Analysis and Identification of Spamming Behaviors in Sina Weibo Microblog

Chengfeng Lin
Shanghai Jiaotong University
alex_lin@sjtu.edu.cn

Yi Zhou
Shanghai Jiaotong University
zy_21th@sjtu.edu.cn

Kai Chen
Shanghai Jiaotong University
kchen@sjtu.edu.cn

Jianhua He
Aston University
j.he7@aston.ac.uk

Xiaokang Yang
Shanghai Jiaotong University
xkyang@sjtu.edu.cn

Li Song
Shanghai Jiaotong University
song_li@sjtu.edu.cn

ABSTRACT
Spamming has been a widespread problem for social networks. In recent years there is an increasing interest in the analysis of anti-spamming for microblogs, such as Twitter. In this paper we present a systematic research on the analysis of spamming in Sina Weibo platform, which is currently a dominant microblogging service provider in China. Our research objectives are to understand the specific spamming behaviors in Sina Weibo and find approaches to identify and block spammers in Sina Weibo based on spamming behavior classifiers. To start with the analysis of spamming behaviors we devise several effective methods to collect a large set of spammer samples, including uses of proactive honeypots and crawlers, keywords based searching and buying spammer samples directly from online merchants. We processed the database associated with these spammer samples and interestingly we found three representative spamming behaviors: aggressive advertising, repeated duplicate reposting and aggressive following. We extract various features and compare the behaviors of spammers and legitimate users with regard to these features. It is found that spamming behaviors and normal behaviors have distinct characteristics. Based on these findings we design an automatic online spammer identification system. Through tests with real data it is demonstrated that the system can effectively detect the spamming behaviors and identify spammers in Sina Weibo.

Categories and Subject Descriptors
H.3.5 [Online Information Services]: Web-based services; J.4 [Computer Applications]: Social and behavioral sciences

General Terms
Design, Experimentation, Security

Keywords
Sina Weibo; proactive honeypots; crawlers; spamming behaviors; automatic spammer identification

The 7th SNA-KDD Workshop '13 (SNA-KDD'13), August 11, 2013, Chicago, United States. Copyright 2013 ACM 978-1-4503-2330-7...$5.00.

1. INTRODUCTION
Spamming has been a long existing problem with Internet applications, especially Email, due to the almost negligible operating cost of spamming. It also penetrated to the new social network platforms. For example, it was reported by Gao et al that there is a large scale of spamming in Facebook [1]. Spamming causes tremendous costs to the public and the Internet service providers. Anti-spamming is an important and active research topic.

Recently microblogging has become a popular means of communication, information diffusion and marketing. For example, Twitter has over 500 million active users and generates over 340 million tweets daily. As microblogs have short size, allowing users to exchange information and post news items quickly, microblogging has become indispensable in the daily life of millions of people. Many events with societal impacts such as man-made or natural disasters, exposure of corruptions and crime tracking are first reported in microblogging services before they are taken up by main stream media. With microblogging's increasing importance as sources of news updates, and information dissemination, it also becomes an attractive platform for spamming, which can be used for example for commercial advertisement, phishing and computer virus propagation. Due to the shortened URL used in Twitter and other microblogs, it is more difficult to discriminate if a microblog is a spam compared to posts in other social network platforms. It has been found by Grier et al from a study in [2] that about 8% of URLs in Twitter microblogs direct to webpages including phishing, malicious software or computer virus contents. And it was reported that the clicking rate of these URLs is about 0.13%, which shows Twitter can be a very effective platform from spammer's point of view. Anti-spamming will play a critical role for microblogging services.

Traditionally anti-spamming has been studied from two directions: detecting spams and detecting spammers. Spams can be detected by the approaches based on the statistics of spams or based on the features of spam content. For example, Huang et al applied statistical method to analyze the features of spams in [3], while Yin et al studied spam detection with the features of content and context information [4]. Due to the huge amount of microblogs generated daily, detection and blocking of spams alone could not effectively prevent spamming. With the interactive features of microblogging detection and blocking of spammers is expected to be more effective. Irani et al analyzed the documents of 1.9 million MySpace users and developed a spammer detection algorithm based on machine learning [5]. Webb et al deployed 51

honeypots and successfully attracted 1570 spammers in MySpace [6]. Lee et al applied similar approaches of honeypots and machine learning to detect spammer in Facebook and Twitter [7]. Wang et al proposed a spammer detection mechanism over multiple social network platforms, in which the features of users' attributes, messages and associated webpages from multiple platforms were formalized and used to detect spammers [8].

