Multidisciplinary Review Journal

Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 1, 2020

Review Article

MACHINE LEARNING TECHNIQUES FOR SPAMMER IDENTIFICATION: STATE OF THE ART AND ANALYSIS

R. KRITHIGA[1*], DR.E. ILAVARASAN[2]

[1]Research Scholar, Department of Computer Science & Engg., Pondicherry Engineering College, Puducherry, India. kriithiga@gmail.com
[2]Professor, Department of Computer Science & Engg., Pondicherry Engineering College, Puducherry, India. eilavarasan@pec.edu

Received: 04.11.2019 *Revised: 12.12.2019* *Accepted: 18.01.2020*

ABSTRACT

Internet offers seamless communication with a high consumption rate. With the increase in the widespread utilization of internet tools such as e-mails, the spammers started to exploit the e-mail network to effect malicious and hazardous activities. As the e-mail spam detection systems became more sophisticated and accurate, the attention of spammers has now turned to the recently emerged social networks due to massive usage and popularity among the people. A range of online social networks (OSN) such as Facebook, Twitter, Instagram are available with each of them offering unique services to the account owners and benefits both professionally and personally. Besides these advantages, the network also houses illicit accounts that disturbs the user experience and destroy the ultimate objective of social networking. The existing techniques employed by these OSN to detect such illicit users are not effective and accurate and also demands manual approaches to spot them. Hence, a number of methods and algorithms were proposed in the literature to identify the spammers concealed in these networks. This work attempts to provide a detailed review of state of the art techniques and methodologies employed in the spam account detection problem along with future research directions.

Keywords: Spam Profile Detection, Classification of Spammers, Twitter Data, Facebook, Instagram, Online Social Networks, Spammer Identification.

DOI: http://dx.doi.org/10.31838/jcr.07.01.87

INTRODUCTION

With the advent of online internet communication, businesses have started adopting various online platforms as a tool for branding, promotions and sales. Diverse fields such as movies, hotels, televisions, photography, arts, and job portals make use of these online medium to satisfy its needs with respect to showcasing of products or services. The internet enables people to stay connected irrespective of time and space.

However, with the coming of recently popular social networks, the people started spending most of their time to simply express themselves, their whereabouts and share information with the network of family and friends. It has almost become an integral part of everyday life.

The way a social media is being used differs from person to person. The expectations and anticipations of users from social media tools are different, and hence possess diverse uses as well. The social media is emerging as a platform where once can socialize, display works and get appreciated and an expression of self.

In this study we have analysed the existing concept and methods for the detection of spammers i.e. the spam accounts on most popular social media sites such as Twitter, Facebook, Instagram and YouTube. Malicious users fool the genuine users by posting advertisements, phishing sites, fraud, pornography, viruses, etc. under the comments section of popular videos of the time. The spammers initiate several hazardous activities to exploit the network. Spams on Twitter not only affect the online social experience, but also threatens the safety of cyberspace.

Therefore, there is an urgent need to effectively combat the spread of spams. The spamming issues have attracted the attention of the research community. Researchers have put many efforts to improve social spam detection in terms of efficiency and accuracy by proposing various novel approaches.

Spam message detection and spam account detection are two different problems that are being researched in the social media domain. There exists a very few works on spam account detection and hence calls for further researching to better the techniques and keep the environment safe and appropriate for the purpose it has been intended to do.

This work aims at throwing a light on the state-of-the-art methods and techniques on the spammers' identification (spam account detection) that also includes the recent advancements and merits and demerits of each of these methods.

The analysis made will give rise to new ideas and concepts to tackle this problem. Spam Detection framework is a binary classification problem that classifies an account as spam or non-spam. A machine learning spam account detection model usually comprises of two phases as follows:

(i) Training phase: It is the first phase in which detection model is trained using classified labelled samples. (ii) Testing phase: It is a follow up of training phase. Unlabelled samples are tested and generate the results by classifying each sample into spam or non-spam class. The rest of the paper is organized as follows: The Section II describes the state- of-the-art of spam detection problem. Section III summarizes the methodology followed in the process of spam account identification and Section IV concludes the work.

SPAM DETECTION METHODS EMPLOYED IN THE LITERATURE

While researching the works done by various authors, it was evident that most of the works concentrate on spam message detection and only a few works exist on spam account or profile detection. Hence, there is a huge research vacuum that prevails in this domain. One of the most challenging parts of the research is the availability of datasets and distribution of the same. Though several social networks exist for day-to-day use, maximum of the researches have been conducted considering the Twitter social network. There are other works on social networks such as fake review identification, fake profile identification, social bot classification, user-sentiment analysis and spam tweet removal.

