Detecting Event Rumors on Sina Weibo Automatically

Shengyun Sun[1], Hongyan Liu[2,*], Jun He[1,*], and Xiaoyong Du[1]

[1] Key Labs of Data Engineering and Knowledge Engineering, Ministry of Education, China
School of Information, Renmin University of China, 100872, China
{sun_shengyun,hejun,duyong}@ruc.edu.cn
[2] Department of Management Science and Engineering, Tsinghua University, 100084, China
hyliu@tsinghua.edu.cn

Abstract. Sina Weibo has become one of the most popular social networks in China. In the meantime, it also becomes a good place to spread various spams. Unlike previous studies on detecting spams such as ads, pornographic messages and phishing, we focus on identifying event rumors (rumors about social events), which are more harmful than other kinds of spams especially in China. To detect event rumors from enormous posts, we studied the characteristics of event rumors and extracted features which can distinguish rumors from ordinary posts. The experiments conducted on real dataset show that the new features are effective to improve the rumor classifier. Further analysis of the event rumors reveals that they can be classified into 4 different types. We propose an approach for detecting one major type, text-picture unmatched event rumors. The experiment demonstrates that this approach is well-performed.

Keywords: rumor, Sina Weibo, rumor detection, social network.

1 Introduction

Over the past several years, online social networks such as Facebook, Twitter, Renren, and Weibo have become more and more popular. Among all these sites, Sina Weibo is one of the leading micro-blogging service providers with eight times more users than Twitter [1]. It is designed as a micro-blog platform for users to communicate with their friends and keep track of hot trends. It allows users to publish micro-blogs including short text messages with no more than 140 Chinese characters, attached videos, audios and pictures to show their opinions and interests. All these micro-blogs published by users will appear in their followers' pages. Also users can make comments on or retweet others' micro-blogs to express their view.

Weibo is a kind of convenient sites for users to show themselves and communicate with others. It is put into service in 2009 by Sina Corporation. A recent official report shows that Sina Weibo has more than 300 million registered users by the end of February 29, 2012 and posts 100 million micro-blogs per day. Unfortunately, the wealth of information also attract interests of spammers who attempt to send kinds of spams

* Corresponding authors.

Y. Ishikawa et al. (Eds.): APWeb 2013, LNCS 7808, pp. 120–131, 2013.

such as rumors, ads, pornographic messages or links to phishing or malicious web sites. Some of these malicious spams may cause significant economic loss to our society and even threaten national security and harmony.

In this paper, we focus on a kind of spam, and we call it ***event rumor*,** which is more harmful than other kinds of spams especially in China. Event rumors are spams that make up untruthful social events and they're harmful to society security and harmony in potential. Figure 1 shows an instance of event rumor which says that the picture is not a scene of a movie or a war, but shows that an officer was on his way to investigate people's life under the protection of several soldiers. It was posted by a Sina weibo user (for the purpose of privacy protection we hide the user's name) at 08:31 on May 11.

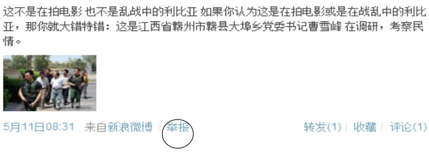

Fig. 1. Instance of event rumor on Sina Weibo

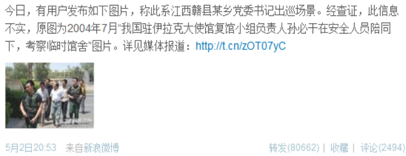

Fig. 2. Instance of Sina Weibo Rumor-Busting

To detect spams, Sina Weibo has tried several ways such as adding a "report as a spam" button to each micro-blog as shown in the left bottom of Figure 1. Any users considering a micro-blog to be suspicious can report it to Sina official Weibo account (with the translated English user name of "Weibo Rumor-Busting" [1]) who specializes in identifying fake micro-blogs manually. Figure 2 shows a micro-blog published by the Sina official account, which points out that the message showed in Figure 1 is a rumor and gives detailed and authentic reasons. Though artificial method is accurate, it costs a lot of human efforts and financial resources. What's more, there is a certain delay by this approach. It cannot identify a spam quickly once it appears. The rumor may have been spread widely and produced bad effects on society before it is identified manually. So an automatic approach to detecting event rumors is essential. To the best of our knowledge, [1] is the first paper that detect rumor automatically on Sina Weibo, but it aims to detect general rumors not event rumors. And the accuracy has big room to be improved.