It is noted that although there have been a relatively wide research on the anti-spamming over Twitter, very little was report on the anti-spamming over Chinese microblogs, such as Sina Weibo. Sina Weibo has more than 300 million registered users and about 100 million messages as are posted in Sina Weibo daily. It is currently the dominant microblogging service provider in China. As the spammers in Sina Weibo are expected to have significantly different features from those in Twitter, we are interested in finding out their behaviors and how to detect them and thereafter detect spammers. In this paper we present a systematic research on the analysis and identification of spamming behavior and spammers in Sina Weibo platform, which is believed to be the first kind of such research work. Different from the main approach of detecting microblog spammers with a set of features to be used by one spammer classifier, our spammer detection approach is based on a group of behavior classifiers, each using a separate set of user features and working jointly as a spammer classifier to detect spammers. Our main contributions are in three folds.

• To analyze spamming behaviors we devised several methods to collect spammer samples, which include using proactive honeypots, using crawlers, keywords based searching and buying spammer samples directly from online merchants. A large set of spammers were effectively captured. Different from traditional honeypots, our honeypots are proactive, e.g., publishing microblogs and interacting with other users. It was revealed spammers are become more cautious and only honeypots with lots of activities can attract spammers.

• With these spammer samples we thoroughly processed the database associated with them and found three representative spamming behaviors: aggressive advertising, repeatedly duplicate reposting and aggressive following. We extract various features and compare the behaviors of spammers and legitimate users with regard to these features. It is found that spamming behaviors and normal microblogging behaviors have distinct characteristics.

• According to the above feature analysis and findings we designed an automatic spammer identification system, which is based on the classifiers of different spamming behaviors. Through tests with real data samples it was demonstrated that the system can effectively detect the above mentioned spamming behaviors and identify spammers in Sina Weibo. It is believed that the spammer identification system can be used to effectively block spammer and spams, and promote research on spamming behaviors.

2. SPAMMER COLLECTION

2.1 Honeypots

As we aim to analyze spammers in Sina Weibo and find out their characteristics, spammer samples need to be collected first for the further study. Stringhini et al found that spammers in Twitter and Facebook would initiatively follow legitimate users to establish social relationships [9]. We assume that spammers in Sina Weibo would take similar strategy. We built 25 proactive honeypot that can publish microblogs and interactive with other users. We launched an eight month long experiment to attract spammers and finally got 517 users [10].

We browsed through the homepages and microblogs of the users attracted by our honeypots to judge whether they were spammers. Initially we filter users with following behaviors as spammers: posting microblogs containing URLs pointing to advertising or phishing web pages posting URLs pointing to web pages containing malwares and viruses; reposting one microblogs repeatedly or reposting microblogs from one user with a high frequency; posting microblogs with similar or same contents and any other behaviors that may disturb others.

Among the 517 users, 114 (22%) were labeled as spammers as shown in Figure 1. We checked the relationship between the spammers and honeypots (Figure 2) and found that only 9% of the spammers were friends of the honeypots. This means 91% of the spammers followed the honeypot initiatively which confirmed our previous assumption.

We are also interested in finding out the impact of honeypot activities on the effectiveness of attracting spammers. We set different activity levels to the proactive honeypots by controlling the numbers of microblogs to publish and users to follow. Interestingly we found that the activity level of a honeypot has a significant impact on the number of spammers it can attract. Table 1 shows the details of the five most efficient honeypots, including the number of published microblogs and the number of spammers attracted. It is observed that the top three most efficient honeypots attracted 84% of the 112 spammers. And each of the remaining honeypots attracted no more than 3 spammers. This finding shows that the spammers are becoming more cautious and it is important to make the honeypots active otherwise no spammers may be attracted.

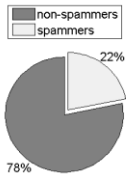


Figure 1. User distribution

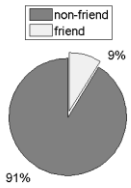


Figure 2. Spammers' Relationship with the honeypots

Table 1. Record of Top Five Honeypots

Micro-blogs	Fans	Following	Attracted Spammer	Non-friend Follower Ratio
1595	192	853	45	50%
1052	95	644	37	41.05%
1919	76	597	12	32.89%
1072	50	355	3	22%
295	6	51	2	50%

2.2 Crawler

It is noted that the honeypots do provide a good number of spammer samples for our research, but the sample size is still small. We expect to get more spammer samples to have a deep analysis of spamming behaviors and develop some effective online spammer identification systems.

