

## Research on the Detection and Tracking of Hot Event in Microblogs Based on the Analysis of Mood Features

Gaofeng Huang<sup>1,2\*</sup>

The Naval University of Engineering  
<sup>1</sup>Information Security Department, <sup>2</sup>Graduate School  
Wuhan, China  
huanggaofeng@163.com\*

Xueguang Zhou

The Naval University of Engineering  
Information Security Department  
Wuhan, China  
zxg196610@hotmail.com

**Abstract**—With the widespread application of microblog, it has become a platform for people to publish their interest topics, to express personal feelings, to participate in a centralized discussion. Aiming at the tracking problem of hot events in microblogs, this paper proposes a new tracking method of hot events in microblogs based on the analysis of the features of mood information that is the total amount of emotion value and its growth rate. The experiment is carried to verify the feasibility of the method by the real-time capture and analysis of microblogs correlating with the hot events in different periods. Experimental results show that the method based on analyzing the features of mood information can effectively identify hot topic in microblogs.

**Keywords**—mood words; grained emotion; microblog; hot events; detection; tracking

### I. INTRODUCTION

After the happening of hot events, people will often firstly obtain, feed back and spread the information of events through the microblog. The spread mode of microblog is of the following features such as open, real-time and free-selective, and the microblog contains users' feelings for events. The emotion analysis and emotion trend analysis of public information which is generated in the microblogs by users can mine well the behavior rules of Internet user groups and identify the potential hot topic.

Shi Daming, Lin Hongfei [1] proposed a method to discover the hot topic which is based on the emotion analysis of comments and the relation analysis of comments and topic. L.Ku, etc [2] proposed a method to reflect the changes of public opinion in a certain moment, by obtaining the information of views to reflect the emotional changes of social groups for the candidates in the process of electing president. Cuneyt Gurcan Akcora, etc [3] proposed a method to discover the public opinion on Twitter by way of Emotion Centroid (EC) and Set Space Model (SSM). Sakaki T, etc [4] designed a monitoring system of event in microblog, which can detect the geographic center such as earthquakes and the space diffusion trajectory of other natural events. Singh V K, etc [5] denoted the users' interest information reflected in the microblog by a form similar to the pixels in the image, and denoted the blogs

related with the event by a dynamic structure similar to a video file by analyzing the time information contained in the microblog, which is used to monitor the changes of the event.

The studies of traditional topic tracking and detection mainly detect and track events from the fact, which mostly have a clustering based on different events, and rarely used the calculation and analysis of the feature of mood information. However, the information in microblogs mostly emerged fragmentation, immediacy, mobility and other features, and most of the content for the microblog users to express their feelings, so the proportion of information of emotional words in microblogs is larger than that of the traditional text, so the calculation method of emotion value in microblog for detecting hot events is feasible. Based on the analysis of emotional words and especially the features of mood words, this paper proposes a tracking method of hot events in microblogs based on the calculation and analysis of the weight value of mood words. The experimental results show that this method can effectively detect and track the hot events in microblogs.

### II. MOOD WORDS AND CALCULATION OF THEIR WEIGHT VALUE

Human emotion is very rich, and the Confucian expressed human emotions as seven sorts of emotions: joy, anger, sadness, fear, love, hate, desire. For example, the words of expressing "joy": delight, happy, excited, joy, etc. the words of expressing "anger": anger, wrath, outrage, indignation, raging, furious, etc. the words of expressing "love": love, cute, care, compassion, affection, compassion, gratitude, etc. the words of expressing "desire": longing, hope, despair, etc. The "desire" in the human "seven sorts of emotions and six sorts of desires" is the desire of seeing, the desire of sound, the desire of smelling, the desire of tasting, the desire of touch, which refers to human physiological needs and desires, unified as one sort of human emotion [6]. The weight value of fine-grained emotion thesaurus based on human emotions or mood vocabulary thesaurus comes from the calculation using the formula of the similarity of words.

The so-called similarity is a value which typically ranges in [0,1]. The semantic similarity of a word itself is 1.