In this paper we study how to detect the special kind of rumor, event rumor, because it is more harmful to our society than others. We consider the problem of rumor detection as a classification problem firstly. To distinguish rumors from ordinary posts, we design special features about event rumors from Sina Weibo and build classifiers based on these features. Then we divide event rumors into 4 types and propose a method to identifying one major type among them. The experimental results indicate that our approaches are effective.

The rest of the paper is organized as follows. In section 2 we give a review of related work. In section 3 we describe our two approaches to identifying event rumors and provide a detailed description of the new features. In section 4 we describe how to collect data and present the experimental results. Finally, we conclude this paper in section 5.

2 Related Work

The previous research about spam detection mainly focuses on detecting spams related to ads, pornography, viruses, phishing and so on. These works can be divided into two types. One is detecting spammers who publish social spams [5-10], while the other is detecting spams directly [1-4].

Lee et al. [6] proposed a honeypot-based approach to uncover social spammers in online social systems including MySpace and Twitter. They deployed social honeypots to harvest deceptive spam profiles from social networking communities firstly. Then they analyzed the properties of these spam profiles and built spam classifiers to filter out spammers. Benevento et al. [11] also used honeypots to collect spam profiles. Benevento et al. [5] extracted attributes related to content of tweets and attributes about user behavior and built a SVM classifier to detect spammers of Twitter. This approach is well-performed in filtering out spammers who focus on posting ads, especially link deception. Stringhini et al. [7] did detailed analysis of spammer on Facebook, MySpace and Twitter. They created honey profiles accepting all requests but sending none request to make friends in order to collect data about spamming activity. Spam bots were divided into four categories Displayer, Bragger, Poster and Whisperer. Random forest algorithm was used to identify spammers.

Wang [3] extracted three graph-based features and three content-based features, and formulated the problem of spam bots detection as classification problem. And then he applied different classification algorithms such as decision tree, neural network, support vector machines, and k-nearest neighbor. Finally they found that Bayes classifier is best well-performed. The three graph-based features were extracted from Twitter's social network graph which represents the following relationship among users. The content-based features contained the number of duplicate tweets, the number of HTTP links, and the number of replies or mentions. Similar to this work [3], in Wang's another work [4], a new content-based feature named trending topics representing the number of tweets that contains the hash tag # in a user's 20 most recent tweets was added to features set. Besides the features studied in previous works, Yang, et al. [1] extracted two new features, client-based feature and location-based feature. The former refers to the client program that user has used to post a micro-blog while the later refers to the actual place where the event mentioned in the micro-blog has happened. They trained a Support Vector Machine classifier to measure the impact of proposed features on the classification performance. A noteworthy point is that the dataset they used is from Sina Weibo. To the best of our knowledge, this is the first paper that studies rumor detection of Sina Weibo. The difference between this work and ours lies in that we focus on event rumors. Gao, et al. [2] studied the wall posts in Facebook, which are usually used to spread malicious content by spammers. They built wall post similarity graph, clustered similar wall posts into a group and then identified spam clusters. The difference from previous works is using clustering method.

As mentioned above, majority of the previous works are about detecting spams such as ads, viruses, phishing, pornography and so on. Unlike previous studies, we focus on identifying event rumors (rumors about social events), which are more harmful to national security and society harmony than other kinds of spams especially in China.

3 Event Rumors Detection

3.1 Problem Statement

Driven by different purposes, spammers spread various kinds of spams such as advertise, pornography, viruses, phishing websites and rumors on Sina Weibo. Most of the previous works focus on detecting spams like ads and pornography as mentioned above. In this paper, we focus on the detection of event rumor which is more harmful than other kinds of spams in China. Other rumors like gossips about stars are not covered by our study. Event rumors are messages which describe fake social events. Usually, an event is described by attributes such as time, location, character and content.

