IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104008 |

Fake Account Detection using Machine Learning

Dr. Mk Jayanthi Kannan[1], Ishan Sikder [2], Vinayak Kini [3], Malathi K [4], Mirunalini C.K[5] , AkashSingh[6]

Professor, School of Engineering and Technology, Jain University Bangalore, India[1]

UG Student, Department of Information Science and Engineering School of Engineering and Technology, Jain University Bangalore, India [2,3,4,5]

ABSTRACT: In the present generation, online social networks (OSNs) have become increasingly popular, people's social lives has become more associated with these sites. They use OSNs to keep in touch with each other's, share news, organize events, and even run their own e-business. The rapid growth of OSNs and the massive amount of personal data of its subscribers have attracted attackers, and imposters to steal personal data, share false news, and spread malicious activities. On the other hand, researchers have started to investigate an efficient technique to detect abnormal activities and fake accounts relying on accounts features, and classification algorithms. However, some of the account's exploited features have negative contribution in the final results or have no impact, also using standalone classification algorithms does not always reach satisfied results. In this paper, a new algorithm, SVM-NN, is proposed to provide efficient detection for fake Twitter accounts and bots, four feature selection and dimension reduction techniques were applied. Three machine learning classification algorithms were used to decide the target account's identity real or fake, those algorithms were deep learning and neural Network (NN), and our newly developed algorithm, SVM-NN, that uses less number of features, while still being able to correctly classify about 98% of the accounts of our training dataset.

KEYWORDS: Energy efficient algorithm; Manets; total transmission energy; maximum number of hops; network lifetime

I. INTRODUCTION

online social network's(osns), such as facebook, twitter, linkedin, google+ have become increasingly popular over last few years. people use osns to keep in touch with each other, share news, organize events, and even run their own e-business. for the period between 2014 and 2018 around 2.53 million u.s. dollars have been spent on sponsoring political ads on facebook by non- profits. the open nature of osns and the massive amount of personal data for its subscribers have made them vulnerable to sybil attacks. in 2012, facebook noticed an abuse on their platform including publishing false news, hate speech, sensational and polarizing, and some others. however, online social networks (osns) have also attracted the interest of researchers for mining and analysing their massive amount of data, exploring and studying users behaviours as well as detecting their abnormal activities. in researchers have made a study to predict, analyse and explain customer's loyalty towards a social media-based online brand community, by identifying the most effective cognitive features that predict their customer's attitude. facebook community continues to grow with more than 2.2 billion monthly active users and 1.4 billion daily active users, with an increase of 11% on a year-over-year basis. in the second quarter of 2018 alone, facebook reported that its total revenue was $13.2 billion with $13.0 billion from ads only. similarly, in second quarter of 2018 twitter has reported reaching about one billion of twitter subscribers, with 335 million monthly active users. in 2017 twitter reported a steady revenue growth of 2.44 billion
u.s. dollars, with 108 million u.s. dollars lower profit compared to the previous year. in 2015 facebook estimated that nearly 14 million of its monthly active users are in fact undesirable, representing malicious fake accounts that have been created in violation of the websites terms of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104008 |

service. facebook, for the first time, shared a report in the first quarter of 2018 that shows their internal guidelines used to enforce community standards covering their efforts between october 2017 to march 2018, this report illustrates the amount of undesirable content that has been removed by facebook, and it covers six categories: graphic violence, adult nudity and sexual activity, terrorist propaganda, hate speech, spam, and fake accounts. 837 million posts containing spam have been taken down, and about 583 million fake accounts have been disabled, facebook also has removed around 81 million undesirable content in terms of the rest violating content types. however, even after preventing millions of fake accounts from facebook, it was estimated that, around 88 million accounts, are still fake. for such osns, the existence of fake accounts lead advertisers, developers, and inventors to distrust their reported user metrics, which would negatively impacts their revenues as recently, banks and financial institutions in u.s. have started to analyse twitter and facebook accounts of loan applicants, before actually granting the loan. attackers follow the concept of having osns user accounts are “keys to walled gardens”, so they deceive themselves off as somebody else, by using photos and profiles that are either snatched from a real person without his/her knowledge, or are generated artificially, to spread fake news, and steal personal information.

