Behavioral Analysis to Detect Social Spammer in Online Social Networks (OSNs)

Somya Ranjan Sahoo[1], B. B. Gupta[1(✉)], Chang Choi[2], Ching-Hsien Hsu[3], and Kwok Tai Chui[4]

[1] Department of Computer Engineering, National Institute of Technology Kurukshetra, Kurukshetra, India
somyaranjan.sahoo@gmail.com, gupta.brij@gmail.com
[2] Gachon University, Seongnam-si, Republic of Korea
enduranceaura@gmail.com
[3] Asia University, Taiwan and CS, Chung Hua University, Hsinchu, Taiwan
robertchh@gmail.com
[4] The Open University of Hong Kong, Kowloon, Hong Kong
jktchui@ouhk.edu.hk

Abstract. The faster and regular usage of Web 2.0 technologies like Online Social Networks (OSNs) addicted to millions of users worldwide. This popularity made target for spammers and fake users to spread phishing attack, viruses, false news, pornography and unwanted advertisements like URLs, images and videos etc. The present paper proposes a behavioral analysis-based framework for classifying spam contents in real time by aggregating machine learning techniques and genetic algorithm. The main procedure of the work is, firstly based on social networks spam policy, novel profile based and content-based features are proposed to facilitate spam detection. Secondly, accumulate a dataset from various social networks like Facebook, Twitter, and Instagram including spam and non-spam profiles. For suitable feature selections, we have used a genetic algorithm and various classifiers for decision making. In order to attest the effectiveness of our proposed framework, we have compared with existing techniques.

Keywords: Online social networks · PSO · Facebook · Machine learning

1 Introduction

Due to the busy schedule of a human being, people use OSNs such as Facebook, Twitter and Instagram for their communication, sharing of thoughts with their friends, post messages, share valuable views and discuss hot topics. These websites play an important role in people's daily life [1–3]. Unfortunately, these activities of social platform become a new gateway for social spammers to achieve their goals such as spreading malware, posting spam content, and doing other illicit activities. Basically, social spammer spotting is a binary classification approach using feature analysis. In order to improvise the performance, suitable feature selections are required. The spreading of malicious content

S. Chellappan et al. (Eds.): CSoNet 2020, LNCS 12575, pp. 321–332, 2020.
https://doi.org/10.1007/978-3-030-66046-8_26

degrades user performance, experience, and various functions at server site such as analysis of user behavior, database server and resource recommendation. Therefore, it becomes desirable to develop a framework for detecting spammer and their activities. Currently, there have been few solutions developed by academicians and industry to detect spammer and their behavior in a social network platform. These solutions are either ineffective due to public feature analysis and manual selection of features [4, 5].

This paper investigates spammer in the social platform by analyzing public and private features by using suitable feature selection based on genetic algorithm and machine learning approach. Meanwhile, in order to improve the performance of the proposed framework, we utilize various social network information and label dataset by using API and crawler to guide the machine learning approach easily. We empirically evaluate the proposed framework on real-world dataset and depict the benefit of the proposed framework. The remaining parts of the paper are organized as follows. In Sect. 2, we reviewed related work for spammer detection. Section 3 describes our proposed framework and suitable feature selection approach. In Sect. 4, we describe an analysis of result and comparative work with others. Finally, in Sect. 5, we conclude our paper and some future research direction.

2 Related Work

Detection of social spammer becomes a hot subject in industry and academic field. Spam is an unwanted message spread through a social network platform. In recent years, many methods and frameworks have been proposed by academician to detect spammer on OSNs including feature analysis, social graph-based analysis and various optimization techniques. In [6], the author used support vector machine to classify the malicious content from a legitimate one. He analyzed app similarity and post name similarity content spread through various users as advertisements. In [7], the author analyzes the user characterization based on the user interaction with their followers. After collecting various features from different profiles, author used a machine learning method to separate spammer contents. The author in [8], identified the spammer content in twitter profile by analyzing the behavior of the user and generated trust score based on profile features. In [9], the author evaluated 4 different features using 16 online learning algorithms and chooses the best-suited algorithm to detect spammer in machine learning environment. The author uses nonnegative matrix factorization based integral framework for spammer detection in social media by implementing collaborative factorization principle [10]. In [11], the author uses extreme learning machine based supervised machine for spammer detection. A set of features are extracted by the crawler and process these datasets using extreme machine learning approach.

