A Hybrid Method for Spammer Detection in Social Networks by Analyzing Graph and User Behavior

Mona Najafi Sarpiri[1], Taghi Javdani Gandomani[2*], Mahsa Teymourzadeh[3], Akram Motamedi[4]

[1] Comp. Dept., Dolat Abad Branch, Islamic Azad University, Dolat Abad, Isfahan, Iran.
[2] Comp. Dept., Boroujen Branch, Islamic Azad University, Boroujen, Iran.
[3] Comp. Dept., Shahrekord Branch, Islamic Azad University, Shahrekord, Iran.
[4] Comp. Dept., Najaf Abad Branch, Islamic Azad University, Najaf Abad, Iran.

*Corresponding author. Tel.: +98-913-5201; email: javdani@ieee.org
Manuscript submitted September 6, 2017; accepted November 12. 2017.
doi: 10.17706/jcp.13.7.823-829

Abstract: With the increasing use of the Internet and social networks, there are many spammers causing security problems and numerous challenges in these services. Detection of spammers has attracted much attention in the recent years and several strategies have been proposed for detection and limitation of their activities by different researchers. However, there are still many challenges and open questions in this area which need further research. Although there are still many problems in this area needs further researches. This study proposes a graph analysis based method for spammer detection by analyzing their behaviors and their relation with the users. Finally, a solution is provided to facilitate the detection process. The aim of this paper, by applying the hybrid graph analysis method and behavior analysis, is to increase the diagnostic accuracy and detection rate with the help of appropriate classification algorithms and the most effective features. So, two scenarios were used to achieve higher accuracy level and lower false positive. The first scenario was based on using the entire data to build and evaluate the model. The results showed that despite the high precision of this approach, due to the high levels of false positive, this approach is not appropriate. In the second scenario, the ratio of the normal users to spammers was considered equal to 2 to 1 which led to satisfactory results. After reviewing the confusion matrix and false positives in different algorithms, the Logistic algorithm was chosen as an appropriate algorithm which meets the objective of this study.

Key words: Spam, spammers, spam detection, user behavior analysis, graph analysis.

1. Introduction

Spams are unsolicited messages that deal extensively with a large number of users and are sent with the aim of fraud, spreading lies, rumors and creating panic, or advertising [1]. These days, about 50% of electronic business letters are spam [2]. Spam causes problems such as the occupation of resources, waste of bandwidth and prolonging the communication time [3]. The people sending spam messages are called spammers. The first spam was sent in 1978 by Gary Fork [4]. Despite many efforts conducted by the previous researches on dealing with the spams, there are still a large number of spams on the social networks and the Internet. This shows that the accuracy and efficiency of these methods are inadequate. One way to deal with spams is to detect the spammer as producers of the spams and prevent their activities.

Spam detection algorithms generally divided into three categories: content analysis, graph analysis and,

user behavior analysis [5]. The methods of text-based content analysis work through identifying features such as keywords, number of words, repetition of content, the text of the message, and the number of links in the message body filtering spam messages. While, graph analysis based methods work through sending messages to users who have no relationship with each other, to detect spammers [6]. In analysis-based methods this is done through behavioral characteristics such as the characteristics of the post and click characteristics [7].

In the next part of this paper a complete review of the previous works and their weaknesses will be provided. In the third part of the paper the proposed approaches and model evaluation are discussed, and finally the conclusion is discussed and the future works are introduced.

2. Related Work

As mentioned previously, several studies have been carried out to detect spams and their resources. In one study, user behavior analysis method was used to discover the spammer [7]. In this work, the user performance while posting (including Posting Count, Posting Standard Deviation, Posting Intensity) and click-based features (including total clicks, click per day, effective average clicks, clicks standard deviation, max clicks) were investigated [5]. The advantage of this method is to benefit from simultaneous use of multiple behavioral characteristics. However, the disadvantage of the mentioned method is the lack of availability of the dataset created by the authors. In addition, only one classification algorithm is used in this study and the training set is done manually.

Cheng and coworkers [7] in 2015, investigated the behavioral analysis capability to detect spam URL in social networks. In [8] with J48 decision tree technique j48 and using the text of Facebook messages, the number of keywords, the average number of the words, the text length, and the number of links the spam messages were detected. The precision and recall rate achieved in this study, are 61% and 63% respectively.

