Time-Space Domain of Group Behavior Research Based on Major Events Influence

TingtingLi[1], LingyuXu[1], JieYu[1],LeiWang[1], YunlanXue[1],YangLiu[2]

[1]School of Computer Engineering and Science, Shanghai University, Shanghai, 200444,China

[2]School of Computer Science and Technology, University of South, Hunan,421001 China

lttsnoopy@sina.com

Keywords: Spatio-temporal data, Group behavior, time series, abnormal detection , k-Medoids

Abstract: When a major event creates a behavior cluster, this cluster's spatio-temporal data is the group behavior time-space domain. The time data of each individual on the space is a time series data. So this article we first to analyze the time series, approximately express the time series with the piecewise linear representation(PLR) method, and then use activity measure method for abnormal detection to get the abnormal time domain. According to the time of the major events occurred, we match individual abnormal time domain, and get the common abnormal groups. Finally, for these abnormal groups, we can get the space domain of the event with k-Medoids clustering algorithm analysis.

Introduction

Network group behavior is generally around the Internet social hot issues, and it has the characteristics of spontaneity, infectious, emotional, and temporary. Therefore, from the network group information, people can be very intuitive see the impact of an event in the network virtual world, movement rule on "time" and "space", as shown in the Figure.1. The stock market [1] is an important part of financial market.XuChuan Wu (2005) in his researchshows that China's stock marketexist the strong herd behavior[2], and it also means the strong consistency trading behavior. The strong consistency of trading behavior is formed a certain effect on stock prices. So we can use these trading behaviors to analyze the stock plates and the cycle on the influence of major events.

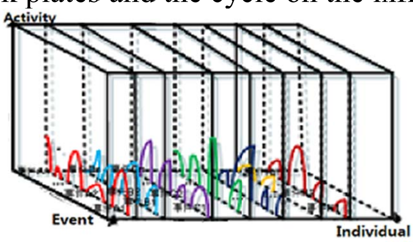


Fig.1 time-space domain of the event

Related Work

The key technologies of the behavior of time and space are ① Time seriesmodel② Abnormal extraction oftime series ③ The clustering of time series.

Time series model

Time series is an ordered set which is made up with record values and record time, refers to $X=<x_1=(v_1,t_1),x_2=(v_2,t_2)\ldots x_i=(v_i,t_i)>$

v_i is value for t_i time records, and t_i is strictly increasing.Usuallywe use follow fourmodes:1)Symbolic[3];2)DiscreteFourierTransformation(DFT);3) Singular Value Decomposition(SVD);4) Points line representation(PLR). Here we focus on discussing PLR, cut the time series into multiple child segments along the time axis, and get each section of the two endpoint coordinates together which is defined as the time series model.[4,5]

Abnormal extraction of time series

Time series of abnormal become a new hot spot in the time series data mining. For time series, we don't care about single sequence points of exceptions, but the abnormal change of time series over a period of time. Ma et al., using support vector regression model for training sequence of historical events, when the new coming time series data deviates from the model, it is defined as the new event of the time series. The TSA - Tree improved algorithm achieve the discovery of Singular Value Decomposition and its singular model was defined as a sudden change in the time series.[6] In this paper, on the basis of these methods, we use activity to extract abnormal pattern.

The clustering of time series

Time series similarity measure is the method to weigh the similarity degree of two time series. Euclidean distance and Dynamic Time Warping are two classic methods for Time series similarity measure, but these two methods have defects when application to financial time series. Euclidean distance [7] are vulnerable to the interference of volatility of financial time series, and the complexity of the dynamic time warping measure algorithm limits its application scope. In this paper, first time, weuse k-Medoids algorithm to clustering analysis.

Time-Space Domain of Behavior Research

The Basic Definition

Definition 1：

1、Behavior spatial domain:$\Omega = \{\Omega_1, \Omega_2, \Omega_3, \dots\}(i = 1,2,3,\dots)$，$\Omega_i$is an individual;

2、Individual behavior time domain:The time range in Ω_i of a behavior;T_i=[t_{i1},t_{i2}]

3、Group behavior time domain:The time range in n individuals;

T=[t_{11},t_{12}]∪[t_{11},t_{12}]∪[t_{i1},t_{i2}]（1）

[t_{i1},t_{i2}] is the scope of the behavior's time in Ω_i

Definition 2：

Behavior time-space domain:All individuals involved a behavior is as the spatial domainΩ, the behavior time domain is T.