In [12], author proposed a trust rank based on URLs posted by various users using direct message principle. An invitation graph scheme proposed for detecting Sybil nodes in various social network platforms to analyze profile characteristics [13, 14]. In [15], the author proposed a model called COLOR + to detect spammer accounts in a social network in mobile devices by analyzing messages shared by the users. The approach proposed in [16, 17] analyze user behavior pattern according to the data interest and user behavior in a different group to detect spammer in a social network. The author identifies

various kinds of anomalies using past behavior that deviates from the current one. In [18], the author observed the model that stores various processes related to information processing in a social platform for detecting spammer. If the observation lie-down below the threshold, it said to be anomalous. After exploring all the above articles, we conclude that spammer on OSNs can be very harmful for social users and their information. They need to be detected and removed at users end. After all, we came to the conclusion that we need some suitable feature extraction algorithm and optimization technique for better feature selections to helps spammer detection.

3 Spammer Detection Framework

Spammer detection framework is depicted in Fig. 1. We collected dataset from various social networks like Facebook, Twitter, and Instagram by using our crawler and API. The dataset divided into two different sets called training data and testing data. Each dataset contains various features associated with different profiles through feature extraction mechanism.

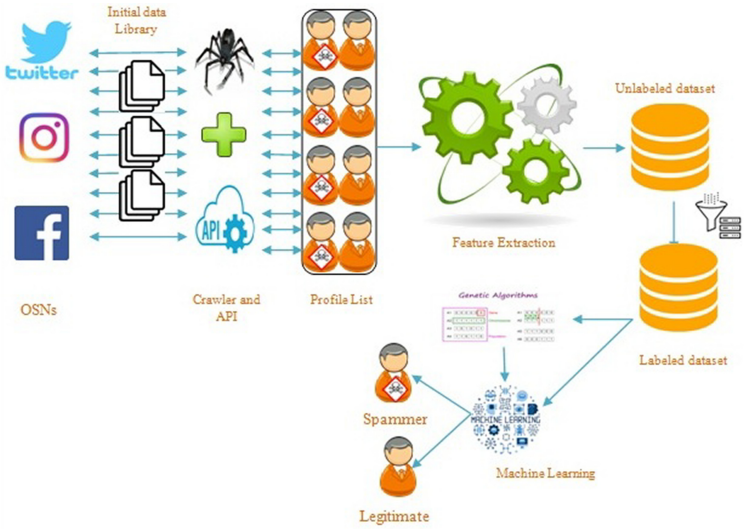

Fig. 1. Spammer Detection Framework

3.1 Data Collection

We collected dataset from various social networks by using crawler and API. Crawler runs on chrome extension to extract profile content and user information related to user profiles. API boosts up the performance of crawler to extract private information about the user. We collected more than 2500 profile information on each social network depicted in Table 1. The dataset contains both spammers as well as legitimate profile information. In addition, we collect some profile activities based on privacy principle applied to social networks.

Table 1. Collected dataset for our framework

Online Social networks	Total Profiles (Spammer + Legitimate)	Spammer	Legitimate
Facebook	3635	1523	2112
Twitter	6558	3123	3435
Instagram	2667	1083	1584

3.2 Training and Test Set Data

The collected dataset separated into two different sets called training set and test set. To obtain the model we have to conduct some experimental analysis of training data, whereas to determine the level of accuracy of the trained model testing data used. For an experimental approach, we used 10-cross-validation technique to separate the dataset into training and test sets.

3.3 Preprocessing

Transforming the raw information into a perceivable format, preprocessing is required in a machine learning approach. To detect the spammer in OSNs, our proposed model needs preprocess the generated content using various approaches like data streaming, folding approach, stopwatch removal, and tokenization.

3.4 Manual Feature Selections

Features are required as a reference to separate spammer from legitimate. Based on related work mentioned above, we select some profile and content-based features for our proposed model. We analyze the most popular features related to user profiles. For the feature extraction, we use web crawler run on chrome extension. Various features used in this research are depicted in Table 2.

