

A Study of Effective Features for Detecting Long-surviving Twitter Spam Accounts

Po-Ching Lin*, Po-Min Huang*

*Department of Computer Science and Information Engineering

National Chung Cheng University, Chiayi, Taiwan 62102

Email: {pclin,hpm99m}@cs.ccu.edu.tw

Abstract—As social networking websites get popular in recent years, the spam accounts on them also increase rapidly. A number of features to detect spam accounts have been proposed in prior studies. In this work, we evaluate the common features to see how effective they are to detect Twitter spam accounts or not. We used the Twitter APIs to collect 26,758 public accounts with 508,403 tweets intermittently over the period from September 2011 to March 2012. Although Twitter has its own method to detect and suspend spam accounts, many long-surviving spam accounts are still on it. We selected 816 out of the long-surviving accounts to persistently observe their activities and recorded their lifespans. We find that some features are not so effective for the detection, and thus select two features, the *URL rate* and the *interaction rate*, to be the detection features in this work. With classification by the J4.8 algorithm, the precision of the detection is estimated to be between 0.829 and 0.885, and the recall is between 0.987 and 0.999. We also present several additional techniques to further improve the detection accuracy.

keyword: Twitter spam, detection, effective features, URL rate, interaction rate.

I. INTRODUCTION

Social networking websites such as Facebook, Twitter, LinkedIn and MySpace are getting popular in recent years. Over one billion persons use them as micro-blogs to record their daily life and share with their friends [1]. As more people are involved than ever, such websites also become a platform for spammers to distribute spam messages including malicious links, advertisements, pornographic content and so on [2]. The spam messages waste the storage space of the websites and annoy ordinary users. Clicking an embedded malicious link can even infect a user's system [3], [4]. Spam on social networking websites, like email spam, has become a problem. Therefore, developing an effective system to classify spam accounts on social networking websites has attracted much attention in recent years.

Most studies detect spammers on social networking websites according to the characteristics in the user profiles with machine-learning methods. The characteristics include number of followers and followings, ratio of followers over followings, tweets per day, hashtags per tweet, and so on [6], [7]. Some of the studies describe the relationship between users in graphs, and use a graph algorithm to extract the features for detection [8], [9]. Moreover, the study in [10] detects automatic activities by analyzing the creation time of the messages. The studies in [11], [12] create honey-profile accounts to be followed by spammers, and then analyze the behavior of the spammers.

Given so many features for detecting spammers on social networking websites, we wonder how effective they are for detecting spammers, i.e., whether or not the features are *significantly different* for normal users and spammers. It is common that dozens of millions of users register at a popular website. The websites may have been equipped with the resources to accommodate the huge number of users, but extracting too many yet probably ineffective features for the detection will bring extra loads to the websites and result in a scalability problem. Therefore, we evaluate the effectiveness of the features in this work, in an attempt to find a lightweight yet effective method to quickly detect and filter out spammers.

Like many prior studies, this work also aims at detecting the spammers on Twitter. Twitter, which was reported to have over 200 million users [1], has its own way to detect and suspend spam accounts [14]. Although Twitter does not clearly describe its detection algorithm, we speculate that it knows the methods presented in the literature, and is likely to have adopted the state-of-the-art methods. However, there are still many long-surviving spam accounts. We are also interested in studying why the spam accounts can survive so long.

We select the *URL rate* and the *interaction rate* as the main features based on the evaluation, and verify their effectiveness with 26,758 public accounts in total collected by the Twitter APIs [5]. The experiments show that the two features are simple yet effective for detecting spammers. Some spam accounts can evade Twitter's detection with a specific strategy, and survive for over one month. Therefore, we will also discuss the possibility of evading the detection in this work.

The main contributions of this paper are as follows:

- 1) We evaluate the effectiveness of existing features for detecting spammers on social networking websites.
- 2) We use two simple yet effective features (i.e., the URL rate and the interaction rate) to classify the Twitter accounts. The experiments based on 26,758 Twitter accounts with 508,403 tweets show that the classification has high recall and precision up to around 0.99 and 0.86.
- 3) We study the long-surviving spam accounts on Twitter and their lifespans. The study is useful to learn how the accounts can evade the detection.