Ref	Network	Methodology	Metrics used for evaluation	Merits	Demerits
[2]	Sina Weibo	Extreme Learning Machine (ELM)	True Positive Rate = 99% for spammers and 99.95 for non-spammers Precision, Recall, F-Measure, Training Time and Testing Time	The proposed methods is efficient and very fast than the existing methods	Very poor dataset size that is inappropriate for generalization
[3]	Twitter	J48	Accuracy = 93% approximately	Proposed new set of strong features which also reduces false positive rates	The method lacks scalability and may fail to spot the evolving spammers.
[4]	Instagram	Random Forest	Accuracy = 96.27% (Accuracy, Execution Time and Throughput)	This paper is the only work that has addressed the spam account detection in Instagram network.	There is a huge room of improvement in terms of rich feature set and fine tuning of models.
[5]	Twitter	J48 classifier with ReliefF feature selection algorithm	Accuracy = 94.4%	Irrespective of the language, the model achieves a fair detection rates	As only publicly available features are considered, it could easily be manipulated by the spammers and disturb the network.
[6]	Facebook	Bayesian Network classifier	Accuracy = 98.4%, Mathew Co-relation Coefficient = 97.7%, F-Score = 98.4%	Introduced a new set of profile and content based features which aid in the effective classification of spammers	The test bed is too small and hence may not be appropriate or certain for real time systems
[7]	Twitter	Deep learning ensemble with Convolutional Neural Networks	Accuracy = 95.7% (Accuracy, Precision, Recall and F-Measure)	The algorithm displays a graceful performance even for imbalanced class	Only the tweets were considered as input for neural networks and no other additional information was passed.
[8]	Youtube	Markov Decision Process	Accuracy = 78.82% (Accuracy, Sensitivity and Specificity)	The video spammers are identified efficiently using video instances	The feature set constructed is very small and calls for enhancement on considering other dimensions.
[9]	Twitter	Random Forest with community and interaction based features	Detection Rate = 97.6% Detection Rate, False Positive Rate and F-Measure	It is difficult for spammers to evade the network the interaction of the node with neighbouring ones are considered	The spammers are not spotted at an early stage and detection after exposing themselves by exhibiting their behaviour.
[10]	Twitter	Rough set theory based attribute selection with	Accuracy = (86.21%, 83.83%, 99.50%, 99.61%, 81.04%) for five different datasets	Efficiently selects a minimal subset of features to deliver an equal or better detection accuracy of spammers	The methodology works only for categorical and to be enhanced for continuous values. The method exhibited significant variation in performance for various social network datasets.
[11]	Sina Weibo	Single Linkage Clustering Scheme with Support Vector Machine	Detection Rate=94.5% (TPR, FPR, ROC, AUC, Detection Rate, F-Measure)	Addressed the problem of evolving spammers considering the temporal evolution factors	The detection system is offline and does not provide any means to work for real time systems
[12]	Indonesian Language Twitter	Logistic Regression	Accuracy=93.67%	Exploratory work of spammer identification in Bahasa Indonesia with a novel set of features	Corpus size is too small that poses challenge to work efficiently on real time.
[13]	Twitter	Random Forest with Information Gain for Feature Selection	Accuracy=91% (Accuracy, Precision, Recall, and F1-score)	Proposed new set of graph and content based features	The model cannot scale and does not work for other types of social networks
[14]	Twitter	Random Forest with Principal Component Analysis for Feature Extraction and K-means for grouping spammers	Accuracy=96.30% (TPR, FPR, Precision, Recall, F-measure and Building Time)	Identified group of Spammers instead of a single entity	Primary grouping of spammers is computationally expensive. And the model cannot scale with the growing size of the Twitter network.

METHODOLOGY

The basic requirement to undergo research in social media is the availability of datasets. Unfortunately, due to the social networks' policy on data sharing, the availability of datasets to pursue research does not exist. Hence, the researchers have crawled the social networks' data using API's and used them for experimentation. A range of features categories have been proposed in the literature as shown in the Fig. 1.

