INTELLIGENT AUTOMATION & SOFT COMPUTING, 2017
http://dx.doi.org/10.1080/10798587.2016.1267247

A Hot Event Influence Scope Assessment Method in Cyber-Physical Space for Big Data Application

Yunlan Xue[a], Lingyu Xu[b], Jie Yu[b] and Gaowei Zhang[b]

[a]Department of Computer Science, Guangdong Polytechnic Institute, Zhongshan, China; [b]School of Computer Engineering and Science, Shanghai University, Shanghai, China

ABSTRACT

The increase of scale and complexity of Internet big data presents unprecedented opportunities on Cyber-Physical Systems (CPS). The incompleteness and incredibility of Internet big data are challenging issues for confirming the event influence scope. To solve the above problem, we propose Cyber-Physical Space Event Model (CPSEM) to analyze event influence in multi-viewer, which maps real data into Cyber Space (CS) and Physical Space (PS). In addition, we propose Event Influence Scope Detection Algorithm (EISDA) to detect the scope of a hot event in Cyber Space and Physical Space.

KEYWORDS

CPS; Cyber space; Physical space; Internet big data; event influence scope

1. Introduction

Cyber-Physical Systems (CPS) is a comprehensive computing, cyber and physical environment for the multidimensional complex systems. CPS is a network physical system with a controllable, credible and scalable function on the basis of environmental perception, depth fusion calculation, and communication and control ability. It realizes depth fusion and increases or extends new function with real-time interaction by the interaction and feedback of calculation process and physical process, and it tests or controls a physical entity in the form of safe, reliable, efficient and real-time. Recent years, due to potential benefits to society, economy, and the environment, more and more scholars take attention on the research related to CPS. CPS as the next generation of engineered systems requires tight integration of computing, communication, and control technologies in many application domains (Rajkumar, et. al. 2010). The CPS as an enabling technology makes many innovative applications to become a reality and give birth to the myriad business model. CPS bring profound changes and challenges to human society, which should be put together with other important new technology review, such as CPS and big data, CPS and artificial intelligence, CPS and cloud computing, CPS and 3D printing, etc. We combine CPS and financial stock big data for business applications, for example, providing reference value to investors, security companies, and investment companies.

In this paper, the specific space of Internet having data mapping of event is called Cyber Space (CS), and Physical Space (PS) is the Internet carriers' data mapping and can reflect the real physical world. Fortunately, CS and PS can make up the limitation of internet data about incompleteness and unreality. When an event happens, the attributes of CS and PS related to the scope of event influence both become abnormal, as shown in Figure 1 (the scope distribution of one event influence). For example, attributes of a hot event (browses, clicks, posts and replies etc.) have changed in the CS. Changes are not only in the CS, however, another attributes of event (the quantity of deal, prices, the activity of stock trade related to the event) do the same in the PS.

When an event happens, an event is mapped to CS and PS as the interaction between CS and PS in the cycle of event. Cyber and Physical Space of Event Model (CPSEM) combine CS with PS to assess event influence scope in CPS. In Cyber Space of Figure 1, it contains some phenomena of Cyber Space for business applications, such as websites, forums, users, etc. In Physical Space of Figure 1, it only contains the real physical world, here uses a finance data to show the real behavior of deal. In Cyber-Physical Space of Figure 1, it means the result of a hot event influence scope assessment based on Cyber Space and Physical Space. The points are members of a hot event and edges are the correlations of members, and the coloured points are members affected by event which means the event influence scope.

The rest of this paper is organized as follows: Section 2 gives the related work. Section 3 defines and constructs CPSEM. Section 4 proposes EISDA algorithm in CPSEM. Sections 5 show experimental results. Section 6 gives conclusions.

2. Related Work

In CPS field, Internet data mining is a hot research subject because of the emergence of large social networks such as Facebook, Flickr and Twitter. In particularly, event influence assessment is one of the frontier subjects (Smith, 2001). Traditional researches of event focus on discovering the hot information in Internet, extraction of hot issues (Hu, et. al. 2007), retrieval of blog tendentiousness (Liu, et. al. 2013), etc. The evaluation methods of information influence are various, such as the minimum energy evaluation of information based on tree structure (Kolmogorov, 2006), subjective stress and experience knowledge (Esteve and Garcia-Haro, 2003), or evaluation of information influence based on hierarchical model (Derek, 1965).

CONTACT Jie Yu jieyu@shu.edu.cn

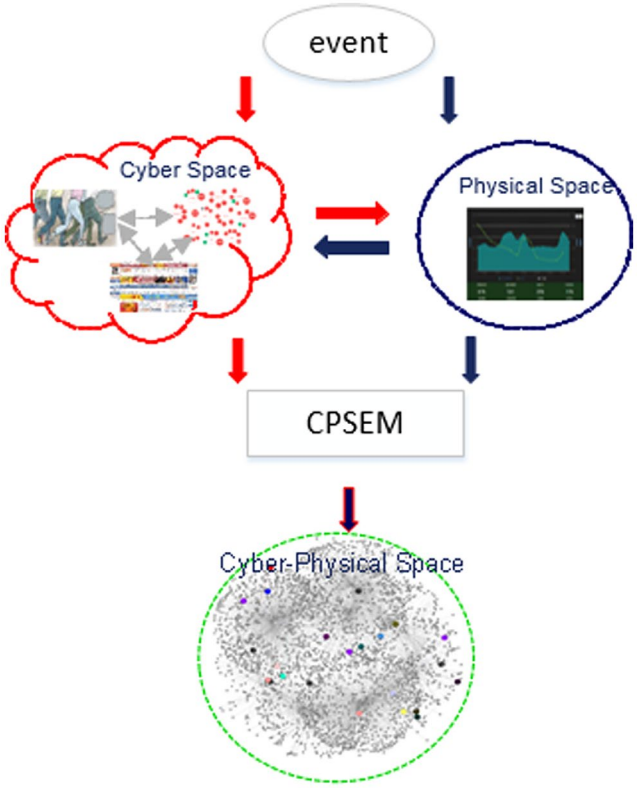


Figure 1. The Cyber-Physical Space for Business Application.

