International Journal of
Applied Mathematics,
Electronics and Computers
ISSN:2147-8228 www.atscience.org/IJAMEC
Advanced Technology and Science
Original Research Paper

Detecting Spammers in Twitter Network

Aso Khaleel Ameen[1], Buket Kaya*[2]

Accepted : 27/12/2017 Published: 31/12/2017 **DOI: 10.18100/ijamec.2017436078**

Abstract: The goal of Twitter is to allow friends communicate and stay connected through the exchange of short messages. However, sometimes, spammers also use Twitter as a platform to post malicious links, send unsolicited messages to legitimate users, and hijack trending topics because of two problems of Twitter. These problems are the possibilities to automatically receive following users' updates and to write on followers' profile pages. For this reason, spam is becoming an increasing problem on Twitter day after day as other online social network sites are. In this article, we present several methods to detect spam tweets on Twitter. For this purpose, we utilize Naive Bayes, Random Forest J48, and IBK algorithms. The experiments conducted on real Twitter accounts demonstrate that the Random Forest algorithm gives us the best result to detect spammers in Twitter.

Keywords: *Spammer Detection, Spam Tweets, Classification Algorithms*

1. Introduction

In recent years, communication has shifted to a different dimension with rapidly changing technology. Today, the most common communication tools are based on social networks. Parallel to the widespread use of the Internet, social media tools have become an indispensable part of our lives. Nowadays, most of users employ social media networks, such as Twitter, Facebook and Linkedin worldwide for promoting their relations and the number of users using these methods is increasing rapidly day after day. However, some of these users, called as Spammers, started to misuse social media platforms by spreading misinformation, malicious links, unsolicited messages, fake news to legitimate users.

Twitter is one of the most well-known social media network. It is continuously under attack by Spammers. According to Twitter, there are seven spammer behaviours [1]. These are: (1) to post malicious links, (2) to exhibit mass following behaviour, (3) to send unwanted messages to users by @ reply or @ mention functions, (4) to send duplicate messages, (5) to construct multiple accounts by manually or automated tools, (6) to send comments repeatedly about hot topics, (7) to post link with unrelated tweets.

In this article, we employ classification methods to detect spammers in Twitter network. For this purpose, we consider Naive Bayes, Random Forest, J48 and IBK algorithms. By using real Twitter accounts, we compare the results of these amgorithms.

We present the approaches used to detect spam Tweets in several domains in Section II. Section III explains to the used data set. Section IV defines user attributes. Section V contains the methods. Section VI gives results and finally Section VI presents conclusion and future work.

[1] *Dept. of Software Engineering, Fırat University, Elazig – TURKEY*
[2] *Dept. of Electronics and Automation, Fırat University, Elazig – TURKEY*
* *Corresponding Author: Email: bkaya@firat.edu.tr*

2. Related Work

The issue of spam within tweet messages and other models is a problem that we face in our daily lives. So many researchers have been interested in offering solutions based on their findings to detect spam. However, the detection of spam in Twitter messages and other networks has not been widely studied so widely and has not been finalized.

Benevenuto et al. [2] listed a set of 62 features for spammer detection and used Support Vector Machine method. Ahmed and Abdulaish [3] used Markov Clustering to detect the features of spam on Facebook social networks. Their studies are based on a real group of good formations and SPAM within Facebook. They found three categories. The first category represents clusters containing purely spam profiles, and the second category includes clusters consisting purely normal profiles, and final groups are a combination of both SPAM and normal features [3]. Experimental results show that majority voting not only reduces the number of clusters to a minimum, but also increases the performance values. Bhat et al. [4] investigated the performance of some ensemble learning methods using community-based structural features for detecting spammers in online social networks. Miller et al. [5] in this area provided three contributions in the detection of these spams on Twitter. Firstly, previous studies have approached spam detection as a classification problem, whereas they view it as an anomaly detection problem secondly, they offer 95 one-gram features from tweet text for detecting these messages, and finally they used a stream of tweets in real time as well as the user's profile data with two stream clustering algorithms, DenStream and StreamKM++. They used the data set for this study consisting of 3239 accounts for users with sample tweets from each account. After several attempts, they saw that these two approaches has achieved 97.1% accuracy and 84.2% F-Measure and 94.0% accuracy and 74.8% F-Measure, respectively [5].

Eshraqi and his colleagues [6, 7] summarize several works prior to the detection of spam within these networks. They used several algorithms, and concluded that Naive Bayes algorithm gave

better results from other algorithms. They also used a clustering algorithm to detect spam messages within frequent Tweets, HTTP links, responses, signals and trending topics in Twitter, and concluded that the proposed approach could determine 89% of the spam tweets available within this data.