In [9], on Twitter social network focusing on five aims, providing real-time filtering, increasing scalability, accurate decision making, ability to retrain the model with new data, and classification, a text independent model was developed. In this study, a dataset containing 500,000 samples was used, in which 400,000 of them were used for educational data and the remaining 100,000 were used for experiment. Training data combined with the ratio of one to one (half normal and half spam), one to four and one to ten were studied. The accuracy of 91% and false positive of 87% were achieved. The advantage of this method is that it is a real-time system that can perform the filtering function with a short delay. This has high scalability and can accurately work for big data. The disadvantage of this research is the lack of availability to the whole dataset. The researchers also did not study the other algorithms. Since in the existing dataset, spammers are lower than normal users, creating and evaluating of a model from all the records in the dataset is not possible. It is mainly because the model considers only characteristics/features of normal users and ignores spammer. To solve this problem, the ratio of different classes has to be considered carefully.

3. The Proposed Method

Social networks and relationships between users can be considered as bi-directional graph vertices in the graph, users and edges of the graph, the relationship between users [6].

Spam messages are played in random order and in a wide range. The possibility that the messages are sent to members of a group is very low. Because of the random sending, messages are sent to members of groups that have no relation to each other.

Formula (1) is used to show clustering coefficient of used detection spam:

$$CC\ (x_i) = n_{xi} / ([Z_{xi}\ (Z_{xi} - 1)] / 2) \quad (1)$$

where n_{xi} number of neighbors of node x_i, Z_{xi} the number of edges (links) between x_i and its neighbors.

The proposed method uses a set of behavioral features in addition to clustering coefficient,

The features that check users' performance when sending posts are the number of posts, Standard Deviation of posts, and intensity of posts. The number of posts shows number of messages sent by a user to show others. Standard Deviation shows the number of posts per day.

Post Intensity can be obtained as follows:

$$\text{Intensity (u)} = \frac{postCount\,(u)}{(std\,(u) * |set\,(postDays(u))|) + 1} \quad (2)$$

where PostCount is number of posts submitted by user u, std is Standard Deviation of submissions of user u, and | set (PostDays) | indicates the number of days the user was sending messages [7].

3.1. Workflow

CRISP-DM methodology was used in this study is a standard model [10]. Fig. 1 shows the adopted steps in this research study.

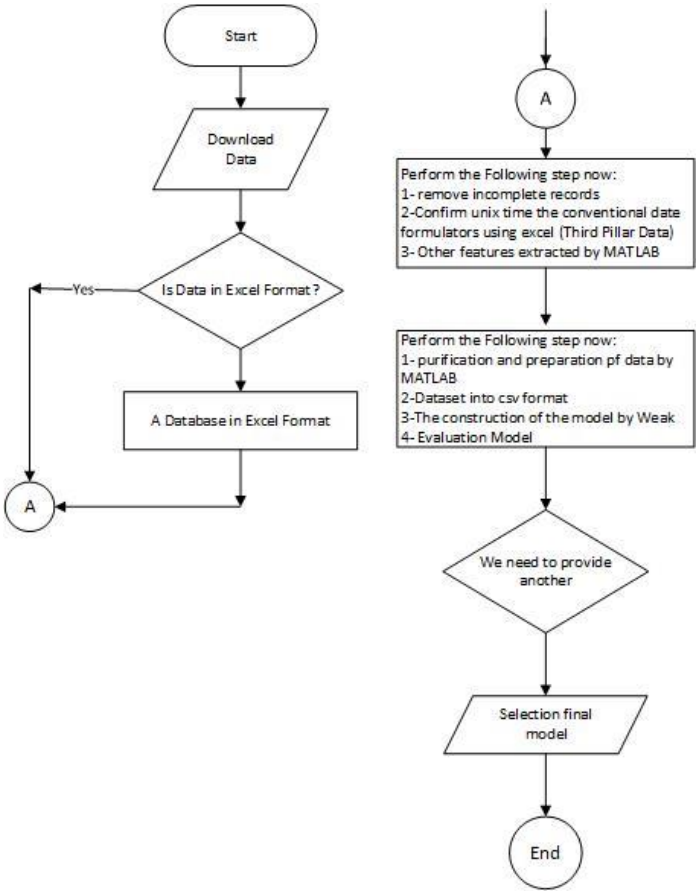

Fig. 1. The steps adopted in this study.

3.2. Create Features

Data used in this study have been received from the Max Planck Institute [11]. This research institute works on various research fields, especially in the areas of cybercrime and social networks crimes. The

available dataset lists all Facebook posts in New Orleans from 2004 to 2009, containing 876,993 posts. This dataset contains only the first three features 1st user ID, 2nd user ID, and time of submission.