The research on the hot event influence scope assessment is another important research subject. The importance of the hot event influence scope assessment on government policy and business application plan attract attention of relevant scholars (Quinn, 2003; Walle, 1995). The subject concept of a hot event influence scope (Getz, 2009; Getz, 2008) and basic framework, research of a hot event influence scope is one of basic content (Dai, 2011). For instance, in 11/11/2012, Chinese Internet happened the most large-scale commercial activities again: Taobao website sales totalled to 19.1 billion yuan. The Taobao can cope with such a great event, because they comprehensively analysed the data of event influence scope that contains past trade volume and super-high concurrency requirements, especially consumption, search and browsing habits of users (Holland, 1997).

Internet records the information of people's living, working, studying, and all other aspects that can almost reflect all kinds of behaviours of people and social development situations. However, the inherent complexity of Internet big data makes its expression, understanding, perception and calculation with enormous challenges. Many traditional data analysis and mining tasks become extremely difficult, such as information retrieve (Zhang, et. al. 2012) and theme discovery (Waldo, et. al. 2015), semantic and emotional analysis (Zhang, 2013). How to accurately describe an event from Internet big data becomes a significant research for CPS.

It is because of the coexistence of CS and PS, we are from two spaces to analyse the characteristics of an event. It is not only beneficial to analysis accuracy, but also more completely to research a hot event for Internet big data processing. In this paper, we propose a Cyber and Physical Space of Event Model (CPSEM), which analyses the trajectory of an event in space of CS and PS. The idea of the CPSEM provides multi-views to describe an event, which is more comprehensive and compensate than incompleteness and incredibility of Internet data. In the space, we propose an Event Influence Scope Detection Algorithm (EISDA) based on event time series. EISDA defines the scope of event influence in cyber space and physical space by CPSEM.

3. The Cyber-Physical Space Event Model (CPSEM)

Based on the idea of event detection and tracking (Allan, et. al. 1998), the definition of events in this paper is given. If event attributes of some members become abnormal in some time and space, and these abnormal values are higher than a certain threshold, then there is an event happening.

When a hot event happens, event attributes in CS and PS will become abnormal. For example, the discussion will become abnormal with the birth, development, climax and decline of an event in BBS of CS, such as attention, participation and interaction. Similarly, the people's behaviours change abnormal in stock market of PS, which is mainly manifested by activity of trade, change of prices, and so on. When we map an event to CS and PS, we can use the data mapping of event attributes from CS and PS to analyse this event.

3.1. Cyber Space for Event Influence Scope Assessment

When an event happens, people freely express views and emotion about this event in Internet. Event influence is mainly reflected by the degree of participation and attention of Internet, such as *Post, Visit, Reply*, and its influence scope is detected by the degree of different affected attributes. The CS records the trajectories of attention, perspective, emotion, etc. psychological change forms of users on Internet about the event.

Definition1. Cyber Space of Event (E_{CS})

In CS, the trajectory of event influence is mainly reflected by the degree of attention, participation and interaction. This paper defines E_{CS} with a group of three units. In this paper, CS denotes Cyber Space, which is the data mapping of psychological change from users on Internet. E_{CS} is Internet public opinion data mapping space of event, which reflects the whole event development lifecycle and contains the trajectories of events in CS. T is the whole work time of event influence in CS. M is a set, which contains one or more members affected by the event in CS. M_K is an element of set M. A_{CS} is the attribute set of event influence in CS based on time series, that is the forms of psychological change of users on Internet in trajectories of event in CS. $A_{CS}(I)$ is the element of set A_{CS}, which has various forms in CS. *Click* is one form of $A_{CS}(I)$, that is click volume from users on Internet about event, and it reflects the degree of attention from users in CS. *Post* is one form of $A_{CS}(I)$- that is the number of posts from users on Internet, and it reflects the degree of participation of users in CS. *Reply* is one form of $A_{CS}(I)$- and it is the number of replies from users on Internet, and it reflects the degree of interaction of users in CS.

$$e_{CS} = \{T, M, A_{CS}\} \tag{1}$$

$$M = < m_k | m_k \in M, 0 \le k \le n > \tag{2}$$

$$A_{CS} = < a_{CS}(i) \in \{Click, Post, Reply\},\ 0 \le i \le n > \tag{3}$$

Where E_{CS} is the data mapping of event in Internet carrier with partial and incomplete features. As can be seen from Definition 1, E_{CS} have the following features:

(1) The form of A_{CS} is diversified in CS, such as data, text and images, and so on. We select *Click, Post, Reply* as three main attributes to reflect event influence in CS, and define event influence scope through their trajectories.

(2) The time and the attributes are not limited in CS. For example, users can make post, click, reply, etc. at any time in stock forum, and the numbers of post, click and reply are not limited.

3.2. Physical Space for Event Influence Scope Assessment

PS reflects the real physical world. For example, trajectories of stock trading volume, price and change rate of price would be recorded by Stock Exchange. When an event happens in society, the data mapping is very limited from PS. How to capture valuable information from real space to compensate noise and lack of information in CS is important. The electricity stores and stock market are miniatures of real behaviours space. Due to the opacity of electricity stores trade data this paper chooses Chinese Stock Exchange as PS. When an event impacts on the stock market, its influence is mainly reflected on the number of buying and selling stocks, price movements, bearish or bullish sentiment of shareholder. Event influence scope is defined through the different degree activity of event attributes in PS.