We intended to find out active spammers by monitoring microblogs of famous people. We designed and implemented a crawler program that helps us to keep an eye on certain microblogs and find out active users who were frequently involved with these microblogs. Among these users, we found a large number of spammers. These spammers repost the microblogs of famous people which would usually become hot microblogs.

With the help of Sina API, we implemented the crawler that can monitor certain users' latest microblogs and collect the reposting record lists. The crawler stored the information of every user who has reposted the microblogs into the local database and pick out the most active ones. We collected the user IDs of the top 100 users in the influence ranking list provided by Sina. These influential users are mostly famous people in real life, all having a huge number of followers. With the help of the crawler program, we got a list of active users. After manually checking these users, we found 879 spammers out of them.

Among them, we found a special "group" of spammers. They tended to repost every microblogs that they were involved with repeatedly for many times. Moreover, the roots of these microblogs pointed to only several, or even one, users most of whom were film stars or popular singers. One typical example was the famous Chinese actor Qilong Wu, who has a huge number of fans in both Sina microblog as well as real life. More than one hundred of these spammers reposted every microblog of Wu for dozens of times with same comment contents, many of which were meaningless or had nothing to do with the original microblog contents. On the contrary, other legitimate fans rarely repost these microblogs for so many times.

It is also interesting that many of the spammers followed at least one of the other spammers, therefore formed a social relationship network as shown in Figure 3. There is a big group in the middle of the whole network, containing a large number of connected spammers who had reposted Wu's microblogs. They seemed to participate in a human controlled spam campaign and collaborate with each other.

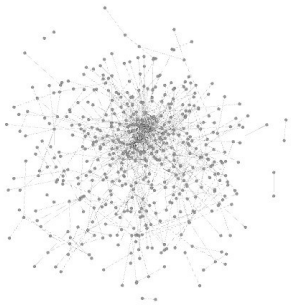


Figure 3. Network of spammers caught by crawlers

2.3 Buying from Online Merchants

In Sina Weibo microblog there is a special kind of spammers which are usually controlled and sold as 'fans' by online merchants. These spammers are controlled to follow a large number of legitimate users. They seldom perform traditional spamming behaviors such as posting spam messages. They are quiet except just following others to maneuver the popularity of the followed users. We bought 8,600 'fans' from several online merchants as spammer samples. Among 7,000 spammers from one merchant, we found 670 ones heavily connected, forming several groups as shown in Figure 4.

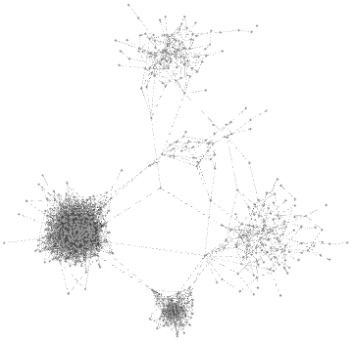


Figure 4. Network of spammers bought online

2.4 Search by Key Words

During the process of manual spammer checking, we found many spammers posted microblogs with similar contents or certain key words. We made use of the microblog content search service provided by Sina to get users posting microblogs containing these key words. We manually checked these users and picked out spammers among them.

3. FEATURE ANALYSIS OF SPAMMING BEHAVIORS

After we collected spammer samples using the approaches described in Section 2, we browsed through their homepages and microblogs and analyzed the features of the samples. We found three representative spammer behaviors among these samples: aggressive advertising behavior (AAB), repeated duplicate reposting behavior (RDRB) and aggressive following behavior (AFB).

Aggressive advertising refers to the behavior of repeatedly posting microblogs containing advertising information for the sake of promotion or propagation. These microblogs are usually related with certain kinds of merchandises or online websites providing fee-based services.

Repeated duplicate reposting refers to the behavior of reposting duplicate microblogs repeatedly. The spammers with this type behavior repost certain user's microblogs with a high frequency. They would repost one single microblog for many times.

Aggressive following refers to the behavior of following a large number of users initiatively. These spammers usually act like dead accounts in terms of microblog posting. Their strong interests in establishing social relationships don't match their inactivity in microblog posting.

We chose some spammers that mainly exhibited one of the above three behaviors, and randomly picked out some legitimate users who were filtered out during the process of manual checking, to form the sample set shown in Table 2 for feature analysis. Next we first present social relationship features and microblog posting features of spammers and legitimate users in Section 3.1, and then analyze the features of the contents of the microblogs resulting from different spamming behaviors in Section 3.2 to Section 3.4, respectively.

