International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 2193-2197 www.ijaem.net ISSN: 2395-5252

Identifying Trolls and Determining Terror Awareness Level in Social Networks Using Data Mining

Mrs. SMITHA P, HARSHITHA G N

Department of information science and engineering east west institution of technology bangalore,karnataka,india

Department of information scienceand engineering east west institution of technology bangalore,karnataka,india

----

Submitted: 15-06-2022 Revised: 25-06-2022 Accepted: 27-06-2022

ABSTRACT - Online social media applications have become an integral part of our everyday life. Not only are they being utilize by individuals and legitimate businesses, but also recently several organized groups, such as activists, hacktivists, and cyber- criminals have adopted them to communicate and spread their ideas. This represents a new source for intelligence gathering for law enforcement for instance, as it allows them an inside look at the behaviour of these previously closed, secretive groups. One possible opportunity with this online data source is to utilize the public exchange of social-media messages to identify key users in such groups.Trolls in social media are 'malicious' users trying to propagate an opinion or distort the general perceptions. Identifying trolls in social media is a task of interest for applications since data cannot be analyzed effectively without eliminating such users from the crowd.In this paper, we utilize Social Network Analysis (SNA) techniques to understand the dynamics of the interaction between users in a Facebook based activist group. Additionally, we aim to identify the most influential users in the group and infer their relationship strength. We incorporate sentiment analysis to identify users with clear positive and negative influences on the group.We used k-Nearest Neighbour (kNN), Naive Bayes, and C4.5 decision tree algorithms.Our tests show that C4.5 has a better performance on troll detection.Our results show that applying such data analysis techniques on users online behaviour is a powerful tool to predict levels of influence and relationship strength between group members. Finally, we validated our results against the ground truth and found that our approach is very promising at achieving its aims.

KEYWORDS--Social Network Analysis;Troll Detection ,KNN, Naive Bayes,C4.5 terrorism awareness.

I. INTRODUCTION

Social media platforms allow people to share and discuss news, ideas, information and other user-generated content in the form of blogs, micro-blogs, forums, video, and audio. This user-generated content on social media is a source of big data for understanding the society or communities by way of analyzing with appropriate methods and tools. Twitter is one of the biggest social media and social network platforms. More than half a billion tweets (micro-blog posts) are posted everyday in average by millions of users on Twitter. Therefore, it is a great source for analyzing big social data.

Social media and network platforms are free to all, therefore we also find many malicious users, also called "trolls" on these platforms. Trolls try to create discord and distort the information flow on the network by false information or inflammation messages [1]. Therefore, their presence and posts prevent the true understanding of the information on the network, and big data analytic on social media requires filtering out trolls in the first place. In this paper we present methods to detect trolls. Trolls can not be exactly determined since they have complex and unstable behaviors. By using several classification algorithms, we have tried to find out the best performing algorithms in order to classify users with troll like behaviors.

Social Network Analysis (SNA) is a method used to investigate social structures by utilizing graph theory concepts. SNA techniques have proven to be particularly useful in studying and analyzing the structure and behaviour of social groups. SNA is typically used to study real-world networks, either using static techniques that analyse the structural properties of the network and/or using dynamic techniques that use statistical methods to model different network processes over time. Furthermore, using SNA metrics such as centrality measures provide insights into the community structure and key players within a network.

The aim of this paper is to investigate SNA metrics that can aid in identifying key players within a given organized group, mainly of activists. National and international activist groups often use web forums to promote movements and distribute propaganda materials. Although some of these activist groups organize peaceful activities, some escalate to hostile movements which may cause disruption and financial losses to targeted organizations. The identification of key players in a given organized group of interest can help authorities save resources spent on investigating the whole network especially when the network is huge and complex. Additionally, this can serve as a proactive measure to predict the occurrence of any potentially disruptive offline action. For instance, last year the Australia and New Zealand Banking Group (ANZ) head office was in lock-down as around 80 people aggressively protest the bank's funding of fossil fuel projects. Another example is when two activists who broke into a coal mine and scaled equipment, obstructed the work of mining equipment resulting in financial losses .

On the other hand, text classification requires fitting the text into a machine learning algorithm in the form of numbers. This text passes through several processes which increases the complexity of the text classification . These processes include prepossessing of text, text visualization, features selection, and features extraction. During the text per-processing, removal of stop words, stemming, lemmatization, and conversion of text into lowercase are carried out. Text per-processing is expected to significantly improve the accuracy and reduce dimensionality of the features space . Text visualization is the conversion of text into numbers; for this technique like a word cloud, bag of words, and term frequency are used. Text visualization converts the text into a high-dimensional feature space.

