International Journal of Hybrid Information Technology
Vol. 9, No.10 (2016), pp. 239-250
http://dx.doi.org/10.14257/ijhit.2016.9.10.24

# Towards Spammer Detection in Microblogging Platforms

Jie Zhao, Yan Liu and Shuhan Liu

*School of Business, Anhui University*
*zj_teacher@126.com*

## *Abstract*

*Spamming on microblogging platforms has been a critical issue in microblog-based applications, because spamming has a significant impact on information quality and credibility. In this paper, we characterize two types of spammers on microblogging platforms, namely advertised spammers and following spammers, and then present preliminary approaches to detect these spammers. We first use a real data set to characterize the features of AS and FS in terms of various aspects such as profile, behavior, and social relationship. Specially, we introduce a new feature named duplication for FS detection, which describes the duplicated behavior of users in sharing information on microblogging platforms. We present a content-sharing graph to model the relationship between users and microblogging contents, and propose an effective algorithm to calculate the duplication feature. We run several classification methods on the characterized features to test the effectiveness of the features in AS and FS detection. The results w.r.t. precision, recall, F-measure, and ROC suggest the effectiveness of our proposed features. In particular, the duplication feature is able to improve the effectiveness of FS detection.*

***Keywords**: microblog, spammer detection*

## 1. Introduction

In recent years, social media services like microblogging have grown to become important media for communication. For example, the Sina weibo has been a dominant microblogging service provider in China. Microblog allows users to share their personal information with friends. A microblog may contain up to 140 characters of text and URL links. The posts published by a user are shared on the news feeds of the user's followers.

As microblog becomes an increasing important source for information dissemination, microblogging platforms become attractive sites for spamming, which can be used for commercial advertisement, crimes, and computer virus propagation. The Sina weibo increasingly have to deal with a wave of spammers that aim at advertising unsolicited messages instead of share information with other people. Spammers in these systems are driven by several goals, such as spreading advertise to generate sales, aggressively following, or simply compromise system reputation.

Spamming may jeopardize the trust of users on the system, furthermore, it waste users' time and energy to filter out spam messages. Maintenance a healthy environment is the critical point for the long-development of social network. In this regard, it is highly desirable to devise techniques and methods for identifying spammers and their behavior in on-line social networks.

In this paper, we aim at detecting spammers in microblog platform. We classify the spammers into two types: adverting intention spammers and following intention spammers. We then analyzed different types of users with different features. For the purpose of finding obscure following intention spammers we also propose a PageRank based duplication score calculating method. The method improves the performance of detecting duplicate posts a lot.

The rest of this paper is organized as follows. In Section 2 we summarize the related work. A number of attributes and their ability to distinguish between advertising intention spammers and following intention spammers are introduced in Section 3. We describe PageRank based duplicate calculating method in Section 4. In Section 5, we discuss the dataset, experiments and results, and finally we conclude the paper in Section 6.

## 2. Related Work

Microblog is one of the new social network platforms boosted in recent years. Generally, microblog has the following properties [1]:

(1) There are a great number of microblog users in the Internet. This number, compared with other types of social communication platforms, is rather a huge one. For example, Twitter has over 100 million users and so do other microblog platforms such as Sina Weibo and Tecent Weibo in China. Those users can generate a large amount of information every day on the microblog platform.

(2) Microblog messages contain rich social network information. This is much different from other types of information, which only present their content. On the contrast, microblog users are connected each other on the microblog platforms via following, reviewing, and reposting behaviors.

(3) Microblog messages are usually very fresh as many users will post microblog message at the first time when they see or meet some special events. Another reason for the real-time property of microblog is that nowadays smart phones are very popular for people to post microblogs in time.

Spam detection has been observed in various social network systems, including YouTube, Twitter, Facebook, and Myspace. And many spamming or spammers detection methods have been proposed in literature.

Feature extraction with machine is one of widely used method. Several studies [2, 3], proposed many features and adopted classical approaches in machine learning to identify spammers, the same as [4, 5]. A naive Bayes to detect spammy names was introduced in [6]. This paper based on the assumption that a member's on-line information reflects his or her real identity. This paper got pretty good result, however, only language features had been taken into consideration. [7] proposed to collect various features could incrementally been updated for SNS users and modified the existing Bayesian network classifiers (BNCs) to customize for SNS features. Lee *et al*. in [8] introduced social honeypots to attract spammers, statistical of the properties of these spam profiles and creating spam classifiers to actively filter out existing and new spammers. Those work got excellent result, however, those methods treat all spammers have similar behaviors, obviously, the assumption is not true in real environment. Lin [9] found three respective users, however the classifier is based on classical features that can't tackle the new escaping mechanism. [10] proposed an unsupervised method for automatic identification of spammers with the method link structure of the network in order to derive a legitimacy score for each account.