In the following we list the notations to be used in the subsequent sections.

- $U = \{u_i\}$: set of users
- $T = \{t_i\}$: set of micro-blogs
- $Followers(u_i) = \{u_j | u_j \in U \text{ } and \text{ } u_j \text{ } is \text{ } u_i\text{'s } follower\}$: a set of u_i's followers
- $Friends(u_i) = \{u_j | \text{ } u_j \in U \text{ } and \text{ } u_i \text{ } is \text{ } u_j\text{'s } follower\}$: a set of u_i's friends
- $t_i.postingtime$: the posting time of micro-blog t_i
- $t_i.text$: the text content of micro-blog t_i
- $t_i.picture$: the picture attached to micro-blog t_i

3.2 Detecting Event Rumors

In this paper, we consider event rumors detection as a classification problem. To identify the event rumors on Sina Weibo automatically, we need to train classifiers and the key point is to extract powerful features. By analyzing micro-blogs' characters we extract 15 features, which are divided into three types: *content-based features, user-based features* and *multimedia-based features*. Some of these features have been studied in previous works [1,3-7], and others are proposed in this paper. Here we introduce them briefly.

Content-Based Features. Content-based features are 8 attributes related to the message content including the number of comments, the number of retweets, the number of "@" tags, the number of URLs, the number of duplications, the client program type, the number of verbs used to describe event, and the message contains strong negative words or not. The former 6 features have been studied already while the latter two are new features proposed by us.

The number of comments and the number of retweets. Event rumors are usually trending topics that attract more comments and retweets than common micro-blogs.

The number of "@" tags. "@" is considered as a mention tag in Sina Weibo similar to that in Twitter. In Sina Weibo, users use "@"+username anywhere in the micro-blog messages as a reply or mention of another user. A user can mention anyone in a micro-blog as long as the user follows him or her. If a user is mentioned in a micro-blog by the mention tag, he or she will see the notification message. The reply and mention function is often utilized by the rumor accounts to draw users' attention to spread rumors.

The number of URLs. Since Sina Weibo only allows users to post a short message within 140 Chinese characters, the URLs are always shown in a shorten format [3]. Shorten URLs can hide the URLs' source and users can't get direct information from the URLs. So URL is an attribute that is often utilized by rumor publisher.

The number of duplications. To spread rumors widely, users may post the same micro-blog several times. So the number of duplications is also a noteworthy feature. Given two micro-blogs named t_i and t_j, we measure the similarity based on Jaccard coefficient as shown in Equation 1.

$$similarity(t_i, t_j) = \frac{|t_i.text \cap t_j.text|}{|t_i.text \cup t_j.text|} \tag{1}$$

Where $t_i.text$ is represented by a set of keywords in the text content of t_i. Keywords are the words of the micro-blog text after removing stopping words. When the value of similarity is bigger than 0.8, we consider the two micro-blogs are duplicates.

Client. It refers to the client program used to post micro-blogs by users. Clients are divided into two types [1]: non-mobile client and mobile client. The non-mobile client program includes Sina Weibo web-app, times posting tools and embedded Sina Weibo's third party applications, while the mobile client program includes mobile-based client and Tablet Personal Computer client. Due to the convenience of non-mobile clients, they are often utilized to post and spread event rumors by users.

The number of event verbs. In order to discriminate event rumors from other micro-blogs, we regard the number of *event verbs* as one of the features. We create an event verb dataset using the following 2-step method. First, we extract all the verbs from news of Sina and Fenghuang. Second, we filter out the common verbs usually used in daily life. Figure 3 compares the number of event verbs in event rumor and non-rumor micro-blogs. In this figure, each point represents a micro-blog message.

The message contains strong negative words or not. We find that most of event rumors are negative social events and contain strong negative sentiment words and opinion words. Figure 4 shows the comparison result. As we expect, event rumors are more likely to contain strong negative words than normal micro-bogs.

