The Motif Detection of Short-term Tendency in Stock Time Series

Wengang CHE

School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, 650051, P.R. China, wgche@yahoo.com

Keywords: stock time series; motif detection; short-term tendency.

Abstract. A stock time series is composed of a sequence of prices measured from a continuous trade day. Stock motifs are approximately similar segments in a group of different stock time series or market dataset. In this paper, at first, a definition of short-term tendency is defined; secondly, a motif of short-term tendency is introduced in the stock time series. And then a simple and effective method of stock motif detection is proposed to predict the short-term tendency in the stock time series analysis. Finally, a new motif detection of short-term tendency is introduced to predict the trend of temporal stock market trend. An empirical analysis is made to show the efficiency of the proposed method.

1. Introduction

A stock time series is a sequence of data points, typically consisting of successive measurements of stock price made over a trade day. Recently, motif detection from large-scale data becomes a hot research topic in data mining research. To the best of our knowledge, the concept of time series motifs was introduced in 2003 [4]. Given such wide applicability, motif and anomaly detection are attracting increasing attention from both researchers and financial market traders.

In many applications, motif discovery and anomaly detection is widely used to find the abnormal case and analyze other kinds of data. For examples, in the applications of network intrusion, credit card fraud, weather forecasting, cardiogram data and stock trading, an abnormal case and motif are more valuable than normal ones. For an example based on stock time series, considering motifs in the context of stock prices, a particular pattern, or called stock motif may be used to predict trends of stock price temporally. Therefore, a method to find a particular pattern and tuning point are more practical and interested. As we know, there are already some results presented in this area, such as the ones based on wavelet analysis [1], density [2], boundary and distance [3], clustering [4] and motif of k-NN clustering algorithm [5] etc.

Previously, we already proposed a singular point detection method in stock time series based on wavelet transform, where singularity using wavelet modulus maxima determination is employed [1]. In [2], we also proposed another method to determine the singularity point from sliding windows. In this paper, we introduce the method of motif detection in abnormal analysis.

In the stock market, short-term tendency are very important signals which is able to indicate the future trend in time series, which are identified as the growth trend or the downward trend. The growth trend is known as the value of stock time series has gained within a specific period and context. The downward trend is also known as the value of stock time series has lost within a specific period and context. The turning point is defined as the time when a growth trend begins to become downward trend or versus. In this tuning point, a trend reversal signal appears because these special potential turning points contain a lot of important information. Practically, the method to find those abnormal or turning point becomes a critic issue which has both theoretic and empirical significance. Furthermore, when we study the behaviours of a group of stock time series, if there are same trends from a group of individual stock picks within the same time period, the short term prediction becomes more confident and practical.

In this paper, a pattern of increasing or decreasing within continues four days are used to detect the tendency of stock time series. This pattern of tendency is defined as stock motif of short-term tendency. We will also propose a new motif detection method in stock time series, where the patterns of growth or downward tendencies in stock time series are used as stock motif.

2. The related definitions

To study stock time series, it is always able to find out that in the certain trading days, there are the same directed movement within in a group of stocks in the same trade period. In this process of stock market volatility, this group of stocks will get gains or losses together during those trade days. We call this stock market volatility as the stock short term tendency. If at least there are two or more stocks having similar behavior in those trade days, the trend is considered as motif of short-term tendency.

Definition 1: Stock Time Series: Given a stock time series X, a close price of stock time series is denoted by $X(t_i, y_i)$ where y_i is used to represent close price of a stock time series at the trade day of t_i.

Definition 2: Short-term Tendency: We denote a segment of X by $S = X[t_{i-1}, t_{i+3}]$, where i-1 and i+3 are the starting and ending day indices respectively. If $y_{i-1} > y_i$ and $y_i < y_{i+1} < y_{i+2} < y_{i+3}$ within duration of $[t_{i-1}, t_{i+3}]$, then time series X is considered having the short-term growth tendency during 4 trade days. If $y_{i-1} < y_i$ and $y_i > y_{i+1} > y_{i+2} > y_{i+3}$ within duration of $[t_{i-1}, t_{i+3}]$, then time series X is considered having the short-term downward tendency during 4 trade days. Point t_i is defined as a turning point of short-term tendency.

Definition 3: Motif of Short-term Tendency: If there are two or more similar segments of $S_k = X_k[t_{i-1}, t_{i+3}]$ for k = 1,2,3, ... n within stock time series of X_k. This similar segment having the same short-term tendency pattern is defined as motif of short-term tendency.

