Event Multiple Influence Calculation and Relationship of Multiple Influence Discovery

Yunlan Xue[#1], Lingyu Xu[#2], Jie Yu[#3], Lei Wang[#4], Gaowei Zhang[#5], Qun Zhuge[#6]
[#]*School of Computer Engineering and Science, Shanghai University*
[1]xueyunlan@i.shu.edu.cn
[2]xly@shu.edu.cn
[3]jieyu@shu.edu.cn
[4]wanglei727@shu.edu.cn
[5]yiqigo0125@163.com
[6]franciszhuge@i.shu.edu.cn

***Abstract*—Due to the development of web technology, the research on Internet big data become more and more popular. However, information presents polymorphism and complexity under the status of the rapid development of Internet big data. Most of the traditional methods of event influence calculation use observable data to evaluate and ignore the decay influence of event. In fact, event influence growths from the moment of event happening to the end, meanwhile, it accompanis with a certain decay. What's more, event influence presents observable influence and decay influence on different network media. This paper proposes a new method to calculate event influence on different network media and find the relationships between observable influence and decay influence on different stages of event. Experiments show that better influence calculation will be achieved by decay algorithm based on Ebbinghaus forgetting curve and information fusion by considering interaction between observable influence and decay influence.**

Keywords—Internet big data; event influence; relationship; decay curve; forgetting curve

I. INTRODUCTION

Currently news event flood which comes from different websites spreads throughout the web. With the development of the techniques of Topic Detection and Tracking (TDT) [1], some web service can gather news information from different websites and structure it into news topics. Because of a lager of news events, reader is difficult to quickly find what they want to read. So ranking news events which can provide the most valuable and influential news events is a valuable research subject.

What's more, the news event is the main source of network and social public opinion place. It is important to judge news event influence for grasping the social public opinion tendency accurately and promptly, and its influence calculation has great significance on social security and other relevant aspects. The news event influence can be understood as composite of a group people that are affected by news event, the size of geographic range, and the force on social factors.

At present, news event influence calculation has the following traditional methods. Some scholars rank news event by efficiency and transfer information of it [2], and others sort web pages by the layout of page and news transfer information [3]. In addition, with the presence of the aging theory [4] for news event life-cycle modelling, the influence of a news event is to decay over time. The biological aging rate is not a constant ratio. In the same way, the decay value of a news event influence should not be a fixed parameter. The rate of decay is actually affected by current and past situations of a news event.

In people's social activities, information production and propagation in the form of topic in most time. Researchers find that different topics have different influence, even if influence of the same topic is also different in different groups [5, 6, 7]. Some scholars rank the events by the frequencies of event reported in time units and the number of consecutive effective time units [8], and a ranking algorithm is proposed to find the most authoritative news sources and identify the most interesting events in the different categories to which news article belongs [9]. A three-step automatic online algorithm for news issue construction is proposed, and time and quantity preference are separately used for ranking news events [10]. In their next work [11], an automatic online news topic rank algorithm is proposed based on inconsistency analysis between media focus and user attention. However, it only considers the constant decay in the aging model.

However, the rapid development of Internet big data under the new status, and information present polymorphism and complexity, event influence presents different trajectories on different stages. The event influence may present observable influence and decay influence, how to quantify the event influence and find the relationships between observable influence and decay influence, which are complementary or alternative relationship. In this paper, we aim to reveal event multiple influence and relationships of them in Eastmoney and Baidu, the biggest financial website and the biggest search engine in China.

The rest of the paper is organized as follows. We introduce preliminary concepts in Section 2. Extensive experimental results are presented in Section 3. We conclude the paper in Section 4.

II. PRELIMINARY CONCEPTS

1) Event Influence Description

The so-called event occurs in a particular time or place by one or more roles, which is composed of one or more actions, it means an action or state change. Event is the unit of

DOI reference number: 10.18293/SEKE2016-006

people understand and experience of the world, and meet people's normal cognition rule.

Event presents a variety of trajectories under the rapid development of network media. We choose stock news events in the field of finance. The news events cause official media and mainstream forums search and comments in Cyberspace. News events move by respective trajectories in different media, the calculation of event influence is not the same on different stages. Event has different observable influence and decay influence on different stages in different media. Our main work is how to calculate the event influence and find the relationships between observable influence and decay influence.