3.5 Suitable Feature Selection Using GA

To detect suspicious profiles in OSNs is a challenging task. By analyzing suitable user profile and content-based features related to a user account is highly necessary for observation. The manual selection of features leads to lower accuracy and higher training time in a machine learning environment. To overcome the above issues, we used GA (Genetic algorithm) for a better selection of features. Genetic algorithms are based on evolution and natural selections to solve different diverse types of problem. The entire process of GA covered 5 different stages called initial population, selection, mating, crossover, and mutation. The algorithm starts with individuals' selection of chromosomes called population. Each chromosome consists of a sequence of genes that could be various characteristics of individual users. In the next phase crossover is used to produce next level chromosomes. At later, mutation is used to find various suitable combinations.

Table 2. Selected features from Facebook, Twitter and Instagram

Features		
Facebook	Twitter	Instagram
#Profile ID	#Profile ID	#Profile ID
#Profile Name	#Number of Tweets	#Profile name
#All friends	#Number of followers	#Number of hashtag
#Number of following	#Number of Followings	#Number of URLs shared
#Number of pages liked	#Replies on Tweets	#Sharing videos
#Number of events	#Media content shard	#Number of live video updates
#Number of participating user group	#Number of re-tweets	#Sharing stories
#Post shared	#Direct message send	#Sharing images
#Photos and #video shared	#Number of URLs shared	#Sharing notifications
#Number of tag and #Hashtag	#Back ground image	#Number of likes
#New post, #Recent post like	#Default profile view	
#Profile with photo guard	#Translator used	
#Current Location share	#Number of Hashtags	

Similar processes are carried out to find final level of features that are suitable and gives better output describes in algorithm 1. The related experimental analysis for selecting various features shows in Fig. 2. Fitness of every individual is calculated using matching percentage of every population with the normal sample.

$$Fitness(X) = \frac{A_i}{A} \tag{1}$$

Where,

A_i = number of chromosomes matching individuals

A = Total size of the chromosomes

We tested Euclidian and Minkowski distance measure formula in genetic algorithm in different trial to calculate the distance between parent and new chromosomes. To detect the malicious content, Euclidian formula also used. The distance between two chromosomes can be calculated by using Eq. (2).