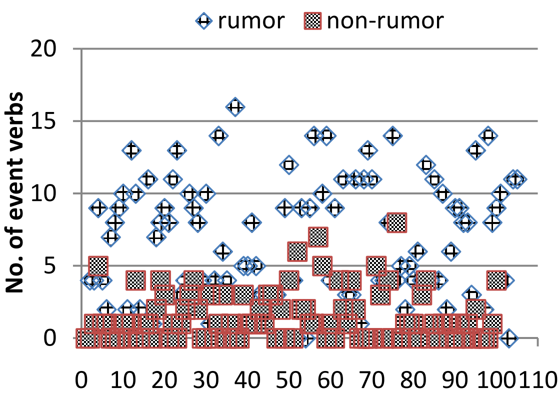

Fig. 3. The number of event verbs used in each micro-blog

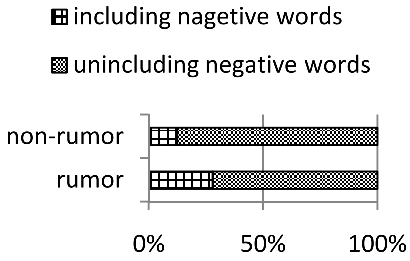

Fig. 4. The distribution of micro-blogs that contains strong negative words

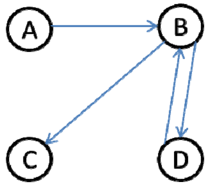

Fig. 5. An example social network graph on Sina Weibo

User-Based Features. User-based features are 6 attributes about users, including the number of user's followers, the number of user's friends, reputation of user, whether the user has VIP authentication, *ratio of micro-blogs containing event verbs*, and *ratio of micro-blogs containing strong negative words*. The former 4 features have been studied already, while the later 2 features are new features proposed by us. Sina Weibo provides a platform to users to create social connections by following others and allowing others to follow them. Figure 5 shows a simple social network graph and the arrows depict the following relationship between users. We can see that user *A* and user *D* are the followers of user *B,* while user *C* and user *D* are friends of user *B*. User *B* and user *D* are mutual followers.

The number of followers. The micro-blogs posted by a user will be seen by all his or her followers. The more followers the user has, the wider the rumors spread.

The number of friends. Anyone can follow a user to be this user's follower without seeking permission. As Sina Weibo will send a private message to you when someone follows your account actively, rumor publishers use the following function to attract attention of users' attention. If you follow other accounts, you can also mention their accounts by using a "@" tag in your micro-blogs, then they will receive your micro-blogs even though they are not your followers.

Reputation. If a user has a small number of followers but a lot of friends, the possibility that he or she posts rumors is higher than others. The reputation of a user is defined as follows by Equation 2.

$$reputation(u_i) = \frac{|Followers(u_i)|}{|Followers(u_i)|+|Friends(u_i)|} \quad (2)$$

Where $|Followers(u_i)|$ means the number of followers and $|Friends(u_i)|$ means the number of friends. The value of reputation is between 0 and 1. Value 1 means a high reputation while 0 means a low reputation.

VIP authentication. If user's identity is verified by Sina Weibo officially, then a VIP tag will occur after the user name. The verified users include famous actors, exporters, organizations, famous corporations and so on. Usually, VIP users are less likely to post rumors.

Ratio of micro-blogs containing event verbs. This attribute means the proportion of micro-blogs that contain event verbs over all micro-blogs posted by the user. The higher this value, the more likely the users post event rumors.

Ratio of micro-blogs containing strong negative words. This attribute means the proportion of micro-blogs that contain strong negative words over all micro-blogs posted by the user. The higher this value, the more likely the users post event rumors.

Multimedia-Based Features. Multimedia-based features are attributes about picture. Sina Weibo allows the micro-blogs containing multimedia like pictures, videos and audios. The most powerful feature we extracted from picture is time span.

