Automatic Fake Followers Detection in Chinese Micro-blogging System

Yi Shen[1,2,⋆], Jianjun Yu[2], Kejun Dong[2], and Kai Nan[2]

[1] University of Chinese Academy of Sciences
[2] Computer Network Information Center, Chinese Academy of Sciences
{shenyi,yujj,kevin,nankai}@cnic.ac.cn

Abstract. Micro-blogging, which has greatly influenced people's life, is experiencing fantastic success in the worldwide. However, during its rapid development, it has encountered the problem of content pollution. Various pollution in the micro-blogging platforms has hurt the credibility of micro-blogging and caused significantly negative effect. In this paper, we mainly focus on detecting fake followers which may lead to a problematic situation on social media networks. By extracting major features of fake followers in Sina Weibo, we propose a binary classifier to distinguish fake followers from the legitimate users. The experiments show that all the proposed features are important and our method greatly outperforms to detect fake followers. We also present an elaborate analysis on the phenomenon of fake followers, infer the supported algorithms and principles behind them, and finally provide several suggestions for micro-blogging systems and ordinary users to deal with the fake followers.

Keywords: Micro-blogging, Fake followers, Classification, Feature extraction.

1 Introduction

Due to its simplicity and rapid velocity, micro-blogging is experiencing tremendous success. However, the micro-blogging services have also encountered several serious troubles during their booming development, one of which is the fake followers problem. The phenomenon of fake followers emerges soon after the birth of micro-blogging systems and now has flooded in the mainstream micro-blogging services such as Twitter[twitter.com] and Sina Weibo[weibo.com]. According to Yahoo reports[1], a considerable part of the followers of celebrities on Twitter are fake, and the proportion may be as high as over 50%. Despite both Twitter and Sina Weibo have made much effort on struggling with the fake accounts, nevertheless, the effect is not very significant.

Generally speaking, people purchase fake followers mainly for two motivates according to the investigation from Theweek[2]. The first one is that people purchase followers just to achieve fame and feed their vanity. They misunderstand that, the more

⋆ This research was supported by NSFC Grant No.61202408 and CAS 125 Informatization Project XXH12503.

[1] http://news.yahoo.com/blogs/prot-minded/10-people-won-t-believe-fake-followers-twitter-215539518.html

[2] http://theweek.com/article/index/243357/how-celebrities-buy-twitter-followings-mdash-and-how-you-can-too

V.S. Tseng et al. (Eds.): PAKDD 2014, Part II, LNAI 8444, pp. 596–607, 2014.
© Springer International Publishing Switzerland 2014

the followers, the greater their influence. The second, some merchants or brands seek a huge number of followers so as to push Ads. The follower merchants take advantage of algorithms and softwares to produce a mass of fake accounts automatically. These accounts act as real followers to meet the needs of their customers. There are already some websites for trading fake twitter followers in public, such as [intertwitter.com] and [fakefollowerstwitter.com]. The prices for 1K fake followers provided by these online merchants are around 5-20$. A recent report from NYTimes pointed out that the fake Twitter followers have become a multimillion-dollar business.

The fake followers are so epidemic in the micro-blogging systems, which have caused plenty of hazards, such as making noise on personalized recommendation and user influence analysis, risking privacy to unknown people, receiving valueless Ads. Therefore, it is very significant to present an effective method to distinguish the fake followers from the legitimate users in micro-blogging systems. Researchers have developed effective tools[3] based on the inactive characteristic and spam-related features to detect fake followers of Twitter. However, the fake followers are distinguished to traditional spammers. The spammers mainly utilize the freedom and rapid nature of micro-blogging platforms to push unsolicited advertisements or malicious information[1], while fake followers aim to follow the users who are urgent to be popular users. Therefore, in many cases, fake followers do not send spam to others, and pretend to be legitimate users. At the same time, fake followers in Sina Weibo seem to be more sophisticated than Twitter's. For they not only appear as active as the legitimate users, but also almost have no spam in their posts. We have checked the tweets of the followers purchased from the four markets (1000 followers each) manually. An account will be labeled as a spammer once there is any spam appear in the most recent 20 tweets of its timeline. Finally, we found about 62% of these fake Twitter followers are belong to spammers. We define the accounts which have less than 10 posts or 10 followers as inactive accounts. In general, 57% of the fake followers for Twitter are inactive. However, according our statistic on fake followers for Sina Weibo, the spammer-ratio and inactive-ratio are just 8% and 6% respectively. As a result, we aim to address the issue of detecting the fake followers for Sina Weibo in this paper. Considering that there are no related public datasets for validating the effects of the proposed method, we just purchased a considerable number of fake followers from different merchants as our datasets.