Gupta and Kaushal [8] studied on detection the spam message within followers, addresses, spam, and replacements and Hashtags in Twitter social network, and used a data set consisting of a log named of 1,064 Twitter users. The data comprises of 62 features containing user specific and tweet specific information. Three algorithms, Naive Bayes, Clustering and Decision Tress was combined to achieve a higher spam detection accuracy. This integrated approach outperforms other classical approaches in terms of overall accuracy and detect Non-Spammers with 99% accuracy with an overall accuracy of 87.9% [8].

Meda et al. [9] applied Random Forest approach for detection of spammers on Twitter. Finally, Xu and his colleagues [10] used a new perspective in this area by combining Twitter and Facebook, each containing an unwanted message, using five algorithms, namely Naive Bayes, J-48, Random Tree, Random Forest and Logistic. Their results show that Random Forest has the best performance with 94.7% accuracy and 66% recall for Twitter spam data set and with 97.7% accuracy and 84.4% recall for Facebook spam message group.

3. Data Collection

Since the role of the contact in general is to save time and the use of technology for the welfare of mankind, a phenomenon is a breakthrough in ways of communication between people. It is available within the communication between people in several ways, including the tweets, which is the transfer of data in the form of text between users.

In our work, we used real Twitter messages for the classification task. Tweets are collected by using Twitter Streaming API [11]. This process is crucial for obtaining a large amount of Tweet contents. Twitter Streaming API returns Twitter reports in JSON format and from here the extraction of the features such as tweets contents, users' ID, tweets date and time is simple task.

4. Identifying Features

Unlike ordinary Twitter users, people who use spam often target commercial intent (such as advertising), ideas, and system reputation [12]. Because non-spammers and spammers have various targets in the system, we look forward to them to also differ from how they behave (for example, who interact with them, how often they interact, etc.) to obtain their targets. Intuitively, we expect to spend junk and more time interacting with other users, doing actions like reply, re-twitter, posting a case without URL, etc. In order to verify this intuition, we analyze returned sets of attributes that reflect the behavior of users within the system, as well as content characteristics published by users.

4.1. Content Attributes

Content attributes are properties of user tweets that capture specific properties related to how users write tweets. Because users typically publish many tweets, we analyze the characteristics of Twitter content based on the age (days) of an account from its creation to the time of sending the most recent tweet, the number of re-tweets for this tweet, the number of followers of this twitter user, the number of following/friends of this twitter user, the number of favorites for this twitter user, the number of user mentions included in this tweet, the number lists added for this twitter user, the number of tweets sent for this twitter user, the number of hash tags included in this tweet, the number of URLs included in this tweet, the number of characters in this tweet and the number of digits in this tweet. Thus, in total we obtain 13 attributes related to content of the tweets.

5. Spam Detection Methods

So far, many classification algorithms have been proposed to detect spammers on online social networks. In this section, we tried four algorithms, Naive Bayes, Random Forest, J48 and IBK, and applied to the dataset we discussed in the section Data Collection. Then, we compared the results obtained from the four algorithms to reach the best result.

5.1. Naive Bayes

It is one of the methods of mathematical inference where we have a society or phenomenon following a probability distribution based on an unknown fixed parameter. We want to estimate the period of this parameter or test a particular hypothesis through random sample data taken from this community where we have preliminary probability information, (Before sampling) for this parameter, which takes different values to be a random variable with a probability distribution obtained using this theory, and the probability distribution produced after the sample is known as the posterior distribution, which is a summary of the data obtained from the sample in addition to the information Tribal than we can estimate this anonymous parameter, it is largely used because it often outperforms to a greater extent on sophisticated classification methods [13].

$$P(C/F) = P(C/F) * P(C) \,/\, P(F) \quad (1)$$

In Eq. (1), F is a vector of the attributes of a user, and C is the class (Spammer/Non-Spammer) of the user. To clarify the application of this rule, we will first lay out four theoretical points [13, 14]:

The first point is that the value of the probability calculation is always a value of zero and one. If the result was zero, this meant that it was an impossible event. If its value is one, it means that it is an inevitable event. But in most cases, it is between these two values.

The second point is that we always calculate the probability of events. This word is an event. The full process for which we calculate the probabilities is called experiment. This is also an official term.

The third point is Bayes himself. Which is to calculate the probability of a first event in terms of a second event that has already occurred, we must divide the value of the probability of occurrence of the first and second events together on the value of the probability of the second event.