$\{\Omega, T\}$=$\{(\Omega_1, T_1), (\Omega_2, T_2), (\Omega_3, T_3), \dots (\Omega_i, T_i)\}$（2）

Where (Ω_i, T_i) is an individual Ω_i's time domain T_i on the influence of the event

Definition3：

Activity:It is the degree of an individual affected by the event, in individual time domain [t_{11},t_{12}],K is the variation of an individual attribution,$\overline{K}$is the average variation value

Activity: =$K/\overline{K}$（3）

Algorithm Description

(1) Determination of individual time domain ofbehavior

In order to discover individual time domain of group behavior,we will analyze the individual time series. First, we should adopt different length of time in the time series for clustering. Second,weuse the PLR(Piecewise Linear Representation) to express the selected time series.

Algorithm 1:

Input: A certain individual period of time series X

X=<$\mathbf{x_1} = (\mathbf{v_1}, \mathbf{t_1})$,$\mathbf{x_2} = (\mathbf{v_2}, \mathbf{t_2})$,…,$\mathbf{x_i} = (\mathbf{v_i}, \mathbf{t_i})$>

Output: Filter the tiny volatility in time series S, Which contains only rise and drop interval,called $\boldsymbol{X}'$.

1)scanX, statistics all the lift range and their corresponding amplitude;

2) take all the absolute amplitude values, then calculate the average as a threshold, the amplitude value is greater than the threshold value is marked as strong, less than the threshold is marked as weak;

3) if (the current range exist small shock range)

we should fuse the range according to all kinds of situations

elsewe can copies the current interval length and fuse the before fused oscillation zone, then calculate the size of each lift range}

4)Finally, return the scope of lift range[$\boldsymbol{t_k}, \boldsymbol{t_{k+i}}$]of $\mathbf{X}'$and related magnitude

Algorithm 2:

Input: the time series $\mathbf{X}'$ made from Step1

Output:Abnormal time interval

1) First,scan X', to calculatethe average rise magnitude $\overline{\mathbf{K}}_{\mathbf{up}}$ and decline magnitude $\overline{\mathbf{K}}_{\mathbf{down}}$byevery month, and we use these values as the background.

2) If ($\boldsymbol{k_i}$>0 &&$\boldsymbol{k_i}$>$\overline{\mathbf{K}}_{\mathbf{up}}$)

3) return $\boldsymbol{k_i/\overline{K}_{up}}$, $[\boldsymbol{t_k}, \boldsymbol{t_{k+i}}]$

4) If ($\boldsymbol{k_i}$<0 &&$\boldsymbol{k_i}$>$\overline{\mathbf{K}}_{\mathbf{down}}$)

return $\boldsymbol{k_i}/\overline{\mathbf{K}}_{\mathbf{down}}$,$[\boldsymbol{t_k}, \boldsymbol{t_{k+i}}]$

By Algorithm 1, time series X=<x_1,x_2,x_3,…， x_n> can be expressed with PLR as follows:

$$X'(t)\begin{cases} k_1 t + b_1, & t \in [1, t_1] \\ k_2 t + b_2, & t \in [t_1, t_2] \\ \quad \dots\dots & \\ k_k t + b_k, & t \in [t_k, n] \end{cases}$$

And the k is the value of interval slope.

By Algorithm 2, each time series'abnormal interval is individual time domain of behavior ,andof thetimeseries and its relative magnitudechanges valueis activity.

(2) Determination of space domain

Algorithm 3:

Input: The all abnormal intervals of stocks and theirs relative magnitudechanges value.

Output: Similar individuals of time series.

1)Find out all the stocks of public abnormal time (or points) according to the abnormal rangethat got at Algorithm2;

2)K-Medoids cluster analysis was carried out on all the individuals which have same abnormal period (point), and divide these individual time series into two clusters,the attribute values of the clustering are the K value that before and after the Abnormal period(point),that is $\boldsymbol{k_{i-1}}, \boldsymbol{k_i}, \boldsymbol{k_{i+1}}$;

3) According to the result of clustering, then combine with the relative changes of abnormal value which are got by Step2 to have similar judgment again.

Experiments

In this paper, the experiment adopts stock data of financial market, including stock trading volume and post amounts in the post bar. Because of stock trading volume is an important index reflecting the behavior of traders, the people trading behavior is driven by significant events, and the change of the stock trading volume shows the degree of influence of the stock affected by the event. Post in the stocks bar is an important index reflecting traders' behavior in the virtual world as well, and people posting behavior is driven by significant events.