2) Observable Influence of Event

With the rise of Internet media, people are used to comment about one event on network carriers, while Internet carriers record information of people's living, working and studying. The current relatively popular network media contain microblog, Twitter, Facebook, forums, etc., which have thousands of users. When a hot event happens, a lot of views and comments about the event have diffusion fast on network to form a powerful network influence. The research and analysis of network events emerge in endlessly [12, 13, 14, 15, 16, 17, 18, 19]. In particular, investors publish their views and emotions when a good or a bad event happens in financial field.

$$Inf_t^{observable} = Post_t \qquad (1)$$

where $Inf_t^{observable}$ is the observable influence of event series. $Post_t$ is the main observable influence of event. It is element of the action, and mean the change process and its characteristics of events, which is the degree of movement, description of the way, method and so on.

3) Decay Influence of Event

Event influence growths from the moment of event happening, which accompany with decay at the same time. The decay curve was first introduced by Ebbinghaus [20] as formula 2. The S is the strength of memory, which controls how fast we forget about events. The R decreases memory while time elapses from the happening of event.

$$R = e^{-\frac{t}{s}} \qquad (2)$$

There are also some variants of forgetting curve such as the S-shaped curve [21] and the power-law curve. Ebbinghaus proposes memory forgetting curve that is memory on a certain event decreases while time elapses from the happening of event, the velocity of the curve from fast to slow. Traditional methods of calculating information influence use different decay schemes to reduce the noise of information. Liang Kong ranks news event by influence decay and information fusion [22], and the interest-forgetting curve for music recommendation use Ebbinghaus memory forgetting curve [23].

In this paper, we use the Baidu search index curve as decay curve to estimate the decay velocity of event influence, because the Baidu search index has the same feature as traditional information, such as twitter, microblog.

We test the similarity of official search index and Ebbinghaus memory forgetting curve. The Fig.1 is the original $SearchIndex$ curve after normalization. The fit curve of $SearchIndex$ of 000799 in Baidu in cycle life of the Plasticizer event (from 19/11/2012 to 16/1/2013) accord with the change rule of Ebbinghaus memory forgetting curve.

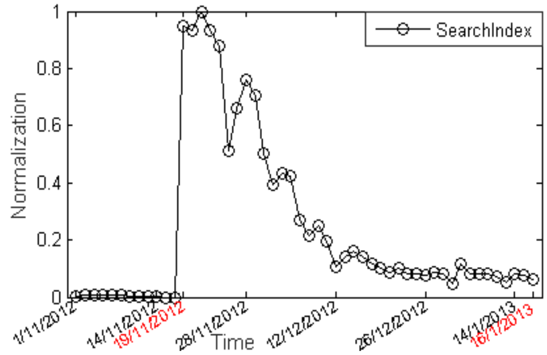

Fig1. Baidu search index

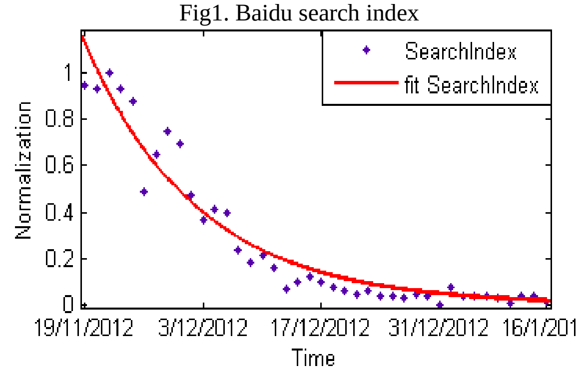

Fig2.The fit curve of Baidu search index

The fit curve of $SearchIndex$ is power law curve as Fig.2, its function is $f(t) = \alpha \cdot e^{\beta \cdot t}$. $\alpha = 1.252$, $\beta = -0.1044$. The 95% confidence bounds of α is between 1.13 and 1.373, and the 95% confidence bounds of β is between -0.118 and -0.09075. Goodness of fit are as follow: SSE is 0.262, R-square is 0.9357, Adjusted R-square is 0.934, RMSE is 0.08303.

This paper combines data of Baidu to form our decay curve to model the effects in people’s comments. Therefore, we bring forward the concept of decaying curve which is expressed as formula 3.

$$Inf_t^{decay} = SearchIndex_t \qquad (3)$$

Where Inf_t^{decay} is the decay influence of event. t is the life cycle from the happening of an event till ending. $SearchIndex$ is the search index of one stock from Baidu, and it means the decay influence of event. The formula 3 characterizes how much influence of a happened event is retained by now.