We frame our contributions as follows:

(1) We examine a number of properties of fake followers in Sina Weibo, and present an effective strategy for automatic detection of fake followers.

(2) We give an in-depth analysis of the profile-evolution of fake followers, and try to understand the principles and algorithms for fake followers' generation.

(3) We provide several suggestions for both the micro-blogging systems and legitimate users on how to deal with fake followers.

The rest of this paper is organized as follows. After a brief review of the related work in Section 2, we introduce datasets used in this paper in Section 3. Next, we analyze the features of fake followers and propose a voting classier for detecting the fake followers

[3] `http://www.socialbakers.com/twitter/fakefollowercheck/methodology/`

automatically in Section 4. Then, we present our experiments in Section 5. Section 6 gives the analysis on the results and the discussion. Finally, we conclude the paper in Section 7 with future work.

2 Related Work

Due to its dual role of social network and news media[2], micro-blogging has become an important platform for people to access information. However, with its rapid development, micro-blogging is plagued by various credibility problems for a long time. The information on micro-blogging systems has been polluted heavily by the rumors [3][4] and the spams [5][6]. As fake followers usually accompany with illusive following-actions, they will also pollute the social connections in the micro-blogging systems.

There are two main methods for the follower merchants to provide fake followers to their customers. One way is to utilize some third-party applications to defraud the authorizations from some legitimate users, then manipulate these compromised users to follow the customers so as to promote their follower-count[7] [8]. Another method is to utilize specific softwares to batch produce a mass of fake accounts that disguised as real users.

The compromised users have a fatal weakness, that is their loyalty to the customers is very low[9]. Since the compromised users are real users, once they discover strangers appeared in their followee-list or receive valueless posts from the customers, they may initiatively remove the customers from their followee-list. In contrast, the fake accounts act as followers are very stable. Moreover, the fake followers have been very tricky, especially in Sina Weibo. As a result, we mainly focus on how to automatically detect the fake followers created by the follower merchants in this paper.

As the spammers in social networks are also created by certain bots [10]. Therefore, the detection of spammers in social networks is related to our work. Thomas et al. [11] analyzed the profiles of suspended spammers list provided by Twitter, then got some significant key points about the techniques of the spammers. [12] presented graph-based methods to analyze the network structure of the spammers. In [13] [14] [15], the researchers proposed feature-based systems to address the spammer-detection problem. All of these works have provided meaningful investigation on the characteristics of the forged accounts in social networks.

Our work is quite different from the works mentioned above, and presented as the first effort on detecting fake followers as far as we know. First, according to our real dataset, only a small part of fake followers in Sina Weibo belong to spammers, and thus the traditional methods for detecting the spammers in social networks are proved not appropriate to identify fake followers. Secondly, previous approaches did not take account of the evolution of fake accounts. We extract several evolutionary features for our detection task, which are proved significantly important to promote detection performance.

3 Data Collection

We notice that fake followers have two kinds of following-behaviors: following the customers who have purchased their service, and randomly following other users so as to disguise themselves. To investigate these two behaviors respectively, we collect two different datasets. We buy 20,000 fake followers in Sina Weibo from 4 different follower merchants, and use these followers on 4 new-created accounts[4].