Definition 2. Physical Space of Event **(E_{PS})**

An event with its members, the change of trade behaviours, prices of event directly reflect the trajectory of event in some time and space. This paper defines E_{PS} with a group of three units. In this paper, PS is Actual Behaviours Data Space, which is the data mapping of real behaviours from persons affected by an event on Internet carriers. E_{PS} is actual behaviours data mapping space of event, which records the trajectories of the event in real trade market. T is the whole effective work time of event influence in PS. M is the member, who has action under the event influence in PS, and the significance of M is the same as formula 2, but the set M is not same. A_{PS} is an attribute set of event influence in PS based on time series. In this paper, we mainly focus on stock events, so the attribute set contains all the original attributes of the stock market. There are *Close, High, Low, Open, Pre Close, Change, %Chg, %Turnover, Vol, Turnover, Market Cap, CSV.* $A_{PS}(I)$ is the element of A_{PS} set, and it has different forms of member in PS. *Close* is the closing price of one stock close the trading time. *High* is the highest price of one stock during one day trading. *Low* is the lowest price of one stock during the trading time. *Open* is the opening price of one stock at the beginning trading time. *Pre Close* is yesterday's closing price. *Change* is the amount money of rise or fall of one stock. *%Chg equal to Change* divide the *Pre Close*, which is the price change rate of one stock. *%Turnover* equal to the percentage of the result of *Vol* divide the total number of circulation shares. *Vol* is the amount of trade shares of one stock. *Turnover* equals to *Vol* multiply by traded shares, which is the amount money of trade shares. *Market Cap* equals to *Close* multiply by the total shares, which is the total market price of stock. *CSV* equals to *Close* multiply by the circulation shares, which is the circulation price of one stock.

$$e_{PS} = \{T, M, A_{PS}\} \tag{4}$$

$$\begin{aligned} A_{PS} =< a_{PS}(i) \in \{&Close, High, Low, Open, \\ &Pre\text{Close}, \text{Change}, \%Chg, \%Turnover, \\ &Vol, Turnover, MarketCap, CSV\}, 0 \le i \le n > \end{aligned} \tag{5}$$

Because of several price present forms are similar, so we use *Amplitude* to reflect the price activity and action difference in PS, which is the combination of a variety of prices, including *High, Low* and *Pre Close*.

$$Amplitude = \frac{High - Low}{Pre\ Close} \tag{6}$$

Due to the liner relationship between many attributes, this paper chose three attributes to calculate event influence in PS based on Pearson correlation coefficient. The correlation coefficients between *%Chg, %Turnover*, and *Amplitude* are smaller.

E_{PS} is real data of social behaviours, which not only filters false data but also adds incompleteness of Internet data. As can be seen from Definition 2, E_{PS} have the following features:

(1) The data of PS from the real trade market, which contains trade data in stock exchange or electricity trading platform (Alibaba, eBay, Amazon, etc.). The forms of attribute are mainly a series of data sets from PS, such as price, volume, price, and so on.

(2) The domain of time and attributes has their specific limitations. For example, the maximum change of rise and drop is 10% in stock market, the opening time of stock market is at 9:30–11:30 and 13:00 - 15:00 on a working day, and 9:00–9:30 is bidding Time.

3.3. The Construction of Cyber-Physical Space Event Model (CPSEM) for Event Influence Scope Assessment

As an event is mapped to CS and PS, its performance is reflected by different A from different M. For example, when an event occurs, the stock trade volume, prices become abnormal in stock exchange because of persons affected by the event. However, the event also affects psychology change of persons, such as the attention, comments, reply, etc. of users in stock forums will become abnormal too, and vice versa.

Definition 3. Cyber-Physical Space Event Model (CPSEM)

The assumption in this paper is that the trajectory of event mapping in CS and PS is constructed by *T, S, M, A*. This paper constructs CPSEM on basis of CS and PS, and defines CPSEM with a group of four units. E_{CPS} means what is happening, which includes time, space, member and attributes of an event, and the trajectory of an event is recorded by the activity of attributes. T is the time stamp attached to each attribute, which shows the post time of messages or action, indicating the approximate time of event. S means where it is happening, which records the location of an event. The space associated with the profile of each event, which is mapped into CS and PS. M means who are involved, which can be one or more. A is the attribute set of event, which is union set of A_{CS} and A_{PS}. It is affected by events, while the trajectory sets of many attributes are mapping of event trajectory.

$$e_{CPS} = \{T, S, M, A\} \tag{7}$$

$$S = \{CS \cup PS\} \tag{8}$$

$$A = A_{CS} \cup A_{PS} \tag{9}$$

Where T and M are the same as Definition 1 and Definition 2 in this paper. As can be seen from the Definition 3, the combination of CS and PS solves incompleteness and unreality with using Internet data based on CPSEM.

Analysis of CS and PS

In the paper we describe an event by both CS and PS comprehensively. There are some common features and characteristics of CS and PS, which are presented as follows:

(1) The common features of CS and PS:

The comments or real trading behavior on Internet of investors once formed, which will leave trajectories on the carriers of public opinion space (CS) or real behavior space carriers (PS), and they can be recorded and stored. A of CS or PS is spontaneously generated, spread, and accepted by investors. The interaction processing is more orientation that is the network evaluation information interacted with the attention of investors, click, and reply. This interaction brings unprecedented influence on investor group behavior due to the interactivity by investor's attention, and vice versa.

(2) The characteristic properties of CS and PS:

The data of CS contains a large number of loose and redundant information, but the data of PS is real data that cannot reflect what the thinking of investor is. The comments of CS reflect user's ideology or psychological changes, but which cannot really reflect the trading behavior. Therefore, we need to combine the CS and PS for the research hot event of CPS.

4. A Hot Event Influence Scope Assessment Method and Algorithm Description

With the development of Internet big data, the characteristics of events not only in physical space has its data mapping, but also has its data mapping in cyber space. However, two spaces are intertwined and impact each other, and their respective data mapping is not able to fully reflect the real feature of event. So we need to extract information of event from two spaces, so as to improve the analysis and research for events.

4.1. Event Influence

In the field of information influence, there are many commercial news search engines, such as Google news, Yahoo news, and a full list of business news engines is given in literature. In addition, there have news information processing prototype systems, such as News In Essence (Radev, et. al. 2009; Radev, et. al. 2001) and QCS (Dunlavy, et. al. 2004). Although in news event retrieval and information processing in ceaseless effort, the influence of network news event involves little sort of academic research. Ranking news based on event influence (Xu, et. al. 2016a; Xu, et. al. 2016b) is mainly based on the timing of news event and news event reproduced information to prioritize. And other scholars use layout of pages and news event transfer information to sort of a web page (Yao, et. al. 2006; Xuan, et. al. 2016).

Traditional information influence calculation of Weibo is based on the different weights of information to accumulate the energy of one Weibo, and a similar calculation of a news event influence based on complex network is on the sum of edge and degree a to measure the influence size. While this article measures the event influence from a different angle, we do not have to integrate the various elements to measure the event influence, but rather discover the change rule of event influence based on the fusion of each element. In addition, we define the scope of event, and give the transmission range of event influence from small to large.