II. RELATED WORK

Inspired by the importance of detecting fake accounts, researchers have recently started to investigate efficient fake accounts detection mechanisms. Most detection mechanisms attempt to predict and classify user accounts as real or fake (malicious, Sybil) by analysing user level activities or graph-level structures. There are several data mining methodologies and approaches that help detecting fake accounts that are described in the following subsections.

(A) Feature Based detection

This approach relies on user-level activities and its account details (user logs and profiles). Unique features are extracted from recent user activities (e.g. frequency of friend requests, fraction of accepted requests), then those features are applied to a classifier that has been trained offline using machine learning techniques. In the authors used a click-stream dataset provided by Ren Ren, a social network used in China, to cluster user accounts into similar behavioral groups, corresponding to real or fake accounts. Using the METIS clustering algorithm with both session and clicks features, such as:

- Average clicks per session
- Average session length
- The percentage of clicks used to send friend requests
- Visit photos
- Share contents

The authors were able to classify the data with 3% false positive rate and 1% false negative rate. Authors used ground-truth provided by Ren Ren to train an SVM classifier in order to detect fake accounts. Using simple features, such as:

- frequency of friend requests
- fraction of accepted requests

The authors were able to train a classifier with 99% true positive rate (TPR) and 0.7% false-positive rate (FPR). Researchers used a ground truth provided by Twitter; the data have been processed using two main approaches:

- Single classification rules
- Feature sets proposed in the literature for detecting spammers

Some features have been used from previous work such as Stateofsearch.com rule set and Social bakers rule set. The authors were able to correctly classify more than 95% of the accounts of the original training set.

(B) Feature Reduction

High dimensional data could be a serious problem for many classification algorithms because of its high computational cost and memory usage. On the other hand, reducing the dimension space would remove noisy (i.e. irrelevant) and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104008 |

redundant features and lead to a better classification model and simple visualisation technique.Feature reduction techniques can be categorized into two types:

- Dimension reduction where the data in high dimensional space is transformed into a space of fewer dimensions
- Feature subset selection where the original features set is disjoint into a selected subsets of features to buildsimpler and faster models, increases the models performance, and gain a better understanding of the data.
- Feature subset selection could be broken down into (filtering methods, and wrapping methods)

(C) Neural network (NN), and Support vector machine (SVM)

Authors extracted the profile features using PCA, and then applied Neural Networks, and Support Vector machines to detect legitimate profiles." Variance maximisation" was selected as a mathematical way of deriving PCA results, which were:

- Number of Languages.
- Profile Summary
- Number of Connections
- Number of Skills
- Number of LinkedIn Groups

III. PROPOSED ALGORITHM

A. *ALGORITHMS USED*

- LOGISTIC REGRESSION
- DECISION TREE CLASSIFIER
- RANDOM FOREST CLASSIFIER
- UML DIAGRAMS
- USECASE DIAGRAM
- CLASS DIAGRAM
- SEQUENCE DIAGRAM
- COLLABORATION DIAGRAM
- ACTIVITY DIAGRAM
- DEPLOYMENT DIAGRA
- COMPONENT DIAGRAM
- ER DIAGRAM
- DFD DIAGRAM

IV. SIMULATION RESULTS

The simulation studies involve the deterministic small network topology with 5 nodes as shown in Fig.1. The proposed energy efficient algorithm is implemented with MATLAB. We transmitted same size of data packets through source node 1 to destination node 5. Proposed algorithm is compared between two metrics Total Transmission Energy and Maximum Number of Hops on the basis of total number of packets transmitted, network lifetime and energy consumed by each node. We considered the simulation time as a network lifetime and network lifetime is a time when no route is available to transmit the packet. Simulation time is calculated through the CPUTIME function of MATLAB. Our results shows that the metric total transmission energy performs better than the maximum number of hops in terms of network lifetime, energy consumption and total number of packets transmitted through the network.

The network showed in Fig. 1 is able to transmit 22 packets if total transmission energy metric is used and 17 packets if used maximum number of hops metric. And the network lifetime is also more for total transmission energy. It clearly shows in Fig. 2 that the metric total transmission energy consumes less energy than maximum number of hops. As the network is MANET means nodes are mobile and they change their locations. After nodes have changed

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104008 |

their location the new topology is shown in Fig .3 and energy consumption of each node is shown in Fig. 4. Our results shows that the metric total transmission energy performs better than the maximum number of hops in terms of network lifetime, energy consumption and total number of packets transmitted through the network.