We keep monitoring these fake accounts for about 4 months(02/03/2013-30/06/2013), and only 17 of them are suspended by Sina Weibo during this period. It means these fake followers are very deceptive and seem to be good at disguising themselves. The remaining fake followers are made as our dataset(DATASET1) for the proposed method, which contains 19,146 records after removing the duplicated ones.

20 new Sina Weibo accounts are registered and used as baits to attract randomly following from fake followers. We keep these baits no actions (no posts and no following behaviors) for a long period(10/03/2013-15/06/2013), and recognize their followers as fake follows since legitimate users only follow the users they are really interested in. Totally, 724 accounts are captured by our baits and used as DATASET2.

In order to compare with fake followers, a dataset of legitimate users is also necessary. 114 volunteers are invited to identify the accounts of their acquaintances from their respective followee-list on Sina Weibo. In this way, 14,873 different accounts of real users are obtained. We also crawl the accounts of 6,472 celebrities whose identity have been officially verified by Sina Weibo(”Big V”). After merging these two parts of datasets, a legitimate dataset(DATASET3) contains 20,211 users. Finally, we utilize Sina Weibo API[5] to access the profiles of all users in DATASET1, DATASET2 and DATASET3 for further research.

4 Proposed Method for Detecting Fake Followers

In this section, we mainly present our approach of detecting fake followers in Sina Weibo. The fake followers detection issue can be considered as a binary classification task. Through comprehensive analysis of the datasets, we extract numerous features with great discrimination between legitimate followers and fake followers to build the classifier. These features can be divided into three types: the post-related features, the user relationship features and the evolutionary features. We randomly sample 2,000 items from DATASET1 and DATASET3 respectively, depict their differences on various features, and provide a voting-SVM classifier.

4.1 The User Relationship Features

The Ratio of Followee Count and Follower Count (RFF). RFF of fake followers is surprisingly high due to their large number of followees and very few followers. According to our statistics, for a typical legitimate user, this ratio is usually within a range of [0.5,3]. Since some celebrities have a huge number of followers, their RFF are often close to zero. RFF is shrunk with a logarithmic function as follows:

[4] All these four merchants do not have public website. We get in touch with them via searching in Sina Weibo with the keyword ”add followers”. The prices of them are all around ¥ 60-80 per 1K followers.

[5] `http://open.weibo.com`

$$RFF(U) = \lg \frac{FolloweeCount + 1}{FollowerCount + 1} \tag{1}$$

Figure 1(a) shows the cumulative distribution function(CDF) of RFF for fake followers and legitimate users. It is very clear that RFF is able to discriminate the two types of users distinctly.

