ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 7, July 2019

Semi-Supervised Spam Detection in Trending Topics Using a Statistical Analysis

Y.C. Navaneetha[1], S.Akhilendranath[2], P.Gangadhara[3]

M.Tech, Dept. of CSE, Shri Shirdi Sai Institute of Science and Engineering, Affiliated to JNTUA, Ananthapuramu, AP, India[1]

Assistant Professor, Dept. of CSE, Shri Shirdi Sai Institute of Science and Engineering, Affiliated to JNTUA, Ananthapuramu, AP, India[2]

Assistant Professor & HOD, Dept. of CSE, Shri Shirdi Sai Institute of Science and Engineering, Affiliated to JNTUA, Ananthapuramu, AP, India[3]

ABSTRACT: Online Social Networks (OSNs), especially Twitter and Facebook, are becoming an integral part of people's daily life around the world. People share their activities in OSNs, at the same time, view updates from friends. However, the popularity of OSNs also attracts spammers. Different to Email spam, OSN spam is more challenging due to its short content and streaming characteristic. Most of the existing studies on Twitter spam focus on account blocking, which is to identify and block spam users, or spammers. In this paper, we propose a semi-supervised framework for spam tweet detection. The proposed framework consists of two main modules: spam detection module operating in real-time mode, and model update module operating in batch mode.

KEYWORDS*:* multi-authority, attribute-based encryption.

I. INTRODUCTION

Online Social Networks (OSNs), like Twitter and Facebook, have become increasingly popular in the last few years. Internet users spend more hours on social network sites than any other website [64]. In addition, social network sites have become the main news source for around 80% of users according to the survey. Moreover, the social networking apps in smart phones make users' access to such sites become ubiquitous. Billions of users spend vast time in OSNs making friends with people who they are familiar with or interested in. After the relation is built, users can receive messages, usually something interesting or recent activities shared by the friends they are connected to, in the terms of tweets, wall posts or status updates. These connected users and information they shared form a huge social graph with tremendous information spreading in it. These large social networks have attracted many researchers' interest. Twitter provides several methods for users to report spam. A user can report a spam by clicking on the "report as spam" link in the their homepage on Twitter. The reports are investigated by Twitter and the accounts being reported will be suspended if they are spam. Another simple and public available method is to post a tweet in the "@spam @username" format where @username is to mention the spam account. I tested this service by searching "@spam" on Twitter. Surprisingly the query results show that this report service is also abused by both hoaxes and spam. Only a few tweets report real malicious accounts. Some Twitter applications also allow users to flag possible spam. However, all these ad hoc methods require users to identify spam manually and depend on their own experience. Twitter also puts effort into cleaning up suspicious accounts and filtering out malicious tweets. Meanwhile, legitimate Twitter users complain that their accounts are mistakenly suspended by Twitter's antispam action. Twitter recently admitted to accidentally suspending accounts as a result of a spam clean-up effort.

In this paper, we propose a Semi-Supervised Spam Detection (S3D) framework for spam detection at tweet-level. The proposed framework consists of two main modules: spam detection module operating in real-time mode, and model update module operating in batch mode. The spam detection module consists of four light-weight detectors: (i) blacklisted domain detector to label tweets containing blacklisted URLs, (ii) near-duplicate detector to label tweets that are near duplicates of confidently pre-labeled tweets, (iii) reliable ham detector to label tweets that are posted by trusted users and

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 7, July 2019

that do not contain spammy words, and (iv) multi-classifier based detector labels the remaining tweets.

II. LITERATURE SURVEY

1) Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell.

We show how users' activity on Facebook relates to their personality, as measured by the standard Five Factor Model. Our dataset consists of the personality profiles and Facebook profile data of 180,000 users. We examine correlations between users' personality and the properties of their Facebook profiles such as the size and density of their friendship network, number uploaded photos, number of events attended, number of group memberships, and number of times user has been tagged in photos. Our results show significant relationships between personality traits and various features of Facebook profiles. We then show how multivariate regression allows *prediction* of the personality traits of an individual user given their Facebook profile. The best accuracy of such predictions is achieved for Extraversion and Neuroticism, the lowest accuracy is obtained for Agreeableness, with Openness and Conscientiousness lying in the middle.