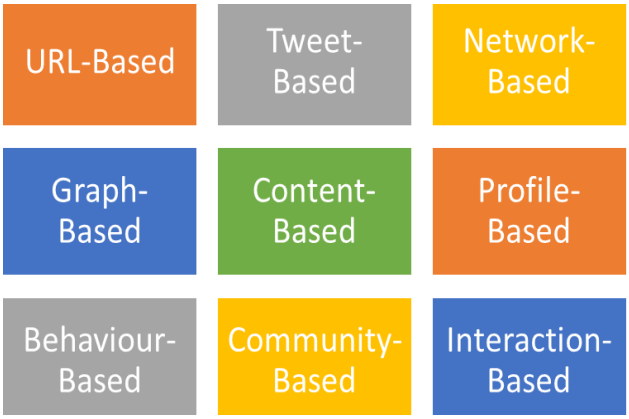

Fig. 1: Categories of Feature Sets

The above is a consolidation of the feature sets employed in the literature. However, not all the features have been used in any of the study. Few or more categories were proposed and newer features were proposed for the feature bunches. Though Precision, Recall, F-Measure, Accuracy, True Positive Rate, False Positive Rate, ROC and AUC have been used for model evaluation, the training and testing time have also been used to check for the quickness in categorizing the profiles.

CONCLUSION AND RESEARCH DIRECTIONS

This paper provided a deep analysis and insights on the state-of-the art techniques existing for the spam profile identification. The paper comprises of the recent works that also include the pros and cons of the methodologies proposed in the literature. The size of the dataset employed for the previous researches were not sufficient to generalize to that of a social network that consists of enormous amount of data. Secondly, the social networks are of growing in nature and so the methods should be devised that could address the scalability of networks. Finally, instead of building models for exclusive social networks, a possibility is that a unified approach could be devised through which the system would be able to detect the spammers irrespective of the kind of social network it belongs to.

REFERENCES

1. Chao Yang, Robert Harkreader, and Guofei Gu,'Empirical Evaluation and New Design for FightingEvolving Twitter Spammers', *IEEE Transactions on Information Forensics and Security,* Vol. 8, No. 8, August 2013, pp. 1280-1293.
2. Xianghan Zheng, Xueying Zhang, Yuanlong Yu, Tahar Kechadi, Chunming Rong,'ELM-based spammer detection in social networks', *Journal of Supercomputing,* 2015, doi 10.1007/s11227-015-1437-5.
3. Abdullah Almaatouq, Erez Shmueli, Mariam Nouh et al, 'If it looks like a spammer and behaves like a spammer, it must be a spammer: analysis and detection of microblogging spam accounts', *International Journal of Information & Security, 2016,* doi 10.1007/s10207-016-0321-5.
4. Wuxain Zhang and Hung-Min Sun, 'Instagram Spam Detection', 22nd Pacific Rim International Symposium on Dependable Computing, *IEEE,* 2017, pp. 227 – 228.
5. Ala' M. Al-Zoubi, Ja'far Alqatawna, Hossam Faris, 'Spam Profile Detection in Social Networks Based onPublic Features', 8th *International Conference on Information and Communication Systems (ICICS), IEEE,* 2017, pp. 130 – 135
6. Shailendra Rathore, Vincenzo Loia, Jong Hyuk Park, SpamSpotter: An Efficient Spammer Detection Framework based on Intelligent Decision Support System on Facebook, *Applied Soft Computing Journal,* 2017, http://dx.doi.org/10.1016/j.asoc.2017.09.032
7. Sreekanth Madisetty and Maunendra Sankar Desarkar, 'A Neural Network-Based Ensemble Approach for Spam Detection in Twitter', IEEE Transactions on Computational Social Systems, 2018, doi:10.1109/TCSS.2018.2878852.
8. Simran Kanodia et al, 'A Novel Approach for Youtube Video SpamDetection using Markov Decision Process', IEEE, 2018
9. Mohd Fazil and Muhammad Abulaish, 'A Hybrid Approach for Detecting AutomatedSpammers in Twitter', Transactions on Information Forensics and Security, 2018
10. Soumi Dutta, Sujata Ghatak, Ratnadeep Dey et al, 'Attribute selection for improving spam classification in online socialnetworks: a rough set theory-based approach', *Social Network Analysis and Mining,* 2018, https://doi.org/10.1007/s13278-017-0484-8
11. Qiang Fu, Bo Feng, Dong Guo, Qiang Li, Combating the evolving spammers in online social networks, Computers & Security (2017), http://dx.doi.org/doi: 10.1016/j.cose.2017.08.014. i,
12. Erwin B. Setiawan, Dwi H. Widyantoro, 'Detecting Indonesian Spammer on Twitter', 6th *International Conference on Information and Communication Technology (ICoICT),*, 2018
13. Zulfikar Alom, 'Detecting spam accounts on Twitter', *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),* 2018, pp. 1191 – 1198.
14. Kayode Sakariyah Adewole, 'Twitter spam account detection based on clustering and classificationmethods', *The Journal of Supercomputing,* 2018.