For the application of data mining techniques according to the formula presented in the previous section, we need to extract new features. The new features for each user include the number of messages, the number of neighbors, the number of links between neighbors, clustering coefficient, standard deviation of posts, intensity of posts, and class.

The new dataset contains 8 columns: 1- User ID 2-Number of Posts/ messages 3- number of neighbors 4- number of links between neighbors 5- Standard Deviation of posts 6- Intensity of posts 7-clustering coefficient 8-class.

4. Model Evaluation

For data preprocessing, model building and evaluation, samples were taken from different classes with different ratios.

In the first scenario, the entire data was used for model building and evaluation. In the second scenario, the ratio of one to two has been used; i.e. number of normal users is twice more than the number of spammers. It is mainly because about 30 to 50 percent of messages are spam messages.

4.1. First Scenario

In this scenario the whole data were used for model building and evaluation. After data processing by MATLAB and extracting the required features, dataset contained 14044 records in which 14044 records were normal users and 257 of them were spammers.

To build the model and test it, 70% of the samples were used as training data to construct a model and 30 percent were used as samples for testing.

From 14301 available records 10010 records were used as training data, and 4290 records were considered as test records. The accuracy of different algorithms is given in Table 1.

Table 1. Comparison of Different Algorithms in the First Scenario

TP	FP	Precision	Recall	Algorithms
0.981	0.981	0.963	0.981	Bayesnet
0.974	0.975	0.964	0.974	NaiveBayes
0.981	0.981	0.963	0.981	Knn100
0.981	0.981	0.963	0.981	Neural Net
0.981	0.981	0.963	0.981	j48
0.981	0.981	0.963	0.981	logistic

According to the above table, the precisions of the algorithms in most cases are equal and the precision is very high as well. Matrix analysis is used to analyze the results in which effectiveness of the respective algorithms are shown. Each column of the matrix shows an example of the predicted value, while each row contains an actual value. The numbers on the diagonal of the matrix display the number of correct classifications. When all the numbers not on the main diagonal are zero, the algorithm has the maximum accuracy. False positive tells how many spammer users are detected as normal. Therefore, the amount of false positives needs to be reduced as far as possible. Table 2 presents the false positive of different algorithms in the first scenario.

Table 2. False Positive of Different Algorithms in the First Scenario

Algorithm	Neural Net	AdaBoost	NaiveBayes	J48	Knn-100
TP	0.981	0.981	0.974	0.981	0.981
FP	0.981	0.981	0.957	0.981	0.981

According to the Table 2, high rate of false positive indicates that the result of the first test does not make sense.

4.2. Second Scenario

In the second scenario, the normal user records were twice more than the spammer's user records. The number of records was 257 for the spammers and 515 for the normal users. Also, 70% of the data were training data and 30% were experimental data. From the 772 available records, 540 records belonged to training and 232 records belonged to test. The accuracy of each of the algorithms are given in Table 3.

Table 3. Comparing the Accuracy of Different Algorithms

Precision	Recall	Accuracy	Algorithm
0.829	0.832	83.18%	BayesNet
0.807	0.754	75.43%	NaiveBayes
0.983	0.983	98.28%	Neural Net-Epoch1000
0.962	0.961	96.12%	J48
0.988	0.987	98.71%	LMT
0.966	0.966	99.60%	Logistic

According to Table 4, Logistic algorithm correctly detected all the spammer users and only one of the normal users was incorrectly detected as a spammer user. According to the accuracy of different algorithms, the Logistic algorithm is proposed as an appropriate algorithm. LMT algorithm has high efficiency since all 77 spammer records are correctly identifies as spammer. This algorithm identified 3 of the normal users as spammers. Neural network algorithm identified only one of the spammer users incorrectly as normal user.

Table 4. Algorithms Confusion Matrix

algorithm	a	b	
NaiveBayes	108	47	a=normal
	10	67	b=spammer
Logistic	154	1	a=normal
	0	7	b=spammer
Neural Net-Epoch1000	152	3	a=normal
	1	76	b=spammer
J48	154	1	a=normal
	8	69	b=spammer
LMT	152	3	a=normal
	0	77	b=spammer

In Table 5, the accuracy of different algorithms in the second scenario with the experimental data are compared.

Table 5. The Accuracy of the Algorithm on the Experimental Data

Algorithm	Test Samples	classification
Logistic	231	Correctly classified samples
	1	Incorrectly classified samples
LMS	229	Correctly classified samples
	3	Samples incorrectly classified
J48	223	Correctly classified samples
	9	Incorrectly classified samples
	232	Total samples

According to Table 6, in the second scenario Logistic algorithm achieved 99.569 %accuracy.