2) F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida

Understanding how users behave when they connect to social networking sites creates opportunities for better interface design, richer studies of social interactions, and improved design of content distribution systems. In this paper, we present a first of a kind analysis of user workloads in online social networks. Our study is based on detailed click stream data, collected over a 12-day period, summarizing HTTP sessions of 37,024 users who accessed four popular social networks: Orkut, MySpace, Hi5, and LinkedIn. The data were collected from a social network aggregator website in Brazil, which enables users to connect to multiple social networks with a single authentication. Our analysis of the clickstream data reveals key features of the social network workloads, such as how frequently people connect to social networks and for how long, as well as the types and sequences of activities that users conduct on these sites. Additionally, we crawled the social network topology of Orkut, so that we could analyze user interaction data in light of the social graph. Our data analysis suggests insights into how users interact with friends in Orkut, such as how frequently users visit their friends' or non-immediate friends' pages. In summary, our analysis demonstrates the power of using clickstream data in identifying patterns in social network workloads and social interactions. Our analysis shows that browsing, which cannot be inferred from crawling publicly available data, accounts for 92% of all user activities. Consequently, compared to using only crawled data, considering silent interactions like browsing friends' pages increases the measured level of interaction among users.

3) Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro

Users increasingly rely on the trustworthiness of the information exposed on Online Social Networks (OSNs). In addition, OSN providers base their business models on the marketability of this information. However, OSNs suffer from abuse in the form of the creation of fake accounts, which do not correspond to real humans. Fakes can introduce spam, manipulate online rating, or exploit knowledge extracted from the network. OSN operators currently expend significant resources to detect, manually verify, and shut down fake accounts. Tuenti, the largest OSN in Spain, dedicates 14 full-time employees in that task alone, incurring a significant monetary cost. Such a task has yet to be successfully automated because of the difficulty in reliably capturing the diverse behavior of fake and real OSN profiles.

We introduce a new tool in the hands of OSN operators, which we call SybilRank. It relies on social graph properties to rank users according to their perceived likelihood of being fake (Sybils). SybilRank is computationally efficient and can scale to graphs with hundreds of millions of nodes, as demonstrated by our Hadoop prototype. We deployed SybilRank in Tuenti's operation center. We found that ~90% of the 200K accounts that SybilRank designated as most likely to be fake, actually warranted suspension. On the other hand, with Tuenti's current user-report-based approach only ~5% of the inspected accounts are indeed fake.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 7, July 2019

III. EXISTING SYSTEM

❖ Many social spam detection studies focus on the identification of spam accounts. Lee et al. [14] analyzed and used features derived from user demographics, follower/following social graph, tweet content, and the temporal aspect of user behavior to identify content polluters.

❖ Hu et al. [13] exploited social graph and tweets of a user to detect spam detection on Twitter. They formulated spammer detection task as an optimization problem. Online learning has been utilized to tackle the fast evolving nature of spammer [12]. They have utilized both content and network information and incrementally updated their spam detection model for effective social spam detection.

❖ Tan et al. [21] proposed an unsupervised spam detection system that exploits legitimate users in the social network. Their analysis shows the volatility of spamming patterns in social network. They have utilized non spam patterns of legitimate users based on social graph and user link graph to detect spam pattern.

❖ Gao et al. [9] identified social spam by clustering posts based on text and URL similarities and detected large clusters with bursty posting patterns. Incremental clustering-based approach has been used to detect spam campaigns on Twitter.

Disadvantages

➢ There is no Semi-supervised learning.

➢ There is no option to find type of different spammers.

IV. PROPOSED SYSTEM

❖ In the proposed system, the system proposes a semi-supervised framework for spam tweet detection. The proposed framework mainly consists of two main modules: 1) four lightweight detectors in the spam tweet detection module for detecting spam tweets in real time and 2) updating module to periodically update the detection models based on the confidently labeled tweets from the previous time window. The detectors are designed based on our observations made from a collection of 14 million tweets, and the detectors are computationally effective, suitable for real-time detection.