The fourth point relates to the calculation of probabilities in the numerator and denominator in the previous law, we are in most cases not to receive it. But we must count it in some way. A technique that enables us to do so is the Tree Planning technique. We write all possible events as branches of a tree. The philosophy of this technology is known through digital stores on the Internet and even in public life in general.

5.2. Random Forest

Randomized forest is a strong and fully automated learning technique. It almost no needs any data preparation, or any

modeling experience, and enables analysts to get amazingly effective models.

In random forest approach, many decision trees are created. Each note is fed to each decision tree [15]. The generality prevalent result is used for every memorandum and final output. New control is fed in all trees and a majority vote is taken for each rating model, the error is estimated in cases that are not used while building the tree. OOB (Out-of-bag) is called prediction error that is received as a percentage [16].

The basilar syntax to the creation of random forests in R is Random Forest (formula, data). Formula is a formula that describes the expected variables and their response. Data is the noun of the collection data existence used.

5.3. J48

J48 is an ID3 extension which is an algorithm utilized to classify data. The input of the algorithm is a collection of data and output. It is a classifier rule that become strong as to group new information that was not previously used as a type of prediction for this new data. This classifier is in the form of a tree structure, they can be converted into a set of rules, so they are also called decision rules [17].

It is based on the idea of the principle of dividing the problem into parts, solving whole of them separately and assembling the solutions.

5.4. IBK

K-Nearest Neighbor method performed in the WEKA tool is the IBK classifier [18], in order to assess any problems generally it considers three important aspects, ease of interpretation of the output, time calculation, and predictive power [19].

The basis of the algorithm is the allocation of membership as a circle of the vector distance from K- nearest neighbor and the membership of that adjacent in the classes possible [20].

6. Experimental Results

As mentioned above, the aim of this paper is to detect spammers using Naive Bayes, Random Forest, IBK and J48 algorithms. In order to test the classification methods, we used Benevenuto et al.'s online dataset [2]. From this dataset, we selected 355 spammers and 828 non-spammers as shown in Table 1. In the experiments, we used WEKA tool for the purpose of applying algorithms to dataset.

Table 1. Dataset

Tool	File Format	# Spammers	# Non-Spammers
WEKA	ARFF	320	760

The dataset was tested using two methods for measuring accuracy. At the first method, 70% of the users is determined as training and the remaining 30% is used as test. As the second method, at k-fold cross validation mode, the data is divided to 10 folds, a fold is used for the test in every run and remaining folds are used as training.

6.1. Measure of accuracy and error rate using split

The results of accuracy and error are as in Table 2.

Table 2. Accuracy and error rate in split

Algorithm	Accuracy (Split 70%)	Error rate (Split 70%)
Naïve Bayes	56.41%	43.58%
J48	89.66%	10.33%
Random Forest	92.95%	7.05%
IBK	80.33%	19.66%

The Random Forest achieves the highest accuracy (92.95%) and the lowest error rate (7.05%) using percentage split. On the other hand the results in J48 classifier less accurate than that in Random Forest, where accuracy (89.66%) and error rate (10.33%). Then comes IBK with accuracy (80.33%) and error rate (19.66%). But, the Naïve Bayes classifier achieves the lowest accuracy (56.41%) and the highest error rate (43.56%) comparing with other three classifiers.

6.2. Measure of accuracy and error rate using cross validation

The results of accuracy and error rate are as in Table 3.

Table 3. Accuracy and error rate in cross validation

Algorithm	Accuracy (10 Folds CV)	Error rate (10 Folds CV)
Naïve Bayes	55.48%	44.48%
J48	89.92%	10.07%
Random Forest	93.13%	6.87%
IBK	81.24%	18.75%

The Random Forest achieves the highest accuracy (93.13%) and the lowest error rate (6.87%) using cross-validation. On the other hand the results in J48 less accurate than that in Random Forest, where accuracy (89.92%) and error rate (10.07%). IBK results are less than from Random Forest and J48 and greater than from Naïve Bayes classifier, where accuracy (81.24%) and error rate (18.75%). But, the Naïve Bayes classifier achieves the lowest accuracy (55.48%) and the highest error rate (44.48%) comparing with other three algorithms.

Lastly, time needed to build the models in seconds in cross validation and split is as in Table 4.

