International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. 02 Issue 02 | 2018

A Survey on Statistical Twitter Spam Detection Demystified: Performance, Stability and Scalability

Rohit V.Adagale [1], Aniket C.Sanap[2], Anil V.Gitte[3], Prof. R. H. Kulkarni[4]
[1,2,3,4] Department of Computer Engineering, JSPM Narhe technical campus,pune-411041, Maharashtra, India.

ABSTRACT

Today, peoples are increasing amount of time in social networks. However, because of the popularity of online social networks, cybercriminals are spamming on these platforms for potential victims. Spams invite users to external phishing sites or malware downloads huge security issue online and undermined the user experience. However, current solutions do not reveal the Twitter spamming accurately and indeed. In this article, we compared the performance of a wide range of conventional machine learning algorithms, with the aim of identify those that offer satisfactory detection and stability performance based on a large amount of true field data. With the objective in order to realize the real-time spam detection capability, we evaluated scalability algorithms. Performance the study evaluates the accuracy of the detection, the TPR / FPR and the F measure; stability analyzes the stability of algorithms using randomly selected training samples of different sizes. Scalability aims to better understand the impact of in reducing training time learning algorithms.

***Keywords**-Machine learning, Twitter, spam detection, parallel computing, scalability*

1. INTRODUCTION

Social networking sites such as Twitter, Facebook, Instagram and some enterprise of online social network have become extremely popular in the last few years. Individuals spend vast amounts of time in OSNs making friends with people who they are familiar with or interested in. Twitter, which was founded in 2006, has become one of the most popular micro blogging service sites. Around 200 million users create around the 400 million new tweets per day the growth of spam. Twitter spam, which is referred as unsolicited tweets containing malicious links that directs victims to external sites containing malware spreading, malicious link spreading etc. has not only affected a number of legitimate users but also polluted the whole platform. Consider the example as during the Australian Prime Minster Election in 2013 published an alert that confirmed its Twitter account @AusElectoralCom was hacked. Many of its followers received direct spam messages which contained malicious links. The ability to sort out useful information is critical for both academia and industry to discover hidden insights and predict trends on Twitter. However, spam significantly brings noise into Twitter. To automatically detect spam, machine learning algorithms have been applied by researchers to make spam detection as a classification problem. Classifying a streaming tweet instead of a Twitter user to spam or non-spam is more realistic in the real world.

2. RELATED WORK

Sr. No.	Author, Title and Journal Name	Advantages	Disadvantage	Refer Points
1	Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in *Proc. Symp. Netw. Syst. Des. Implement. (NSDI)*, 2012, pp. 197–210.	1. Identifying highly suspicious accounts by ranking users. 2. Low computational cost.	1. This work is carried out manually so it is time consuming and expensive based on CAPTCHA.	1. SybilRank, an effective and efficient fake account inference scheme, which allows OSNs to rank accounts according to their perceived likelihood of being fake. 2. It works on the extracted knowledge from the network so it detects, verify and remove the fake accounts.
2	G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *Proc. 26th Annu. Comput. Sec. Appl. Conf.*, 2010, pp. 1–9.	1. To improve the security. 2. To detect spammers on Twitter this based on the machine learning algorithm.	1. Mainly require the historical information to build the social graph.	1. Help to detect spam Profiles even when they do not contact a honey-profile. 2. The irregular behavior of user profile is detected and based on that the profile is developed to identify the spammer.
3	J. Song, S. Lee, and J. Kim, "Spam filtering in Twitter using sender receiver relationship," in *Proc. 14th Int. Conf. Recent Adv. Intrusion Detection*, 2011, pp. 301–317.	1. The spam filtering systemWill be more powerful. 2. The accuracy is better. 3. Caching technique will help both client-side and server-side to reduce computing overhead.	1. The relation feature approach is very difficult to calculate.	1. A spam filtering method for social networks using relation information between users. 2. System use distance and connectivity as the features which are hard to manipulate by spammers and effective to classify spammers.
4	K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots + machine learning," in *Proc. 33rd Int. ACM SIGIR Conf. Res.Develop. Inf. Retrieval*, 2010, pp. 435–442.	1. The deployment of social Honey pots for harvesting deceptive spam profiles from social Networking. 2. Statistical analysis of these spam's profiles.	1. Mainly Time consuming and resource consuming for the system.	1. System analyzes how spammers who target social networking sites operate. 2. To collect the data about spamming activity, system created a large set of "honey-profiles" on three large social networking sites.