If two words are not replaceable in any context, their semantic similarity is 0. In this paper, the calculation of similarity between emotional words draws out weight value reflecting a tendency to emotional intensity to distinguish the fine-grained emotional words. The word similarity in HowNet is based on the calculation of original meaning of the word. The words of same original meaning in HowNet compose a hierarchy tree. The hierarchy of HowNet is shown in Figure 1, and thus the calculation of similarity of the original meaning translates into the calculation of the distance between the two original meanings in the hierarchy tree.

Suppose that the distance of original meaning  $s_1, s_2$  of the two words  $W_1, W_2$  in the hierarchy tree is  $dist(s_1, s_2)$ , the similarity calculation of the two original meanings is formula (1):

$$Sim(W_1, W_2) = \frac{\partial}{\partial + dist(s_1, s_2)} \quad (1)$$

$S1$  and  $S2$  represent two original meaning of two different words;  $dist(s_1, s_2)$  expresses their distance in the hierarchy tree of original meanings;  $\partial$  is a tuning parameter, which means the distance of the two words in the tree when their similarity being 0.5, generally take the value 1.6. Sometimes there is more than one original meaning in a word, in the calculation of the similarity between the words with many of the original meanings, taking the maximum similarity as the similarity between the original meanings of the words. For the two Chinese words  $W1, W2$ , assuming that they have more than one original meaning, the number of original meanings of  $W1$  is  $m$ , which is  $s_{11}, s_{12}, \dots, s_{1m}$  respectively; the number of original meanings of  $W2$  is  $n$ , which is  $s_{21}, s_{22}, \dots, s_{2n}$  respectively; and thus the calculation of their similarity is formula (2):

$$Sim(W_1, W_2) = \text{Max} \left[ Sim(s_{1i}, s_{2j}) \right]_{i=1,2,\dots,m; j=1,2,\dots,n} \quad (2)$$

The weight value of mood words correlates with the similarity degree among original meaning of the word itself and original meanings of reference words which refers to those words whose expression of emotion is very obvious, most common, and representative. The contact with the reference words is closer, and the tendency of word is stronger. Therefore, we can calculate the similarity of original meanings between the word and every one word of the reference words to distinguish the different fine-grained emotional vocabularies.

### III. ENERGY OF MICROBLOG AND ENERGY OF TOPIC

The spread of microblog is different in nature from the spread of traditional news or online media. It is different from the linear spread of traditional media, but also from the mesh spread of network media, and it is a kind of fission spread [7]. When an original microblog is released, it is usually the first to be found by the original author's fans and may be forwarded, and the forwarded posts may be found and forwarded by fans of the fans, spreading layer by layer and forming the fission spread. This fission model of spread

makes a microblog spread rapidly in the community with common interests, in which the users discuss and exchange views together, and so the microblog generates a hot topic. If the spread path on the microblog platform is abstracted out, and remove the spread ring between two users' nodes, then the spread path of a microblog can be abstracted into a tree. The discussions, transfers and replies around a microblog are also the same topic. So The microblogs relevant to the discussing, forwarding and replying the same microblog post can be merged into a tree called the microblog discussion tree [8], shown in Figure 1.