From trading volume trend of a single stock A, as shown in figure 3, monthly trading volume trend is similar with annual trading volume.As shown in figure 4, from the various stocks volume trend, every stock's trading volume trend is also similar. The abnormal trend in the figure is affected by the significant events. The consistency of each stock in abnormal time reflects the consistency of people ehavior in a certain environment, at the same time, these stocks also constitute behavior space domain of event impact.

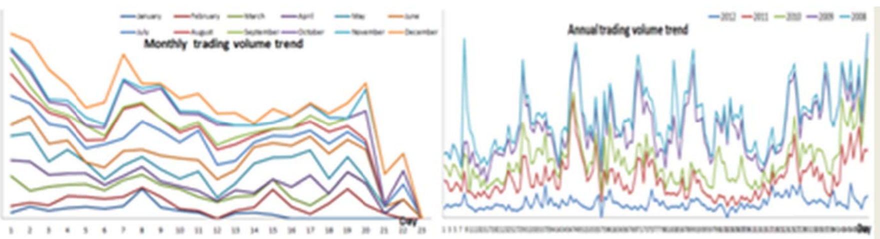


Fig.2 Trading volume trend

In the experiment, we analyze 26 stocks' trading volume data of wine industry in November to December 2012,Firstly,we analyze each stock abnormal periods, and these stocks which appear abnormal periods at the same time constitute a behavior spatial domain. As shown in figure3,The bright red area is the abnormal area of each stock.

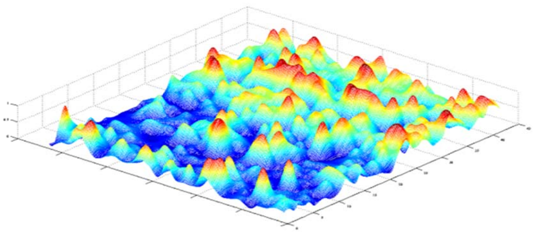


Fig.3 Stock trading volume of 26 kinds wine industry

The abnormal period of stocks.

Starting from each stock (individual), with historical trading data (that isthe background domain of time) as the background, extract the unusual period of stocks by the above algorithm 2.As shown below.

Table1Abnormal time periods of each stock

Stock	Abnormaltime period
Jiugui	[11/19,11/29]
Moutai	11/19,12/11,12/24
Wuliangye	[11/16,11/19],[12/3,12/5],[12/13,12/17]
Yanghe	[11/9, 11/12], 11/19,12/24,12/27
Shanxi Fenjiu	[11/16,11/19], [11/30, 12/5],12/14,[12/26, 12/27]
Swellfun	[11/15,11/19],12/10,[12/24,12/27]

The space-time model of group behavior

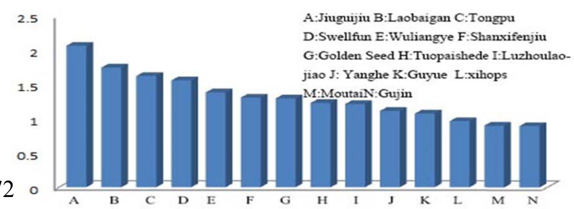
Through the previous analysis of abnormal time interval, we preliminarily establish the contact between each stock.AS shown in figure 5, we also need to use curvesimilarityalgorithm for related behavior spatial domain.

Tracing to the plasticizer events in November 19, 2012 as the source, we can find out the abnormal interval of all the stocks in November 19, 2012, and then through the previous algorithm 3,weuse line segment to trading volume time series of each stock. Table 2 shows The K value between the two intervals of periods before and after.

Table2 K values

Stock	k_1	k_2	k_3	Stock	k_1	k_2	k_3
Jiugui	-0.150189855	0.494768533	-0.404274797	Wuliangye	-0.18184553	0.266665426	-0.327298877
Moutai	-0.216869755	0.432170976	-0.230014863	Yanghe	-0.17267975	0.300421043	-0.155979363
Shanxi Fenjiu	-0.054322847	0.477183145	-0.559589328	Swellfun	-0.349288851	0.613708822	-0.259637235

With k - Medoids algorithm to clustering analysis to the K value, We can divide the stock to 2 class,asshown in figure 4. Plasticizerevents impact on the stocks can be measured by ratio ofthe K valueof the abnormal interval and the average K value, and when the ratio is less than 1, the stock is not affected by the event, as shown in figure 5.