This paper puts forward a new concept of event influence calculation inspired by news information such as Weibo influence calculation.

$$\begin{cases} inf_{CS}(m_k) = \bigcup_{i=1}^{n} a_{CS}(i), 1 \le k \le m, 1 \le i \le n \\ inf_{PS}(m_k) = \bigcup_{i=1}^{n} a_{PS}(i), 1 \le k \le m, 1 \le i \le n \end{cases} \tag{10}$$

Where $INF_{CS}(M_K)$ is to calculate event influence of the *kth M* in CS; $INF_{PS}(M_K)$ is the function to calculate event influence of the *kth M* in PS.

4.2. Event Influence Scope Assessment Method

The scope of information influence is mainly based on complex networks, such as using the greedy algorithm to calculate the influence maximization problem in social network (Richardson et al., 2002), and the information transmission range based on center degree of heuristic rules (Kempe, et. al. 2003). But the events in this paper did not form a clear network structure between members, and most members are parallel without connected edges, so traditional methods of the scope of event influence do not apply in this paper.

Because of simple and fast features of the clustering algorithm (Waldo, et. al., 2015; Liu and Lampinen, 2002), it is widely used in various fields. In this paper, we use clustering to extract influence scope of an event based on the activity degree of event attributes. Based on the idea of Jaccard similarity measure (Zhang, 2013), the two nodes have the common number of neighbours is more, and the relationship between the two nodes is closer, so the influence is also bigger between them. We calculate the *AScope* of event influence from the overlapping degree of event influence distributions in CS and PS; on the other hand, we calculate the *SScope* of event influence from the merger degree of event influence distribution in CS and PS. We incorporate multidimensional attributes characteristics of event to improve the accuracy of the *SScope* and the *AScope* of event influence. In addition, we calculate difference of the *SScope* and the *AScope* of event influence to determine the boundaries of event. The smaller boundary means the spreading of event influence more concentrated, and the bigger means it more dispersed. The scope of event influence has various.

In formula 11, $Scope_{CS}(I)$ is the scope of event influence of one M scope in CS, which is the set of abnormal A affected by event in CS. $Scope_{PS}(I)$ is the scope of event influence of one M scope in PS, which is the set of abnormal A affected by event in PS. *Cluster()* is a cluster function, which gains abnormal M based on K-means cluster idea. If the total of A or the value of A becomes abnormally higher than a certain threshold in one M, we believe this M is a part of event influence scope. In CS or PS, we put those affected M in one set, this set is $Scope_{CS}(I)$ or $Scope_{PS}(I)$.

$$\begin{cases} Scope_{cs}(i) = Cluster(inf_{CS}(m_k)), 1 \le i \le n, 1 \le k \le m \\ Scope_{ps}(i) = Cluster(inf_{PS}(m_k)), 1 \le i \le n, 1 \le k \le m \end{cases} \tag{11}$$

In formula 12, $Ascope_{PS}$ is the absolute scope of event influence in PS, which means all attributes of one M affected by event in PS. $SScope_{PS}$ is the support scope of event influence in PS, which means partial attributes affected by events in PS. $BOUNDARY_{PS}$ is the difference of $Ascope_{PS}$ and $Sscope_{PS}$, which equals to $Sscope_{PS}$ minus $Ascope_{PS}$, and it means the boundary of event influence in PS.

$$\begin{cases} AScope_{PS} = \bigcap_i Scope_{ps}(i), 1 \le i \le n \\ SScope_{ps} = \bigcup_i Scope_{PS}(i), 1 \le i \le n \\ Boundary_{PS} = SScope_{ps} - AScope_{PS} \end{cases} \quad (12)$$

In formula 13, $Ascope_{CS}$ is the absolute scope of event influence in CS, which means all attributes of one M affected by event in CS. $Sscope_{CS}$ is the support scope of event influence in CS, which means partial attributes affected by event in CS.$BOUNDARY_{CS}$ is the difference of $Ascope_{CS}$and $Sscope_{CS}$, which equal to $Sscope_{CS}$ minus $Ascope_{CS}$, and it means the boundary of event influence in CS.

$$\begin{cases} AScope_{CS} = \bigcap_i Scope_{CS}(i), 1 \le i \le n \\ SScope_{CS} = \bigcup_i Scope_{CS}(i), 1 \le i \le n \\ Boundary_{CS} = SScope_{CS} - ascope_{CS} \end{cases} \quad (13)$$

In formula 14, $Ascope_I$ is the absolute scope of the event influence. $SScope$ is the support scope of event influence. $Boundary$ is the difference of $AScope$ and $SScope$, which equals to $SScope$ minus $AScope$, and it means the boundary of event influence.

$$\begin{cases} AScope_i = \bigcap_i (Scope_{PS}(i) | Scope_{CS}(i)), 1 \le i \le n \\ SScope = SScope_{PS} \cup SScope_{CS} \\ Boundary = SScope - AScope \end{cases} \quad (14)$$

4.3. Data Pre-processing

Because this paper uses the trading data from the stock market as data of event attributes in PS, there is most linear relationship between them. In order to eliminate redundancy and improve efficiency, this paper calculates the correlation coefficients between each A to reduce multi-dimensions of data based on the idea of Pearson correlation coefficient. This paper chooses A for experimental analysis whose correlation coefficients are smaller.

$$\rho_{x,y} = \frac{\sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\sigma(x - \bar{x})^2 \sigma(y - \bar{y})^2}} \quad (15)$$

Where X and Y are the difference A. $\rho_{x,y}$ is the correlation coefficient of X and Y, which means the similarity of X and Y.

There are a lot of big redundancies of internet data, and data source of CS is crucial to the analysis of experimental results. We select two professional websites for comparative analysis and calculate the Pearson correlation coefficients between them. If the Pearson correlation coefficient is higher than 0.6 (experience), then we choose one data source as the analysis object, which could effectively improve the efficiency of data processing. Due to the particularity of events (stock of events) for this paper, generally, the stock of events is both good and bad emotion. We need to judge the emotion (positive or negative) of events before the data processing, and the emotion will be removed which is opposite to events. Above processing not only achieves the effect of data compression, but also improves the accuracy of the experimental data based on the noise removing.