❖ More importantly, our detectors utilize classification techniques at two levels, tweet level and cluster level. Here, a cluster is a group of tweets with similar characteristics. With this flexible design, any features that may be effective in spam detection can be easily incorporated into the detection framework. The framework starts with a small set of labeled samples and updates the detection models in a semi-supervised manner by utilizing the confidently labeled tweets from the previous time window. This semi-supervised approach helps to learn new spamming activities, making the framework more robust in identifying spam tweets.

Advantages

➢ **Confidently Labeled Tweets**- Tweets that are labeled by the first three detectors (i.e., blacklisted domain, near duplicate and reliable ham tweet) are considered as confidently labeled tweets.

➢ **Near-Duplicate Cluster Labeling** - Recall that the near duplicate detector computes a signature for each tweet to check if the tweet is a near duplicate of a labeled cluster. If the signature of a tweet does not match any relabeled cluster, then the tweet is passed to the next level detectors.

V. IMPLEMENTATION

➢ **Admin Server**

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as View All Users And Authorize, View Friend Request And Responses, View All Users Tweets, View All Retweets Details, Add Spam Filter, View Spam Detection in Twitter Stream, View Tweet Score Results, View Spam Detection Results.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 7, July 2019

➢ **User**

In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like Search Friend And Find Friend Request ,View All My Friends, Create Your Tweet, View All Your Tweets, View All Your Friends Tweets and Retweet.

VI. CONCLUSIONS

In this paper, a semi-supervised spam detection framework is implemented which is called as S3D. S3D utilizes four lightweight detectors to detect spam tweets on real-time basis and update the models periodically in batch mode. The experiment results demonstrate the effectiveness of semi-supervised approach in the spam detection framework. In our experiment, we found that confidently labeled clusters and tweets make the system effective in capturing new spamming patterns. Tweet-level spam detection is a fine-grained approach which can be used to detect spam tweets in real time. However, for a given tweet only limited information can be obtained. In contrast, more discriminative features can be derived from user account, historical tweets of the users, and social graph. However, by the time a malicious user is detected, the user might affect many other users. We believe that tweet-level spam detection complements user-level spam detection. Due to the limited user information in our data set, we have used the simple technique to deal with user-level spam detection.

REFERENCES

1. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," ACM SIGCOMM Computer Communication Review, vol. 39, no. 1, pp. 50–55, 2008.
2. iCloud. (2014) Apple storage service. [Online]. Available: https://www.icloud.com/
3. Azure. (2014) Azure storage service. [Online]. Available: http://www.windowsazure.com/
4. Amazon. (2014) Amazon simple storage service (amazon s3). [Online]. Available: http://aws.amazon.com/s3/
5. K. Chard, K. Bubendorfer, S. Caton, and O. F. Rana, "Social cloud computing: A vision for socially motivated resource sharing," Services Computing, IEEE Transactions on, vol. 5, no. 4, pp. 551–563, 2012.
6. C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacypreserving public auditing for secure cloud storage," Computers, IEEE Transactions on, vol. 62, no. 2, pp. 362–375, 2013.
7. G. Anthes, "Security in the cloud," Communications of the ACM, vol. 53, no. 11, pp. 16–18, 2010
8. Shamir, "Identity-based cryptosystems and signature schemes,"in *Advances in Cryptology*. Berlin, Germany: Springer-Verlag, 1985, pp. 47–53.
9. Sahai and B. Waters, "Fuzzy identity-based encryption," in *Advances in Cryptology*. Berlin, Germany: Springer-Verlag, 2005, pp. 457–473.
10. V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proc. 13thCCS*, 2006, pp. 89–98.
11. K. Yang, X. Jia, K. Ren, and B. Zhang, "DAC-MACS: Effective data access control for multi-authority cloud storage systems," in *Proc. ICSE INFOCOM*, Apr. 2013, pp. 2895–2903.
12. W.-G. Tzeng, "Efficient 1-out-of-n oblivious transfer schemes with universally usable parameters," *ICSE Trans. Comput.*, vol. 53, no. 2, pp. 232–240, Feb. 2004.
13. M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute based encryption," *ICSE Trans. Parallel Distrib. Syst.*, vol. 24, no. 1, pp. 131–143, Jan. 2013.