Table 6. Logistic Algorithm

rate	Characteristic
0.996	True Rate
0.002	False Rate
231	The number of records classified correctly
1	The number of misclassified records
77	The number of spammer records classified correctly
0	Misclassified spammer records
154	The number of normal records classified correctly
1	The number of normal records incorrectly classified
232	Total records
99.569%	Algorithm accuracy

Table 7 compares the proposed algorithm with the other methods on the adopted dataset.

Table 7. Comparing the Proposed Algorithm with other Methods

Research	Year	Method	Accuracy
Deylami [6]	2013	Graph analysis	80.60%
This study	2017	The proposed method	99.57%

5. Conclusion

In this study, a method has been proposed and described for classification of social network users and required features for data mining and extraction were explained. In this study, two scenarios were used to access higher accuracy and lower false-positive. In the first scenario, the entire data were used to build and evaluate the model and the best achieved accuracy was 98.13%. Despite the high accuracy of the test, the method was not able to detect any spammer users which were identified by the high value of false positive (0.981). Analyzing the confusion matrix showed that this model cannot be applicable. In the second scenario, the ratio of two to one for spammer to normal users was used. The results were evaluated for different algorithms. After reviewing the confusion matrix and false positive, finally LMT algorithm was shown as an appropriate algorithm. However, research in the field of spam detection in social networking and the spammer users still requires further research.

6. Future Work

In addition to the proposed method, more features can be used for spammer detection, especially the user-based features such as the number of clicks, the average number of clicks, and the maximum number

of clicks. Studying different tests by combining different ratios can lead to higher precession. This type of classification is also a new method that has been addressed in recent researches.

References

[1] Jinda, N., & Bing, L. (2008). Opinion spam and analysis. *Proceedings of the 2008 International Conference on Web Search and Data Mining*.

[2] Symantec Intelligence Report. (2015). Retrieved from intelligence-report-06-2015.en-us.pdf

[3] Heydari, A., Tavakoli, M., Salim, N., & Heydari, Z. (2015). Detection of review spam: A survey. *Expert Systems with Applications*, *43(7),* 3634-3642.

[4] Shanjani, A. A. (2010). *Legal Issues Pornography in Cyber Space.* Tehran: A New Way.

[5] Spirin, N., & Han, J. (2012). Survey on web spam detection: Principles and algorithms. *ACM SIGKDD Explorations Newsletter, 13(2)*, 50-64.

[6] Deylami, H. M., & Singh, P. Y. (2013). Cybercrime detection techniques based on support vector machines. *Artificial Intelligence Research, 2(1)*, 1-12.

[7] Cao, C., & Caverlee, J. (2015). Detecting spam URLs in social media via behavioral analysis. *Proceedings of Advances in Information Retrieval*, Springer, 703-7014.

[8] Soiraya, M., Thanalerdmongkol, S., & Chantrapornchai, C. (2012). Using a data mining approach: Spam detection on facebook. *International Journal of Computer Applications, 58(13),* 26-31.

[9] Thomas, K., Grier, C., Ma, J., Paxson V., & Song, D. (2011). Design and evaluation of a real-time url spam filtering service. *Proceeding of IEEE Symposium on Security and Privacy (SP)*.

[10] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-Step Data Mining Guide.*

[11] Viswanath, B., Alan, M., Meeyoung, C., & Krishna, G. P. (2009). On the evolution of user interaction in facebook. *Proceedings of the 2nd ACM Workshop on Online Social Networks*.

Mona Najafi Sarpiri currently is a master of science candidate in the software engineering program. She is working on working on spammer detection in social networks and agile project management.

Taghi Javdani Gandomani is an assistant professor in Islamic Azad University which holds a Ph.D in software engineering from Universiti Putra Malaysia (UPM). He obtained his B.S and M.S from Isfahan University of Technology, Iran and Isfahan University, Iran, respectively. He is also leader of a research group working on various aspects of software engineering. His main research interests are agile software development, empirical software engineering, human aspects of software engineering, and software cost estimation.

Mahsa Teymourzadeh received her B.S from Islamic Azad University, Khorasan Branch, Iran, and her M.S degree from Islamic Azad University, Shahrekord Branch, both in software engineering. Her research interests are software engineering, software quality, and mobile application development. Currently, she works as a freelance researcher.

Akram Motamedi holds a M.S in software engineering from Islamic Azad University, Najaf Abad Branch. She works as a freelance researcher working on spam and spammer detection, genetic algorithms, and data mining.