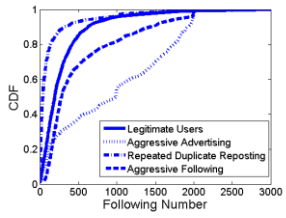
Table 2. Composition of the Spammer sample set

Main Spamming Behavior	Amount
Aggressive Advertising	716
Repeated Duplicate Reposting	710
Aggressive Follwoing	1000
Legitimate Users	3198

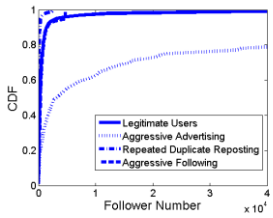
3.1 Social Relation Features and Microblog Posting Features

In order to find out the differences between spammers and legitimate users in social activity and microblog posting activity, we collected the following information of each sample user: the number of followings; number of followers; the number of friends; the number of microblogs and the age of the account.

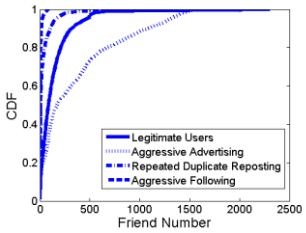
We drew the CDF curves of following number, follower number and friend number of the spammers and legitimate users in Figure 5. Spammers with AAB have more followings, followers and fans. 50% of AAB spammers followed more than 1,000 people while more than 80% of users other than AAB spammers had a following number less than 1,000. Over 20% of AAB spammers had more than 40,000 followers, much higher than those of the other spammers as well as legitimate users. This means spammers with AAB are much more active in establishing social relationships. They followed others and expanded their friend networks so as to gain more followers and make their advertisement seen by more people.



(a) Number of followings



(b) Number of followers



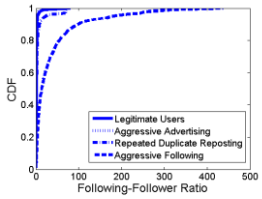
(c) Number of friends

Figure 5. Social Relationship Features.

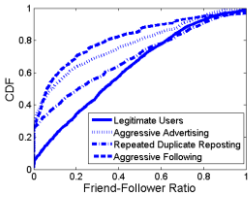
On the other hand, spammers with RDRB are less active. They focused on reposting specific users' microblogs and do not need excessive followings and followers. As a result, they would have less followers and friends.

Spammers with AFB are quite different from the above two kinds of spammers. They followed a lot of users while having few followers or friends. Figure 6 shows the CDF curves of following-follower ratio and friend-follower ratio. The following-follower ratios of spammers with AFB are much higher while the friend-follower ratios are quite low. They do nothing except following others so it is quite difficult for them to attract followers.

As shown in Figure 6b, the friend-follower ratio of legitimate users was higher than these of spammers, because followers of legitimate users were more likely to be their real life friends. In other words, the 'purity' of followers of legitimate users is much higher than spammers.



(a)



(b)

Figure 6. Following-follower ratio and friend-follower ratio

Figure 7 shows the CDF curves of the daily microblog output. Spammers with AAB posted microblog more frequently. On the other hand, spammers with AFB have quite low posting frequency since they seldom posted microblogs.

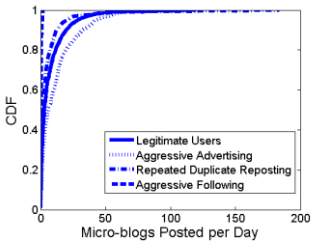


Figure 7. Number of microblogs posted per day

3.2 Content Analysis: Aggressive Advertising

An advertising microblog usually contains URLs which points to a web page related with the microblog content. Besides that, most advertising microblogs come along with a picture which helps to attract users. We assume that spammers with AAB will post more microblogs of such kind. So we picked out all the spammers with AAB and extracted the following features for comparison: average number of URLs in one's microblogs; average number of URLs one posts every day; proportion of microblogs containing at least one picture and average number of sign @ in one's microblog.

The CDF curves of two URL related features are shown in Figure 8 and Figure 9. AAB spammers tend to post more URLs than legitimate users since including URLs is probably the most common advertising method. Besides that, there are more pictures in spammers' microblogs as shown in Figure 10.

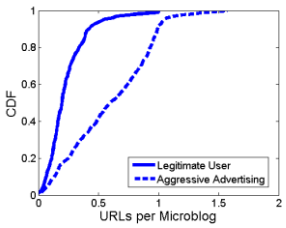


Figure 8. Average number of URLs in one microblog

Considering the average number of sign '@' in microblogs as shown in Figure 11, spammers with AAB use fewer '@' than legitimate users. Users will use '@' in two conditions: 1) when they repost someone's microblog, all the users in the reposting chain will be mentioned with a sign of '@'; 2) when they intend to push a microblog to specific users, they would add '@' signs in the microblog. Advertising spammers are more likely to post advertisements themselves instead of reposting. Even if they advertise by reposting, they tend to repost root microblogs so few users would be involved with a sign of '@'. Few spammers used '@' to push spams to others as doing so would risk of being suspended.