3. Short-term tendency detection and algorithms

In order to detect the present tendency of a stock time series, two traditional analysis methods called long term moving average versus short-term moving average are widely used to predict the trends of a stock time series. They are well known as the indicators of "Golden Cross" or "Dead Cross" of the stock trade process. However, it is difficult to clearly define a motif pattern to describe the behaviour of stock time series using those two traditional methods.

In this paper, we use new method the counts of the each increasing or decreasing days to present the short-term tendency. Those indicator are also well known by stock traders as "Three Red" versus "Three Green" shown in Fig. 1(a) and Fig. 1(b).
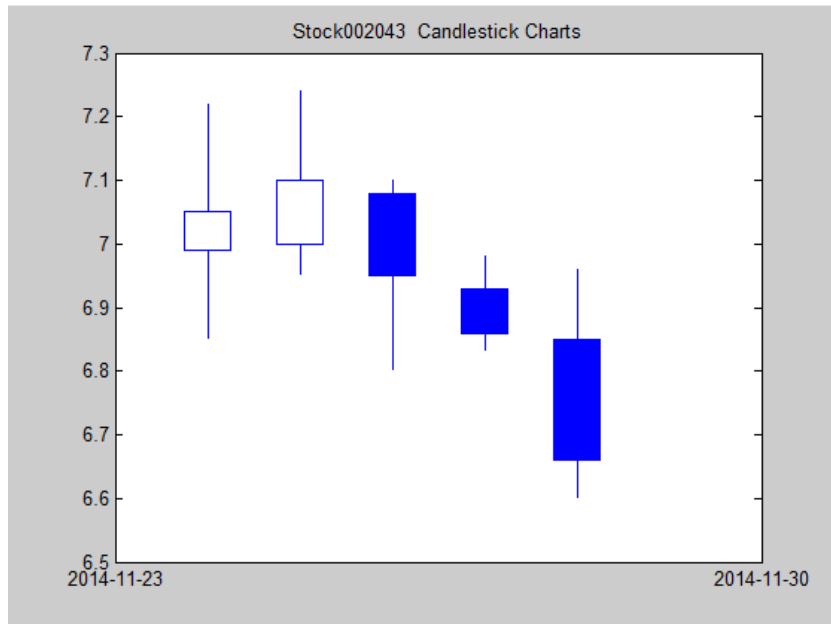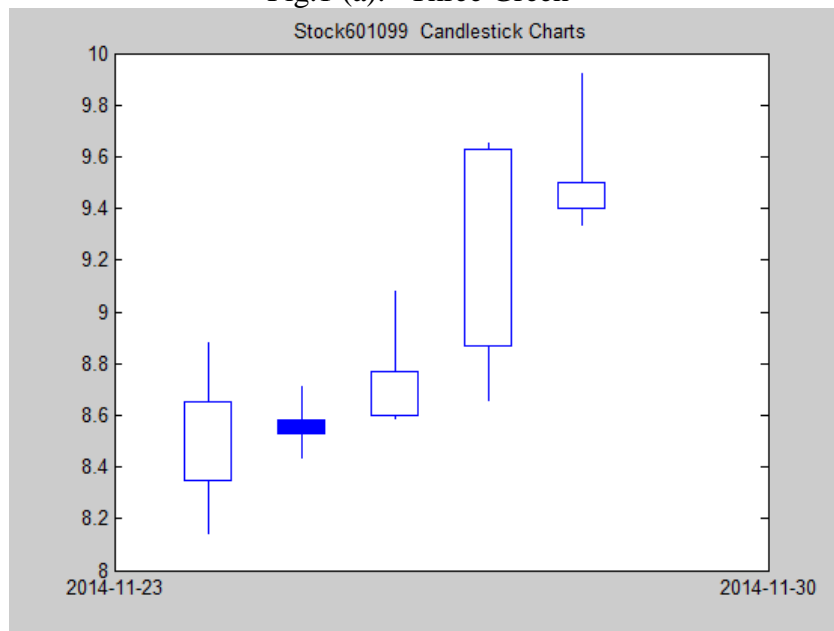
Fig.1 (a): "Three Green"

Fig.1 (b): "Three Red"

Using the definition 2, we develop algorithms to detect the short-term tendency.

Algorithm 1: Short-term Growth Tendency

Step 1: Take a stock time series of $X(t_i, y_i)$, from any trade day of i, compare the stock price of day i and day i-1, if $y_{i-1} > y_i$, go to step 2; otherwise increase to next day and go back to step 1.