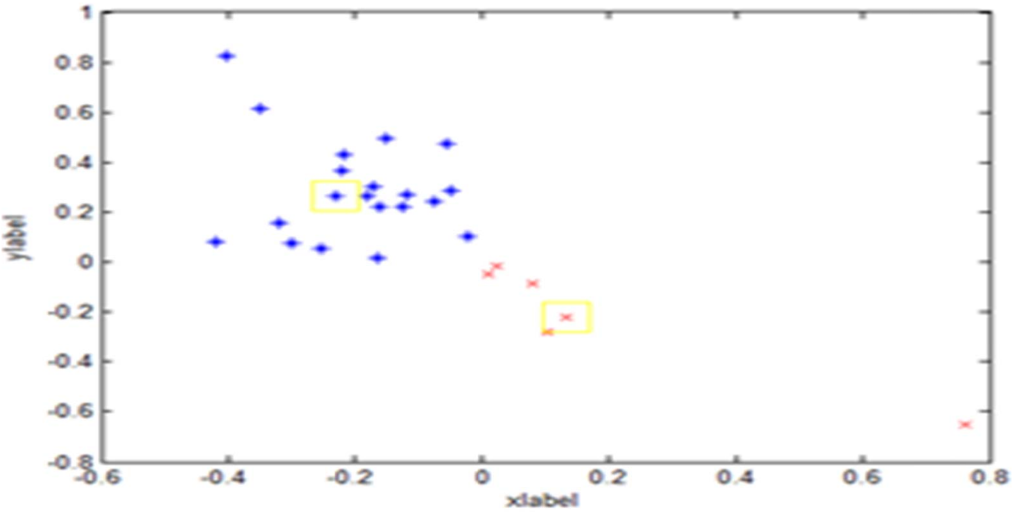
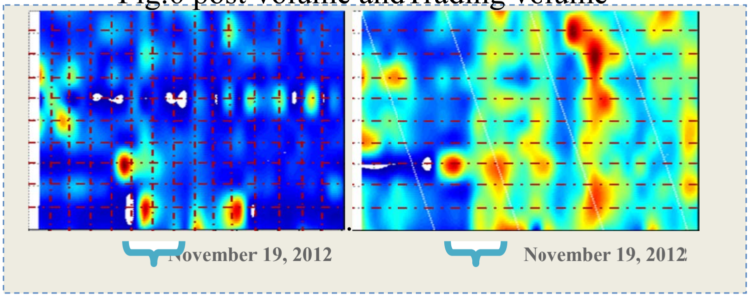


Fig.5The influence degree of plasticizer

Finally we get the stocks affected by the plasticizer event, they constitute the behavior spatial domain of plasticizer event. In order to verify the experiment results, we adopt the same method to analyze post amounts in related post bar in this period, we also obtain the same individual stocksaffected by the event. From the experimental results, the trading volume changes and post amountschanges have strong consistency during the plasticizer event, as shown in figure 6

Fig.6 post volume andTrading volume



Conclusion

Through the above experiment, we get a time-spatial domain of the plasticizer event, since November19in2012,theeventin liquor-making field affects the stock includingJiuguijiu,Laobaigan,Tongpu,Swellfun,Wuliangye,ShanxiFenjiu,Golden seed and so on, These experiments illustrate not only the trading behavior in the real world,but also the posting behavior in virtual world is affected by the event, theirs influence also has the strong consistency of field.

Acknowledgement

This work is supported by The Ocean Public Welfare Project of the Ministry of Science and Technology (No.201105033) and The National Natural Science Foundation of China. (No.40976108)

References

[1] MuGuoHua, ZhouWeiXing, ChenWei, Kertesz Janos. Order flow dynamics around extreme Price exchanges on an emerge in stock market.NewJoumalofPhysies,2010,12:07503

[2]WuXuChuan, HePeng. China open-end fund herd behavior analysis[J]. Journal of Financial

Research, 2005 (5): 60-69.

[3]LIU Yi, BAO De-pei, YANG Ze-hong, ZHAO Yan-nan.Research of New Similarity Measure Method on Time Series Data[J].Application Research of Computers, 2007, 24(5): 112-114.

[4] Li Guo'ai, CaoZheng. On-Line Segmentation of Time-Series Data [J].Journal of Software, 2004, (11).

[5]LiuliJiaPengtaohua-can he, etc.Overview of time series data mining [J]. Application Research of Computers, 2007, (11):15-18.

[6] Hung N Q V, Anh D T. Combining SAX and Piecewise Linear Approximation to improve similarity search on financial time series[C]//Information Technology Convergence, 2007. ISITC 2007.International Symposium on. IEEE, 2007: 58-62.

[7]LUO Xin OUYANG Yuan-Xin XIONG Zhang YUAN Man.The Effect of Similarity Support in K -Nearest-Neighborhood Based Collaborative Filtering[J].Chinese Journal of Computers, 2010.