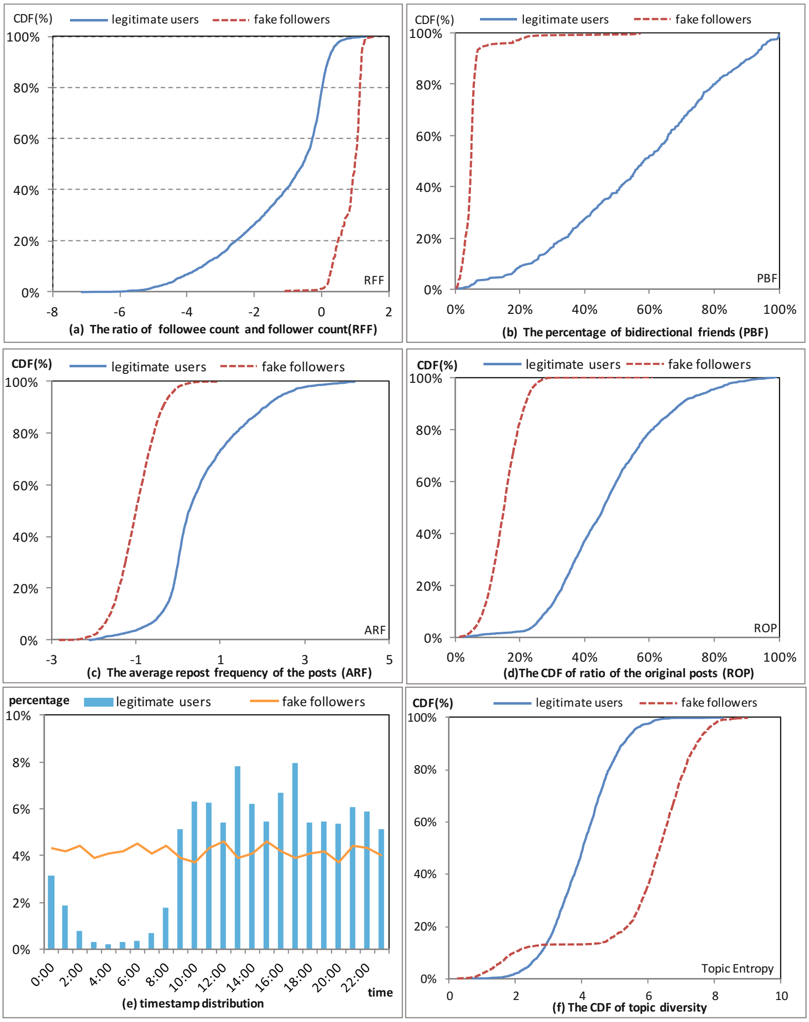

Fig. 1. The comparison between legitimate users and fake followers

The Percentage of Bidirectional Friends (PBF). PBF is calculated with Equation (2). In real scenarios, people have their own social communities, and often the legitimate users belong to the same community always follow each other in micro-blogging systems. Certainly, fake followers have no friends, no classmates and no colleagues, therefore, their PBF should be lower than legitimate users'. The CDF curves of PBF in Figure 1(b) have proven this assumption.

$$PBF\left(U\right) = \frac{CountOf(Followee \bigcap Follower) + 1}{FolloweeCount + 1} \tag{2}$$

4.2 The Post-related Features

Average Repost Frequency of the Posts (ARF). ARF can reflect the influence of the user[16]. The ARF of a fake follower is usually very low for two reasons. One reason is that a fake follower has hardly any high quality followers. The other is that its post behavior is manipulated by softwares, as a result, their posts are meaningless for others in most cases. According to our datasets, the average ARF of normal users is much higher than fake accounts. The CDF in Figure 1(c) shows that the ARF of most normal users is over 0, which means the posts from normal users will be reposted at least once in average. In contrast, it is obvious that over 80% of the fake followers have a low ARF between [-2,-1].

$$ARF\left(U\right) = \lg \frac{\sum_{P \in Posts(U)} RepostCount(P)}{TotalPostsCount(U) + 1} \tag{3}$$

Ratio of the Original Posts (ROP). We find the vast majority of posts from fake followers are belong to repost. In Figure 1(d), we find that ROP for almost all fake followers are less than 20%, while ROP for legitimate users are in the range of(20%-60%).

$$ROP\left(U\right) = \frac{OriginalPostsCount + 1}{TotalPostsCount + 1} \tag{4}$$

Proportion of Nighttime Posts (PNP). We make a comparison of the post-timestamp distributions of users in DATASET1 and DATASET3. We notice that legitimate users rarely publish posts during the nighttime. However, the timestamp distribution of fake followers obeys an approximative uniform distribution, which is contrary to the common sense since we all know that people need to sleep and with a very low probability to publish posts in the middle of the night. Consequently, we suspect that the post-creation behavior of fake followers is controlled by a periodic creation algorithm. As the distributions in Figure 1(e) have shown a remarkable discrimination between fake followers and legitimate users, we are confident to employ the proportion of nighttime (1:00am-7:00am) posts as an indicator for classification.

$$PNP\left(U\right) = \frac{NightPostsCount + 1}{TotalPostsCount + 1} \tag{5}$$

Topic Diversity. We apply author topic model[17] to mine the themes of the accounts in DATASET1 and DATASET3. Normally, a legitimate user usually has countable interests of topics, while the fake followers usually present variable interests since they repost from others randomly. As a result, we assume that the topics of fake followers should be more diverse than legitimate users. We use topic entropy [18] to measure the topic diversity as Equation 6. Correspondingly, fake followers present higher topic entropy than legitimate users'.

$$H(u) = -\sum_{i=1}^{K} P(z_i|u) \log_2 P(z_i|u) \tag{6}$$

Where z_i denotes the topics generated by author topic model. K is the count of topics.