II. PROPOSED SYSTEM

The aim of this paper is to investigate SNA metrics that can aid in identifying key players within a given organized group, mainly of activists. National and international activist groups often use web forums to promote movements and distribute propaganda materials. Although some of these activist groups organize peaceful activities, some escalate to hostile movements which may cause disruption and financial losses to targeted organization. The identification of key players in a given organized group of interest can help authorities save resources spent on investigating the whole network especially when the network is huge and complex. Additionally, this can serve as a proactive measure to predict the occurrence of any potentially disruptive offline action. For instance, last year the Australia and New Zealand Banking Group (ANZ) head office was in lock-down as around 80 people aggressively protest the bank's funding of fossil fuel projects. Another example is when two activists who broke into a coal mine and scaled equipment, obstructed the work of mining equipment resulting in financial losses.

The main contributions of our work are summarized as follows:

- Understand the dynamics of the interactions in potentially suspicious activist networks.
- Apply SNA techniques to identify the key players in organized activist groups. This includes the most active and most influential.
- Inference of trust relations between actor pairs within a social network, based on structural properties and sentiment analysis information to gain further group insights.
- Perform a temporal analysis of the network posting structure and compare it over time. In addition, we investigate the correlation of the amount of online activity with related real-world events.

III.METHODOLOGY

It is very different to detect opinion manipulators (trolls) and troll-like behaving users. Todoretal. proposed a method to solve this issue with different variations of a troll definition. A classifier has been trained to distinguish trolls from non-trolls. There were also several studies targeting detection of trolls accurately. To classify users, Kumaretal. developed a way of troll detection which was faster than many past algorithms proposed in the literature.Many different machine learning algorithms are used for classifying and analyzing tweets in the literature. Naive Bayes, k-Nearest Neighbour (kNN) and C4.5 are commonly used algorithms as is the case in However, the performances of the algorithms are different from each other depending on the dataset and the problem at hand. Manikandanetal. found that Naive Bayes technique performs better than kNN and C4.5 Decision tree methods.Vijayaranietal.analyzed the performance of Bayesian and Lazy classifiers. From their experimental results, they observed that the Lazy classifier is more efficient than Bayesian classifier. Hence, in our study, to able to get the best result we compared commonly used supervised learning algorithms. In contrast to other studies, our goal is to analyze tweets and detect "troll" users first and then analyze the terrorism awareness among social media users.

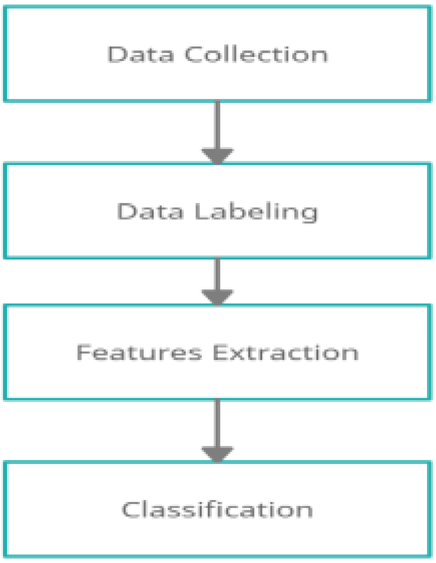


Fig 1.The flow chart of the proposed methodology.

TROLL DATASET

Troll, or malicious user, detection in social media is an important problem for social media analysis since these users distort the true message in the media and the elimination of these users and their content from the dataset before analysis is vital. Here we consider three different machine learning algorithms to classify Twitter users as troll or non-troll. We also present the dataset we collect for the evaluation.

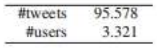
DATASET

In order to collect test data, Twitter REST API is used. There is a usage limit of 180 queries per 15 minutes in the API. Using the upper limits on the number of queries per minute and the maximum number of tweets per query, we were able to collect 18.000 tweets per 15 minutes. We collected 95.578 tweets belonging to 3.321 users on the topic of terrorism by using the Twitter REST API.

We also collected meta data about these users including follower/following counts, tweet counts, profiile picture, and retweet counts. The size of the collected data is approximately 50 MB, and had 95.578 tweets in English and Turkish belonging to 3.321 unique users. The most frequently used terms were chosen as keywords to collect tweets. Some of the keywords are "teror, terror, terrorism, canlıbomba('suicide bomber'), PKK, DAES". We used the WEKA4 tool to apply kNN, Naive Bayes, and C4.5 decision tree algorithms on the dataset.

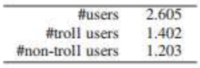
Table I shows the summary of collected datasets, part of which is divided into a training set that has the properties listed in Table II. 2.605 users out of 3.321 users were considered as training data and remaining 716 users out of 3.321 users were considered as test data.