Another mainstream of anti-spam countermeasures is social network or graph based detection algorithm [11-14]. Lee *et al*. [15] considered the correlated redirect chains of URL in a number of tweets and trained a statistical classifier with features derived from correlated URLs, however, the same with [13], this work has regardless tweets that do not contains URLs. [16] proposed to build a sender and receiver based graph, and find the spammers according to the distance which cannot handle the spammers who seldom send post to others [14] proposed a behavior based profile graph ,and then community detection algorithms is applied on the graph to identify various communities. Butah found spammers on the degree of community instead of each account. Yin Zhu *et al*.[17]proposed a supervised matrix factorization method with social regularization (SMFSR) for spammer detection in social networks that exploits both social activities as well as user's social relation in an innovative and highly scalable manner. However, this work is based on the assumption that normal users perform similarly with their neighbors.

Some authors proposed new point of view or abstraction on this issues [18][19][20]. In [21], the author view spammer detection as an anomaly detection problem, and modified two stream clustering algorithms (StreamKM++ and DenStream) to facilitate spam identification. This work gave a good solution on Computational efficiency without consider the hypothesis that normal twitter users with all outliers being treated as spammers. In [22] the authors focus on the problem of identifying potential social spammers who copy pieces of information from others, and an improved locality-sensitive hashing based method is used for detecting duplicated tweets. Clearing this work only consider this particular type spammer.

Plenty of valuable and instructive work of spam microblogging had also done. [23] proposed and evaluated three types of robust features to detect spammers, which gave us a good source of inspiration. Weibo message content features specifically lexical features were proposed in [24] for Chinese character.

In our work, unlike treat all spammers equally we aimed at detecting specify spammers (advertising intention spammers and following intention spammers), different from previous works we introduced new approaches to overcome the escaping mechanism.

Before we start our investigation, it is important to have a clear definition of spammers. The define of spammers is cite from twitter which include but not limits to the follow rules :using feeds of third-party content to update and maintain accounts under the names of those third parties, mass invitations, publish or link to malicious content intended to damage or disrupt another user' browser or computer or to compromise a user's privacy. Behavior is one-sided or includes threats. In this paper we aim at find two type spammers' users: advertising intention users and following intention users.

We set the spammers into two categories: advertising intention spammers (AS) and following intention spammers (FS). We define advertising intention spammers as the users whose goal is to publish production advertisement or promote other activity. The definition of following intention spammers is the users whose goal is to following unrelated accounts randomly.

## 3. Feature Analysis

In order to find the difference between various types of users, we chose some spammers as well as legitimate users to form the sample set. Different types of spammers have different behaviors and their social roles. The following part explains the reason by analyzing users' features.

### 3.1. Data Set

In order to evaluate our approach to detect spammers on Sina weibo, we need a labeled collection of users, which pre-classified into spammers and non-spammers. To the best of our knowledge, no such collection is publicly available. In this paper we collect real data of our own.

**Seed User**. The start of the collection is from some seed users. In order to get various users we collection seed users from many sources, which include crawling from famous and verified person, ordinary person and spammers brought from merchants. In Sina Weibo there is a special kind of spammers which are usually controlled and sold as fans by on-line merchants. These spammers are controlled to follow a large number of accounts. They seldom perform traditional spamming behaviors such as posting spam messages. They just follow others to maneuver the popularity of the followed users. We bought 200 fans from one on-line merchant as spammers' samples seed.

**Crawl Strategy.** Totally we collected 19,033 used accounts and about 6,832,804 posts posted by the collected users. To obtain posts for labeling, we write a program to interact with Sina weibo's public API, we also crawl HTML pages of weibo users and posts. We could collect user profiles, social relationship (*i.e.,* social follower and following list) and user's historical posts.

**AS and FS Set.** In this paper we label users into four categories: advertising intention spammers, following intention spammers, legitimate users and user which are not belong to the above categories. We determine the user types by log in user's home page manually one by one. The principle of a user treated as AS is a user's homepage contains many URLs linked to E-commerce sites and the post have rarely interactive with followers would be labeled as AS. The principle of a user treated as FS is a user's posts totally re-tweeted or have been posted by other users, nearly all posts have no comments or interactive, the user's friends and followers ratio is extreme unbalanced would be labeled as FS. If a user is satisfy both the rules of AS and FS, it will be marked as AS priority.

