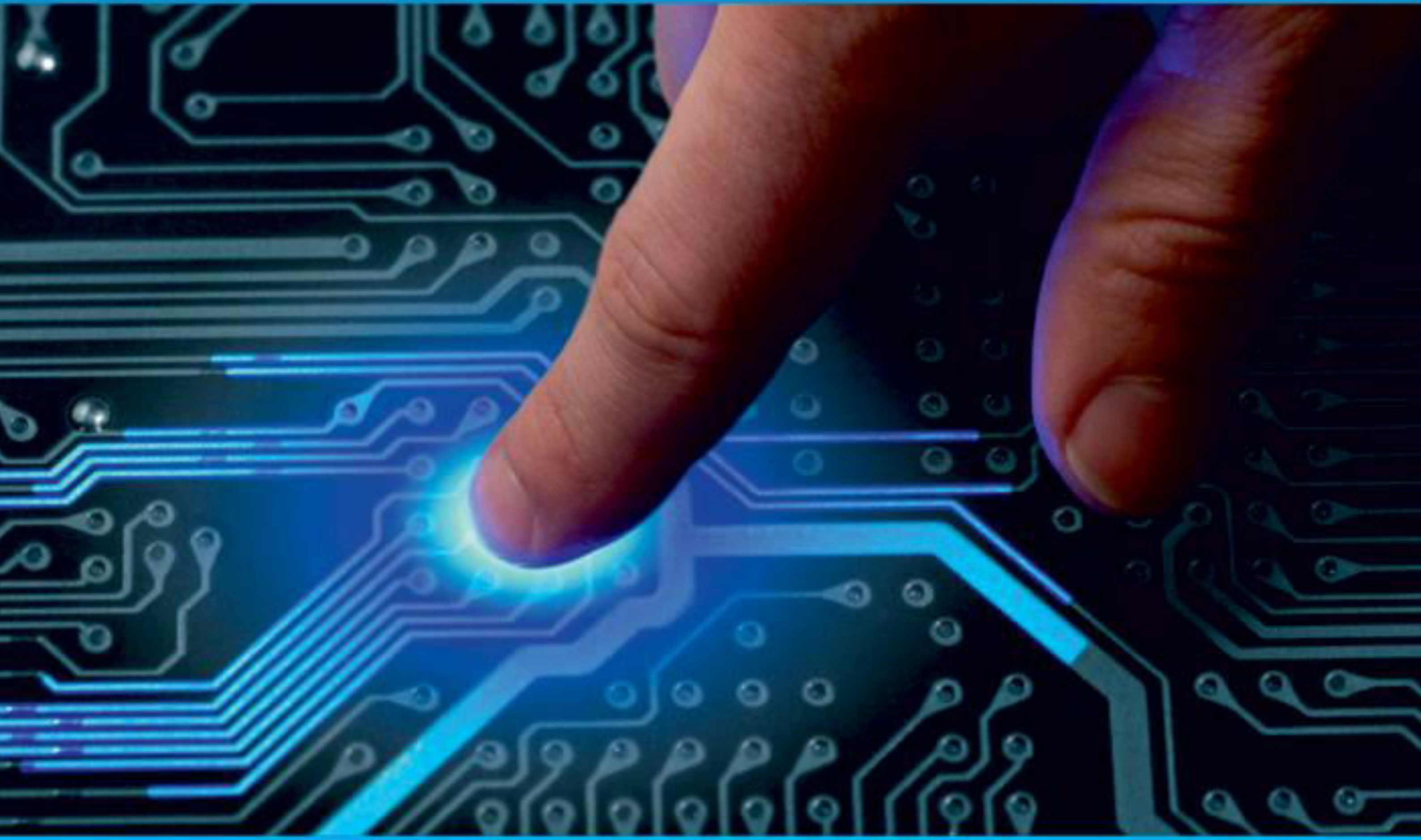
IJIRCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 3, March 2022