Timespan. To make the rumors look like normal micro-blogs, users usually attach corresponding pictures to the micro-blogs. These pictures often come from Internet and were published before. We call them outdated pictures. Users utilize the outdated pictures and new messages to make up a new micro-blog related to event rumors. The timespan function is defined by Equation 3.

$$timespan(t_i.text, t_i.picture) = \begin{cases} 0, & t_i.picture = \emptyset \\ 1, & time(t_i.text) - time(t_i.picture) > tw \\ 2, & time(t_i.text) - time(t_i.picture) \le tw \end{cases} \quad (3)$$

Where $time()$ is the time function, $time(t.text) = t.postingtime$, and tw is a timespan threshold. In our experiment we set the threshold value as a week.

The time of picture is calculated by the following steps. First, we use the picture-to-action function provided by Baidu[1] search engine which is used to identify similar pictures to find out the attached picture on the Internet. If we submit a picture shown in Figure 6 which is an attached picture of the rumor shown in Figure 1 as a query, the search engine will output lots of result records and each record includes the corresponding picture, title, description, URL, and time as shown in Figure 7. Second, according to the time order, we identify the oldest picture and consider this picture's time as the attached picture's time. If a micro-blog doesn't contain any picture, then we consider the value of timespan to be 0. If the timespan between the text and the picture is bigger than the threshold, then the value is 1. Otherwise the value is 2.

[1] http://shitu.baidu.com/

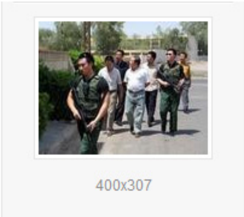

Fig. 6. An attached picture

Fig. 7. An result record

Figure 8 depicts the distribution of timespan of micro-blogs and Figure 9 depicts the distribution of timespan of micro-bogs that contain pictures. The two distributions indicate that rumors are more likely to contain an outdated picture than normal micro-blogs indeed.

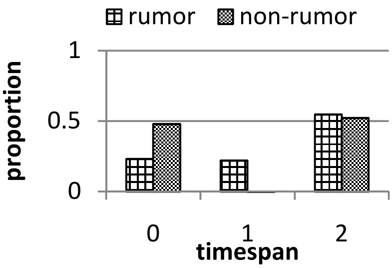

Fig. 8. The distribution of times pan of micro-blogs on Sina Weibo

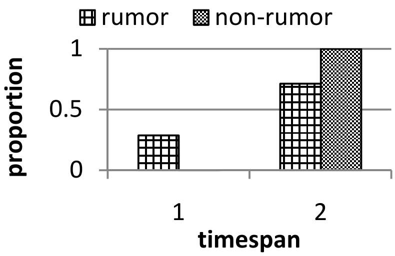

Fig. 9. The distribution of timespan of micro-blogs containing pictures on Sina Weibo

Based on these features, classifiers are built and used to predict whether a micro-blog is an event rumor or not.

3.3 Detecting Text-Picture Unmatched Rumors

Further analysis of sample event rumors reveals that event rumors can be classified into different types. In the following discussion, we provide our analysis of several common types of event rumors and propose an effective method to one major type, detect text-picture unmatched rumors.

Based on our observation, most event rumors belong to one of the following types:

- Complete fiction: The social event itself is a pure fabrication.
- Time-sensitive: First, time-sensitive event rumors were true messages in the past, but the user published it again later and made it like just happened, so these messages are out of date and become time-sensitive event rumors.
- Fabricated details: As mentioned in section 3.1, an event contains the following attributes: time, location, character and content. This kind of rumor is made up by modifying the time, location or characters of another real event happened before.

The difference between fabricated details type and time-sensitive type is that the latter was true but is out of date while the former is fake.

- Text-picture unmatched rumors: As shown in Figure 1, to increase event rumors' credibility, users often attach a picture to the text, but there is no relationship between the picture and the text.

From Figure 8 we can see that about 80% event rumors contain a picture and a majority of them are text-picture unmatched rumors according to our observation. Therefore, we focus on study how to detect this kind of event rumors. Figure 1 shows an instance of text-picture unmatched rumor, whose text describes an officer's investigation of people's life under the protection of soldiers, but the picture actually describes the team leader of Chinese Embassy in the Republic of Iraq visiting temporary embassy accompanied by security personnel in fact. The rumor account utilizes the unrelated picture and new text to make up a new micro-bog and we call it *text-picture* unmatched rumor. We propose a 5-step method to identify this kind of event rumors.

