i + us'_i + ls'_i} < 0.25$	Pattern 6: 1. $body'_i = 0$ 2. $us'_i \neq 0$ 3. $us'_i > ls'_i$
Pattern 7: 1. $YY_i = 1 \cap YY_{i-1} = 0$ 2. $body_{i-1} < body_i$ 3. $gapu_i > 0$ 4. $gapl_i < 0$ 5. $flag = 1$	Pattern 8: 1. $YY_i = 1 \cap YY_{i-1} = 0$ 2. $body_{i-1} < body_i$ 3. $gapu_i > 0$ 4. $gapl_i < 0$ 5. $flag = 1$
Pattern 9: 1. $YY_i = 1 \cap YY_{i-1} = 0$ 2. $gapl_i > 0$ 3. $body_{i-1} + gapl_i \geq body_i / 2$ 4. $flag = 1$	Pattern 10: 1. $YY_i = 1 \cap YY_{i-1} = 0$ 2. $gapl_i > 0$ 3. $gapu_i < 0$ 4. $body_{i-1} > body_i$ 5. $flag = 1$

Table 2. Data Representation for Candlestick Patterns

For the i^{th} RBFNN corresponding to the Pattern i , input features are computed as shown in Table 2. The output of each base RBFNN is the classification of “Real Bullish Candlestick” or “False Bullish Candlestick”. The Real Bullish Candlestick indicates a strong up-trend of the stock price in future stock trading. Each hidden neural of a RBFNN corresponds to a prototype of a candlestick pattern. The optimal number of a candlestick pattern’s prototypes is selected via a minimization of L-GEM which greatly improves the generalization ability of RBFNN in stock decision support problem.

4. Experiments for Hong Kong Equity Market

In the section, we carry out stock investment simulation experiments using 10 years’ stock price data from 42 component stocks of Hong Kong’s Hang Seng Index. For each stock, first half of the historical data is used to train the MCS. The rest of the data is used to evaluate the generalization capability of each method when dealing with new trading days after training.

4.1. Introduction to the simulation experiment

We select the architecture of the RBFNN by minimizing the L-GEM. Each base RBFNN corresponds to a particular candlestick pattern. In the training data set, if a given candlestick pattern appears at that day, it is labeled as a positive sample, otherwise a negative sample. For each base RBFNN, We collected all positive samples (N_{pos}) and randomly select $3N_{pos}$ negative samples for training. This is to relieve the unbalance between positive and negative samples. Then, a MCS is trained to fuse the classification results from these 10 base RBFNNs by a weighted sum. Finally, we calculate the average profit of each stock earned by the proposed method via using the samples in the testing dataset. Since the training samples to train the base RBFNN for each candlestick pattern are selected randomly, so we perform 10 times for each stock.

In this work, we buy the stock when the MCS yields a buy signal (indicating the occurrence of an up-trend) and then sell it with the 3-day investment strategy. We will compare the proposed method with a Random Investment strategy (RI) and the Ordinary Candlestick Pattern Trading (OCPT) strategy. Owning to the different times of the buy signals yielding from the MCS, we will randomly buy the stock with the number of buying being the same times as the proposed method does. At the same time, the RI method will be performed 30 times. For the OCPT method, we buy a stock whenever a candlestick pattern appears and sell it using the same 3-day investment strategy.

4.2. Experimental Results

Table 3 lists the average earning of each method for 42 different stocks in the first three columns. The 4th (5th) column shows the t-test value comparing L-GEM with RI (OCPT) method. A t-test value > 1.96 indicates the proposed method yielding a statistical significant improvement over the comparing method with 95% confidence.

From table 3, The proposed method outperforms the RI method with 95% confidence for 20 out of 42 stocks. In contrast, there are only 6 out of 42 stocks that the RI method outperforms the proposed method. The total earning of the proposed method is HKD167.32 while both the RI and OCPT methods yield a negative value in earning, i.e. -HKD47.56 and -HKD184.57 respectively.

Experimental results show that the MCS based on RBFNN training by minimizing the generalization error enhances the forecasting accuracy of the candlestick patterns and reduces the loss on stock investment greatly.