### 3.2. Profile Analysis

Profile information mainly includes level and live days. Level indicate a user's active days, the more time a user spend on weibo the higher level he/she would be. The live days only indicate how many days since the account was created.

The CDF (Cumulative Distribution Function) curve of user live and level day. Level is an indication of how activity the user is in weibo, live day is the age of user. It could be found that even many spammers have long live days, they are not activity in social network. For the two types of spammer users their peak level is level-3 which represents the user's activity is about 33 days. But from the live days we could find that many users especially FS legitimate have created over 20 month, it is suppose that compared with legitimate users, and spammers have a lower activity.

### 3.3. Behavior Analysis

In this part, we analyze user's behavior which includes posts count, distribution of type of posts, distribution of posts create time, average number of mentions of their posts, number of posts shared by other users.

Depending on whether the post is retweeted or contains pictures, we divided posts into four types. They are original post without picture, original post with picture, repost post contains picture, and repost post without picture. We investigate the user's post type distribution. The results show that reposting posts that contain pictures is the most popular posting action. For advertising spammers, it is more likely to send original post with picture. On the contrary, following spammers like to send original posts with picture. The reason may be that advertising spammers need pictures to attract users clicking their URLs, and for following spammers just need to pertinent to be a legitimate user without considering whether the post is attractive. We summarize a user's post time distribution by splitting the time into 24 periods per day. It shows that all users share analogous distribution, and AS tends to send more posts than other users.

Regarding the CDF curves of average number of hashtag in a post, our study shows that the ratio of advertisement spammers is significantly higher than following spammers and legitimate user. That means AS is more likely to send posts that contain hashtag to attract other users' attention. For the CDF curves of the ratio of a user containing specify domain URLs, we consider the specific domain URLs such as *taobao* and *mogujie*. It shows that the ratio of advertisement spammers was higher than that of following spammers and legitimate users. In another word, advertise spammers are more likely to send posts that contains URL which link to E-commerce sites.

### 3.4. Social Relationship

In social relationship we mainly analyze the relationship between friends, followers and bi-followers (user following each other) number. The CDF values of relationship among following number, follower number and bi-follower number show that 80% advertise spammers spammers have a bi-follower ratio which is less than 10, and the ratio of more than 90% FOLLOWING INTENTION SPAMMERS bi-follower are less than 10, which is much higher than legitimate users. As spammers do nothing except following others, it is quite difficult for them to attract followers. In addition, the friend-follower ratio of legitimate users was higher than that of spammers, because followers of legitimate users are more likely to be their real life friends.

## 4. Content Share Graph

The spammers that bought from merchant are mainly following intention spammers. Through investigation, we found many spammers are created and controlled by the same account. According to the same content shared by characteristics, we proposed some new features for spammer detection.

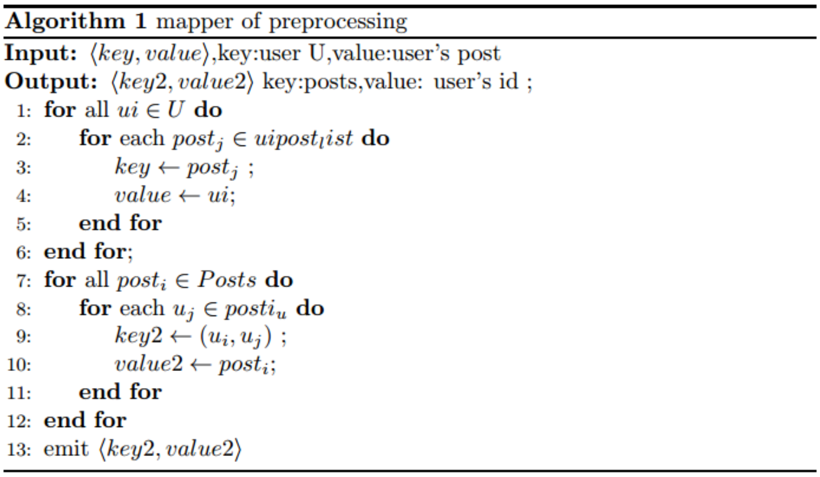
### 4.1. Posts with Same Content

Many spammers send posts frequently in order to pertinent to be legitimate users. Some spammers even use escaping mechanism by sending many original posts instead of reposting posts from other accounts. For example, if we consider text alone, it is hard to detect spamming text in many cases. But when we took the text as a query and made a search in Sina Weibo, we noticed that many original posts with the query text could be found. We can get the conclusion that there must be some correlations among those users which published posts with the same text content. Take full advantage of this relation will help us find the implied spammers.