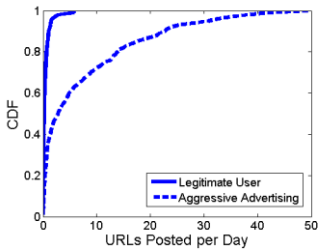


Figure 9. Average number of URLs posted per day

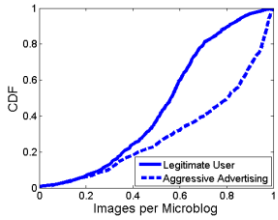


Figure 10. Average number of images in one microblog

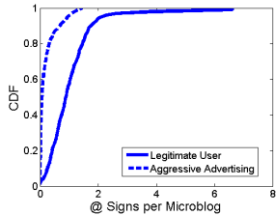


Figure 11. Average number of @ signs in one microblog

3.3 Content Analysis: Repeated Duplicate Reposting

Spammers with RDRB keep reposting microblogs. They usually repost a single microblog for many times while legitimate users rarely do so. Besides that, most of them focus on one or several root users and seldom repost microblogs from other users. They seem to be customized for these root users.

We picked out all these "re-posters" among the spammer samples and extracted the following features for comparison:

• Proportion of duplicated reposting microblogs: the proportion of microblogs sharing the same root. For example, if one reposted microblog Alpha twice and microblog Beta three times, then the proportion is computed as 3/5.

• Average number of reposting (for single microblog): the average number of reposting one performed for one microblog. It is 2.5 for the above example.

• Maximum number of reposting (for single microblog): the maximum number of reposting one performed for a single microblog. It is 3 for the above example.

• Number of different source users: Source users are the authors of the source microblogs of the reposting microblogs, not the authors of root microblogs.

• User focusing metric: the metric is used to measure whether one is focused on reposting a single user. We calculate the proportions of microblogs from different source users and the metric equals to the highest one.

The CDF curve of duplicated reposting proportion is shown in Figure 12. About 60% spammers had more than 80% duplicated reposting microblogs while the duplicate proportion of 50% legitimate users is less than 60%. The CDF curves in Figure 13 and Figure 14 show the average and maximum number of reposting actions one user performs to a single microblog and it can be clearly seen that spammers generally perform more reposting actions to a single microblog than legitimate users do. The three figures altogether proved that spammers are more likely to repost microblogs repeatedly. A very little number of legitimate users had relatively high maximum number of reposting; however, their average number of reposting is much lower than those of the spammers.

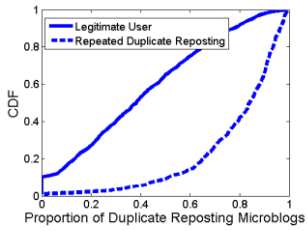


Figure 12. Proportion of duplicate reposting mciroblogs

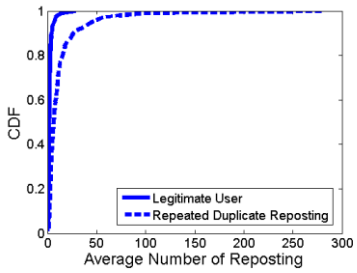


Figure 13. Average number of reposting

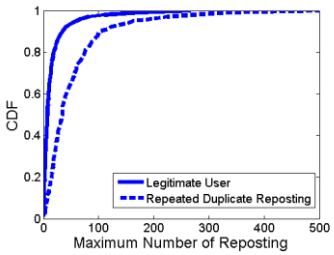


Figure 14. Maximum number of reposting

Considering the source of the reposting microblogs, legitimate users' microblogs have a lot of different sources while spammers' sources are limited to relatively few users as shown in Figure 15. Spammers tend to concentrate on several users who may be their "customers" and repost their microblogs while legitimate users browse through a much larger scope of users. Moreover, the user focusing metrics of spammers, shown in Figure 16, are higher than that of legitimate users. Spammers with RDRB may follow many users; however, most of their reposting microblogs come from one user. We found that many of these spammers reposted microblogs of the same user which implies that they might be used for popularization campaigns.