Table 3. Investment Returns of 42 stocks

Stock Code	LGEM avg	RI avg	OCPT	t-Test of LGEM vs RI	t-Test of LGEM vs OCPT
0001	19.22	-4.18	-82.72	3.48	1.58
0002	4.41	0.15	-10.00	2.42	0.67
0003	-1.83	2.34	7.73	-5.48	-1.20
0004	11.61	9.72	16.08	0.63	-0.33
0005	6.98	-12.98	4.88	3.72	0.06
0006	14.89	10.74	23.95	2.51	-0.38
0011	18.74	-18.81	0.48	6.02	0.41
0012	-2.60	-9.29	-4.12	1.87	0.10
0013	-11.76	-15.65	-31.56	1.07	0.64
0016	30.25	-1.84	-3.24	4.09	0.50
0017	4.23	-0.52	13.79	4.20	-0.43
0019	16.93	14.36	-17.96	0.22	0.69
0023	8.95	-1.00	4.40	3.57	0.14
0066	1.27	1.37	3.37	-0.03	-0.18
0083	9.19	-4.92	-3.94	9.32	0.95
0101	1.89	3.26	-23.99	-0.96	1.34
0144	0.22	-3.42	-13.73	1.57	0.53
0267	6.24	-15.90	-29.21	8.59	1.74
0291	4.89	1.97	-14.78	1.61	1.06
0293	-6.65	-6.59	-5.12	0.19	0.14
0330	-39.96	-28.97	-42.95	-1.63	0.27
0386	-0.81	0.61	1.92	-2.80	-0.45
0388	61.49	14.84	43.20	4.73	0.30
0494	10.59	3.84	4.06	3.31	0.33
0688	2.74	1.47	-13.30	1.11	1.48
0700	27.72	21.21	1.28	1.10	0.66
0762	-8.53	0.17	-0.70	-8.82	-1.10
0836	2.24	-0.30	1.38	4.20	0.03
0857	0.66	0.75	6.56	-0.14	-0.60
0883	6.80	0.22	-5.50	7.57	1.53
0939	-0.57	-0.68	-3.62	0.46	0.90
0941	-8.64	16.55	34.84	-3.68	-0.79
1088	-1.63	-8.24	-16.45	2.66	1.13
1199	-4.73	-6.59	-10.57	1.51	0.44
1398	1.56	0.05	0.02	4.40	0.43
2038	-2.26	-11.04	-9.41	8.75	0.94
2318	2.42	-0.45	-1.40	4.47	-0.79
2388	-8.34	-2.96	-13.02	-6.21	0.10
2600	0.38	-3.23	3.70	3.02	-0.36
2628	-11.51	6.06	2.81	-8.64	-0.79
3328	0.11	0.12	0.06	0.23	-0.06
3988	0.21	0.06	-2.14	1.19	1.48
Total	167.32	-47.65	-184.57		

5. Conclusion

In this work, we use a MCS with 10 base RBFNNs trained by L-GEM to generate a buy signal and apply it to 42 component stocks of Hong Kong's Hang Seng Index for stock trading. The results demonstrate that our work can achieve higher returns than the random investment strategy and simply buy in whenever a candlestick pattern occurs. More powerful candlestick patterns can be added into the RBFNN to enhance the predicting accuracy of stock's up-trend in future works.

Acknowledgements

This work is supported by National Natural Science Foundation of China (61003171 and 61003172) and the Fundamental Research Funds for the Central Universities 2009ZZ0050 and 2011ZM0066.

References

- [1] W. Leigh, C.J. Frohlich, S. Hornik, R.L. Purvis and T.L. Roberts, "Trading With a Stock Chart Heuristic", IEEE Trans. on SMC (A), pp. 93 – 104, 2008
- [2] John J. Murphy, *Intermarket Technical Analysis*, New York, 1992
- [3] John J. Murphy, *Technical Analysis of Financial Markets: a Comprehensive Guide to Trading Methods and Applications*, New York Institute of Finance, New York, 1999
- [4] E. Fama, "The Behavior of Stock Market Price", *Journal of Business*, Vol.38, pp.34-105,1965
- [5] B.G. Malkiel, "A Random Walk Down Wall Street", 8th edition, New York, NY: W.W. Norton, 2003
- [6] C.R. Harvey, "Predictable Risk and Returns in Emerging Markets", *The Review of Financial Studies*, Vol.8, pp. 773-816, 1995
- [7] B.R. Marshall, M.R. Young and Lawrence C. Rose, "Candlestick Technical Trading Strategies: Can They Create Value for Investors?", *Journal of Banking and Finance*, pp.2303-2323, 2006
- [8] Feng Li, Cheng Liu, "Application Study of BP Neural Network on Stock Market Prediction", *Hybrid Intelligent Systems*, pp.174-178, 2009
- [9] Y. Chen, S. Mabu, K.Hirasawa, "A genetic network programming with learning approach for enhanced stock trading model", *Expert Systems with Applications*, pp.12537-12546, 2009
- [10] L. Yu, H. Chen, S. Wang and K.K. Lai , "Evolving Least Squares Support Vector Machines for Stock Market Trend Mining", *IEEE Transactions on Evolutionary Computation*, pp.87-102, 2009
- [11] S. Nison, "Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Investment Technique of the Far East", New York Institute of Finance, 1991
- [12] Daniel S. Yeung, Wing W.Y. Ng, Defeng Wang, Eric C.C. Tsang and Xizhao Wang, "Localized Generalization Error and Its Application to Architecture Selection for Radial Basis Function Neural Network", *IEEE Trans. on Neural Networks*, vol. 18, no. 5, pp. 1294 – 1305, 2007
- [13] Wing W.Y. Ng, A. Dorado, Daniel S. Yeung, Witold Pedrycz and E. Izquierdo, "Image Classification with the use of Radial Basis Function Neural Networks and the Minimization of Localized Generalization Error", *Pattern Recognition*, pp. 19 - 32, 2007
- [14] Daniel S. Yeung, Wing W.Y. Ng, Aki P.F. Chan, Patrick P.K. Chan, Michael Firth and Eric C.C. Tsang, "Multiple Intelligent Agent System for Credit Risk Prediction via an Optimization of Localized Generalization Error with Diversity", *Journal of Systems Science and Systems Engineering*, vol. 16, no. 2, pp. 166 – 180, 2007
- [15] Huili Li, Wing W.Y. Ng, John W.T. Lee, Binbin Sun, Daniel S. Yeung, "Quantitative Study on Candlestick Pattern on Shenzhen Stock Market", *IEEE International Conference on Systems, Man and Cybernetics*, pp.54-59, 2008
- [16] L-GEM based MCS aided candlestick pattern investment strategy in the Shenzhen stock market