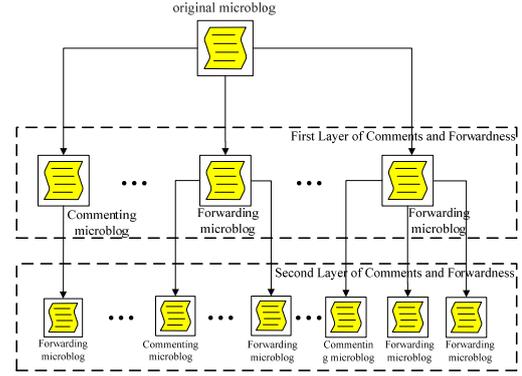


Figure 1 Microblog discussion tree

The concern level of microblog depends largely on the number of microblog fans. At the same time, the number of users to discuss is more, the range of spread is broader and the attained responses are greater. A key factor of microblog to be commented and forwarded is highly correlated with the number of existing fans, and the correlation coefficient is 0.704, and is a low correlated with that whether the user is an authenticated user, but not correlated with the number of microblog posts released [9]. Meanwhile, if a microblog can attract a large number of microblog users to discuss, its content should be resonant certainly. Therefore, a factor of author's microblog to be concerned by a large number of microblog users is his having a certain number of fans on the microblog platform. If we define the energy value of microblog, we can think its energy is equal to the number of the commenting and forwarding fans, and a microblog nobody to comment and forward is no energy. Thus, giving a microblog user (denoted by  $A$ ) and a microblog  $MB$  released by  $A$ , there is a definition of microblog energy, seeing the definition 1.

Definition 1: The energy of microblog  $MB$  is equal to the number of user  $A$ 's fans who comment or forward the microblog, denoted by  $ActiveFans(A)$ .

$$Energy(MB) = ActiveFans(A) \quad (3)$$

Because a microblog discussion tree is composed of  $n$  pieces of microblogs, so a microblog discussion tree is denoted by  $MBT$ , its energy is:

$$Energy(MBT) = Energy(MB_1) + Energy(MB_2) + \dots + Energy(MB_n) \quad (4)$$

Definition 2: A topic may be composed of a number of microblog discussion trees, given a moment denoted by  $T$ , the energy of topic denoted by  $T_p$  is the summation of energy

of microblog discussion trees which compose the topic. The energy of the topic is denoted by  $Energy(T_p)$ , that is:

$$Energy(T_p) = \sum_{i=1}^n Energy(MBT_i) \quad (5)$$

Microblog is highly real-time, and the number of microblogs consisting of a topic varies with time. Given a period of time  $\Delta T$ , the energy of microblog discussion tree is denoted by  $\Delta Energy(MBT_i)$ , so the energy of the topic is:

$$\Delta Energy(T_p) = \sum_{i=1}^n \Delta Energy(MBT_i) \quad (6)$$

With the increase of the number of people participating in the topic, the topic energy of microblog is also increasing. When the energy reaches a peak, because of the development of event itself or the arising of the new topics, microblog users would reduce the intensity of concern and transfer the attention, therefore, the energy of topic also declines correspondingly.

#### IV. THE DETECTION AND TRACKING OF HOT EVENT

People express their views or attitudes by microblog, when a hot event occurs on the microblog platform, as will inevitably lead to a wide range of comments and forwards, is simultaneously bound to generate a large amount of mood vocabulary, that is say the total emotion value also increases. This phenomenon provides an important basis for the research work of this article. Therefore, we can identify potential hot events by detecting the total emotion value of microblog.

It is worth noting that although the response of some microblog topic which is able to arise people's reactions would not be strong enough at some time, but if the time is too long, it would lead to a large number of comments and forwardness, so, that may lead to false estimate. In order to avoid the appearance of such a false estimate, this paper introduces the concept of growth rate of emotion value in the microblogs correlating with the topic.

The so-called growth rate of emotion value in microblogs refers to the growth rate of the total emotion value of microblog discussion tree correlating with the topic within a specific period. First detecting the microblogs whose emotion value ranks more front, then detecting the growth rate of total emotion value of microblog discussion trees correlating with the topic in a special period of time, thus we can find the hot topics. The topic with the growth rate of total emotion value of microblog discussion trees being larger is the hot topic. As long as the microblog topics are able to provoke people's strong reaction, their comments and forwardness must be able to generate a large number of emotional words in a short time, leading to a rapid increase of the total emotion value, as is an important reference to determine whether the topic being a hot microblog topic.

With "Event of Granny extorting three children in Sichuan" for example, we start up immediately the tracking of the microblogs correlating with this topic after the news of this event releases. The growth changes of emotion value of the microblog discussion tree correlating with this topic is shown in Figure 2 .