TABLE I. DATASET COLLECTED



#tweets	95.578
#users	3.321

TABLE II. TRAINING DATASET



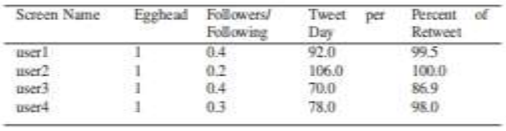
#users	2.605
#troll users	1.402
#non-troll users	1.203

In the proposed model, the awareness of Twitter users regarding terrorism and the detection of trolls were investigated through the processing Twitter data. The data collection was performed by a mechanism in order to enable us to collect Twitter data for several days belonging to a group of users. The mechanism also collected data giving ideas on several features of the users to determine if they are trolls or not. These features are: the average number of tweets a user sends per day, the number of 'followers' a user, has the number of users a user was 'following' and the ratio of re-tweets in tweets sent by a user. These were the key data for troll detection. For the purpose of training the system, 2.605 twitter users were manually examined one by one. As a result of the examinations, the following rules were extracted in order to decide if a user is troll:

- Users sending more than 50 tweets a day,
- Users having follower/following rate 0.4 and below,
- Users exceeding 70% retweet rate (70 retweets per 100 tweets sent)
- Default egghead profile images refer to 0 and other profile images refer to 1 (this gives some hint on the users but not a core differentiator)

Hence, we extracted general features and tried to detect obvious trolls such as those listed in Table III. Our training dataset included data from 2.605 users formed from both troll and non-troll users. This dataset was training data and metric for the classifier.

TABLE III. SAMPLE FEATURE VALUES FOR TROLL USERS



Screen Name	Egghead	Followers/ Following	Tweet per Day	Percent of Retweet
user1	1	0.4	92.0	99.5
user2	1	0.2	106.0	100.0
user3	1	0.4	70.0	86.9
user4	1	0.3	78.0	98.0

During the detection process of trolls several machine learning algorithms are employed.

TERROR AWARENESS DETECTION

A. Methods

1) Classifying Users Using the Mahout:

Generally, classification algorithms can be used to automatically classify documents and images in many domains. The Apache Mahout6 platform is a machine-learning library that is run on Hadoop in distributed manner [6]. We use Mahout in order to classify trolls with the Multinational Naive-Bayes classifier. This Bayesian-algorithm works for text data by using a training data set, that is a set of tweets each of which is associated with a subject category ("Terror and War" and "Other" categories in this case). Training dataset is manually labeled for categorizing tweets by their subject contents as "Other" and "Terrorism and War". This set includes 400 users out of 3.321 users and their 400 tweets

(one tweet per user). Training dataset was converted to the Hadoop sequence file format. After uploading this file to HDFS, Mahout was run to transform the training set into vectors using TF-IDF weights with Multinational Naive Bayes classifier. The TF(Term Frequency) is in general defined the number of times a given term t appears in a document d. In practice, the term frequency is often normalized by dividing the raw term frequency by the document length.

$$TF(t) = \frac{(Number\ of\ times\ term\ t\ appears\ in\ a\ document)}{(Total\ number\ of\ terms\ in\ the\ document)}$$

The Inverse Document Frequency(IDF) is number of information the word provides, that is, whether the term is common or rare across all documents.

$$IDF(t) = log_e(\frac{Total\ number\ of\ documents}{Number\ of\ documents\ with\ term\ t\ in\ it})$$

TF-IDF approach also assumes that the importance of a word is inversely proportional to how often it occurs across all documents.

$$TF - IDF(t) = TF(t) * IDF(t)$$

After the TF-IDF weight calculation, the training set was used to train the classifier. The classifier was worked properly

on the testing set. Finally, in line with our purpose, we classi-

fified users belonging to each category "Other" and "Terrorism

and War".

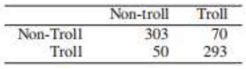
IV.RESULTS

A. Results for classification using three supervised machine learning methods

First, we applied kNN algorithm to 716 users out of 3.321 users for the test dataset by using training dataset in WEKA. We tested by selecting ’Supplied test set’ option and uploading

arff fifile. The confusion matrix is generated for T-NT (Troll Non-Troll) classes having two possible outcome values "troll" or "non-troll".

TABLE IV. CONFUSION MATRIX FOR KNN ALGORITHM

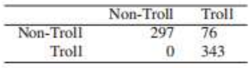


	Non-troll	Troll
Non-Troll	303	70
Troll	50	293

Correctly Classified Instances 83%.