The remainder of the paper is organized as follows. The Twitter basics and the prior detection methods are reviewed in Section II. We describe collecting the Twitter accounts, evaluating existing detection features, and selecting the effective

tive features in Section III. We explain how the accounts are classified, study the cases of false positives and negatives, and evaluate the accuracy of the classifier in Section IV. Section V concludes this work.

II. BACKGROUND AND RELATED WORK

In this section, we will introduce the main Twitter functions used in this work and the detection methods in the literature.

A. Twitter Basics

Twitter provides the functions summarized below for users to interact with each other.

- *Tweets*: Tweets are the messages posted by the users on Twitter. Every tweet has a 140-character restriction, so users tend to use short URLs when they post URLs.
- *Followers & Followings*: If a user A follows a user B, then A becomes B's *follower*, and B is the *following* of A. A user can receive the tweets from his or her followings.
- *Mention*: A *mention* is a Twitter update containing an *@username* in the tweet body, and a reply is also considered a mention. If A mentions B in a tweet, B can see the tweet, even though B is not A's follower. Therefore, spammers can use mention to send their spam messages, even though they have rare followers.
- *Retweet*: A *retweet* is someone else's tweet that one chooses to share with all of the followers.

Twitter provides official APIs for programmers to retrieve or modify data. We use two API functions, `statuses/public_timeline` and `statuses/user_timeline`, to collect the account information. The `public_timeline` function will return 20 most recent statuses from the public timeline, including the user identifier, screen name, tweet creation time, tweet identifier, tweet content, and so on. The `user_timeline` function will return 20 most recent statuses posted by an authenticated user, and the return value is similar to that of the `public_timeline` function. The two functions will reveal only the information of public accounts, but not that of protected accounts. The Twitter APIs allow a client to make only a limited number of 150 calls in an hour [15].

B. Related Work

The works in [6] and [7] used the characteristics in the user profiles as the detection features. The work in [6] analyzed more than 500,000 accounts by observing the number of followers and followings, ratio of followers over followings, frequency of tweets, account registration date, and tweeting device make-up. The work then classified the users into three main categories: human, bot, and cyborg. The work in [7] collected around 80 million accounts, and used the content attributes and the behavior attributes for classification. The content attributes include the text of tweets, such as number of URLs per words, number of words in each tweet, number of hashtags in each tweet, number of users mentioned in each tweet, and so on (totally 39 attributes). The behavior attributes

include the user behavior, such as number of followers, number of followings, fraction of followers per followings, number of tweets, and so on (totally 23 attributes). Both works used the SVM classifier for classification.

The works in [8] and [9] combined the graph algorithms with part of the aforementioned features for the detection. They viewed each Twitter account as a node and the relations with the followings as directed edges. The authors assumed a spammer usually has many followings but fewer followers (causing the ratio of bi-directional edges low), and his/her followings rarely follow each other. They developed graph algorithms to detect the relationships, since the edges for spam accounts will differ from those for legitimate user accounts.

The works in [11] and [12] are based on a honeypot approach. They created honey-profiles on Facebook, MySpace, and Twitter, and then waited spammers to contact the accounts for analyzing their behavior. The authors in [11] developed six features to detect whether an account is spammer or not, i.e., FF ratio (ratio of followers over followings), URL ratio, message similarity, choices of friends, messages sent, and number of friends. The authors in [12] considered four features to analyze the data, i.e., the demographics, contributed content, activities, and connections of users.

The work in [10] evaluated 106,573 distinct accounts, and detected accounts with automated activities by analyzing the tweet timestamps. The study assumed that the timestamps for humans should be randomly drawn from a uniform distribution, while those for an automated account should be drawn from a non-uniform distribution. The work in [13] identified over 1.1 million accounts suspended by Twitter, and examined their behavior. The authors categorized spam accounts into five sets according to their activities. The spammers in each set have their own way to post spam messages, and some of them use Twitter as their spam-as-a-service market to make money.

The aforementioned studies used dozens of features to detect spam accounts, but some features do not have a notable effect on detecting spam accounts, according to our discussion in Section III-B. Therefore, we will review the features in this work, and find the effective ones for efficient classification.