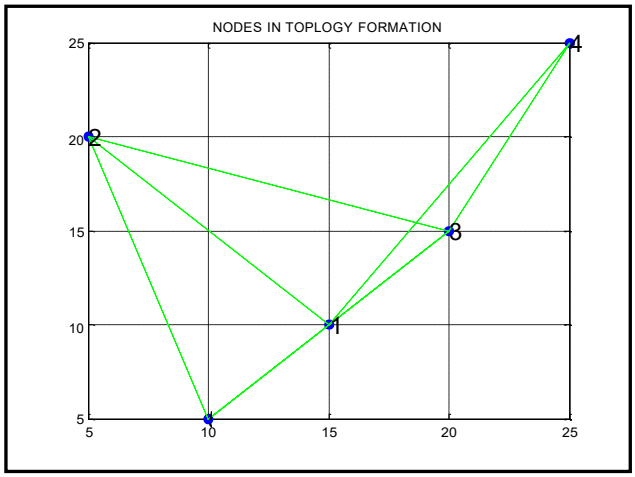

Fig.1. Ad Hoc Network of 5 Nodes

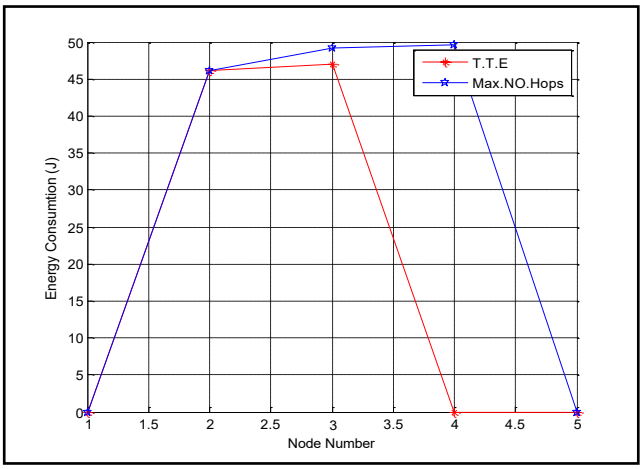

Fig. 2. Energy Consumption by Each Node

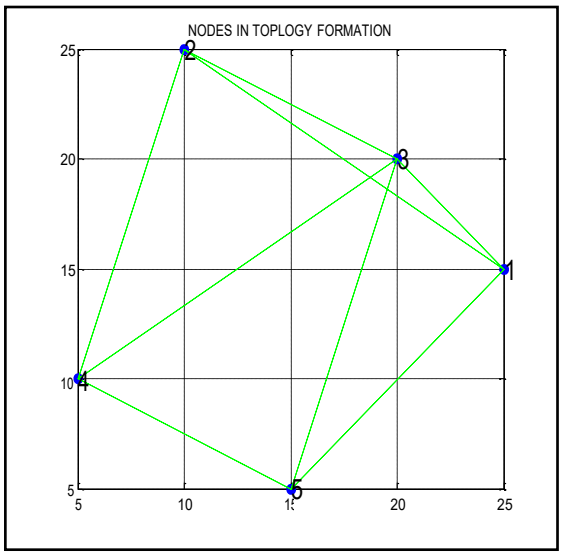

Fig. 3. Ad Hoc Network of 5 Nodes

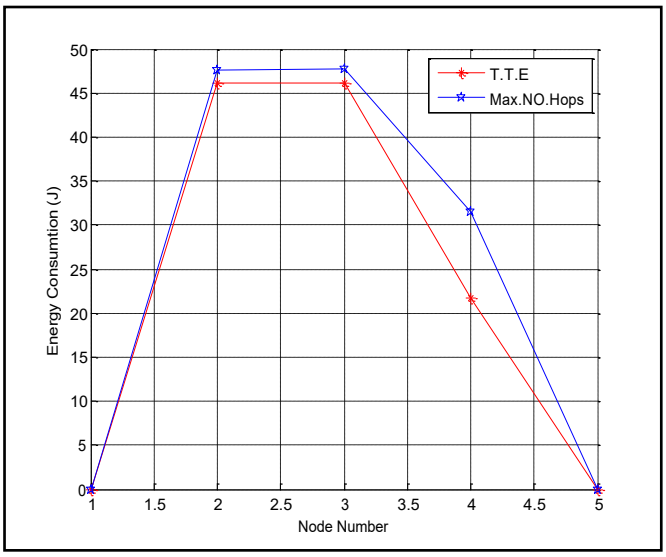

Fig 4. Energy Consumption by Each Node

V. CONCLUSION AND FUTURE WORK

The simulation results showed that the proposed algorithm performs better with the total transmission energy metric than the maximum number of hops metric. The proposed algorithm provides energy efficient path for data transmission and maximizes the lifetime of entire network. As the performance of the proposed algorithm is analyzed between two metrics in future with some modifications in design considerations the performance of the proposed algorithm can be compared with other energy efficient algorithm. We have used very small network of 5 nodes, as number of nodes increases the complexity will increase. We can increase the number of nodes and analyze the performance.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104008 |

REFERENCES

1. Anjum Asma and Gihan Nagib,'Energy Efficient Routing Algorithms for Mobile Ad Hoc Networks–A Survey', International Journal of Emerging Trends & Technology in computer Science, Vol.3, Issue 1, pp. 218-223, 2012.
2. Hong-ryeol Gil1, Joon Yoo1 and Jong-won Lee2 ,'An On-demand Energy-efficient Routing Algorithm for Wireless Ad hoc Networks', Proceedings of the 2nd International Conference on Human. Society and Internet HSI'03, pp. 302-311, 2003.
3. S.K. Dhurandher, S. Misra, M.S. Obaidat, V. Basal, P. Singh and V. Punia,'An Energy-Efficient On Demand Routing algorithm for Mobile Ad-Hoc Networks', 15 th International conference on Electronics, Circuits and Systems, pp. 958-9618, 2008.
4. DilipKumar S. M. and Vijaya Kumar B. P. ,'Energy-Aware Multicast Routing in MANETs: A Genetic Algorithm Approach', *International Journal of* Computer *Science and Information Security* (IJCSIS), Vol. 2, 2009.