We notice that there is a step appears in the beginning of the CDF curve of fake followers as shown in Figure 1(f). This is because a small part of fake followers publish spams frequently, which causes their lower topic entropy.

4.3 The Evolutionary Features

As we observed, the fake followers always evolve when time changes. To better capture the pattern of their evolution, we keep tracking the accounts in DATASET1 for a period of 60 days (07/04/2013-06/06/2013), and several important parameters(i.e. the count of posts, followees, followers) have been recorded every day. We then display the evolution model of typical fake accounts in Figure 2 comparing with legitimate users[6].

We can find that the evolution of post-count of fake followers is consistent with gradient trend every day, which means they share the same post-frequency. This also implies that they are manipulated by the same algorithm. The evolutionary curves of post-count from legitimate users in Figure 2(b) are flatter than fake followers'. This is because most of legitimate users publish fewer posts than fake followers and are with low probability of mutation.

Because people rarely remove excessive amount of users from their followee-lists during a short time[19], their followee count will tend to increase in general, whereas the followee-count of five fake followers fluctuate violently for many times as shown in Figure 2(c). According to this surprising investigation, we are more confident to confirm that the fake followers are manipulated by software. In Sina Weibo, an account can only follows 2,000 users at most, the software has to make the fake followers to remove some followees who do not purchase their service in order to release the room for their new customers.

The Figure 2(e) and Figure 2(f) demonstrate that follower-count of both fake followers and legitimate users maintain a relatively stable trend. However, the fake followers seem to decrease more frequently.

Base on the analysis above, we extract 6 features to model the evolutionary characteristics:

[6] Due to the space limitation, we just plot 15-day evolution(23/05/2013-06/06/2013) in Figure 2.

(1) the standard deviation of post-count(σ_{post}).
(2) the general slope of post-count(g_{post}).
(3) the standard deviation of followee-count($\sigma_{followee}$).
(4) the decrease frequency of followee-count($DF_{followee}$).
(5) the standard deviation of follower-count ($\sigma_{follower}$).
(6) the decrease frequency of followee-count($DF_{follower}$).

The following Equations give the calculation of these features.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2} \tag{7}$$

$$DF = \frac{\sum_{i=2}^{N} sigmoid(X_i - X_{i-1})}{N} \tag{8}$$

$$g_{post} = \frac{(X_N - X_1)}{N} \tag{9}$$

Among these features, the three standard deviations are used to model the degree of fluctuation. The decrease frequency can capture the exceptions of followee-count and follower-count. In addition, g_{post} is able to reflect the rate of increase of posts during a period of time. We believe that these evolutionary features are also significant for the classification task. For each account, we calculate these features above based on the profiles in the latest 30 days, and integrate them into our classifier.

4.4 Classifier for Detecting Fake Followers

Due to its prominent fame on solving multidimensional classification problem, we utilize Support Vector Machine(SVM) as our basic classifier. We exploit LibSVM[20] to implement a SVM classifier with RBF (Radial Basis Function) kernel function. Since we do not know the actual proportion of fake followers in Sina Weibo, the classifier directly trained from the original dataset may be biased. To this end, we adopt "bagging"[21] strategy to overcome this problem. DATASET1 and DATASET3 are merged as the final dataset(DATASET4), therefore, it contains 13,873 fake followers and 20,211 legitimate users. 3,400 accounts in DATASET4 are stochastically selected as test dataset, and the left is used as training dataset. We randomly sample 2,000 accounts from DATASET4 for 15 times, then use 15 subsets to train a SVM classifier respectively. Finally, the category for an account in test dataset will be determined by the voting results of these 15 SVM classifiers. A 10-fold cross validation on DATASET4 shows that our voting classifier works well. The average accuracy is about 98.7%, and the average false positive rate is very close to 0%.