III. SPAM DETECTION ON TWITTER

In this section, we first describe data collection in this work, and review the effectiveness of the features presented in prior studies. We then select two robust features and discuss their effectiveness for detecting Twitter spam accounts.

A. Data Collection

Twitter provides various API functions for programmers to crawl specific data [5] related to user accounts. Among the API functions, we used the `publicTimeline` and `userTimeline` functions (see Section II-A) to intermittently collect 26,758 accounts with 508,403 tweets over the period from September 2011 to March 2012. We first called the former function to get the 20 most recent statuses, extracted the user identifiers from the statuses, and then used the identifiers

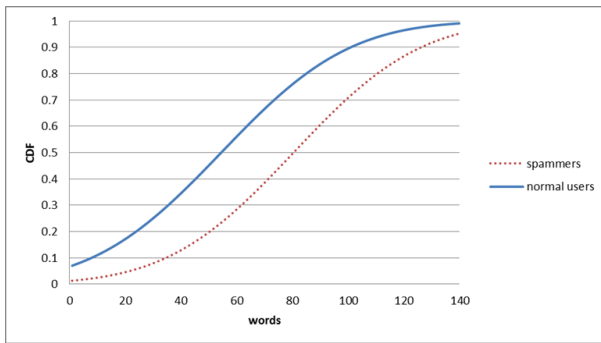


Fig. 1. The CDFs of words in tweets from both the spammers and the normal users.

to get each user's recent status by the `userTimeline` function. We stored each user's identifier, screen name, following & follower count, tweet content, the tools for tweeting, account creation time, and tweet creation time in the database for later analysis.

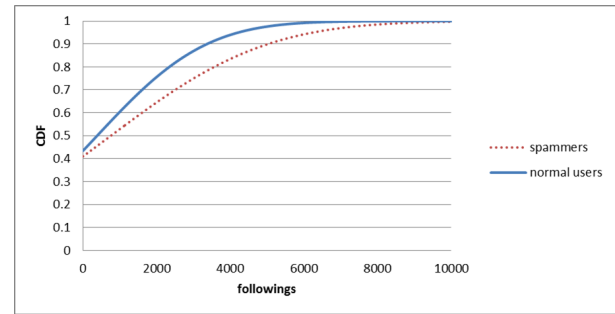
Although Twitter strives hard to protect its users from spam and abuse [14], and is supposed to adopt the state-of-the-art methods to detect the spammers, numerous long-surviving spam accounts are still on it. To find the reasons of their long survival, we manually selected 816 spam accounts from the database to persistently collect their tweets for six months, and recorded their lifespans in days. The collection process is achieved by calling the Twitter's OAuth API [16] through the PHP library. The collected data are also stored in the database.

B. Feature Analysis

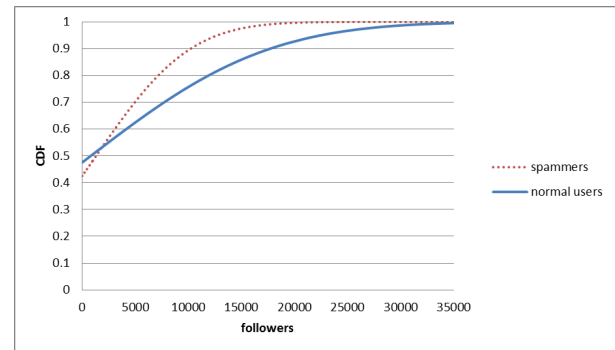
We select and compare the common features in prior studies for both normal users and spammers, and then study how effective the features are in detecting spam accounts. The comparisons with graph-based features are left to the future work, since they take more time to implement. The total accounts in this work involve 23,322 normal user accounts and 3,436 spam accounts collected by the method described in Section III-A.

Figure 1 presents the CDFs of the words in tweets from the spammers and the normal users. The spammers have an average of 80.45 words per tweet, while the normal users have an average of 54.49 words per tweet. The spammers have more words in the tweets than the normal users on average, since their tweets may detail the product information in the advertisement or post more URL links than the normal users. Nevertheless, the distinction of this feature between both types of users is insignificant. Neither the normal users nor the spammers tend to post short or long tweets. Moreover, spammers can easily evade this feature by rewording the tweets with short terms.