5. Experiment and Analysis

5.1. Feature Extraction

When an event happens, the scope of its influence is reflected by the strength of A, and we define the scope of event influence by fusing multiple A from CS and PS. In this paper, the data source of PS is Chinese Stock Exchange, and its features are unique. However, data source of CS is more complex, because each website has all kinds of comments and analysis of events. We chose Sina and Eastmoney websites of Chinese authority sites, and analyse posts from January 2013 to May 2014 as shown in Figure 2. The Pearson correlation coefficient between them is 0.877, so they have a strong linear relation. Therefore, a professional Website (Eastmoney) as data source of CS is selected in this paper.

We choose the data of Chinese Stock Exchange as the object of PS, which has a linear relationship between them. Since there is redundancy and noise in attributes of PS's data, the Pearson correlation coefficient is used to calculate the correlations between each attribute pair to reduce of dimensions of data. There is smaller correlation from *Change, %Chg, Vol, %Turnover, Close, High, Low, Open, Pre Close*. We choose *%Chg* of the normalized attribute based on the correlation is 0.986

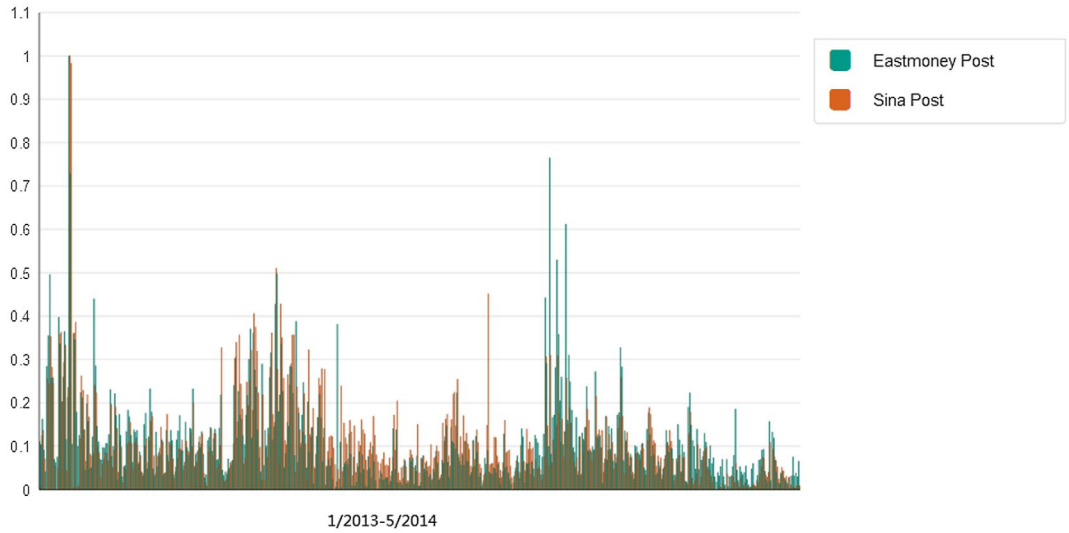


Figure 2. Sina and Eastmoney Post from Jan 2013 to May 2014.

Table 1. All A of Plasticizer Events.

AM	%Chg	%Turnover	Amplitude	Click	Post	Reply
	000,799600,779600,702600,559600,199000,568000,858600,809600,197600,616000,596600,519	000,799600,779600,702600,197600,559600,199	000,799600,779600,702600,197600,559600,199002,304000,568000,596000,858600,809600,616600,519600,059	000,799600,779600,197000,858600,199000,729200,869	000,799600,779600,702000,858002,304600,559	000,799600,779600,702200,869000,799600,059600,543600,809600,616

Table 2. The Different Scope of Plasticizer Event Influence.

ScopeM	$Ascope_{PS}$	$Sscope_{PS}$	$Boundary_{PS}$	$Ascope_{CS}$	$Sscope_{CS}$	$Boundary_{CS}$	$Ascope_{MAX}$	Ascope	Boundary
	000,799600,779600,702600,197600,559600,199	000,799600,779600,702600,197600,559600,199002,304000,568000,596000,858600,809600,616600,519600,059	002,304000,568000,596000,858600,809600,616600,519600,059	000,799600,779	000,799600,779600,197000,858600,199000,729200,869002,304600,702600,559600,059600,779600,059600,543600,809600,616	600,197000,858600,199000,729200,869002,304600,702600,559600,059600,543600,809600,616	000,799600,779	000,799600,779600,702600,197600,559600,199002,304000,568000,596000,858600,809600,616600,519600,059000,729200,869600,543	600,702600,197600,559600,199002,304000,568000,596000,858600,809600,616600,519600,059000,729200,869600,543

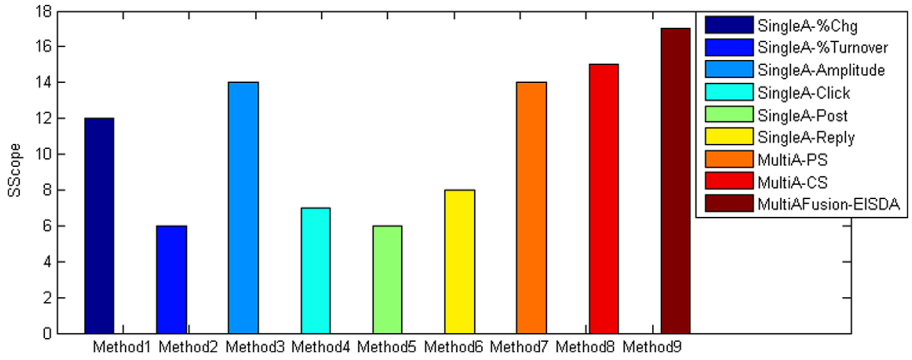


Figure 3. SScope of every A and PS, CS, EISDA in Plasticizer Event.