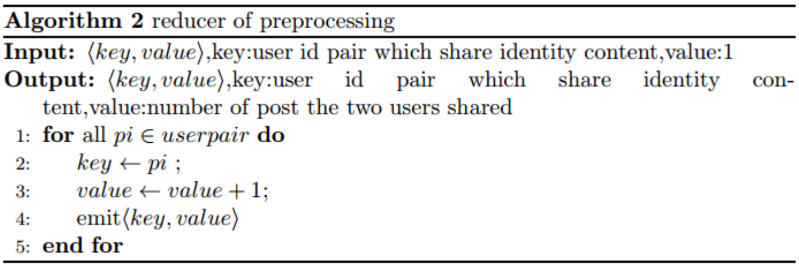
### 4.2. Build Graph

The graph is aimed at describing users sharing posts with same text content. If user$_i$ and user$_j$ used to send a post with same content it means there is an edge with weight 1 between the two users. Obviously, the graph is an undirected, symmetry weighted graph. To formulate the above ideas, we treat the graph as G=(V, E), where V is the set of users, and E is the set of shared edges in the graph.

We combined the users that share the same original posts into a set. We use map-reduce method to get the common shared graph which could be expanded to larger scale dataset. During the process we find that the same account will change the content by mixture with unrelated characteristics or missing some words, to tackle with this issue we use edit distance to calculate the similarity. The pseudo code of the implementation is shown in Algorithm 1 and 2.



```
Algorithm 1 mapper of preprocessing
Input: $\langle key, value \rangle$,key:user U,value:user's post
Output: $\langle key2, value2 \rangle$ key:posts,value: user's id ;
for all $ui \in U$ do
    for each $post_j \in uipost_list$ do
        $key \leftarrow post_j$ ;
        $value \leftarrow ui$;
    end for
end for;
for all $post_i \in Posts$ do
    for each $u_j \in posti_u$ do
        $key2 \leftarrow (u_i, u_j)$ ;
        $value2 \leftarrow post_i$;
    end for
end for
emit $\langle key2, value2 \rangle$
```



```
Algorithm 2 reducer of preprocessing
Input: $\langle key, value \rangle$,key:user id pair which share identity content,value:1
Output: $\langle key, value \rangle$,key:user id pair which share identity content,value:number of post the two users shared
for all $pi \in userpair$ do
    $key \leftarrow pi$ ;
    $value \leftarrow value + 1$;
    emit$\langle key, value \rangle$
end for
```

### 4.3. Page Rank Based Spam Detection

As the number of common shared post of each user pair had been calculated, next we will build graph and calculate the duplicates score. In this part, we introduce our PageRank-based method in calculating this score. PageRank produces a static ranking of web pages in the sense that a value is represents the importance of the pages.

For a single account, the duplicate score could be calculated by the following equation.

$$s_i = \sum_{u_i, u_j i \in pairs} (c_{ij} \times s_j) \tag{4.1}$$

Si Is the score of *user_i*, a higher score indicates the user having a higher degree of identify posts with other users. *C_ij* is the number of posts with same content shared by *user_i* and *user_j*. Naturally, we define *S* as a user's score, and the score equation is:

$$S = \sum (C \times S) \tag{4.2}$$

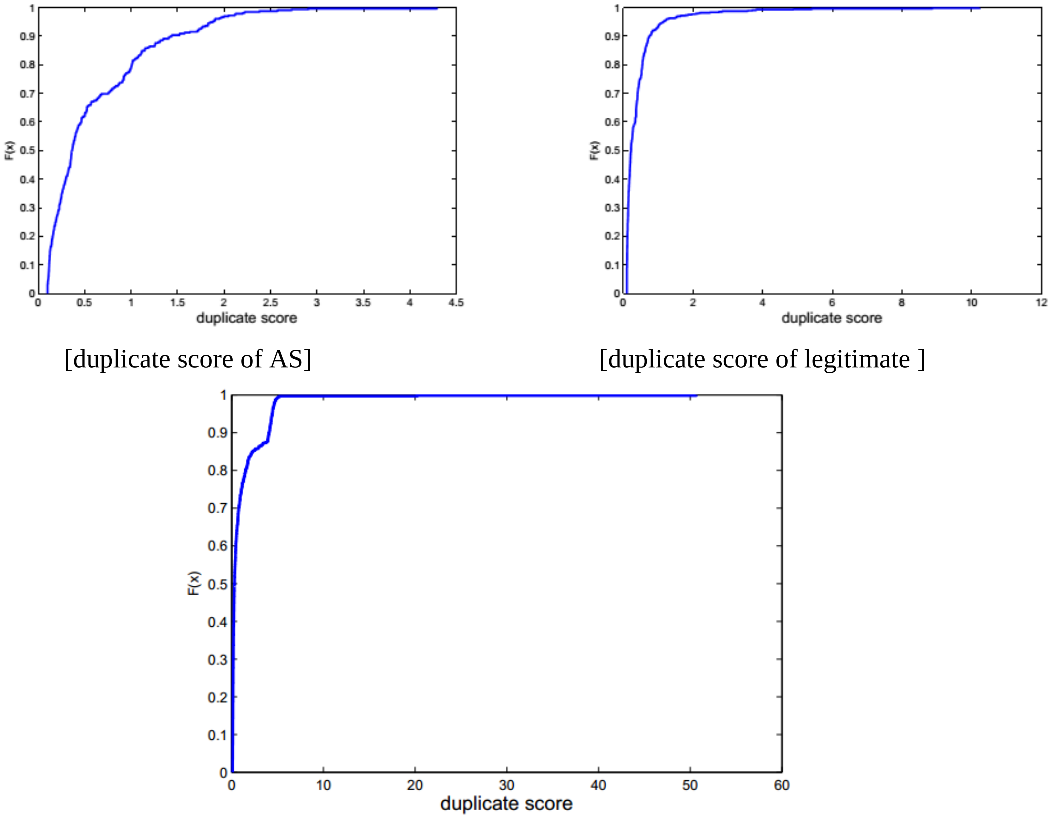
S is the edges matrix, C is the number of identified posts shared by *user_i* and *user_j*.

(1) Edge between two users is an implication of post with same content shared by each other. Thus, the more links that a user shared with other user, the more higher the score is.

(2) Users also have their own scores. A user with a higher prestige score linked to is more influential than a user with lower scores.

However sometimes we can't find all users sharing posts with the same content. So just like PageRank, we thought that each user would have a probability to have post with the same content with other users. We could adjust the probability by setting different parameters. The final equation is

$$S = E \times p + (1 - p) \times \sum (C \times S) \qquad (4.3)$$



[duplicate score of AS]

[duplicate score of legitimate ]

[Duplicate score of FS]

**Figure 1. Duplicate Score**

Figure 1 shows the duplicate score distribution of AS, FS and legitimate users. It indicates that the FS score distribution ranges up to 60, which is much larger than AS's score 4.5 and legitimate users' score 12.

In summary, a user is more likely to be spammers if he/she shares more posts with other potential spammers. As the follow graph shows. If a user shares more common posts with other users the more likely he would be a spammer, and the more posts shared with a potential spammer the more like he/she is a spammer as well. So we could calculate a user spammers score according to the content share graph.

## 5. Experiments

### 5.1. Experimental Setting

In order to evaluate our approach to detect spammers on Sina weibo, we need a labeled collection of users, which pre-classified into spammers and non-spammers. To the best of our knowledge, no such collection is publicly available. In this paper we collect real data of our own.

The start of the collection is from some seed users. In order to get various users we collection seed users from many sources, which include crawling from famous and verified person, ordinary person and spammers brought from merchants. In Sina Weibo 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 accounts. They seldom perform traditional spamming behaviors such as posting spam messages. They just follow others to maneuver the popularity of the followed users. We bought 200 fans from one on-line merchant as spammer samples.

Totally we collected 19,033 used accounts and about 6,832,804 posts posted by the collected users. To obtain tweets for labeling, we write a program to interact with Sina weibo's public API, we also crawl HTML pages of weibo users and posts. With the help of Sina API, we could collect user profiles and their lasted posts. Through python crawler we could download a user's history posts as well as their social follower and following list.

Once the data was gathered, our next task is to develop a collection of posts labeled into spam and non-spam categories which could be used to train our classifier. In this paper we label users into four categories: advertisement intention spammers, following intention spammers, legitimate users and users which are not belong to the above categories.

### 5.2. Results

We designed a spammer classifier which is based on the features that were analyzed before. Here we use Weka as the classification tool, and we use two different models with different features to implement the classification.

In this paper we defined two types of spammers, and we use two-classifiers. Each time we select one type of spammers as the positive samples, another spammer combination with legitimate user as the negative sample. For each kind of spamming behavior, we tested many algorithms provided by Weka with 10-fold cross-validation. Finally we choose the best model for each spammer according to their performance. Table 1 shows the train set.