The numbers of followings and followers for both types of users are compared in Figure 2(a) and Figure 2(b). The spammers have an average of 757 followings and 1317.14 followers, while the normal users have an average of 384.04 followings and 819.17 followers. Note that over 40% of the



(a)



(b)

Fig. 2. The CDFs of the numbers of followings and followers.

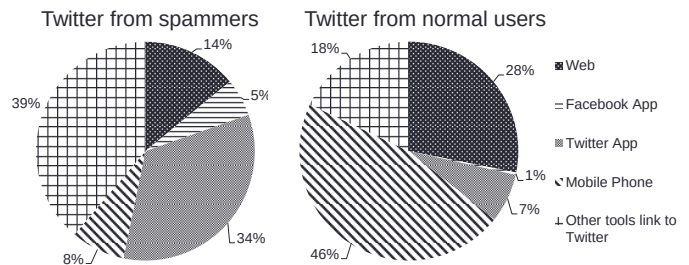


Fig. 3. The percentages of the Tweeting tools.

accounts have fewer than 100 followers or followings, so the average numbers are not so high as they look. As presented in Figure 2, the numbers for both types of users are close, and the features are unable to distinguish both types well.

Figure 3 presents the types of tweeting tools of the normal users and the spammers. Web means directly posting the tweets on the Twitter homepage. Facebook App means sharing the Facebook content to the Twitter account, and Twitter App means feeding a user's blog to Twitter. Other tools denote miscellaneous tools other than the preceding ones for accessing Twitter. It is clear to see that the normal users tend to post the tweets by mobile phones and web, but the spammers tend to post the tweets by the Twitter App and other tools instead. Nevertheless, since none of the tools dominate the usage and are significantly distinct in either type of users, using the feature is insufficient for the classification.

We also studied the average of (1) the time gaps between tweets and (2) the number of tweets per day between both

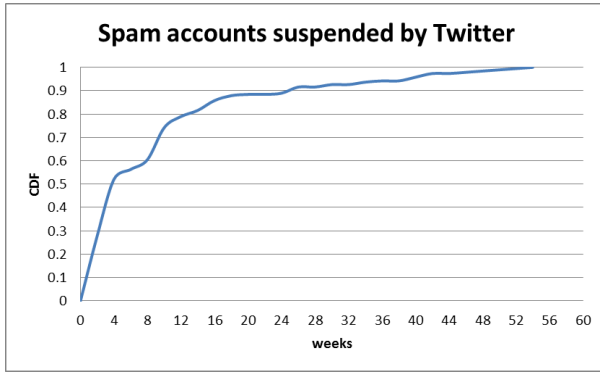


Fig. 4. The CDF of the days before the long-surviving spam accounts were suspended by Twitter.

types of users. The former is 16.49 and 7.37 for the spammers and the normal users, and the latter is 20.51 and 11.07 for both types of users. The spammers have higher values than the normal users, but the two features are easy to evade, since the spammers can pretend to be normal users by reducing the number of tweets or the time gaps.

For the average ratios of (1) the number of hashtags over that of tweets and (2) the number of followers over that of followings between both types of users, the former ratio is 0.186 and 0.117 for the spammers and the normal users, and the latter is 1.74 and 2.133 for both types of users. The ratios do not differ obviously for both types of users, so the two features are not effective for the classification.

Combining the above features for the detection may help to increase the accuracy, but as there are usually hundreds of millions of accounts in a popular social networking website like Twitter, involving more features means increasing the system load and space to detect spam accounts. Therefore, we would like to select the most effective features in this work for distinguishing between both types of users. We will discuss this issue later in Section III-D.

C. Observations on Long-surviving Spam Accounts

We observed that there are still many long-surviving spam accounts, which have not been suspended by Twitter yet, or were not suspended until a long time after their first appearance. We therefore selected 816 such accounts to persistently observe their activities, and recorded their lifespans in days. Figure 4 presents the CDF of the surviving days of the suspended spam accounts from their birth to suspension. Although about 28% accounts suspended by Twitter in fewer than two weeks, over 47% accounts can survive for over one month till being suspended by Twitter. Moreover, we also notice that over 50% of such accounts can survive for longer than one year.