Step 2: Compare the stock prices of day i+1 and day i, if $y_{i+1} > y_i$, go to step 3; otherwise increase i to next day and go back to step 1.

Step 3: Compare the stock prices of day i+2 and day i+1, if $y_{i+2} > y_{i+1}$, go to step 4; otherwise increase i to next day and go back to step 1.

Step 4: Compare the stock prices of day i+3 and day i+2, if $y_{i+3} > y_{i+2}$, go to step 5; otherwise increase i to next day and go back to step 1.

Step 5: A short-term growth tendency is detected.

Algorithm 2: Short-term Downward Tendency

A similar algorithm with above one is used where the large than calculation " $>$ " substituted by less than calculation" $<$ ".

Step 1: Take a stock time series of $X(t_i, y_i)$, from any trade day of i, compare the stock price of day i and day i-1, if $y_{i-1} < y_i$, go to step 2; otherwise increase to next day and go back to step 1.

Step 2: Compare the stock prices of day i+1 and day i, if $y_{i+1} < y_i$, go to step 3; otherwise increase i to next day and go back to step 1.

Step 3: Compare the stock prices of day i+2 and day i+1, if $y_{i+2} < y_{i+1}$, go to step 4; otherwise increase i to next day and go back to step 1.

Step 4: Compare the stock prices of day i+3 and day i+2, if $y_{i+3} < y_{i+2}$, go to step 5; otherwise increase i to next day and go back to step 1.

Step 5: A short-term downward tendency is detected.

4. The motif detection of short-term tendency

Generally, time series motif is a pattern appearing frequently in a time sequence which is very useful to forecast the stock short-term tendencies and prices. In this section, we propose an algorithm to find those motifs of short-term tendency in stock time series.

Algorithm 3: Motif detection of Short-term tendency

Step 1: Take a stock time series of $X(t_i, y_i)$, detect the short-term tendency and remember the time duration.

Step 2: Take another stock time series of $X(t_i, y_i)$, detect the short-term tendency and compare if there are similar segment of time series during same time period. If yes, got to step 3, otherwise go to step 2.

Step 3: Remember the number of similar segments detected and then go back to step 2.

Step 4: The motifs of the stock time series are detected.

5. Empirical Analyses

The close stock prices of A-shares in Shanghai among May 2010 are used as the experimental data for short-term tendency detection. Using the algorithm we defined above, the number of similar segments having same short-term tendency is counted as the result of motif detection. The result is shown in Table 1 and the observed time series shown in Fig.1.

Table 1 The observed data and the segments found

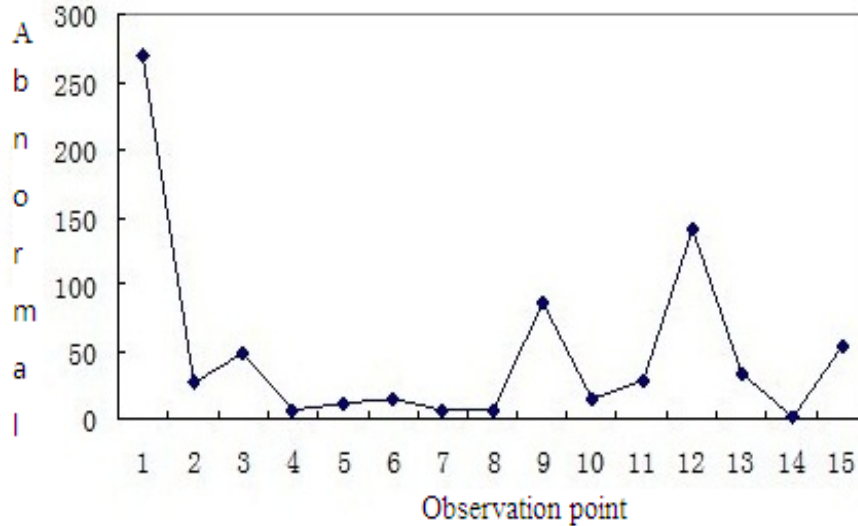| Observed day	segments Found	Observed day	Segments Found	Observed day	Segments Found
2010-05-05	269	2010-05-06	27	2010-05-07	50
2010-05-10	6	2010-05-11	12	2010-05-12	15
2010-05-13	7	2010-05-14	7	2010-05-17	87
2010-05-18	16	2010-05-19	28	2010-05-20	141
2010-05-21	34	2010-05-24	2	2010-05-25	55