5 Experiments

In this section, we first make a comparison between our voting-SVM classifier to several baselines. Next, we utilize our proposed classifier to evaluate the accounts in DATASET2

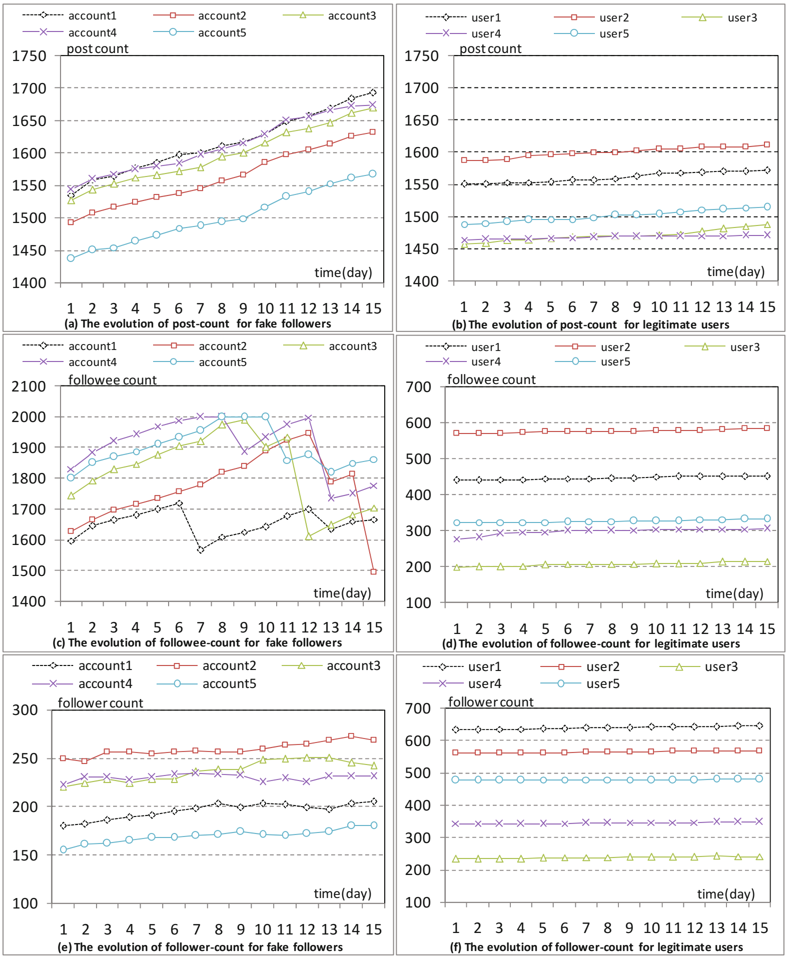

Fig. 2. The evolution of typical fake followers and legitimate users

which are attracted by our baits. Finally, we adopt our voting-SVM classifier to detect fake followers in the wild.

To examine the importance of evolutionary features, we implement a version of voting-SVM classifier with no evolutionary features. A normal SVM classifier is employed as a baseline to validate the improvement of the ”bagging” strategy. We also implement the spammer classifier introduced in [13] as another baseline. Our voting-SVM outperforms others with all the metrics as shown in Table 1. The evolutionary features and the ”bagging” strategy both play an important role in the final results. The accuracy of spammer classifier is much lower than others’. It is because most fake followers in Sina Weibo do not have obvious spam-related characteristics.

As we observed, the suspicious accounts attracted by our baits in DATASET2 are very similar to fake followers in DATASET1 in many aspects(i.e. the features

Table 1. The percentage of fake followers for different groups

Classifier	Accuracy	False Positive Rate	F1
Spammer Classifier	63.4%	0.7%	0.712
SVM(no evolution)	91.3%	3.8%	0.834
SVM	95.1%	1.2%	0.933
Voting-SVM(no evolution)	93.9%	3.5%	0.917
Voting-SVM	**98.7%**	**0.4%**	**0.964**

discussed above), we guess that they have a great probability to be fake followers. To prove our conjecture, we apply our trained voting-SVM classifier to classify the accounts in DATASET2. Up to 95.6% account in DATASET2 are judged as fake followers, which indicates that the accounts in DATASET1 and DATASET2 are extremely homogeneous. Therefore, we confirm that most followers of our baits are indeed caused by randomly following from fake followers.