Twitter is effective to quickly suspend an account that posts many malicious links or a large number of tweets in a short period of time, or that randomly mentions other accounts to force them to receive its spam messages, but the long-surviving spam accounts have their own strategies to evade Twitter's

detection. For example, an account can repeatedly post three tweets at a time, wait for a couple of minutes, and then go on. This account can survive for several months with this simple strategy. In summary, we find that if an account has many followers, does not post many tweets in short period of time, and does not post malicious links, then they can survive for a long time.

D. Selecting Effective Features

We selected two effective features from prior studies, and studied the difference of the features between the spammers and the normal users. The detection method based on the features will be described in the end of this subsection.

1) *URL Rate*: Posting URL links is common for spammers to spread commercial, harmful or pornographic messages. We crawled each account's recent 20 tweets by the `userTimeline` function, and then checked whether the tweets contain URL links or not. Let $tweet_{URL}$ denote a tweet with a URL. The URL rate is calculated by

$$URL\ rate = \frac{\text{number of } tweets_{URL}}{\text{total number of tweets}}. \quad (1)$$

We observed that the average URL rate for the spammers is 95%, and that for the normal users is only 7% in the collected tweets. It is clear to see that spammers have a much larger proportion of tweets with URL links than the normal users. It is hard for spammers to evade the detection with this feature, unless they just tweet the spam messages without any URL links, but that will seriously restrict the amount of information that spammers can carry in the tweets.

2) *Interaction Rate*: We use the interaction rate as an effective feature because normal user accounts usually have interactions with their friends, including followings and followers. In contrast, most spam accounts do not have such interactions, and usually just post URL links.

Some spam accounts may randomly mention other accounts to force them to receive spam messages, even though those accounts are not their friends. We examined the `in_reply_to_user_id` field in the return value from the `userTimeline` function to check whether the user of a tweet reply is a friend of the observed account or not. If the user is in the friend list, then a tweet receives an interaction. Some normal users can use the `retweet` function to interact with their friends. In this case, the `in_reply_to_user_id` field is null, and the interaction rate will be underestimated. Therefore, when calculating the interaction rate, we also included the `retweet` case to eliminate the underestimation. Let $tweet_{interaction}$ denote a tweet interacting with friends. The interaction rate is defined to be

$$Interaction\ rate = \frac{\text{number of } tweets_{interaction}}{\text{total number of tweets}}. \quad (2)$$

Compared with the normal users, only around 1% of the tweets from the spam accounts interact with their friends in our observation. These few tweets are primarily replies to the customer's questions. In contrast, around 29% of the

tweets interact with the friends. Around 8% of the tweets with interactions are from the *retweet* function, and should not be neglected. The distinction of the interaction rate between both types of users is significant. In summary, it is effective to classify an account with a low URL rate and a high interaction rate as a normal user account.

IV. CLASSIFICATION AND EVALUATION

In this section, we first explain the detection method, and evaluate the accuracy of this method. We then discuss an ambiguous case of the accounts with zero URL link & zero interaction, and finally discuss possible evasion strategies.

A. Training Set & Test Set

Like prior studies, we use supervised learning to classify the Twitter accounts. We selected 400 labeled accounts collected in Section III-A as the training set, in which 200 were normal users and 200 were spammers, and calculated the URL rate and the interaction rate of each account. The accounts in the training set were manually inspected to ensure that no accounts were misclassified. The J4.8 algorithm implemented in Weka, a machine learning tool [17], was selected to perform the training process and generate the decision tree.

We used the rest of the collected accounts as the test set, and fed them into Weka for classification. Each account was classified by its recent 20 tweets derived from just one API call, so the classification is very efficient. In some cases, however, we have to keep tracking an account to collect more tweets for further analysis (see Section IV-C). In the tracking, we regularly called the `userTimeline` function to fetch the recent tweets, checked whether their identifiers had been in the database or not, and then stored new ones.