4) Patterns of Time Series

This paper gives some definitions of time series pattern due to the idea of paper [24].

DEFINITION 1. (***Uptrend Pattern***), given a time series $P=<p_1, p_2, ..., p_w>$ with the window threshold w, while $p_1 = min < p_1, p_2, ..., p_w >$ and $p_w = max < p_1, p_2, ..., p_w >$, this time series P as uptrend pattern, which is P_U.

DEFINITION 2. (***Downtrend Pattern***), given a time series $P=<p_1, p_2, ..., p_w>$ with the window threshold w, while $p_1 = max < p_1, p_2, ..., p_w >$ and $p_w = min < p_1, p_2, ..., p_w >$, this time series P as downtrend pattern, which is P_D.

5) Event Influence

We calculate event influence combine observable influence and decay influence. The observable influence and decay influence effect event interaction on the beginning stage of the life cycle when their curve patterns are different; while, only one of observable influence or decay influence effect event on the ending stage when their time series patterns are same.

$$inf_t^{event} = \begin{cases} inf_t^{observable} + inf_t^{decay}, & P(inf_t^{observable}) \neq P(inf_t^{decay}); \\ inf_t^{observable}, & P(inf_t^{observable}) = P(inf_t^{decay}). \end{cases} \qquad (4)$$

where inf_t^{event} is the event influence, the complementary or alternative relationships between observable influence and decay influence under the constraint of pattern. When $P(inf_t^{observable}) \neq P(inf_t^{decay})$, the observable influence and decay influence is complementary relationship, we combine $inf_t^{observable}$ and inf_t^{decay} to calculate the event influence. And $P(inf_t^{observable}) = P(inf_t^{decay})$, the observable influence and decay influence is alternative relationship, we choose the observable influence as the event influence.

III. EXPERIMENT AND ANALYSIS

1) Detection the cycle life of event

Traditional methods to detect and track events contain the keyword of event detection [1]. C. Chen proposes an aging theory to model the life cycle of event [4]. The thought of TDT and aging theory are employed in our work. If we calculate event influence, we need to know the life cycle of event. In this paper, we define the cycle life of event that is from the begin to the end of event. The begin is from when the topic keyword appear in the context last Δt days, and the end is when the topic keyword disappear last Δt days.

This paper, we choose the Plasticizer event of Jiuguijiu in 19/11/2012. We crawl the posts of Jiuguijiu (000799) from Eastmoney website from 2012 to 2013. We extract the topic keyword of Plasticizer event by TF-IDF method, and the time cycle of Plasticizer in the post is from 19/11/2012 to 16/1/2013 based on our definition of the cycle life of event. So the life cycle of the Plasticizer event is from 19/11/2012 to 16/1/2013 as Fig.3.

2) The event influence calculation fusion decay curve

In this paper, we crawl the post from Eastmoney website from 2012 and 2013. The observable curve fuse decay curve to calculate event influence, the observable curve uses number of post and the decay curve uses $SearchIndex$. Fig.4 is the original data about $SearchIndex$ and number of $Post$ after normalization.

We need to judge the patterns of observable curve and decay curve in the cycle life of the Plasticizer event by definition 1 and definition 2. We define $w=3$ according with experience value. Fig.5 is the patterns of $SearchIndex$ and $Post$. The $Pattern(SearchIndex) = P_D$ and $Pattern(Post) = P_U$ are from 19/11/2012 to 29/11/2012, while $Pattern(SearchIndex) = P_D$ and $Pattern(Post) = P_D$ are from 30/11/2012 to 16/1/2013.

We calculate event influence according to the formula 4. When the patterns of curve are different, we need to consider the acting force of decay curve of information. This time, we fuse observable influence and decay influence to calculate event influence.

The condition of $P(inf_t^{observable}) \neq P(inf_t^{decay})$ is from 19/11/2012 to 29/11/2012, the Plasticizer event effect $SearchIndex$ and $Post$, and the event influence is equal to the sum of $inf_t^{observable}$ and inf_t^{decay}. The condition of $P(inf_t^{observable}) = P(inf_t^{decay})$ is from 30/11/2012 to 16/1/2013, and the event influence can use $inf_t^{observable}$ or inf_t^{decay}, because of $inf_t^{observable}$ and inf_t^{decay} are alternative. As is shown in Fig.6.