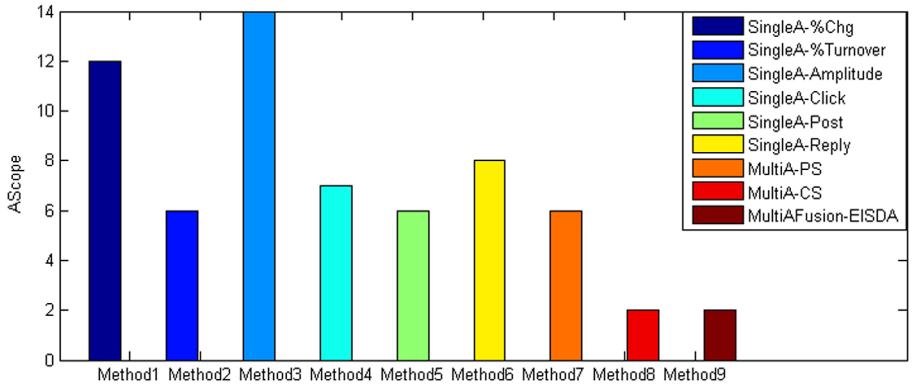


Figure 4. AScope of every A and PS, CS, EISDA in Plasticizer Event.

between *%Chg* and *Change*; and the correlation coefficient is larger in these attributes, such as *Close, High, Low, Open, Pre Close*, and *Amplitude* is constructed by *High, Low, Pre Close* by equation 6. Since *Amplitude* is normalized and conducive to experimental analysis, we choose *Amplitude* as the third attribute of PS.

5.2. Event Influence Scope Assessment

The research of this paper is stock event in the financial field for business applications, and the research data-set has 15 GB. The experimental data of this paper is chosen from Chinese stock market data as PS data and the corresponding stock forum data (www.eastmoney.com) as CS data. In terms of Plasticizer event, we filter out the five stocks (000,752, 000,869, 000,929, 002,461, 600,090) (positive) which are opposite to the emotion of event (negative) from Shanghai Stock Exchange 27 stocks of wine plate. We use the remaining 22 stocks (000,568, 000,596, 000,729, 000,799, 000,858, 000,995, 002,304, 200,869, 600,059, 600,132, 600,197, 600,199, 600,365, 600,519, 600,543, 600,559, 600,573, 600,600, 600,616, 600,702, 600,779, and 600,809) as the experimental data through the data cleaning. Through the clustering and integration of event attributes (*%Chg, %Turnover, Amplitude, Click, Post, Reply*), we choose

the Plasticizer event *AScope* (absolute scope) and *SScope* (support scope) and *Boundary*.

Table 1 is the affected *M* of the Plasticizer event, which contains six attributes of *%Chg, %Turnover, Amplitude, Click, Post,* and *Reply*. We can see the larger scope of event influence focusing on *%Turnover* than *%Chg* and *Amplitude* in PS from the Table 1, and in CS, the Event influence scope is larger in *Post* than *Click* and *Reply*. The bold *M* is affected by every *A*, and they have the greatest impact by the event. We find a propagation rule that is event influence size of *M* affected by how many *A*. The *M* are greater affected with nearer distance to main *M* (000,799) of event, such as 600,779.

We calculate $Ascope_{PS}$, $Sscope_{PS}$, $BOUNDARY_{PS}$ in PS, and $Ascope_{CS}$,$Sscope_{CS}$,$BOUNDARY_{CS}$ in CS, and $Ascope_{MAX}$, *SScope*, *Boundary* combination PS and CS. The scope of Plasticizer event influence is as shown in Table 2.

In our experiment, method1, method2, method3, method4, method5, method6, calculate event influence of *M* by the single *A*; method7 and method8 calculate event influence of *M* by multiple *A* respectively with PS and CS; Method9 calculate event influence of *M* by multi-factor fusion based on Cyber-Physical Space. We find some rules of the event from financial field based on the multiple *A* test analysis. When we determine the *SScope* (Figure 3) of event influence, the scope based on PS and CS is larger than the scope of single *A* (*%Chg, %Turnover, Amplitude, Click, Post, Reply*), and the scope of EISDA is the largest. When we determine the *AScope* (Figure 4) of event influence, the scope based on PS and CS is smaller than the scope of single *A* (*%Chg, %Turnover, Amplitude, Click, Post, Reply*), and the scope of EISDA is the smallest. Lastly, when we determine the *Boundary* (Figure 5) of event influence, the *Boundary* of EISEA is wider than the others methods of CS and PS.

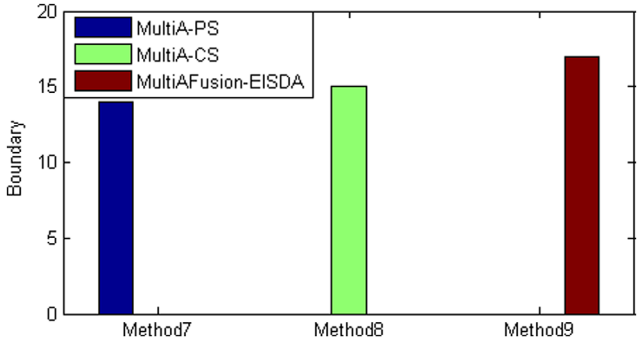


Figure 5. Boundary of every A and PS, CS, EISDA in Plasticizer Event.

In this paper, we make the tendency division of the boundary of event influence on the basis of the above experiments, which are then conducive to the research of diffusion direction of event influence. The result of the scope of Plasticizer event influence is as shown in Figure 6. We can see the diffusion direction of Plasticizer event influence, which is from strong too weak as shown below. 600,779,000,799→600,702→600,197,600,559,600,199→000,858,600,809,600,616→000,568,002,304,600,059,200,869,600,519→000,729,000,596,600,543.

$$AScope_5 = \bigcap_{i=5}(Scope_{BS}(i)|Scope_{OS}(i)) = \{600, 702\}$$

$$AScope_4 = \bigcap_{i=4}(Scope_{BS}(i)|scope_{OS}(i)) = \{600, 197, 600, 559, 600, 199\}$$

$$AScope_3 = \bigcap_{i=3}(Scope_{BS}(i)|Scope_{OS}(i)) = \{000, 858, 600, 809, 600, 616\}$$

$$AScope_2 = \bigcap_{i=2}(Scope_{BS}(i)|Scope_{OS}(i)) = \{000, 568, 002, 304, 600, 059, 200, 869, 600, 519\}$$

$$AScope_1 = \bigcap_{i=1}(Scope_{BS}(i)|Scope_{OS}(i)) = \{000, 729, 000, 596, 600, 543\}$$