**Table 1. Train Data**

| spammers | positive | negative |
|---|---|---|
| AS | 586 | 1114 |
| FS | 637 | 1063 |

In this paper, we use Precision, Recall, ROC (receiver operating characteristic curve) and F-measure as the metrics. The precision and recall is reference to the positive samples.

Table 2 shows the main features that used in classifier. The features divided into four categories, and the duplicate only used for FS detecting.

**Table 2. Features**

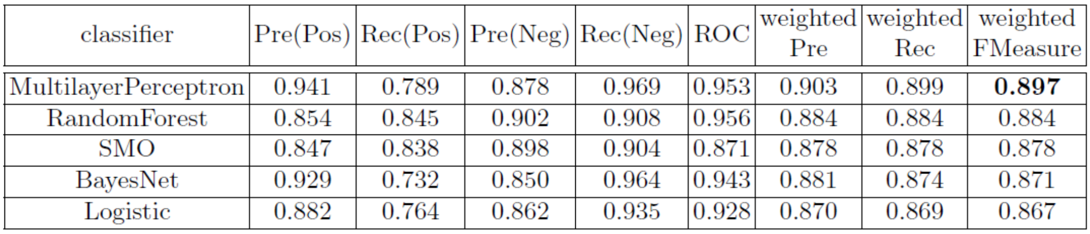
| categories | Features | AS | FS |
|---|---|---|---|
| profile | level | * | * |
| | live days | * | * |
| | statuses Count | * | * |
| | favorite Count | * | * |
| | average post | * | * |
| behavior | mention num | * | * |
| | average mentions | * | * |
| | hashtag number | * | * |
| | four types | * | * |
| | top source ratio | * | * |
| | time during | | * |
| | average interval | | * |
| | max comments | * | |
| | comments_ratio_0 | * | |
| | average comments number | * | * |
| | per average comments | * | * |
| | retweet ratio | | * |
| | retweet_ratio_0 | * | |
| | specify domain | * | |
| social network | friendsCount/followersCount | * | * |
| | biFollowers/followersCount | * | * |
| | follower followers | * | * |
| duplicate | same post number | | * |
| | same post person | | * |
| | duplicate score | | * |

For the advertising intention spammers, we choose the features marked by a star symbol in the AS column in Table 2. For following-intention spammers, we choose the features marked by a star symbol in the FS column in Table 2. The results are shown in Table 3 and 4.

**Table 3. AS Classifier Performance**

| classifier | Pre(Pos) | Rec(Pos) | Pre(Neg) | Rec(Neg) | ROC | weighted Pre | weighted Rec | weighted FMeasure |
|---|---|---|---|---|---|---|---|---|
| MultilayePerceptron | 0.906 | 0.854 | 0.926 | 0.950 | 0.950 | 0.919 | 0.919 | **0.918** |
| RandomForest | 0.915 | 0.827 | 0.914 | 0.960 | 0.954 | 0.914 | 0.914 | 0.913 |
| SimpleLogistic | 0.888 | 0.854 | 0.924 | 0.943 | 0.960 | 0.912 | 0.912 | 0.912 |
| SMO | 0.877 | 0.850 | 0.922 | 0.937 | 0.894 | 0.907 | 0.907 | 0.907 |
| NaiveBayes | 0.855 | 0.816 | 0.905 | 0.927 | 0.941 | 0.888 | 0.889 | 0.888 |