B. Performance of Classification

Manually verifying whether each account in the test set (more than 20,000) is a spammer or not for the ground truth is rather time consuming. We therefore randomly sampled 1,000 accounts for verification from the test set, divided them into 20 groups, each of which had 50 accounts, and calculated the precision and recall in each group. The 95% confidence intervals of the precision and the recall can be calculated with the equation: $[\bar{x} - z_{0.025}(\frac{\sigma}{\sqrt{n}}), \bar{x} + z_{0.025}(\frac{\sigma}{\sqrt{n}})]$, where $\bar{x}$ and σ are the mean and the variance of the precision or the recall, n is the number of groups, and z is the standard normal distribution $N(0, 1)$. The precision of the classifier is between 0.829 and 0.885, and the recall is between 0.987 and 0.999. The recall is rather high, meaning that classification with the two simple features alone can detect spammers effectively.

The precision is lower than the recall because some normal users behave rather like spammers. We categorize the cases of false positives into four major types, as presented in Figure 5. They are discussed in the order of their significance as follows:

- The first (48%) results from tweeting photo or video links without interactions with the friends in the recent 20 tweets. The false positives can be reduced by extracting

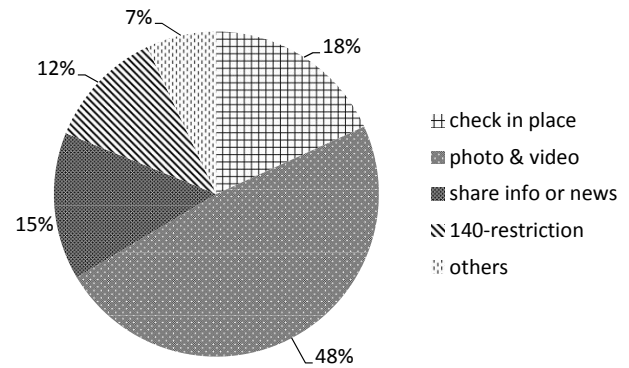


Fig. 5. Categories of false positives.

more recent tweets from the suspicious accounts to see whether they are really spammers. The Twitter administration can easily track more than 20 tweets of the accounts in the history, so it is not a problem.

- The second (18%) results from continuously posting check-in information. These accounts like to share their locations with their friends, and generate the URL links that lead to false positives. This problem can be solved by checking the API return value, `geo` column, which contains the latitude and longitude if the information exists. If the column contains the information, then we can ignore the tweets to reduce false positives.
- The third (15%) results from sharing news or specific information with the friends. These accounts can post many URL links with few interactions with their friends. Like the first type, the accounts can be tracked with more recent tweets to reduce the false positives.
- The fourth (12%) results from using tools such as `writelonger` and `twitterfeed` to evade the 140-word restriction on the tweet length. Once the length is over 140 characters, Twitter will generate a link to direct the reader to see the full tweet. This problem is to be solved in the future work.

C. Tweets with Zero URL Link and Zero Interaction

Some accounts have their tweets without any URL links and do not interact with their friends. The ambiguous accounts complicate the detection, since some of them are truly spammers, but the others are just normal users. The accounts always share a specific piece of information (e.g., weather news, daily horoscope, local traffic flow, etc) with others, or tweet their daily trivia, but nobody else interacts with them.

Due to the ambiguity of the accounts, we leverage the thought in [10] for the detection. That is, if an account posts the tweets periodically, it is likely to be from an automatic program, and is considered a spammer; otherwise, it is likely to be a normal user. We found some spammers may pause posting tweets for a period of time, say several hours, and resume the posting, so we add the *change detection* method to detect the abrupt posting time. The steps in the detection method are as follows :

TABLE I
CLASSIFICATION RULES.

Statistic of χ^2 test \ variance	small	large
small	spammer	tracking list
large	impossible	normal user

- 1) Calculate the time gaps between the consecutive tweets.
- 2) Apply the CUSUM algorithm [18] to find the abrupt tweeting time, and ignore that abrupt time gap.
- 3) Calculate the variance of the time gaps, and also do the Pearson's χ^2 test for them [10].
- 4) An account can be classified by both the variance of the time gaps and the statistic of the χ^2 test according to the rules in Table I.

Table I presents the classification rules for the ambiguous accounts. The critical region of the χ^2 test is set to be $\chi_{0.05}^2(4-1) = 7.815$ because the entire duration of 20 tweets is divided into four equal time slots, and the significance level of the test is 0.05. Since the data of just 20 tweets are slightly insufficient for the χ^2 test, we add the variance of the time gaps in the rules. The threshold is set to be 20 minutes for the variance of the time gap. If the χ^2 test for an account indicates uniformity, but the variance is large, the account will be put into the tracking list, and will be tracked for more tweets until it can be classified with the rules.