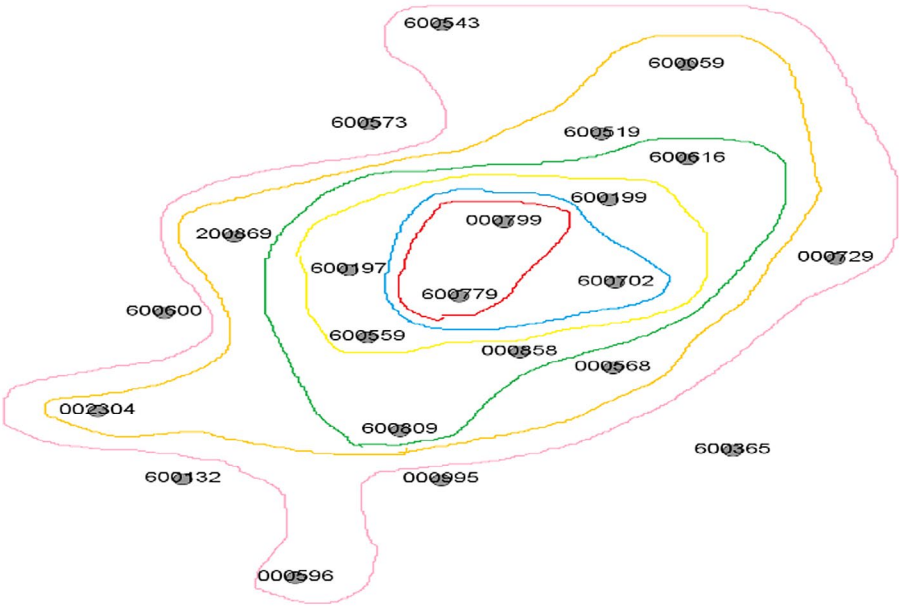


Figure 6. The Scope of Plasticizer Event Influence Decay Distribution.

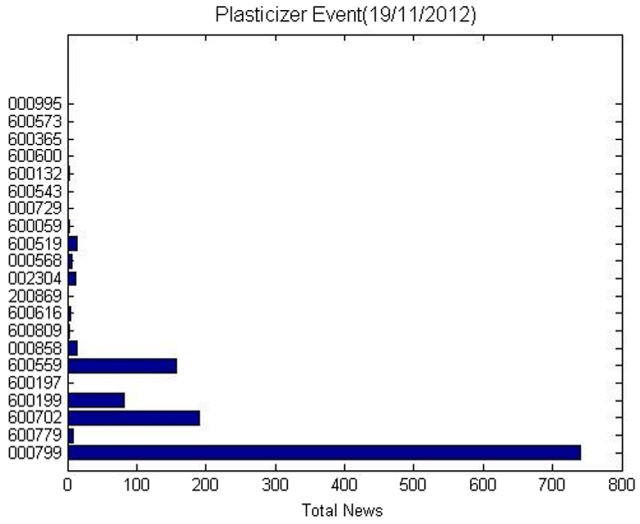


Figure 7. News of Plasticizer Event from Baidu Website.

5.3. Experimental Analysis

We measure the distribution scope of event influence through the total news of affected member in the same day of event occurred, and the news data from Baidu search in 19/11/2012. If the key words co-occurrence of member and event in one new, we think that the member is affected by the event. The stocks (000,779, 600,779, 600,702, 600,199, 600,197, 600,559, 000,858, 600,809, 600,199, 200,869, 002,304, 000,568, 600,519, 600,519, 600,059, 000,729, 600,543, 600,132, 600,600, 600,365, 600,573, 000,995) use key words (Jiuguijiu, Shuijinfang, Tuopaishede, Jinzhongzijiu, Yilite, Laobaiganjiu, Wuliangye, Shanxifenjiu, Jinfengjiuye, ZhangyuB, Yanghegufen, Luzhoulaojiao, Guizhoumaotai, Guyuelongshan, Yanjingpijiu, Mogaogufen, Chongqinpijiu, Qingdaopijiu, Tongpugufen, Huiquanpijiu, Huangtaijiuye). The key word of event is Plasticizer. Figure 7 is the result of affected members in Plasticizer event in 19/11/2012.

From Figure 7, we can see the distribution scope of Plasticizer event influence accord with the scope of event influence by EISDA. From down to up the event influence becomes smaller in Figure 7, and from center to outer the event influence also becomes smaller in Figure 6.

6. Conclusions

In this paper, we proposed a Cyber-Physical Space Event Model (CPSEM) and an Event Influence Scope Detection Algorithm (EISDA) from Cyber-Physical Space, and made relevant experiments on the basis of formal description. The CPSEM analysed events from Cyber Space (CS) and Physical Space (PS) which effectively made up the sidedness only from one space. In addition, EISDA was proposed based on CPSEM in the further, which identified a hot event influence scope on the basis of two spaces. Experiments verified that CPSEM solved incompleteness and unreality of internet data, and EISDA identified scope and diffusion direction of event influence from a new perspective and have improved the experiment accuracy for business application.

Disclosure statement

No potential conflict of interest was reported by the authors.

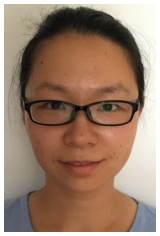
Funding

This work is supported by The Ocean Public Welfare Project of the Ministry of Science and Technology (No. 201,105,033) and the National Natural Science Foundation of China. (No. 40,976,108).

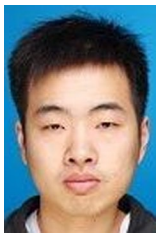
Notes on contributors

Yunlan Xue is a lecturer in the Department of Computer Science and Guangdong Polytechnic Institute. She received a B.S. degree in Computer Science in Hefei University of Technology, an M.S. degree in Computer Science in Inner Mongolia University and a Ph.D. degree from the School of Computer Engineering and Science at Shanghai University. Her research interests include data mining and financial big data analytics.