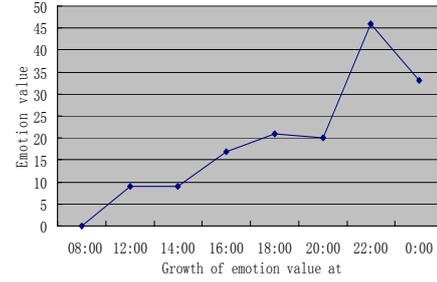


Figure 2 Growth chart of "Event of Granny extorting three children in Sichuan"

As can be seen from Figure 2, the event triggered a lot of comments and forwardness at the start. As time passes, the growth of emotion value in microblogs is also constantly changing. The growth rate associates with the period. The peak time of users in the Internet is closer, and the growth rate is more obvious.

#### V. EXPERIMENTAL RESULTS AND ANALYSIS

In order to verify the accuracy of the method, the research group has the real-time tracking of the microblogs correlating with the hot topics in different periods, such as "The bus bombing in Xian," "Poisoning case at the college of Fudan in Shanghai" and "Event of Granny extorting three children in Sichuan", having a real-time capture of Sina microblog correlating with the event. The statistical results of tracking of three hot topics are following.

Once the microblogs and especially the microblog of large V (a microblog user has a large number of fans) are involved with the hot events, they would inevitably lead to being commented and forwarded by a large number of fans, resulting in the increasing of energy of microblog, and followed by a large amount of emotional vocabulary contained in a large number of comments, there is a proportional relationship between the two. We have had an experiment of statistical comparison between the total emotion value of microblog and energy of microblog, and the data is shown in Figure 3.

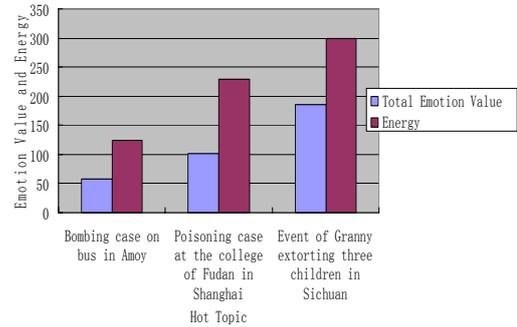


Figure 3 Comparison chart between the energy of microblog topics and their total emotion value

We can conclude in Figure 3 that the total energy of microblog discussion tree related to the topic would be in direct proportion to its total emotion value. Educuing inversely the energy of microblog from the topic with

higher energy, we found that the energy of user was greater, and the probability of generating of influential events was higher.

With "Event of Granny extorting three children in Sichuan" for example, we start up immediately the tracking of the microblogs correlating with this topic after the news of this event releases.

On the day or the next day of the hot news released, we start up immediately the tracking and crawling of the Sina microblogs, from 8:00 am to midnight, every two hours to conduct a statistic, drawing out the changing trend chart of the total emotional value of microblog discussion trees related to the topics. It is shown in Figure 4.

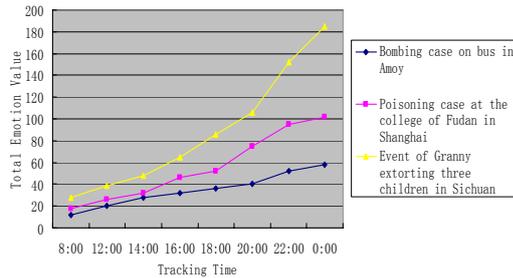


Figure 4 Changing trend chart of the total emotional value of microblog discussion trees

Figure 4 shows the total emotion value of microblog topic increases with the passage of time. By further analysis of the rate of change, we drew the growth chart of the total emotion value of microblog. It is shown in Figure 5.