5	Nathan Aston, Jacob Liddle and Wei Hu*, “Twitter Sentiment in Data Streams with Perceptron,” in *Journal of Computer and Communications*, 2014, Vol-2 No-11.	1. Suitable for unbalanced classes 2. Simple computation 3. Suitable for incremental learning	1. Independence assumption for computing P_c often invalid 2. Conservative estimate	1. The implementation feature reduction we were able to make our Perceptron and Voted Perceptron algorithms more viable in a stream environment. 2. In this paper, develop methods by which twitter sentiment can be determined both quickly and accurately on such a large scale.
6	K. Thomas, C. Grier, D. Song, and V. Paxson, “Suspended accounts in retrospect: An analysis of Twitter spam,” in *Proc. ACM SIGCOMM Conf. Internet Meas.*, 2011, pp. 243–258.	1. Fledgling spam-as-a-service market - Affiliate programs - Account providers	1. Low barrier to creating accounts 2. Weak defenses, slow response	1. The behaviors of spammers on Twitter by analyzing the tweets sent by suspended users in retrospect. 2. An emerging spam-as-a-service market that includes reputable and not-so-reputable affiliate programs, ad-based shorteners, and Twitter account sellers.
7	K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, “Design and evaluation of a real-time URL spam filtering service,” in *Proc. IEEE Symp. Sec. Privacy*, 2011, pp. 447–462.	1. It provides 90.78% accuracy for identifying web service spam. 2. Run-time performance is high as 5.54 seconds.	1. Expensive	1. Monarch is a real-time system for filtering scam, phishing, and malware URLs as they are submitted to web services. 2. Monarch’s architecture generalizes to many web services being targeted by URL spam, accurate classification hinges on having an intimate understanding of the Spam campaigns abusing a service.
8	X. Jin, C. X. Lin, J. Luo, and J. Han, “Socialspamguard: A data mining based spam detection system for social media networks,” *PVLDB*, vol. 4, no. 12, pp. 1458–1461, 2011.	1. Automatically harvesting spam activities in social network by monitoring social sensors with popular user bases; 2. Introducing both image and text content features and social network features	NA	1. Proposed *SocialSpamGuard*, a scalable and online social media spam detection system based on data mining for social network security. 2. GAD clustering algorithm for large scale clustering and integrate it with the designed active learning algorithm

		to indicate spam activities; 3. Integrating with our GAD clustering algorithm to han- dle large scale data; 4. Introducing a scalable active learning approach to identify existing spams with limited human efforts, and Perform online active learning to detect spams in real-time.		
9	S. Ghosh *et al.*, “Understanding and combating link farming in the Twitter social network,” in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 61–70.	1. Vast amounts of information and real-time news 2. Twitter search becoming more and more common 3. Search engines rank users follower-rank, Pagerank to decide whose tweets to return as search results 4. High indegree (#followers) seen as a metric of influence	NA	1. Search engines rank websites / webpages based on graph metrics such as Pagerank - High in-degree helps to get high Pagerank 2. Link farming in Twitter -Spammers follow other users and attempt to get them to follow back
10	H. Costa, F. Benevenuto, and L. H. C. Merschmann, “Detecting tip spam in location-based social networks,” in *Proc. 28th Annu. ACM Symp. Appl. Comput.*, 2013, pp. 724–729.	1. High accuracy	NA	1. Identifying tip spam on a popular Brazilian LBSN system, namely Apontador. 2. Based on a labeled collection of tips provided by Apontador as well as crawled information about users and locations, we identified a number of attributes able to distinguish spam from non-spam tips.

3. PROPOSED SYSTEM APPROACH

In proposed system, we evaluate the spam detection performance on our dataset by using machine learning algorithms. The process of Twitter spam detection by using machine learning algorithms. Before classification, a classifier that contains the knowledge structure should be trained with the prelabeled tweets. After the classification model gains the knowledge structure of the training data, it can be used to predict a new incoming tweet. The whole process consists of two steps: learning and classifying. Features of tweets will be extracted and formatted as a vector. The class labels i.e. spam and non-spam could be get via some other approaches. Features and class label will be combined as one instance for training. One training tweet can then be represented by a pair containing one feature vector, which represents a tweet, and the expected result, and the training set is the vector. The training set is the input of machine learning algorithm, the classification model will be built after training process. In the classifying process, timely captured tweets will be labeled by the trained classification model.

3.1 Advantages

- Extraction of features and categories as Tag based features and URL based features.
- The system implements a method which will use ML mechanism to detect whether the post is spam or not.
- The system implements application can also be hosted online for its use and the data will be stored and fetched from server.
- User with maximum number of spam can be blocked from the system.