First, we create a list, including 60 news websites consisting of almost all the major domestic media, and foreign media[2]. Second, we submit the picture attached to a micro-blog as a query to search engine to look for similar pictures. If the output result is nothing, then we consider that the picture matches the text, and the micro-blog is non-rumor. Otherwise, we order the result records according to their website's reliability and their posting time descending. The websites in the list created in the first step are credible and their credibility is the same while others are not credible. Third, we crawl the main content of the top ranked website. Finally, we use Jaccard coefficient to calculate the similarity between micro-blog's text and the crawled content after removing stopping words. If they are similar, we consider the micro-blog is non-rumor. Otherwise, they are text-picture unmatched rumors.

4 Experiments

4.1 Collection of Dataset

In order to evaluate the proposed approach for detecting event rumors, we need a labeled dataset of micro-blogs. To the best of our knowledge, no such collection is publicly available, so we need to build one. Labeling manually costs lots of human labor. Besides, the result may not be accurate by the errors in human judgment. Fortunately, Sina Weibo has an official account named "Weibo Rumor-Busting" for publishing micro-blogs relevant to rumor topics. All the announced rumors are verified by authoritative sources. Based on this official rumor busting service, we build a high quality dataset.

The micro-blogs posted by "Weibo Rumor-Busting" are messages explaining which micro-blog is rumor and why it is rumor, but they are not the original rumors themselves, so we need to construct queries manually to find out the original rumors

[2] `http://news.hao123.com/wangzhi`

by the search function provided by Sina Weibo. We select the top one micro-blog from the search results. We collect 104 event rumors matching the keywords of micro-blogs published by "Weibo Rumor-Busting" from May 2011 to December 2011. Also we extract the profiles (number of followers, number of followings and VIP authentication) of users who published the event rumors correspondingly and all their micro-blogs. The profiles of the followers and followings and their micro-blogs are also included in our dataset. Finally we create a high quality dataset that consists of 1943 users and 26972 micro-blogs, which includes 104 event rumors and 26868 non-rumors.

4.2 Evaluation

To assess the effectiveness of our methods, we use the standard information retrieval metrics of precision, recall and F1 [12]. The precision is the ratio of the number of rumors classified correctly to the total number of micro-blogs predicted as rumors. The recall is the ratio of the number of rumors correctly classified to the total number of true rumors. The F1 is the harmonic mean between both precision and recall, and is defined as $F1 = \frac{2 \times precision \times recall}{precision + recall}$. We train 4 different classifiers used in previous works [1, 3-5] including Naïve Bayes, Bayesian Network, Neural Networks and Decision Tree. 10-fold cross validation strategy is used to measure the impact of those features introduced in Section 3.2 on the classification performance.

Table 1. The comparison of different classifiers using different set of features

Classifier	*Precision*		*Recall*		*F-measure*	
	Without	With	Without	With	Without	With
Naïve Bayes	0.375	0.176	0.606	0.827	0.463	0.29
Bayesian Network	0.5	0.850	0.058	0.654	0.103	0.739
Neural Network	0.5	0.783	0.163	0.692	0.246	0.735
Decision Tree	0.553	0.750	0.202	0.663	0.296	0.704

The experimental results are shown in Table 1. The experimental results indicate that before adding the 5 new features the precision, recall and F-measure are 0.555, 0.202 and 0.269 respectively. There are two reasons why the performance of these old features in our experiment is worse than that in previous works. One is that the datasets are different, and the other is that the purposes are different. Previous studies focus on identifying spams such as ads, while we focus on detecting event rumors. After introducing the 5 new features to the classification process, the precision, recall and F1 are improved to 0.85, 0.654 and 0.739 respectively. As shown in Figure 10, the precision, recall and f1-measure have increased 0.475, 0.048 and 0.276 separately, demonstrating the effectiveness of our proposed new features.