5. AlGabri Malek, Chunlin LI, Z. Yang, Naji Hasan.A.H and X.Zhang ,' Improved the Energy of Ad hoc On-Demand Distance Vector Routing Protocol', International Conference on Future Computer Supported Education, Published by Elsevier, IERI, pp. 355-361, 2012.
6. D.Shama and A.kush,'GPS Enabled E Energy Efficient Routing for Manet', International Journal of Computer Networks (IJCN), Vol.3, Issue 3, pp. 159-166, 2011.
7. Shilpa jain and Sourabh jain ,'Energy Efficient Maximum Lifetime Ad-Hoc Routing (EEMLAR)', international Journal of Computer Networks and Wireless Communications, Vol.2, Issue 4, pp. 450-455, 2012.
8. Vadivel, R and V. Murali Bhaskaran,'Energy Efficient with Secured Reliable Routing Protocol (EESRRP) for Mobile Ad-Hoc Networks', Procedia Technology 4,pp. 703- 707, 2012.
9. Nobuo Ezaki, Marius Bulacu Lambert , Schomaker , "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons" , Proc. of 17th Int. Conf. on Pattern Recognition (ICPR), IEEE Computer Society, pp. 683-686, vol. II, 2004
10. Mr. Rajesh H. Davda1, Mr. Noor Mohammed, " Text Detection, Removal and Region Filling Using Image Inpainting", International Journal of Futuristic Science Engineering and Technology, vol. 1 Issue 2, ISSN 2320 – 4486, 2013
11. Uday Modha, Preeti Dave, " Image Inpainting-Automatic Detection and Removal of Text From Images", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622 Vol. 2, Issue 2, 2012.
12. Muthukumar S, Dr.Krishnan .N, Pasupathi.P, Deepa. S, "Analysis of Image Inpainting Techniques with Exemplar, Poisson, Successive Elimination and 8 Pixel Neighborhood Methods", International Journal of Computer Applications (0975 – 8887), Volume 9, No.11, 2010

BIOGRAPHY

Anjum Asma Mohammed is a Research Assistant in the Information Technology Department, College of Computer and Information Sciences, King Saud University. She received Master of Computer Application (MCA) degree in 2005 from BAMU, Aurangabad, MS, India. Her research interests are Computer Networks (wireless Networks), HCI, Algorithms, web 2.0 etc.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details