$$D(X, Y) = \sqrt{(X_1 - Y_1) + (X_2 - Y_2) + \ldots (X_N - Y_N)} \tag{2}$$

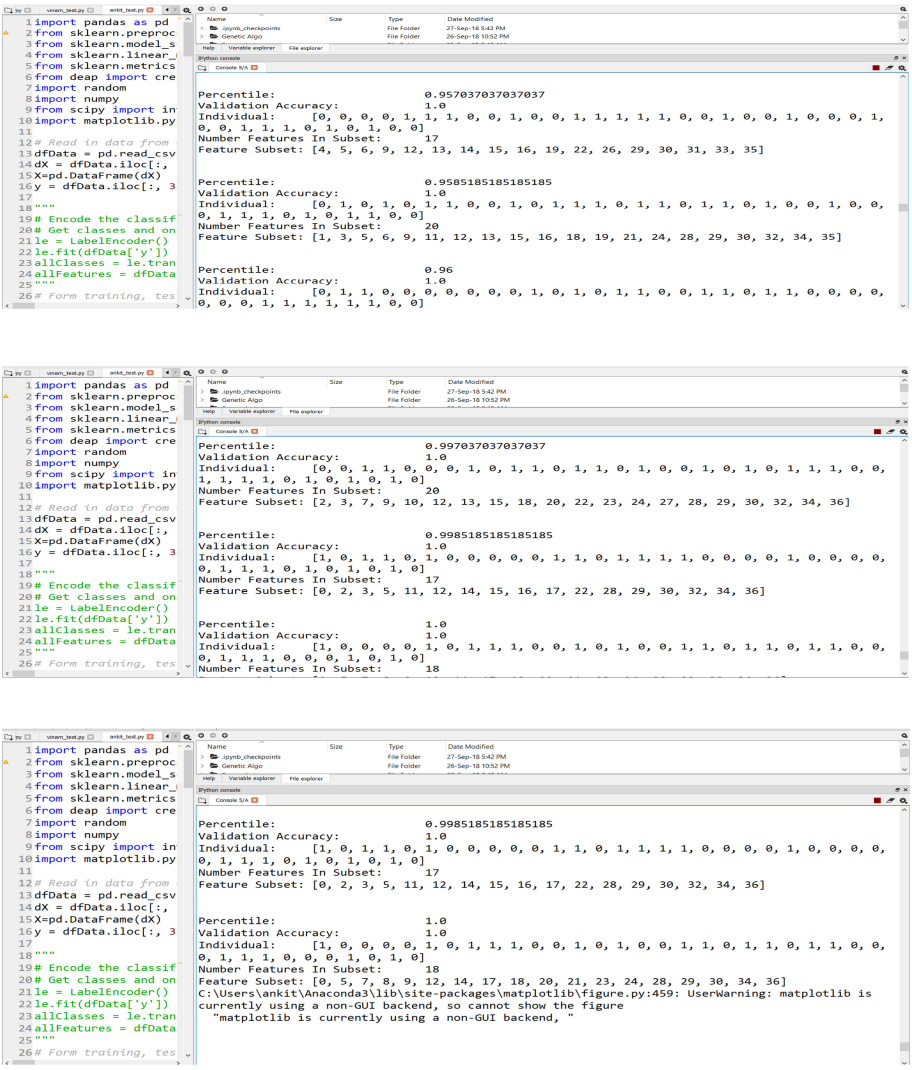

Fig. 2. Feature selection based on Genetic algorithm

To calculate the power value between chromosomes, we use Minkowski distance measure formula using p-norm dimension depicted in Eq. (3).

$$\mathrm{D(X,\ Y)} = \left(\sum_{i=0}^{N} \left(|X_i - Y_i|^p\right)\right)^{\frac{1}{2}} \tag{3}$$

Algorithm: Feature selection using Genetic Algorithm (GA)

1. Choose 'N' number of different individuals from training set
2. On continuous features run binning algorithm
3. Select random individuals from 'N' for initial population
4. For specific number of generations do
5. For choose size of population
6. Select two individuals (parents) i.e. N_1 and N_2
7. Apply crossover to produce new individual called N_3
8. Use mutation over crossover
9. Compute distance $d^1$and d^2 i.e. from (N_3 to N_1) and (N_3 to N_2)
10. Compute fitness of old and new populations as F_1, F_2 and F_3 respectively.
11. If ($d_1<d_2$) and ($F_1>F_2$) then
12. Replace N_1 with N_3
13. else
14. If ($d_2<=d_1$) and ($F_1>F_3$) then
15. Replace N_2 with N_3
16. End if
17. End if
19. End for
20. End for
21. Extract the best for operation.

3.6 Classification Based on Machine Learning Approach

The targeted classifier produced the outputs as spammer and legitimate using various features extracted by our crawler. We used various classifiers, namely Support vector machine (SVM), Random forest, bagging, J48, decision tree and Logistic Regression. To evaluate predictive models, we use 10-fold cross validation by partitioning original sample into training set and test set. The evaluation result in the form of precision, recall, true positive rate, false positive rate and ROC area observed for decision making.

4 Experiment and Result Analysis

We use different social networks dataset, which contains more than 2500 user information's. Our crawler run on chrome extension extracts profile information along with date and time of every activities posted by the user. In the evaluation process, we consider confusion matrix for spammer detection. The proposed approach is evaluated various metrics, namely true positive rate, true negative rate, precision, recall and F-score related

to classifiers. Accuracy is the ratio of total correctly classified instances of both classes over total instances in the dataset and expressed as,

$$\text{Accuracy} = \frac{True\ positive + True\ negative}{True\ positive + True\ negative + False\ positive + False\ Negative} \quad (4)$$

$$\text{Precision} = \frac{True\ positive}{True\ positive + False\ Positive} \quad (5)$$

$$\text{True positive rate (TPR)} = \frac{True\ positive}{True\ positive + False\ negative} \quad (6)$$

4.1 Data Analysis

We observed that, by using various characteristics analysis, the follower of legitimate users is more as compared to the spammers in twitter account. But, the number of likes by the user for any event is more by spammers. As expected, spammers spread more advertisements and fraudulent information's in different social network to attract users. After all, Random forest classifier gives higher accuracy as compared to other classifications. But, in Logistic regression, false positive rate is less in Twitter dataset depicted in Table 3, Table 4 and Table 5.