Table 4. Time needed to build the models in seconds

Algorithm	Split 70%	10 folds CV
Naïve Bayes	0.04	0.12
J48	1.01	1.77
Random Forest	8.21	8.71
IBK	0.01	0.01

In second set of the experiments, we decreased the training data to 60%. The results of IBK classifier with accuracy (80.63%) and error rate (19.37%) in Fig. 1 and with accuracy (80.89%) and error rate (19.11%) in Fig. 2 are less than the results of Random Forest and J48 classifier. From Fig. 3, we can say that the IBK algorithm is superior to three algorithms Naive Bayes, J48 and Random Forest in time need.

In addition, Naïve Bayes comes after IBK on execution time, but achieve the lowest results in Fig.2 with accuracy (55.48%) and the highest error rate (44.48%) using cross-validation. also lowest results with accuracy (57.23%) and highest error rate (42.77%) using percentage split in Fig. 1.

J48 comes pre-final- on time build, but in Fig.1, it comes after Random Forest for achieving the result where accuracy (89.66%) and error rate (10.33%) using percentage split, also in Fig.2, it comes after from Random Forest with accuracy (89.92%) and error rate (10.07%) using cross-validation.

Finally, Random Forest comes in the last order of time building in Fig. 3 and Fig. 1 with the highest (92.95%) and the lowest error rate (7.05%) comparing with other three classifiers using percentage split. It achieves the highest accuracy (93.13%) and the lowest error rate (6.870) comparing with other three classifiers using cross-validation in Fig. 2.

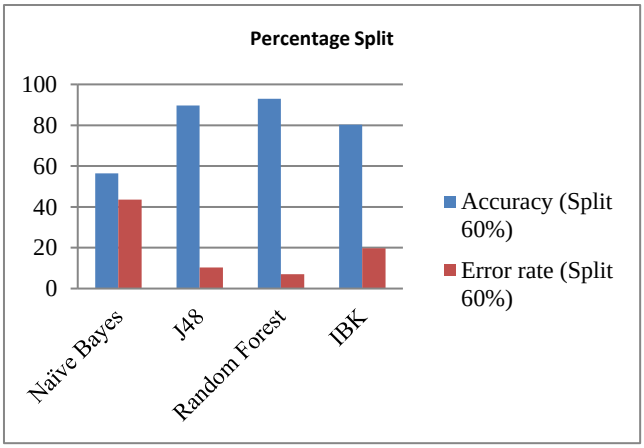

Fig. 1. Results of percentage split (accuracy and error rate)

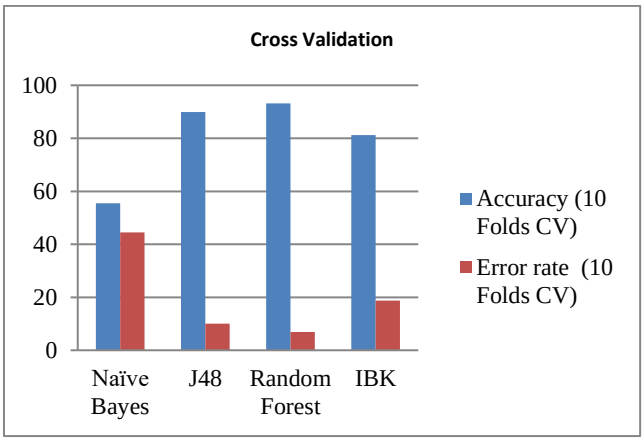

Fig. 2. Results of cross validation (accuracy and error rate)

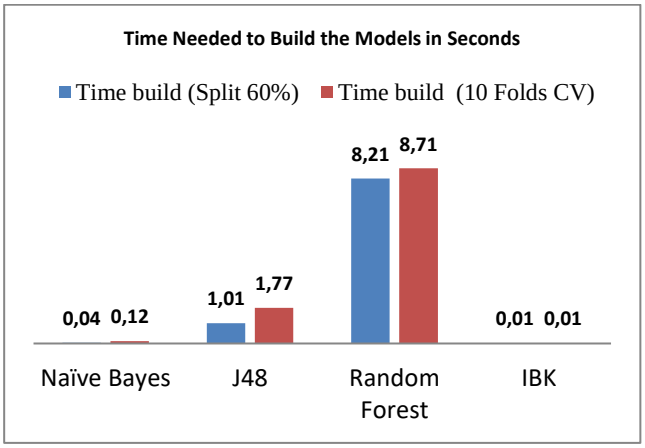

Fig. 3. Results of time build (percentage split and cross validation)

7. Conclusions and Future Work

In this paper, we compared four algorithms Naive Bayes, J48, IBK, and Random Forest to detect spammers which consists of dataset of 1183 users: 355 spammers and 828 non-spammers for the first experiment. Random Forest showed the best result compared to the other three algorithms J48, Naive Bayes, and IBK. The Random Forest classifier has the highest accuracy (92.95%) and the lowest error rate (7.05%) in percentage split and it achieves in cross validation the highest accuracy (93.13%) and the lowest error rate (6.87). But IBK achieves the highest score in execution time, where on percentage split (0.01) second and on cross validation (0.01) second. So, we can say that usage of these algorithms are significantly effective for detecting spammers in Twitter network.