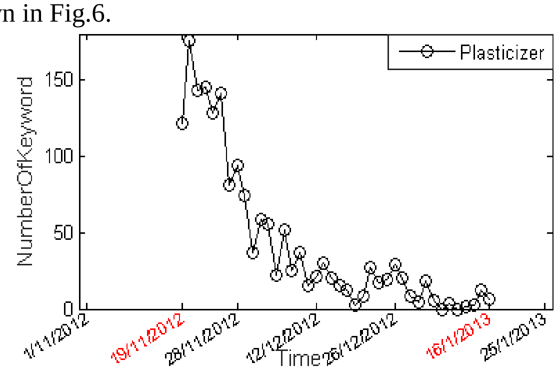

Fig3. Keyword of Plasticizer is from 19/11/2012 to 16/1/2013

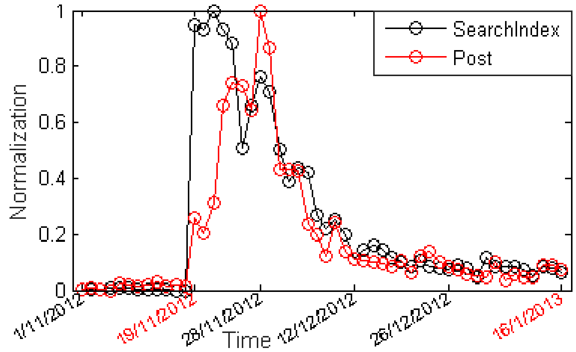

Fig4. The time series of *SearchIndex* and *Post*

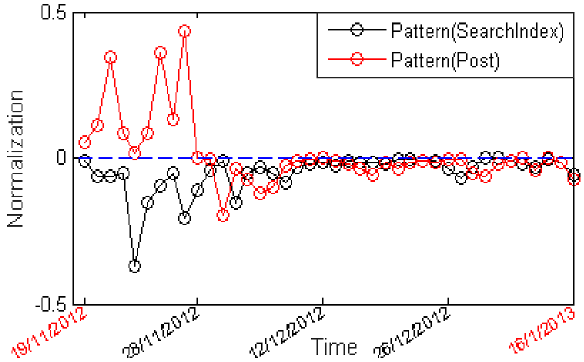

Fig5. The *Pattern*(*SearchIndex*) and *Pattern*(*Post*)

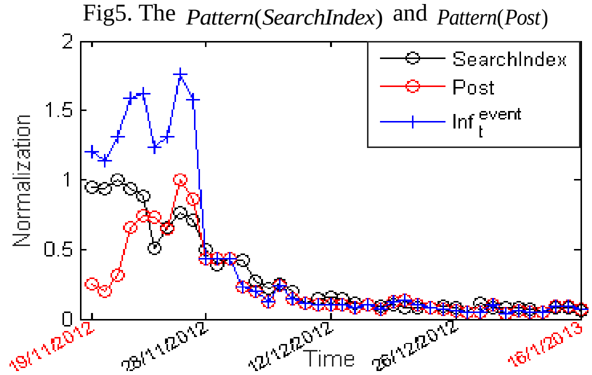

Fig6. The time series of event influence

IV. CONCLUSIONS

In this paper, we calculated event influence in a new method and found the relationships between observable influence and decay influence on different stages in the life cycle of event. We not only took both observable influence and decay influence into consideration, but also brought in the interaction between them to our framework. Experimental results showed that our method as a typical scheme using both observable influence and decay influence impact, and found that observable influence and decay influence work together on event in the beginning time and they are alternative relationship in the ending of event. At the same time, their cut-off point is very apparent.

As a future work, we will consider multiple spaces feature of event in the condition of Internet big data to improve our method, since such features have been widely used in hot news event.

REFERENCES

[1] J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y. Yang. Topic Detection and Tracking Pilot Study: Final Report. In Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, volume 1998, pages 194–218, 1998.

[2] Del Corso GM, Gullí A, Romani F. Ranking a stream of news. In: Ellis A, Hagino T, eds. Proc. of the WWW 2005. New York: ACM, 2005. 97−106.

[3] Yao JY, Wang J, Li ZW, Li MJ, Ma WY. Ranking Web news via homepage visual layout and cross-site voting. In: Lalmas M, ed. Proc. of the ECIR 2006. Heidelberg: Springer-Verlag, 2006. 131 −142.