Table 3. Experimental analysis of Facebook dataset

Resultant								
Different Classifications	TP rate	FP Rate	Precision	Recall	F-measure	MCC	ROC Area	PRC Area
Random Forest	.994	.011	.989	.994	.992	.983	.999	.999
Bagging	.994	.012	.989	.994	.991	.982	.998	.998
JRip	.992	.014	.986	.992	.989	.978	.991	.986
J48	.991	.013	.988	.991	.990	.979	.991	.987
PART	.987	.012	.988	.987	.988	.975	.993	.991
Random tree	.985	.017	.984	.985	.984	.968	.984	.977
Logistic	.957	.008	.992	.957	.975	.955	.995	.995
SVM	.893	.017	.896	.893	.890	.898	.899	.898

Table 4. Experimental analysis of Twitter dataset

Resultant								
Different Classifications	TP rate	FP Rate	Precision	Recall	F-measure	MCC	ROC Area	PRC Area
Random Forest	0.993	0.003	0.997	0.993	0.995	0.989	1.000	1.000
Bagging	0.990	0.006	0.995	0.990	0.993	0.984	0.998	.997
JRip	0.995	0.004	0.996	0.995	0.996	0.990	0.995	.995
J48	0.993	0.006	0.995	0.993	0.994	0.986	0.995	.994
PART	0.992	0.010	0.991	0.992	0.991	0.981	0.994	.991
Random tree	0.996	0.007	0.994	0.996	0.995	0.989	0.995	.992
Logistic	0.985	0.014	0.986	0.985	0.986	0.980	0.985	.985
SVM	0.896	0.010	0.891	0.896	0.893	0.885	0.896	.893

Table 5. Experimental analysis of Instagram dataset

Resultant								
Different Classifications	TP rate	FP Rate	Precision	Recall	F-measure	MCC	ROC Area	PRC Area
Random Forest	.973	.018	.963	.961	.971	.968	.962	.962
Bagging	.942	.019	.946	.939	.949	.943	.959	.942
JRip	.962	.011	.959	.956	.960	.962	.932	.953
J48	.981	.011	.980	.981	.982	.986	.989	.984
PART	.967	.012	.968	.977	.958	.952	.963	.961
Random tree	.975	.016	.974	.975	.974	.978	.964	.967
Logistic	.967	.012	.962	.957	.965	.965	.965	.965
SVM	.873	.016	.872	.879	.870	.878	.879	.878

4.2 Performance Analysis

We evaluate our proposed framework by using various classifications and compared the analysis with some existing approaches. Particularly our experimental approach in the form of accuracy is higher as compared to other state of art techniques. It reaches higher accuracy above 99% in all social network platforms. Likewise, the precession of various analyses is higher as compared to other approaches . Especially, by using

genetic algorithm, accuracy in every case increases by 12% to 15% as compared to normal feature selection. The selection of suitable features from group of all features by GA achieved higher detection rate. In all experimental approach, SVM produces lower accuracy due to structured dataset. Comparative analysis of various classifications in different social platforms with other existing approaches like Ameen et al., (2017) [19], Ala'm et al., (2017) [20] and Herzallah et al., (2017) [21] depicted in Fig. 3, Fig. 4, Fig. 5 and Fig. 6.