In order to understand the global state in the whole Sina Weibo system, we pick up 100 legitimate users. Among the sampled accounts, 50 of them are registered by ordinary people with different occupations. Another half are those celebrities in different fields. We exploit our classifier to analyze the quality of their followers, and the average percentage of fake followers is about 37.2%. As listed in Table 2, the celebrity accounts contain more fake followers than ordinary users. We guess that most celebrities have the suspects on purchasing fake followers.

Table 2. The percentage of fake followers for different groups

Ordinary Users	Celebrity Users	Fake Followers	Baits	Average
14.4%	42.3%	94.1%	95.6%	37.2%

6 Discussion

6.1 The Principle Behind Fake Followers

Based on all analysis mentioned above, we give an investigation on the behaviors and characteristics of fake followers, which helps to find several key points of principal control algorithms behind fake followers in Sina Weibo:

(1) The fake follower merchants exploit certain register tools to produce massive accounts in Sina Weibo.

(2) Then they make fresh fake accounts to follow each other, as a result, both the followee-count and the follower-count will increase rapidly within a short time.

(3) Next, the fake accounts will send random following-action to many legitimate users. On the one hand, this operation may attract some back-following from legitimate users, which can improve their level of camouflage; for the other hand, the fake accounts would get the source of the reposts.

(4) Since it is very complicated to automatically generate diverse posts with high quality, so reposting the posts from legitimate users is a conservative method, especially

for those Chinese posts. The control algorithm manages fake accounts to repost many posts every day, which not only ensures the quality of post-content, but also makes fake followers appear to be much more realistic.

(5) When someone has purchased the fake-follow service, the merchants will manipulate fake accounts to follow this customer, so as to make he/she appears to have many followers.

(6) Once the followee-count of a fake follower is close to the limit(2,000 in Sina Weibo), the control software will remove some legitimate users or other fake accounts from the followee-list of fake followers.

6.2 How to Struggle with Fake Followers

It is a comprehensively difficult work for micro-blogging systems to identify fake followers. The micro-blogging services should make crucial measure to deal with fake followers in a fair and just manner. We believe that our work provides quite good suggestions for micro-blogging systems. As normal users, we need to be cautious to follow others in micro-blogging systems, or we may receive many valueless reposts from fake followers. We discover that fake followers prefer to choose to follow some novice users. Due to their less experience, many novice users tend to follow back when they are followed by fake followers. Also it is easy for a user to utilize the features above to judge whether a follower is fake manually through PNP, RFF and OPR, since these features of an account are public.

In fact, the count of followers alone means very little about the influence of a user[16]. As a result, it's unnecessary for us to purchase followers. We should share this fact with the customers of fake followers, so as to fundamentally undermine the follower market.

7 Conclusion

The fake followers have severely hurt the credibility of micro-blogging systems. In this paper, we mainly focus on automatic detection of fake followers in Sina Weibo. We extract many discriminative features especially several evolutionary features to build a classifier detecting fake followers. The proposed classifier performs satisfactorily on the standard metrics for classification. We also summarize several principles behind fake followers. Finally, we give suggestions for combating with fake followers.

Still there is much work to do in the future. We believe that the camouflage algorithm of fake followers has huge space for improvement. The manipulator can adjust the characteristics of fake followers to evade our detection algorithm. The race between the detection algorithms and the camouflage strategies will exist for a long time. As a result, it is necessary for us to keep tracking the evolution of fake followers and constantly sum up new features to deal with them.