Experimental results confirmed the effectiveness of the considered features. In the future, we plan to build a multifunctional framework that can be efficiently classifying various kind of spam within various networks, and also working on modifying algorithms or dataset for best results.

References

[1] Reporting Spam on Twitter, https://help.twitter.com/tr/safety-and-security/report-spam, (Last accesed 29 December 2017).

[2] F. Benevenuto, G. Magno, T. Rodrigues and V. Almedia, "Detecting Spammers on Twitter, 7th Annual Collaboration, Eletrocnic Messasging", Anti-Abuse and Spam Conference, Washington, USA, 2010.

[3] F. Ahmed and M. Abulaish,. "An mcl-based approach for spam profile detection in online social networks". In Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on (pp. 602-608). IEEE, 2012.

[4] S. Y. Bhat, M. Abulaish and A.A. Mirza., "Spammer classification using ensemble methods over structural social network features". In Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on (Vol. 2, pp. 454-458). IEEE, August 2014.

[5] Z. Miller, B. Dickinson, W. Deitrick, W. Hu, and A.H. Wang, ,. "Twitter spammer detection using data stream clustering". Information Sciences, 260, pp.64-73, 2014.

[6] N. Eshraqi, M. Jalali and M.H. Moattar. "Spam detection in social networks: A review". In Technology, Communication and Knowledge (ICTCK), 2015 International Congress on (pp. 148-152). IEEE, 2015.

[7] N. Eshraqi, M. Jalali and M.H. Moattar. “Detecting spam tweets in Twitter using a data stream clustering algorithm”. In Technology, Communication and Knowledge (ICTCK), 2015 International Congress on (pp. 347-351). IEEE, 2015.

[8] A. Gupta, and R. Kaushal. “Improving spam detection in online social networks”. In Cognitive Computing and Information Processing (CCIP), 2015 International Conference on (pp. 1-6). IEEE, 2015.

[9] C. Meda, F. Bisio, P. Gastaldo, and R. Zunino. “A machine learning approach to Twitter Spammers Detection”, 2014 International Carnahan Conference on Security Technology (ICCST), Italy, 2014

[10] H. Xu, W. Sun, and A. Javaid. “Efficient spam detection across Online Social Networks”. In Big Data Analysis (ICBDA), 2016 IEEE International Conference on (pp. 1-6). IEEE, 2016.

[11] Twitter Streaming API, https://dev.twitter.com/docs/api/streaming.

[12] P. Heymann, G. Koutrika, and H. Garcia-Molina. “Fighting spam on social web sites: A survey of approaches and future challenges”. IEEE Internet Computing, 11(6), 2007.

[13] http://www.saedsayad.com/naive_bayesian.htm, Naive Bayesian, acssessd on May 2, 2017.

[14] T.R. Patil and S.S. Sherekar, “Performance analysis of Naive Bayes and J48 classification algorithm for data classification”. International Journal of Computer Science and Applications, 6(2), pp.256-261, 2013.

[15] L. Naidoo, M.A. Cho, R. Mathieu and G. Asner. “Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment”. ISPRS Journal of Photogrammetry and Remote Sensing, 69, pp.167-179, 2012.

[16] P.K. Korir, P. Geeleher and C. Seoighe. “Seq-ing improved gene expression estimates from microarrays using machine learning”. *BMC bioinformatics*, *16*(1), p.286, 2015.

[17] G. Kaur and A. Chhabra. “Improved J48 classification algorithm for the prediction of diabetes”. International Journal of Computer Applications, 98(22), 2014.

[18] D.W. Aha, D. Kibler and M.K. Albert.”Instance-based learning algorithms”. Machine learning, 6(1), pp.37-66,1991.

[19] T. Srivastava, “Introduction to k-nearest neighbors: Simplified”https://www.analyticsvidhya.com/blog/2014/10/introduction-k-neighbours-algorithm-clustering/ [Accessed: Nov. 3, 2017].

[20] J.M. KellerGray, M.R. and J.A. Givens. “A fuzzy k-nearest neighbor algorithm”. IEEE Transactions on Systems, Man, and Cybernetics, (4), pp.580-585, 1985.