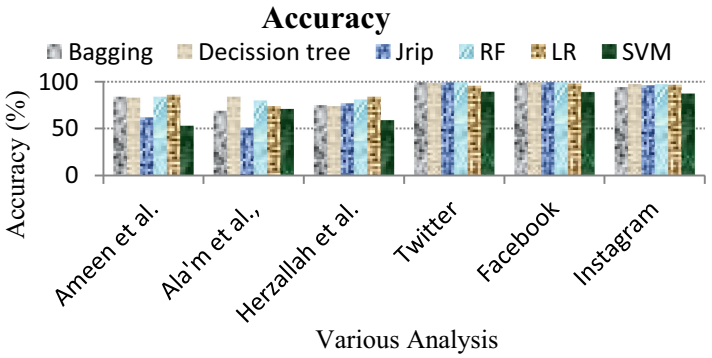

Fig. 3. Comparative analysis of Accuracy

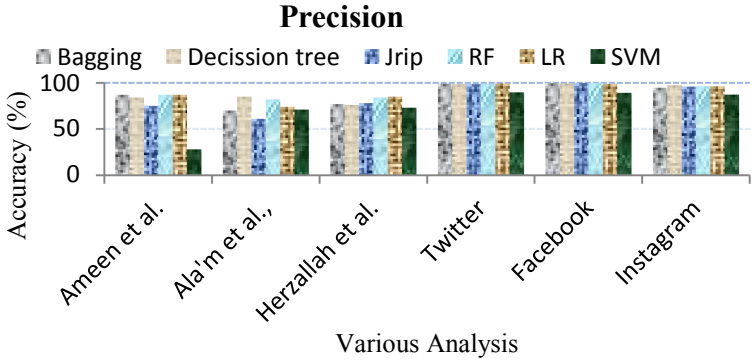

Fig. 4. Comparative analysis of Precision

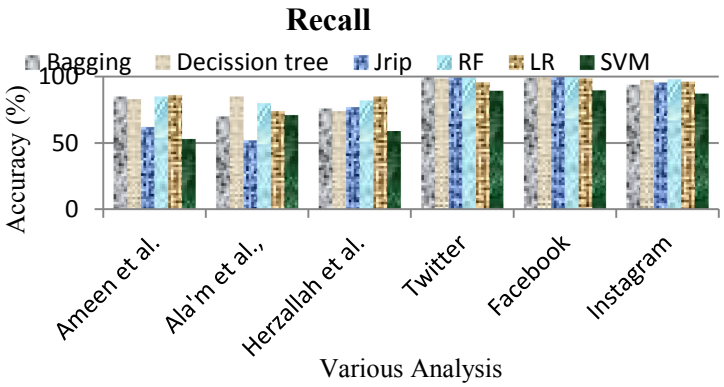

Fig. 5. Comparative analysis of Recall

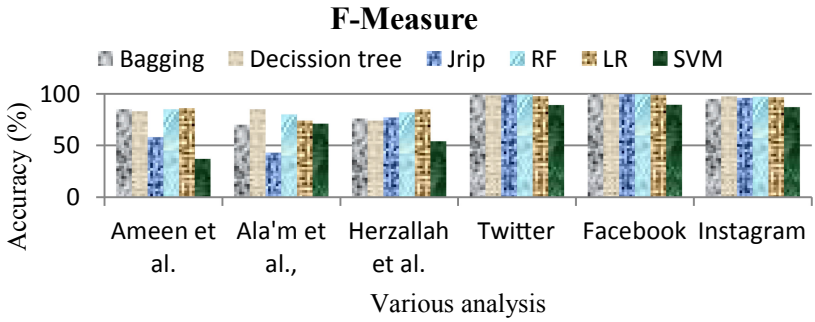

Fig. 6. Comparative analysis of F-Measure

5 Conclusion and Future Work

The paper presents a Genetic algorithm-based feature selection approach with machine learning classifier to detect spammers in social network platform. A set of content and behavioral features are collected from Twitter, Facebook and Instagram using our crawler. By investigating various user behaviors, we provided a detection mechanism to detect spammer content in OSNs. Through a set of experiment and rating with a real-world dataset, proposed framework produces better accuracy and detection rate as compared to other frameworks. Next, we plan to extend our proposed framework in the following aspects. Firstly, we consider other private features related to the users account to detect spammer. Secondly, we wish to improve the detection rate by using other optimization approaches. Finally, design an online detection mechanism, which automatically detect the spammer behavior in social network platform.