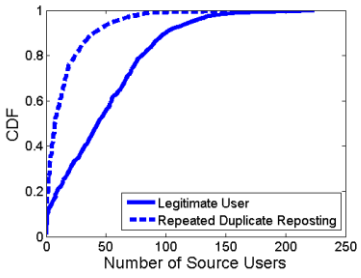


Figure 15. Number of source users

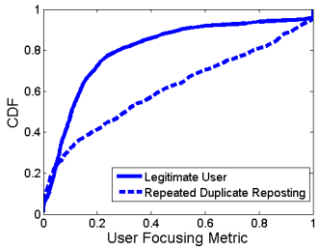


Figure 16. User focusing Metric

3.4 Content Analysis: Aggressive Following

Spammers with AFB will follow a large number of users. However, they are quite inactive in microblog posting and communication: they seldom post microblogs or communicate with others through comments. We extracted the following features of the users for comparison:

- Proportion of microblogs being reposted.
- Proportion of microblogs being commented.
- Average number of comments in microblogs that have been commented.

Microblogs of legitimate users are more frequently reposted and commented as shown in Figure 17 and Figure 18, respectively. 40% of microblogs from the spammers are never reposted and 60% of them have no comments. Even if their microblogs may be commented, they rarely reply to these comments. This means they rarely communicate with others through comments. Figure 19 shows the average number of comments for the commented microblogs. The metric is calculated by dividing the sum of comments with the number of commented microblogs, which means the value will be bigger than one if the user replied to any comment. The figure shows that 80% of the spammers with AFB never communicated with others by replying comments. The three figures show the difference between spammers with AFB and legitimate users in social activities. Spammers with AFB have few social interactions while legitimate users tend to maintain communications with others.

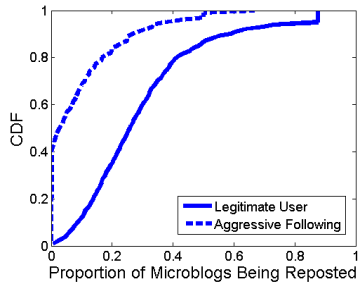


Figure 17. Proportion of microblogs been reposted

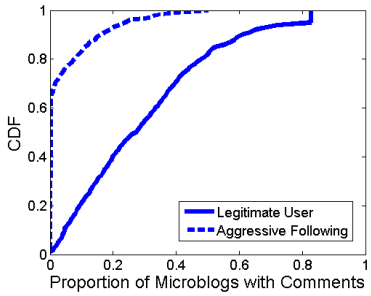


Figure 18. Proportion of microblogs been commented

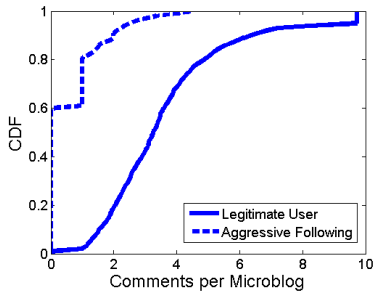


Figure 19. Average number of comments in microblogs being commented

3.5 Summary

In summary, we studied the user features of spammers with different spamming behaviors, and found some characteristics associated with these behaviors.

For aggressive advertising behavior, spammers focusing on advertising are passionate in social activity. They have more followings, followers and friends. They post microblogs more frequently with many URLs and pictures but seldom use '@' signs.

For repeated duplicate reposting behavior, the spammers have fewer followers and friends than legitimate users. They often repost one microblog for many times thus having higher average number of reposting and maximum number of reposting. The sources of their reposting microblogs are limited to one or several users.

For aggressive following behavior, the spammers follow many users but have few followers. They are very inactive in posting microblogs and replying comments. Their microblogs are seldom reposted or commented.

Among the collected spammers, we found some spammers performed more than one kinds of spamming behaviors and showed the features of different spamming behaviors. For example, some spammers not only reposted microblogs containing advertising information from several fixed users, but also posted microblogs containing advertising URLs. Therefore the spammer identification system to be designed should take different spamming behaviors into account in order to detect spammers.

4. SPAMMER IDENTIFICATION

According to the observed characteristics of different spamming behaviors, we designed and built a spammer identification system which can distinguish spammers apart from legitimate users automatically.

We built a classifier for each kind of behaviors based on the feature analysis presented in Section 3. Machine learning techniques are used for the classifiers in our work as they have been widely used and proved to be reliable and effective for classification applications. Three classifiers work together to form the core of the identification system. The system works by collecting features of the target user and feeding these features to the classifiers. The classifiers then determine whether a user has any kind of spamming behaviors and output the results. The system then labels those with detected spamming behaviors as spammers.