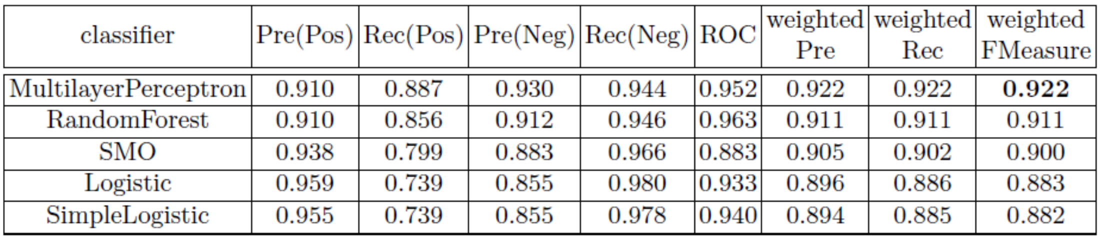
**Table 4. FS Classifier Performance**



| classifier | Pre(Pos) | Rec(Pos) | Pre(Neg) | Rec(Neg) | ROC | weighted Pre | weighted Rec | weighted FMeasure |
|---|---|---|---|---|---|---|---|---|
| MultilayerPerceptron | 0.941 | 0.789 | 0.878 | 0.969 | 0.953 | 0.903 | 0.899 | **0.897** |
| RandomForest | 0.854 | 0.845 | 0.902 | 0.908 | 0.956 | 0.884 | 0.884 | 0.884 |
| SMO | 0.847 | 0.838 | 0.898 | 0.904 | 0.871 | 0.878 | 0.878 | 0.878 |
| BayesNet | 0.929 | 0.732 | 0.850 | 0.964 | 0.943 | 0.881 | 0.874 | 0.871 |
| Logistic | 0.882 | 0.764 | 0.862 | 0.935 | 0.928 | 0.870 | 0.869 | 0.867 |

As discussed in Section 4, we consider to compute the duplicate score of microblogs to improve the performance of spammer detection. The results with duplicate scores are shown in Table 5. By comparing the results in Table 4 and Table 5, we found that the duplicate-score featureis helpful in enhancing precision and recall. For example, the precision of randomforest classifier had been improved from 0.884 to 0.911.

**Table 5. FS Classifier Performance with Duplicate Scores**



| classifier | Pre(Pos) | Rec(Pos) | Pre(Neg) | Rec(Neg) | ROC | weighted Pre | weighted Rec | weighted FMeasure |
|---|---|---|---|---|---|---|---|---|
| MultilayerPerceptron | 0.910 | 0.887 | 0.930 | 0.944 | 0.952 | 0.922 | 0.922 | **0.922** |
| RandomForest | 0.910 | 0.856 | 0.912 | 0.946 | 0.963 | 0.911 | 0.911 | 0.911 |
| SMO | 0.938 | 0.799 | 0.883 | 0.966 | 0.883 | 0.905 | 0.902 | 0.900 |
| Logistic | 0.959 | 0.739 | 0.855 | 0.980 | 0.933 | 0.896 | 0.886 | 0.883 |
| SimpleLogistic | 0.955 | 0.739 | 0.855 | 0.978 | 0.940 | 0.894 | 0.885 | 0.882 |

## 6. Conclusion

In this paper, we analyze different types of spammers' behaviors in Sina Weibo platform. We processed the database associated with these spammers and found two representative spamming behaviors: advertising intention spammers and following intention spammers. We analyze various features and compared the behaviors of spammers and legitimate users as well as two types of spammers and found that spamming behaviors and legitimate microblogging behaviors have distinct characteristics. By analyzing the potential relationship among following intention users we introduce duplication score. We test the performances using real data samples and it is demonstrated that our system is effective in detecting the above mentioned spamming behaviors and identifying spammers.

There is much room for improvement in performance of FS. Social relationships could be taken into consideration. A more comprehensive and abundance dataset is needed. More flexible, robust, and low cost system is needed to be designed to detect spammers, which is considered as our future work.

## Acknowledgments

This paper is partially supported by the National Science Foundation of China (No. 71273010) and the Doctor Start-up Fund of Anhui University.

## References

[1] L. Zhang, Z. Zhang and P. Jin, "Predicting Retweet Action over Microblog Dataset via Classify Techniques", The 13th International Conference on Web Information System Engineering (WISE'12), Challenge track, X.S. Wang (Eds.): WISE 2012, LNCS 7651, Springer, **(2012)**, pp. 771—776.

[2] C. Yang, R. C. Harkreader and G. Gu, "Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers", in Proceedings of the 14th International Conference on Recent Advances in Intrusion Detection, RAID'11, **(2011)**, pp. 318-337.

[3] K. Thomas, C. Grier, D. Song and V. Paxson, "Suspended accounts in retrospect: An analysis of twitter spam", in Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC '11, **(2011)**, pp. 243-258.

[4] Z. Chu, I. Widjaja and H. Wang, "Detecting social spam campaigns on twitter", in Applied Cryptography and Network Security - 10th International Conference, ACNS 2012, **(2012)**, pp. 455-472.

[5] Y. Zhou, K. Chen, L. Song, X. Yang and J. He, "Feature analysis of spammers in social networks with active honeypots: A case study of Chinese microblogging networks", in ASONAM, **(2012)**, pp. 728-729.

[6] D. M. Freeman, "Using naive bayes to detect spammy names in social networks", in AISec'13, Proceedings of the 2013 ACM Workshop on Artificial Intelligence and Security, Co-located with CCS 2013, Berlin, Germany, **(2013)**, pp. 3-12.