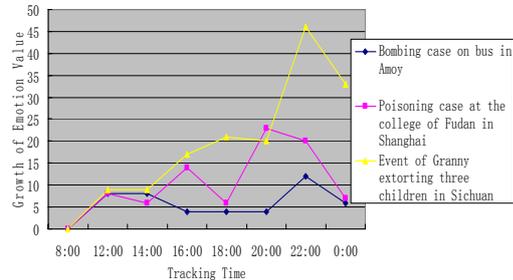


Figure 5 One-day growth chart of the total emotion value of microblog discussion trees

Figure 5 shows the growth rate of the total emotion value of microblog discussion tree related to the topic was with a different speed of change in the different periods. In figure 5 we found in the peak of Internet the total emotion value increased rapidly, but not obvious.

So, conducting a tracking of the microblogs for the next three days, we drew out the three-day growth chart of the total emotion value. It is shown in Figure 6.

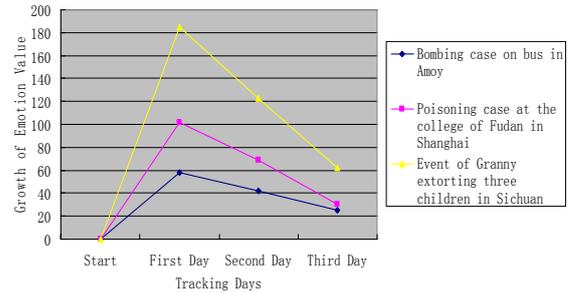


Figure 6 Three-day growth chart of the total emotion value of microblog discussion trees

Seen from Figure 6, the growth rate of the emotion value was greater obviously in the approaching period of the outbreak of event, and then it is slowing down. Thus, the conclusions of the article are verified: According to the growth of the total emotion value of microblog topic, we can predict whether the microblog topic might evolve into a hot event.

## VI. CONCLUSIONS

The total emotion value of microblog topic reflects the degree of concern for the topic and emotional state of people, but also on behalf of the microblog topic own energy. Through experiments the following conclusions are arrived:

(1) The growth rate of the total emotion value is in direct proportion to the degree of concern for the topic. The growth rate of total emotion value is faster, and the number of persons concerned to comment and forward the topic is more.

(2) The total emotion value of microblog topic is in direct proportion to the energy of microblogs correlating with the topic. The emergence of new hot topic is bound to promote a large number of comments and forwardness, so the total emotion value increases rapidly, and meanwhile the energy of microblog also simultaneously increases. With the passage of time or the emergence of new hot topic, the hot microblog gradually fade from people's field of vision, so its energy gradually tends towards zero, and its growth of emotion value also tends towards zero.

Experimental results show that the method can accurately identify hot microblog topic by analyzing the changes of emotion value of microblog topic. These conclusions are of important reference value for further research on the detection and tracking of hot event in microblogs.

## REFERENCES

- [1] Shi Daming, Lin Hongfei Blog Hot Post Discovery and Author Reputation Degree Research [D] Dalian University of Technology, 2007
- [2] L Ku, Y Liang, and H Chen. Opinion extraction, summarization and tracking in news and blog corpora [C] //Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs. California, USA, 2006, 100-107

- [3] C Akcord, M Bdyir, M Demirbas, H Ferhaosmanoglu. Identifying Breakpoints in Public Opinion [C]//Proceedings of KDD Workshop on social Media Analytics, Washington, July 2010
- [4] Sakaki T, Okazaki M, Matsuo Y. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors[C]// Proceedings of the 19th International World Wide Web Conference. Raleigh, NC (USA), 2010:851-860
- [5] Singh V K, Gao M, Jain R. Situation Detection and Control using spatio-temporal Analysis of Microblogs [C] //Proceedings of the 19th International World Wide Web Conference. Raleigh, NC (USA), 2010:1181-1182
- [6] “Seven sorts of emotions and six sorts of desires”, <http://www.baidu.com.cn>
- [7] Jiang Hongmei. Microblog characteristics and influence public opinion [J]. Networks spread, 2011,5:85-86
- [8] LiuJia. Microblog-oriented Hot Topics Discovery and Tracking [D] South China University of Technology, Guangzhou, China, 2012
- [9] Yuan Yi. Microblog path information dissemination structure and its influencing factors [J] Library and Information Service, 2011,55 (12) :26-30