4.1 Building Behavior Classifiers

We used Weka Toolkits to train behavior classifiers. For each kind of spamming behavior, we tested more than 30 algorithms provided by Weka with 10-fold cross-validation. We chose different algorithms for different behaviors according to their performance.

Table 3. Number of Users in the Training Sets

Training Set	Legitimate Users	Spammers
Aggressive Advertising	698	716
Repeated Duplicate Reposting	1500	710
Aggressive Following	1000	1000

Table 3 shows more details of the training sets. Both the positive (spammers) and negative (legitimate users) samples came from the dataset in Section 3 for feature analysis. For each kind of spamming behavior we chose different user features.

For AAB we choose the following features: follower number, following number, friend number, friend-follower ratio, number of microblog posted per day, number of URLs posted per day, average number of URLs in one microblog, the proportion of microblogs containing a picture and average number of '@' signs in one microblog.

For RDRB we choose the following features: follower number, friend number, proportion of duplicate reposting microblogs, average number of reposting, maximum number of reposting, number of different source user and the user focusing metric.

For AFB we choose the following features: follower number, friend number, following-follower ratio, number of micro-bog posted per day, proportion of microblogs being reposted, proportion of microblogs being commented and the average comment number below the microblogs being commented.

For each kind of spamming behavior, the algorithm with best performance among the candidate algorithms provided by Weka was used to train the classifiers as shown in Table 4. The best algorithm is Random Committee for AAB, AD Tree for RDRB and Random Forest for AFB, respectively.

Table 4. Results of 10-Fold Cross-Validation

Behavior Classifier	Algorithm	Precision	Recall	F1
Aggressive Advertising	Random Committee	0.937	0.936	0.936
	Random Forrest	0.934	0.934	0.934
	Decorate	0.924	0.923	0.923
Repeated Duplicate Reposting	AD Tree	0.842	0.843	0.842
	simple logistic	0.842	0.842	0.842
	smo	0.840	0.839	0.840
Aggressive Following	random forrest	0.963	0.963	0.963
	Classification via Regression	0.961	0.961	0.961
	Decorate	0.958	0.958	0.958

4.2 Testing the Identification System

After training the behavior classifiers in the spammer identification system, we test it with real test data. We invited some volunteers to collect spammer samples with the above spamming behaviors as well as samples of trusted users from their real life friends. The test dataset is shown in Table 5. All these user samples were collected and verified manually by our volunteers. They were asked to label every spammer they collected according to the major spamming behavior(s) they've observed: AAB, RDPB and/or AFB.

The evaluation results of each behavior classifier are shown in Table 6. The evaluation metrics were calculated based on the result of tests on part of the test set. For each kind of classifier, only spammers with behavior of this kind and all the legitimate samples were involved. The AAB classifier ranks first in both true positive rate and accuracy while RDPB classifier is a little bit worse in true positive rate. However, it has a relatively high precision since it didn't classify any legitimate users as spammers.

Table 7. Test result of the identification system

Label	Amount	Labeled as AA	Labeled as RDR	Labeled as AF	Performing any spamming behavior
Legitimate	811	26	0	24	48
Advertising	336	295	5	67	314
Repeated Duplicate Reposting	435	135	279	138	366
Aggressive Following	812	75	18	580	619

Table 5. Composition of the Test Set

User Label	User Amount
Legitimate	811
Aggressive Advertising	336
Repeated Duplicate Reposting	435
Aggressive Following	812

Table 6. Evaluation of the classifiers on the Test Set

Classsifiers	TP Rate	FP Rate	Accuracy
AAB classifiers	0.8779	0.0321	0.9415
RDPB classifiers	0.6414	0	0.8747
AFB classifiers	0.7143	0.0296	0.8423
Classsifiers	**Precision**	**Recall**	**F1**
AAB classifiers	0.919	0.878	0.898
RDPB classifiers	1	0.6414	0.7815
AFB classifiers	0.9603	0.7143	0.8192

After we apply the classifiers to the whole test set, we get the results shown in Table 7. Some of the spammers with one behavior label are detected by other kinds of behavior classifiers. And some of the spammers are detected by more than one behavior classifiers. We manually checked these spammers and found that they performed more than one kind of spamming behaviors. For example, some spammers, who are initially labeled "repeated duplicate reposting", added URLs for advertising when reposting.

Although single classifier for single spamming behavior may not have an excellent performance with these complex spammers, combining several spamming behavior classifiers together to build the identification system is expected to improve detection performances. Considering the overall identification performance, the system works quite well. It detected 82.06% of the spammers while only 5.92% of the legitimate users were classified as spammers thus having an accuracy of 86.13%.