[7] D.-H. Park, E.-A. Cho and B.-W. On, "Social spam discovery using Bayesian network classifiers based on feature extractions", in 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2013 / 11th IEEE International Symposium on Parallel and Distributed Processing with Applications, ISPA-13 / 12th IEEE International Conference on Ubiquitous Computing and Communications, IUCC-2013, Melbourne, Australia, **(2013)**, pp. 1808-1811.

[8] K. Lee, J. Caverlee and S. Webb, "Uncovering social spammers: social honeypots+ machine learning", in Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, **(2010)**, pp. 435-442.

[9] C. Lin, J. He, Y. Zhou, X. Yang, K. Chen and L. Song, "Analysis and identification of spamming behaviors in sina weibo microblog", in SNAKDD, **(2013)**, pp. 5.

[10] M. Bouguessa, "An unsupervised approach for identifying spammers in social networks", in IEEE 23rd International Conference on Tools with Artificial Intelligence, ICTAI 2011, Boca Raton, FL, USA, **(2011)**, pp.832-840.

[11] C. Yang, R. Harkreader, J. Zhang, S. Shin and G. Gu, "Analyzing spammers' social networks for fun and profit: a case study of cyber-criminal ecosystem on twitter", in Proceedings of the 21st international conference on World Wide Web, **(2012)**, pp. 71-80.

[12] X. Zhang, S. Zhu and W. Liang, "Detecting spam and promoting campaigns in the twitter social network", in Proceedings of the 2012 IEEE 12th International Conference on Data Mining. IEEE Computer Society, **(2012)**, pp.1194-1199.

[13] E. Tan, L. Guo, S. Chen, X. Zhang and Y. Zhao, "Unik: unsupervised social network spam detection", in Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM, **(2013)**, pp. 479-488.

[14] F. Ahmed and M. Abulaish, "Identification of Sybil communities generating context-aware spam on online social networks", in Web Technologies and Applications - 15th Asia-Pacific Web Conference, APWeb 2013, Sydney, Australia, **(2013)**, pp. 268-279.

[15] S. Lee and J. Kim, "Warningbird: Detecting suspicious urls in twitter stream", in NDSS, **(2012)**.

[16] J. Song, S. Lee, and J. Kim, Spam filtering in twitter using sender-receiver relationship", in RAID, **(2011)**, pp. 301-317.

[17] Y. Zhu, X. Wang, E. Zhong, N. N. Liu, H. Li and Q. Yang, "Discovering spammers in social networks", in AAAI, **(2012)**.

[18] E. Tan, L. Guo, S. Chen, X. Zhang and Y. E. Zhao, "Spammer behavior analysis and detection in user generated content on social networks", in 2012 IEEE 32nd International Conference on Distributed Computing Systems, Macau, China, **(2012)**, pp. 305-314.

[19] K. S. Xu, M. Kliger, Y. Chen, P. J. Woolf and A. O. Hero, "Revealing social networks of spammers through spectral clustering", in Proceedings of IEEE International Conference on Communications, ICC 2009, Dresden, Germany, **(2009)**, pp. 1-6.

[20] V. Sridharan, V. Shankar and M. Gupta, "Twitter games: how successful spammers pick targets", in 28th Annual Computer Security Applications Conference, ACSAC 2012, Orlando, FL, USA, **(2012)**, pp. 389-398.

[21] Z. Miller, B. Dickinson, W. Deitrick, W. Hu and A. H. Wang, "Twitter spammer detection using data stream clustering", Information Sciences, vol. 260, **(2014)**, pp.64-73

[22] Q. Zhang, H. Ma, W. Qian and A. Zhou, "Duplicate detection for identifying social spam in microblogs", in IEEE International Congress on Big Data, BigData Congress 2013, **(2013)**, pp. 141-148.

[23] M. R. Karim and S. Zilles, "Robust features for detecting evasive spammers in twitter", in Advances in Artificial Intelligence - 27th Canadian Conference on Artificial Intelligence, Canadian AI 2014, Montreal, QC, Canada, **(2014)**, pp. 295-300.

[24] L. Liu and K. Jia, “Detecting spam in Chinese microblogs - A study on sina weibo”, in Eighth International Conference on Computational Intelligence and Security, CIS 2012, Guangzhou, China, **(2012)**, pp. 578-581.

[25] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank citation ranking: Bringing order to the web”, Technical Report 1999-66, Stanford InfoLab, **(1999).**