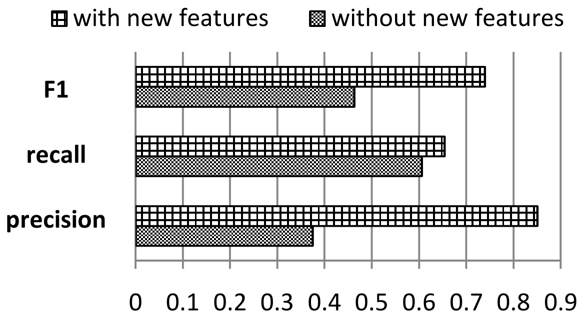

Fig. 10. The comparison between effect of feature set without and with new features

To evaluate the effectiveness of our approach to detecting text-picture unmatched event rumors, we create a dataset including 40 messages, while 17 of them are text-picture unmatched event rumors and 23 non-rumors. All these micro-blog messages contain a picture. Our experiment result shows that the approach is well-performed and the precision, recall and F-measure are 0.833, 0.823 and 0.857 respectively. The disadvantage of this method is that the effectiveness of the picture search engine affects the experimental results to some extent.

5 Conclusion

In this paper we focus on detecting event rumors on Sina Weibo, which are more harmful than other kinds of spams especially in China. To distinguish event rumors from normal messages, we propose five new features and use classification technique to predict which one is event rumor. Further, we divide event rumors into four types and propose a method to identifying text-picture unmatched event rumors with the help of picture search engine. The experiments show the proposed approaches are effective. In the future we will focus on study how to detect other three forms of event rumors.

Acknowledgement. This work was supported by the HGJ PROJECT 2010ZX01042-002-002-03 and National Natural Science Foundation of China under Grant No. 71272029, 70890083, 71110107027 and 61033010.

References

1. Yang, F., Liu, Y., Yu, X., Yang, M.: Automatic detection of rumor on Sina Weibo. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, Beijing, China, pp. 1–7. ACM (2012)
2. Gao, H.Y., Hu, J., Wilson, C., Li, Z.C., Chen, Y., Zhao, B.Y.: Detecting and characterizing social spam campaigns. In: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, Melbourne, Australia, pp. 35–47. ACM (2010)

3. Wang, A.H.: Don't follow me: Spam detection in Twitter. In: Proceedings of the 2010 International Conference on Security and Cryptography (SECRYPT) (2010)
4. Wang, A.H.: Detecting spam bots in online social networking sites: A machine learning approach. In: Foresti, S., Jajodia, S. (eds.) Data and Applications Security and Privacy XXIV. LNCS, vol. 6166, pp. 335–342. Springer, Heidelberg (2010)
5. Benevenuto, F., Magno, G., Rodrigues, T., Almeida, V.: Detecting Spammers on Twitter. In: Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, Redmond, Washington, US (2010)
6. Lee, K., Caverlee, J., Webb, S.: Uncovering social spammers: social honeypots + machine learning. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, pp. 435–442. ACM (2010)
7. Stringhini, G., Kruegel, C., Vigna, G.: Detecting spammers on social networks. In: Proceedings of the 26th Annual Computer Security Applications Conference, Austin, Texas, pp. 1–9. ACM (2010)
8. Yang, C., Harkreader, R.C., Gu, G.: Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers. In: Sommer, R., Balzarotti, D., Maier, G. (eds.) RAID 2011. LNCS, vol. 6961, pp. 318–337. Springer, Heidelberg (2011)
9. Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J., Goncalves, M.: Detecting spammers and content promoters in online video social networks. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Boston, MA, USA, pp. 620–627. ACM (2009)
10. Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J., Zhang, C., Ross, K.: Identifying video spammers in online social networks. In: Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web, Beijing, China, pp. 45–52. ACM (2008)
11. Webb, S., Caverlee, J., Pu, C.: Social Honeypots: Making Friends with a Spammer near You. In: Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, Mountain View, California, US (2008)
12. Yang, Y.: An Evaluation of Statistical Approaches to Text Categorization. Journal of Information Retrieval 1(1/2), 67–88 (1999)