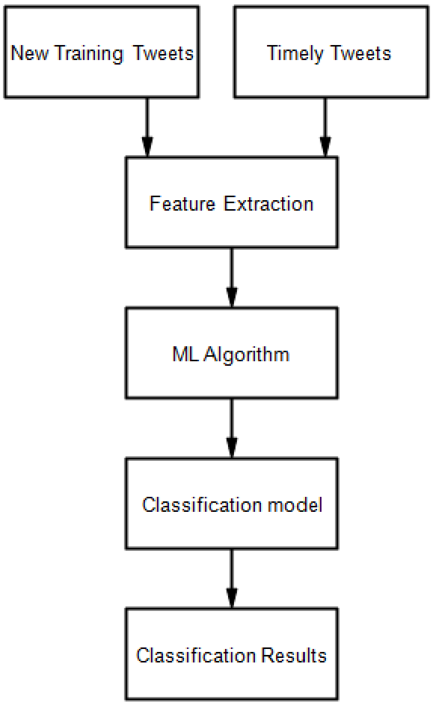

Fig 1. Proposed System Architecture

4. SYSTEM ANALYSIS

TABLE I Performance Evaluation on Datasets I and II

Unit:%	Dataset I			Dataset II		
Classifier	**TPR**	**FPR**	**F-measure**	**TPR**	**FPR**	**F-measure**
Naive Bayes	97.3	77.1	70.9	97.3	78.8	11.5

TABLE II Confusion Matrix of Random Forest on Both Datasets

Classified	Spam	Non-spam	Spam	Non-spam
Spam	4645	355	4645	355
Non-spam	282	4718	6766	88234
	Dataset I		**Dataset II**	

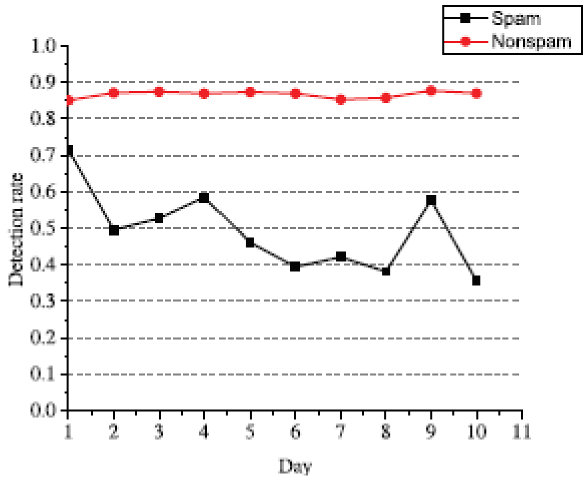

Fig.2. Spam Detection Rate

5. CONCLUSION

In this Project, System found that classifiers ability to detect Twitter spam reduced when in a near real-world scenario since the imbalanced data brings bias. System also identified that Feature discretization was an important preprocess to ML-based spam detection. Second, increasing training data only cannot bring more benefits to detect Twitter spam after a certain number of training samples. System should try to bring more discriminative features or better model to further improve spam detection rate.

6. REFERENCES

[1] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, “Aiding the detection of fake accounts in large scale social online services,” in *Proc. Symp. Netw. Syst. Des. Implement. (NSDI)*, 2012, pp. 197–210.

[2] G. Stringhini, C. Kruegel, and G. Vigna, “Detecting spammers on social networks,” in *Proc. 26th Annu. Comput. Sec. Appl. Conf.*, 2010, pp. 1–9.

[3] J. Song, S. Lee, and J. Kim, “Spam filtering in Twitter using sender receiver relationship,” in *Proc. 14th Int. Conf. Recent Adv. Intrusion Detection*, 2011, pp. 301–317.

[4] K. Lee, J. Caverlee, and S. Webb, “Uncovering social spammers: social honeypots + machine learning,” in *Proc. 33rd Int. ACM SIGIR Conf. Res.Develop. Inf. Retrieval*, 2010, pp. 435–442.

[5] Nathan Aston, Jacob Liddle and Wei Hu*, "Twitter Sentiment in Data Streams with Perceptron," in *Journal of Computer and Communications,* 2014, Vol-2 No-11.

[6] K. Thomas, C. Grier, D. Song, and V. Paxson, "Suspended accounts in retrospect: An analysis of Twitter spam," in *Proc. ACM SIGCOMM Conf. Internet Meas.*, 2011, pp. 243–258.

[7] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," in *Proc. IEEE Symp. Sec. Privacy*, 2011, pp. 447–462.

[8] X. Jin, C. X. Lin, J. Luo, and J. Han, "Socialspamguard: A data mining based spam detection system for social media networks," *PVLDB*, vol. 4, no. 12, pp. 1458–1461, 2011.

[9] S. Ghosh *et al.*, "Understanding and combating link farming in the Twitter social network," in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 61–70.

[10] H. Costa, F. Benevenuto, and L. H. C. Merschmann, "Detecting tip spam in location-based social networks," in *Proc. 28th Annu. ACM Symp. Appl. Comput.*, 2013, pp. 724–729.