We selected 100 accounts with zero URL link and zero interaction from the dataset. The accounts were classified into 53 normal user accounts, 17 spam accounts, and 30 accounts were in the tracking list. We found neither false positives nor false negatives in the detection, so the precision and recall are both 1 for the ambiguous accounts.

D. Possible Evasion

A spammer may post many tweets without a URL, or deliberately make the spam accounts interact with each other to affect the URL rate or the interaction rate for evasion. We have not found such two strategies from the accounts observed in this work. The former will dilute the density of useful spam tweets, and is not a good option for the spammers. The interactions from known spam accounts (e.g., those have been detected) can be just ignored to avoid the latter strategy.

V. CONCLUSION

This work studies the effective features for detecting long-surviving Twitter spam accounts. We select the common features from prior studies, and analyze how effective the features are for the detection. We find two simple yet effective features, the URL rate and the interaction rate, are useful for detecting spam accounts with high accuracy, where the precision is between 0.829 and 0.885, and the recall is between 0.987 and 0.999. Although Twitter strives hard to assist normal users in avoiding the disturbance from spammers, and it has suspended many active spam accounts, there are still many long-surviving spam accounts. The presented method can detect the accounts well with the features in this work.

ACKNOWLEDGMENT

This work is supported by the National Science Council (NSC) Project 101-2219-E-194-001-.

REFERENCES

- [1] Population of social networks world wide, <http://technified.net/internet/social-networks-world-wide/555>.
- [2] S. Abu-Nimeh, T. Chen, and O. Alzubi, "Malicious and spam posts in online social networks," IEEE Computer Society, 44(9), Sep. 2011.
- [3] K. Thomas, and D. M. Nicol, "The koobface botnet and the rise of social malware," In Proceedings of the 5th International Conference on Malicious and Unwanted Software, Oct. 2010.
- [4] G. Yan, G. Chen, S. Eidenbenz, and N. Li, "Malware propagation in online social networks: nature, dynamics, and defense implication," In Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, Mar. 2011.
- [5] REST API resource, <https://dev.twitter.com/docs/api>.
- [6] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Who is tweeting on Twitter: human, bot, or cyborg?" In Proceedings of the 26th Annual Computer Security Applications Conference(ACM), Dec. 2010.
- [7] F. Benevenuto, G. Magno, T. Rodrigues and V. Almeida, "Detect spammers on Twitter," In Proceedings of the 7th Annual Collaboration, Electronic messaging, AntiAbuse and Spam Conference(CEAS), July 2010.
- [8] 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 Symposium on Recent Advances in Intrusion Detection(RAID), Sept. 2011.
- [9] J. Song, S. Lee, and J. Kim, "Spam Filtering in Twitter using Sender-Receiver Relationship," In Proceedings of International Symposium on Recent Advances in Intrusion Detection (RAID), Sept. 2011.
- [10] C. M. Zhang, and V. Paxson, "Detecting and analyzing automated activity on Twitter," In Proceedings of the 12th international conference on Passive and active measurement(PAM), Apr. 2011.
- [11] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," In Proceedings of the 26th Annual Computer Security Applications Conference(ACM), Dec. 2010.
- [12] 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, July 2010.
- [13] K. Thomas, C. Grier, V. Paxson, and D. Song, "Suspend account in retrospect: an analysis of Twitter spam," In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, Nov. 2011.
- [14] The definition of Spam on Twitter Help Center, <https://support.twitter.com/groups/31-twitter-basics/topics/114-guidelines-best-practices/articles/18311-the-twitter-rules>.
- [15] Twitter's rate limiting, <https://dev.twitter.com/docs/rate-limiting>.
- [16] The PHP library for working with Twitter's OAuth API, <https://github.com/abraham/twitteroauth>.
- [17] Weka, the machine learning tool, <http://www.cs.waikato.ac.nz/ml/weka>.
- [18] CUSUM, <http://en.wikipedia.org/wiki/CUSUM>.
- [19] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," In Proceedings of the 2011 IEEE Symposium on Security and Privacy, May 2011.