References

1. Zhang, Z., Gupta, B.B.: Social media security and trustworthiness: overview and new direction. Fut. Generation Comput. Syst. **86**, 914–925 (2018)

2. Gupta, B.B., Gupta, S., Gangwar, S., Kumar, M., Meena, P.K.: Cross-site scripting (XSS) abuse and defense: exploitation on several testing bed environments and its defense. J. Inf. Priv. Secur. **11**(2), 118–136 (2015)
3. Zhang, Z., Sun, R., Zhao, C., Wang, J., Chang, C.K., et al.: CyVOD: a novel trinity multimedia social network scheme. Multimedia Tools Appl. **76**(18), 18513–18529 (2017)
4. Brezinski, K., Guevarra, M., Ferens, K.: Population Based Equilibrium in Hybrid SA/PSO for Combinatorial Optimization: Hybrid SA/PSO for Combinatorial Optimization. Int. J. Softw. Sci. Comput. Intell. (IJSSCI) **12**(2), 74–86 (2020)
5. Harrath, Y., Bahlool, R.: Multi-objective genetic algorithm for tasks allocation in cloud computing. Int. J. Cloud Appl. Comput. (IJCAC) **9**(3), 37–57 (2019)
6. Sahoo, S.R., Gupta, B.B.: Classification of various attacks and their defence mechanism in online social networks : a survey. Enterp. Inf. Syst. pp. 1–33 (2019). http://doi.org/10.1080/17517575.2019.1605542
7. Sahoo, S.R., Gupta, B.B.: Classification of spammer and non-spammer content in online social network using genetic algorithm-based feature selection. Enterprise Inf. Syst. 710–736 (2020). http://doi.org/10.1080/17517575.2020.1712742
8. Singh, M., Bansal, D., Sofat, S.: Who is who on twitter–spammer, fake or compromised account? a tool to reveal true identity in real-time. Cybern. Syst. **49**(1), 1–25 (2018)
9. Sahoo, S.R., Gupta, B.B.: Fake profile detection in multimedia big data on online social networks. Int. J. Inf. Comput. Secur. 303–331 (2020). http://doi.org/10.1504/IJICS.2020.105181
10. Yu, D., Chen, N., Jiang, F., Fu, B., Qin, A.: Constrained NMF-based semi-supervised learning for social media spammer detection. Knowl.-Based Syst. **125**, 64–73 (2017)
11. Sahoo, S.R., Gupta, B.B.: Popularity-based detection of malicious content in facebook using machine learning approach. In: First International Conference on Sustainable Technologies for Computational Intelligence, pp. 163–176. Springer, Singapore (2020)
12. Gyongyi, Z., Garcia-Molina, H., Pedersen, J.: Combating Web spam with trust rank. In: Proceedings of the Thirteeth International Conference on Very Large Data Bases, vol. 30, VLDB 2004, pp. 576–587 (2004)
13. Xue, J., Yang, Z., Yang, X., Wang, X., Chen, L., Dai, Y.: Votetrust: leveraging friend invitation graph to defend against social network sybils. In: Proceeding of the 32nd IEEE International Conference on Computer Communications, INFOCOM 2013 (2013)
14. Alweshah, M., Al Khalaileh, S., Gupta, B.B., Almomani, A., Hammouri, A.I., Al-Betar, M.A.: The monarch butterfly optimization algorithm for solving feature selection problems. Neural Comput. Appl. 1–15 (2020)
15. Sahoo, S.R., Gupta, B.B.: Hybrid approach for detection of malicious profiles in twitter. Comput. Electric. Eng. **65–81**, 2019 (2019). https://doi.org/10.1016/j.compeleceng.2019.03.003
16. Ahmed, M., Mahmood, A.N., Hu, J.: A survey of network anomaly detection techniques. J. Network Comput. Appl. **60**, 19–31 (2016)
17. Jain, A.K., Gupta, B.B.: Towards detection of phishing websites on client-side using machine learning based approach. Telecommun. Syst. **68**(4), 687–700 (2018)
18. Kaur, R., Kaur, M., Singh, S.: A novel graph centrality based approach to analyze anomalous nodes with negative behavior. Procedia Comput. Sci. **78**, 556–562 (2016)
19. Ameen, A.K., Kaya, B.: Detecting spammers in twitter network. Int. J. Appl. Math. Electron. Comput. **5**(4), 71–75 (2017)
20. Ala'M, A.-Z., Faris, H. et al.: Spam profile detection in social networks based on public features. In: 2017 8th International Conference on Information and Communication Systems (ICICS), pp. 130–135. IEEE (2017)
21. Herzallah, W., Faris, H., Adwan, O.: Feature engineering for detecting spammers on twitter: Modelling and analysis. J. Inf. Sci. 0165551516684296 (2017)