[4] C. Chen, Y. Chen, Y. Sun, and M. Chen. Life Cycle Modeling of News Events using Aging Theory. In ECML'03, pages 47–59, 2003.

[5]Yunlan Xue, Lingyu Xu, Jie Yu, Lei Wang, Gaowei Zhang, An Estimate Method of Event Influence Scope based on Special Field， 2015 International Conference on Identification, Information & Knowledge in the Internet of Things， Pages: 39 - 44

[6] Gruhl D, Guha R, Liben-Nowell D, Tomkins A. Information diffusion through blogspace // Proceedings of the 13th International Conference on World Wide Web. New York, USA, 2004: 491-501.

[7] Degan Zhang, Guang Li, Ke Zheng. An energy-balanced routing method based on forward-aware factor for Wireless Sensor Network. IEEE Transactions on Industrial Informatics, 2014,10(1):766-773

[8] Yunlan Xue, Lingyu Xu, Jie Yu, Lei Wang, Gaowei Zhang, A Detection Method of Stock Event Source， The 11th International Conference on Semantics, Knowledge and Grids on Big Data, Pages: 174 - 179

[9] Zheng Xu et al..Knowle: a Semantic Link Network based System for Organizing Large Scale Online News Events. Future Generation Computer Systems, 2015, 43-44, 40-50.

[10] C. Wang, M. Zhang, S. Ma, and L. Ru. Automatic Online News Issue Construction in Web Environment. In WWW'08, pages 457–466, 2008.

[11] C. Wang, M. Zhang, L. Ru, and S. Ma. Automatic Online News Topic Ranking using Media Focus and User Attention based on Aging Theory. In CIKM'08, pages 1033–1042, 2008.

[12] Wang Lei， Yu Jie， Xu Lingyu， Xue Yunlan， Zhang Gaowei， Consistency research between virtual space and real space based on an event-driven condition， Source: Journal of Computational Information Systems, v 11, n 7, p 2345-2354, April 1, 2015

[13] C. C.Aggarwaland K. Subbian.Event detection insocial streams. InSDM'12, pages 624-635, 2012.

[14] H. Becker, M. Naamanand L.Gravano. Beyondtrending topics: Real-world event identification ontwitter. InICWSM'11, pages 438-441, 2011.

[15] R. Li, K. H. Lei, R. Khadiwala, and K. C.-C. Chang.Tedas: A twitter-based event detection and analysissystem.Data Engineering, International Conferenceon, 0:1273-1276, 2012.

[16] S. Petrovic, M. Osborne, and V. Lavrenko.Streaming first story detection with application to twitter. InHLT'10, pages 181-189, 2010.

[17] J. Weng and B.-S. Lee. Event detection in twitter.InICWSM'11, 2011.

[18] Q. Zhao, P. Mitra and B. Chen. Temporal andinformation flow based event detection from socialtext streams. In AAAI'07, pages 1501-1506, 2007.

[19] Xiangmin Zhou, Lei Chen， Event detection over twitter social media streams， The VLDB Journal — The International Journal on Very Large Data Bases， Volume 23 Issue 3， June 2014.

[20] H. Ebbinghaus. Memory; A Contribution to Experimental Psychology. New York by Teachers College, Columbia University, 1885.

[21] L. Averell and A. Heathcote. The form of the forgetting curve and the fate of memories. Journal of Mathematical Psychology, 55:25–35, 2011.

[22] Liang Kong , Shan Jiang , Rui Yan , Shize Xu , Yan Zhang, Ranking News Events by Influence Decay and Information Fusion for Media and Users, CIKM'12 , October 29–November 2, 2012, Maui, HI, USA. Copyright 2012 ACM 978-1-4503-1156-4/12/10.

[23] Jun Chen Chaokun Wang Jianmin Wang, Modeling the Interest-Forgetting Curve for Music Recommendation, MM'14, November 03 - 07 2014, Orlando, FL, USA. Copyright 2014 ACM 978-1-4503-3063-3/14/11.

[24] Fu-Lai Chung, Tak-Chung Fu, Vincent Ng, and Robert W. P. Luk, An Evolutionary Approach to Pattern-Based Time Series Segmentation, IEEE transactions on evolutionary computation, vol. 8, October 2004.