Secondly, C4.5 algorithm was applied on the test dataset by using training dataset. We tested by selecting ’Supplied test set’ option and uploading interfile in WEKA. The confusion matrix was generated for T-NT (Troll- Non-Troll) class having two possible values "troll" or "non-troll".

TABLE V. CONFUSION MATRIX FOR C4.5

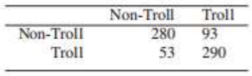


	Non-Troll	Troll
Non-Troll	297	76
Troll	0	343

Correctly Classified Instances 89%.

Thirdly, Naive Bayes algorithm was applied on the test dataset by using training dataset as the other algorithms used. We tested by selecting ’Supplied test set’ option and uploading file in WEKA. The confusion matrix was generated for TNT (Troll- Non-Troll) class having two possible values troll or non-troll.

TABLE VI. CONFUSION MATRIX FOR NAIVE BAYES



	Non-Troll	Troll
Non-Troll	280	93
Troll	53	290

Correctly Classified Instances 79%.

V.CONCLUSION AND FUTURE WORK

In this paper, the detection and elimination of the effects of trolls generates complicated structure which needed extreme computation power and sophisticated architectural design in order to obtain expected results. Luckily, we achieved to detect ’obvious trolls’. During the experiments, we categorized users as troll or not about terrorism by using three supervised algorithms namely kNN, C4.5 Decision Tree and Naive Bayes.

And we analyzed the algorithms by using the classification accuracy. From the results, it is observed that the C4.5 algorithm performs better than the other algorithms.

Social media is a platform, allowing people to share, discuss and modify user-generated content. For this reason, there may be many users called ’trolls’ that creates trouble about a delicate issue like terrorism. In this study, we detected these kind of users and demonstrated how these users aware of terrorism in Turkey and around the world by using HIVE. From the results, it is observed that the United States has higher terror awareness level than the other countries on the world and Istanbul has higher terror awareness level than the other cities in Turkey.

In the future, the study can be extended by using sentiment or opinion analysis in order to determine every kind of trolls more precisely. And different classifying algorithms can be used to compare their performance and accuracy.

REFERENCES

[1]. T. Mihaylov, G. D. Georgiev, P. Nakov, "Finding Opinion Manipulation Trolls in News Community Forums," in Proceedings of the Nineteenth Conference on Computational Natural Language Learning, CoNLL, July, 2015, Vol. 15, pp. 310-314.

[2]. S. Kumar and F. Spezzano, "Accurately detecting trolls in Slashdot Zoo via decluttering," in Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on, August, 2014, pp.188-195.

[3]. S.A. Paul, L. Hong,E.H. Chi, "What is a Question? Crowdsourcing Tweet Categorization," in Workshop on Crowdsourcing and Human Computation at the Conference on Human Factors in Computing Systems (CHI), 2011.

[4]. P. Andre, M.S. Bernstein,K. Luther, "Who Gives A Tweet? Evaluating Microblog Content Value," in Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. ACM, February, 2012, pp.471-474.

[5]. C. Wanger, S. Asur,J. Hailpern, "Religious Politicians and Creative Photographers: Automatic User Categorization in Twitter," in Social Computing (SocialCom), 2013International Conference on, September, 2013, pp. 303-310.

[6]. E. Jain and S.K. Jain, "Categorizing Twitter Users on the basis of their interests using Hadoop/Mahout Platform," in 9th International Conference on Industrial and Information Systems (ICIIS), December, 2014, pp.1-5.

[7]. Galan-Garcia, Patxi, Jose Gaviria de la Puerta, Carlos Laorden Gomez, Igor Santos, and Pablo Garcia Bringas, "Supervised machine learning for the detection of troll profifiles in twitter social network: Application to a real case of cyberbullying," in International Joint Conference SOCO’13-CISIS’13-ICEUTE’13. Springer International Publishing, 2014, pp. 419-428.

[8]. P. Manikandan and D. Ramyachitra, "Naive Bayes classification Technique for Analysis of Ecoli Imbalance Dataset," in International Journal of Computational Intelligence and Informatics, July â ˘ A ¸S September, 2014, Vol. 4.

[9]. Ms S. Vijayarani and Ms M. Muthulakshmi, "Comparative Analysis of Bayes and Lazy classification Algorithms," in International Journal of Advanced Research in Computer and Communication Engineering, August,2013, Vol. 2.8: 3118-3124.

[10]. Elijah Olusayo Omidiora,Ibrahim Adepoju Adeyanju and Olusayo Deborah Fenwa, "Comparison of Machine Learning classifiers for Recognition of Online and Offlfline Handwritten Digits," in Computer Engineering and Intelligent Systems, 2013, Vol.4, No.13.