Fig.2: The number of time series having abnormal points and motif segments

Using the close price of the trade data for one year of 2010, motifs of the short-term tendency are detected and shown in Table 2

Table 2: Motif detection of Shanghai and Shenzhen A-share in 2010

Turning Location	Motifs Found	Turning Location	Motifs Found	Turning Location	Motifs Found
2010-01-07	490	2010-05-27	372	2010-08-25	663
2010-03-15	736	2010-06-01	295	2010-09-29	536
2010-03-25	326	2010-07-05	733	2010-10-18	902
2010-04-19	513	2010-07-15	611	2010-11-03	280
2010-04-28	424	2010-07-22	77	2010-11-17	1008
2010-05-05	342	2010-07-27	257	2010-12-09	853
2010-05-20	769	2010-08-12	630		

Conclusions

In this paper, a motif detection of short-term tendency is introduced and the algorithms to detect the short-term tendency are proposed. An empirical data shows that the proposed algorithm has high accuracy and scalability. Since the turning point detection is based on the motif detection of short-term tendency, those turning points have a lot of information for trend change and are significant for researchers and practical for the stock market traders.

References

[1]. Xiaodi Liu, Wengang Che, and Qingjiang Zhao, "Analysis of outliers and public information arrivals using wavelet transform modulus Maximum," The 2nd IEEE International Conference on Information and Financial Engineering, September 2010, pp. 176-179.
[2]. Qiang Fu, Wengang Che, "Study on isolated singularity data mining algorithm based on sliding

window," JINGXI SCIENCE, vol. 29, pp. 273-276, April 2011(in Chinese).

[3]. Yi Lin, Michael D. McCool, Ali A. Ghorbani, "Time Series Motif Discovery and Anomaly Detection Based on Subseries Join", IAENG International Journal of Computer Science, 37:3, IJCS_37_3_08.

[4]. B. Chiu, E. Keogh, and S. Lonardi, "Probabilistic Discovery of Time Series Motifs," Proceedings of the 9th International Conference on Knowledge Discovery and Data mining (KDD'03), pp. 493-498, 2003.

[5]. Yu-Feng Jiang ; Chun-Ping Li ; Jun-Zhou Han "Stock temporal prediction based on time series motifs," Machine Learning and Cybernetics, 2009 International Conference on (Volume:6).

[6]. Feng Jiang, Junwei Du, and Yuefei Sui, "Outlier detection based on boundary and distance," ACTA ELECTRANICA SINICA, vol. 38, pp. 700-704, March 2010(in Chinese).

[7]. Yue Zhang, Jie Liu, and Hang Li, "An outlier detection algorithm based on clustering analysis," The 1st International Conference on Pervasive Computing, Signal Processing and Applications, pp. 1126-1128, September 2010.

[8]. Juntao Wang, Xiaolong Su, "An improved K-Means clustering algorithm," The 3rd IEEE International Conference on Communication Software and Networks, May 2011, pp. 44-46.

[9]. Tao Jiang, Xiaoli Chen, Yufang Zhang, "Improved KNN using clustering algorithm," Computer Engineering and Applications, vol. 45, pp. 153-155, 2009(in Chinese).

[10]. Jing Sun, Chenghu Zhang, Shanxin Chen, "Research of suspicious foreign exchange deals recognition based on outlier detection of time series," STATISTIC AND DECISION, vol. 18, pp. 26-29, 2010(in Chinese).

[11]. Faloutsos, C., Ranganathan, M., & Manolopoulos, Y. (1994). Fast Subsequence Matching in Time-Series Databases. In proceedings of the ACM SIGMOD Int'l Conference on Management of Data. May 24-27, Minneapolis, MN. pp 419-429.

[12]. Chan, K. & Fu, A. W. (1999). Efficient Time Series Matching by Wavelets. In proceedings of the 15th IEEE Int'l Conference on Data Engineering. Sydney, Australia, Mar 23-26. pp 126-133.

[13]. Keogh, E., Chakrabarti, K., Pazzani, M. & Mehrotra, S. (2001). Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases. In proceedings of ACM SIGMOD Conference on Management of Data. Santa Barbara, CA, May 2124. pp 151-162.

[14]. Geurts, P. (2001). Pattern Extraction for Time Series Classification. In proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery. Sep 3-7, Freiburg, Germany. pp. 115-127.