References

1. Stringhini, G., Kruegel, C., Vigna, G.: Detecting spammers on social networks. In: ACSAC 2010 Conference Proceedings, pp. 1–9 (2010)

2. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: WWW 2010 Conference Proceedings, pp. 591–600 (2010)
3. Qazvinian, V., Rosengren, E., Radev, D.R., Mei, Q.: Rumor has it: identifying misinformation in microblogs. In: EMNLP 2011 Conference Proceedings, pp. 1589–1599 (2011)
4. Mendoza, M., Poblete, B., Castillo, C.: Twitter under crisis: Can we trust what we rt? In: Proceedings of the First Workshop on Social Media Analytics, pp. 71–79 (2010)
5. Grier, C., Thomas, K., Paxson, V., Zhang, M.: @spam: the underground on 140 characters or less. In: CCS 2010 Conference Proceedings, pp. 27–37 (2010)
6. McCord, M., Chuah, M.: Spam detection on twitter using traditional classifiers. In: Calero, J.M.A., Yang, L.T., Mármol, F.G., García Villalba, L.J., Li, A.X., Wang, Y. (eds.) ATC 2011. LNCS, vol. 6906, pp. 175–186. Springer, Heidelberg (2011)
7. Egele, M., Stringhini, G., Kruegel, C., Vigna, G.: Compa: Detecting compromised accounts on social networks. In: Symposium on Network and Distributed System Security, NDSS (2013)
8. Gao, H., Hu, J., Wilson, C., Li, Z., Chen, Y., Zhao, B.Y.: Detecting and characterizing social spam campaigns. In: IMC 2010 Conference Proceedings, pp. 35–47 (2010)
9. Stringhini, G., Egele, M., Kruegel, C., Vigna, G.: Poultry markets: On the underground economy of twitter followers. In: WOSN 2012 Conference Proceedings, pp. 1–6 (2012)
10. Ghosh, S., Viswanath, B., Kooti, F., Sharma, N.K., Korlam, G., Benevenuto, F., Ganguly, N., Gummadi, K.P.: Understanding and combating link farming in the twitter social network. In: WWW 2012 Conference Proceedings, pp. 61–70 (2012)
11. Thomas, K., Grier, C., Song, D., Paxson, V.: Suspended accounts in retrospect: an analysis of twitter spam. In: IMC 2011 Conference Proceedings, pp. 243–258 (2011)
12. Yang, C., Harkreader, R., Zhang, J., Shin, S., Gu, G.: Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In: WWW 2012 Conference Proceedings, pp. 71–80 (2012)
13. Benevenuto, F., Magno, G., Rodrigues, T., Almeida, V.: Detecting spammers on twitter. In: Collaboration, Electronic Messaging, Anti-abuse and Spam Conference (CEAS), vol. 6 (2010)
14. Lee, K., Caverlee, J., Webb, S.: Uncovering social spammers: social honeypots+ machine learning. In: SIGIR 2010 Conference Proceedings, pp. 435–442 (2010)
15. Zhu, Y., Wang, X., Zhong, E., Liu, N.N., Li, H., Yang, Q.: Discovering spammers in social networks. In: AAAI 2012 Conference Proceedings, pp. 171–177 (2012)
16. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring user influence in twitter: The million follower fallacy. ICWSM 10, 10–17 (2010)
17. Rosen-Zvi, M., Griffiths, T., Steyvers, M., Smyth, P.: The author-topic model for authors and documents. In: UAI 2004 Conference Proceedings, pp. 487–494 (2004)
18. Chang, J., Gerrish, S., Wang, C., Boyd-graber, J.L., Blei, D.M.: Reading tea leaves: How humans interpret topic models. In: Advances in Neural Information Processing Systems, pp. 288–296 (2009)
19. Kwak, H., Chun, H., Moon, S.: Fragile online relationship: a first look at unfollow dynamics in twitter. In: CHI 2011 Conference Proceedings, pp. 1091–1100 (2011)
20. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2, 1–27 (2011)
21. Quinlan, J.R.: Bagging, boosting, and c4. 5. In: AAAI/IAAI, vol. 1, pp. 725–730 (1996)