5. CONCLUSION

In this paper, we studied different spamming behaviors in Sina Weibo community. We took several approaches to collect spammer samples, which include uses of proactive honeypots and crawlers, keywords based searching and buying spammer samples from online merchants. A large set of spammers were effectively captured for further analysis.

We processed the database associated with these spammers and found three representative spamming behaviors: aggressive advertising (AAB), repeated duplicate reposting (RDRB) and aggressive following (AFB). We extracted various features and compared the behaviors of spammers and legitimate users and found that spamming behaviors and normal microblogging behaviors have distinct characteristics. Spammers with AAB are more successful in building social relationship network and post microblogs containing a lot of URLs and pictures. Spammers with RDPB have fewer followers and friends and tend to repost microblogs from one user repeatedly. Spammers with AFB tend to follow a large number of users but have few followers. They seldom post microblogs and are rarely reposted or commented.

According to the above feature analysis and findings, we designed an automatic spammer identification system which is based on the classifiers of different spamming behaviors. We tested its performances using real data samples and it was demonstrated that the system is effective in detecting the above mentioned spamming behaviors and identifying spammers.

We also found some special spammers with cautious spamming strategies. For example, some of the spammers posted spams during a specific short period of time and acted like legitimate users in the rest of time. Our system was not very effective in detecting such spammers. Moreover, the detection depends on certain user features which may be avoided by improved spammers. For example, spammers with AFB may choose to follow each other to gain a friend network and make their social features more "legitimate". More flexible and robust system need to be designed to detect spammers, which is left as our future work.

6. ACKNOWLEDGMENTS

The work is partially supported by National Natural Science Foundation of China (Grant No. 61025005, 61129001, 61201384), and the National Grand Fundamental Research 973 Program of China (Grant No.2010CB731406), and Shanghai Key Lab of Digital Media Processing and Transmissions STCSM (12DZ2272600).

7. REFERENCES

[1] Hongyu Gao, Jun Hu, Christo Wilson, Zhichun Li, Yan Chen, and Ben Y. Zhao. 2010. Detecting and characterizing social spam campaigns. In *Proceedings of the 17th ACM conference on Computer and communications security* (CCS '10). ACM, New York, NY, USA, 681-683. DOI=http://doi.acm.org/10.1145/1866307.1866396.

[2] Chris Grier, Kurt Thomas, Vern Paxson, and Michael Zhang. 2010. @spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security* (CCS '10). ACM, New York, NY, USA, 27-37. DOI=http://doi.acm.org/10.1145/1866307.1866311.

[3] Huang, C., Jiang, Q., & Zhang, Y. 2010. *Detecting comment spam through content analysis*. In Web-Age Information Management (pp. 222-233). Springer Berlin Heidelberg.

[4] Yin, D., Xue, Z., Hong, L., Davison, B. D., Kontostathis, A., & Edwards, L. 2009. *Detection of harassment on web 2.0*. Proceedings of the Content Analysis in the WEB, 2.

[5] Irani, D., Webb, S., & Pu, C. 2010. *Study of static classification of social spam profiles in myspace*. In Proceedings of the 4th International Conference on Weblogs and Social Media.

[6] Webb, S., Caverlee, J., & Pu, C. 2008. *Social honeypots: Making friends with a spammer near you*. In Proceedings of the Fifth Conference on Email and Anti-Spam (CEAS 2008), Mountain View, CA.

[7] Kyumin Lee, James Caverlee, and Steve Webb. 2010. Uncovering social spammers: social honeypots + machine learning. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (SIGIR '10). ACM, New York, NY, USA, 435-442. DOI=http://doi.acm.org/10.1145/1835449.1835522.

[8] De Wang, Danesh Irani, and Calton Pu. 2011. A social-spam detection framework. In*Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference* (CEAS '11). ACM, New York, NY, USA, 46-54. DOI=http://doi.acm.org/10.1145/2030376.2030382.

[9] Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. 2010. Detecting spammers on social networks. In *Proceedings of the 26th Annual Computer Security Applications Conference* (ACSAC '10). ACM, New York, NY, USA, 1-9. DOI=http://doi.acm.org/10.1145/1920261.1920263

[10] Zhou, Y., Chen, K., Song, L., Yang, X., & He, J. 2012. Feature Analysis of Spammers in Social Networks with *Active Honeypots: A Case Study of Chinese Microblogging Networks. In Advances in Social Networks Analysis and Mining*. (ASONAM), 2012 IEEE/ACM International Conference on (pp. 728-729). IEEE.