Lingyu Xu is a professor in the School of Computer Engineering and Science at Shanghai University. He received a B.S. degree, an M.S. degree and Ph.D. degree from the Northeastern University. His research interests include information fusion, data mining, marine and financial big data analytics.



Jie Yu is an associate professor in the School of Computer Engineering and Science at Shanghai University. She received a B.S. degree in Computer Science in Hefei University of Technology and an M.S. and Ph.D. degree in Computer Science at University of Science and Technology of China. Her research interests include information fusion, data mining, and financial big data analytics.



Gaowei Zhang is a Ph.D. student in the School of Computer Engineering and Science at Shanghai University. He received a B.S and M.S degree from Nanjing University of Science and Technology, Shanghai University in 2010, 2014, all of the department of Computer science. His research interests include data mining, machine learning and financial big data.

References

Allan J., Carbonell J., Doddington G., Yamron J., & Yang Y. (1998, February). Topic detection and tracking pilot study final report [C]. Proceedings of the DARP A Broadcast News Transcription and Understanding Workshop.

Dai, G. (2011). Event tourism in China: Developments problems and prospects [J]. In *Tourism planning and design: Event, City, Tourism* (pp. 6–14). Beijing: China Building Industry Press.

Derek J. (1965). Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front [J]. Science, *149*, July 30, 510–515.

Dunlavy, D.M., Conroy, J., & O'Leary, D.P. (2004). QCS: A tool for querying, clustering, and summarizing documents [J]. *Information Processing and Management, 43*, 1588–1605.

Esteve, P.-S., & Garcia-Haro, J. (2003). Fluid-flow approach to evaluate the information loss probability in a finite buffering switching node under heterogeneous ON/OFF input traffic sources [J]. *Performance Evaluation, 51*, 153–169.

Getz, D. (2008). Event tourism: Definition, evolution, and research [J]. *Tourism Management, 29*, 403–428.

Getz, D. (2009). Event studies: Theory, research and policy for planned events [J]. Tourism Management, *47*, 1665—1666.

Holland, J. (1997). *Emergence: From Chaos to order [M]*. Redwood City, California: Addison-Wesley.

Hu Y., Zhichun, X., & Panda, P. (2007). Modeling deceptive information dissemination using a holistic approach[C]. ACM, SAC '07 Proceedings of the 2007 ACM symposium on Applied computing, pp. 1591–1598.

Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence in a social network [C]. Proceedings of the 9th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, USA, pp. 137–146.

Kolmogorov, V. (2006). Convergent Tree-reweighted message passing for energy minimization[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*, 1568–1583.

Liu, J.-h., & Lampinen, J. (2002). A fuzzy adaptive differential evolution algorithm [C]. Proceedings of the 10th IEEE Region Conference on Computers, Communications, Control and Power Engineering, pp. 606–611.

Liu, P., Tang, J., & Wang T. (2013). Information current in Twitter: Which brings hot events to the world[J]. ACM, WWW '13 Companion Proceedings of the 22nd International Conference on World Wide Web, May 13–17, pp. 111–112.

Quinn, B. (2003). Symbols, practices and myth-making: Cultural perspectives on the Wexford Festival Opera [J]. *Tourism Geographies, 5*, 329–349.

Radev, D.R., Blair-Goldensohn, S., Zhang, Z., & Raghavan, R.S. (2001). *Newsinessence: A system for domain-independent, real-time news clustering and multi-document summarization [C]* (pp. 1–4). Morristown: Association for Computational Linguistics.

Radev, D.R., Blair-Goldensohn, S., Zhang, Z., & Raghavan, R.S. (2009). Interactive, domain-independent identification and summarization of topically related news articles [C]. *Volume 2163 of the series Lecture Notes in Computer Science,* 225–238.

Rajkumar, R., Lee, I., Sha, L., & Stankovic, J. (2010). Cyber-physical systems: The next computing revolution[C]. *Strasbourg,* 731–736.

Richardson, M., Domingos, P., & Glance, N. (2002). Mining knowledge sharing sites for Viral Marketing [C]//. Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Canada, pp. 61–70.

Smith, S.L.J. (2001). Measuring the economic impact of visitors to sports tournaments and special events (book review)[J]. *Annals of Tourism Research, 28*, 829–831.

Waldo, J. Cervantes-Solis, Chris, Baber, Ahmad, K., & Roman, M. (2015). Rule and theme discovery in human interactions with an 'internet of things' [C], British HCI '15: Proceedings of the 2015 British HCI Conference, July.

Walle, A.H. (1995). Festivals and mega-events: Varying roles and responsibilities [J]. *Festival Management and Event Tourism, 3*, 115–119.

Xu, Z., Zhang, H., Hu, C., Mei, L., Xuan, J., Choo, K.-K. R., Sugumaran, V. … Zhu, Y. (2016a). Building knowledge base of urban emergency events based on crowdsourcing of social media. *Concurrency and Computation: Practice and Experience, 28*, 4038–4052.

Xu, Z., Zhang, H., Sugumaran, V., Choo, K.-K. R., Mei, L., & Zhu, Y. (2016b). Participatory sensing-based semantic and spatial analysis of urban emergency events using mobile social media. *EURASIP Journal on Wireless Communications and Networking, 2016*, 40.

Xuan J., Luo X., Zhang G., Lu J., & Xu Z. (2016). Uncertainty analysis for the keyword system of web events. IEEE Transactions on Systems, Man, and Cybernetics: Systems, *46*, 829–842.

Yao, J.Y., Wang, J., Li, Z.W., Li, M.J., & Ma, W.Y. (2006). Ranking web news via homepage visual layout and cross-site voting [C]. *Volume 3936 of the series Lecture Notes in Computer Science*, 131–142.

Zhang L. (2013, January). Contextual and active learning-based affect-sensing from virtual drama improvisation [J]. Transactions on Speech and Language Processing (TSLP), *9*:8.

Zhang C., Shou L., Chen K., & Chen G. (2012). See-to-retrieve: efficient processing of spatio-visual keyword queries [C]. SIGIR'12: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, August