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165

Cyberbullying Detection: Identifying Hate Speech using Machine Learning

Shriram Kulkarni, Atharva Gornal, Nishit Shah, Raunak Singh

PCP Students, Dept. of CO, Pimpri Chinchwad Polytechnic, Akurdi, Pune, Maharashtra, India

ABSTRACT: Bullying has been dominant since the beginning of time, It's just the habits of bullying which have changed over the years, from physical bullying to cyberbullying. According to Williard(2004), there are eight types of cyberbullying such as harassment, denigration, impersonation, etc. It's been around 2 decades since social media sites came into the picture, but there hasn't been a lot of effective measures to curb social bullying and it has become one of the alarming issues in recent times. In this paper, we present a systematic review of some published research on cyberbullying detection approaches and review methods to detect hate speech in social media, while distinguishing this from general profanity. We aim to establish verbal baselines for this task by applying supervised classification methods using a manually annoted open source dataset for this purpose. This paper does a comparative study of various Supervised algorithms, including standard, as well as ensemble methods. The evaluation of the result shows that Collective supervised methods have the potential to perform better than traditional supervised methods. A number of guidelines for future work are also discussed.

KEYWORDS: Machine Learning, Cyberbullying, Supervised, Ensemble, Hate Speech, Natural Language Processing

I. INTRODUCTION

Hate speech refers to words intending to create hatred towards a particular group, that group may be a community, religion or race. This speech may or may not have meaning, but is likely to result in violence. Hate speech online has been linked to a global rise in violence toward subgroups, including mass shootings, lynching's, and ethnic cleansing. Due to the gigantic rise of user-generated web content, particularly on social media networks, the amount of hate speech is also gradually increasing. Over the past years, research into cyberbullying detection has increased, due in part to the proliferation of cyberbullying across social media and its detrimental effect on the younger generation. A growing body of work is emerging on automated approaches to cyberbullying detection. These approaches utilise machine learning and natural language processing techniques to identify the characteristics of a cyberbullying exchange and automatically detect cyberbullying by matching textual data to the identified traits. Natural language processing focusing specifically on this phenomenon is required since basic word filters do not provide a sufficient remedy: What is considered a hate speech message might be influenced by aspects such as the domain of an utterance, its speech context, as well as context consisting of co-occurring media objects (e.g. images,videos, audio), the exact time of posting and world events at this moment, identity of author and targeted recipient. This paper provides a comprehensive and structured overview of automatic hate speech detection, and compares few of its current approaches in a systematic manner, along presenting an insightful review of some published research on cyberbullying detection approaches.

II. RELATED WORK

For Detecting Cyberbullying, numerous approaches have been developed, majorly using Natural Language Processing and Information Retrieval which are then used to classify textual data by extracting it's features by using TF-IDF, Sentiment Analysis, Dimensionality Reduction etc. and they have received commendable accuracies.

[2] tries ways to detect nastiness on social media using NLP techniques to detect and deter cyberbullying eventually. NLP techniques are used in such a way that they can even detect when profanities in the data are used in an insulting way or in a neutral way. Annotations used in the paper are iteratively revised using in lab annotations and crowdsourcing. Data was crawled from English posts on social media sites even including semi-anonymous social media sites such as ask.fm. A ranked list of profanities along with NLP helped in crawling in an effective way. To classify the data, modified linear SVM was used to distinguish bad words in a casual way, multiple other features that could have gone unnoticed were also considered such as, Question answer posts and Emoticons. In the end, F1-Score came out to be 0.59 (which although is less than the Kaggle's winner but considering the fact that this study didn't use customized data and a new and better dataset, F1-Score of 0.59 still looks promising). Challenges faced in this study

were - • In ask.fm, comments are question-answer pairs which are shorter in other datasets and both question-answer may contain only one word making it hard for the algorithm to classify without understanding the full context. • People use informal language and slang on social media which are full of misspellings and abbreviation, making processing them very difficult Acknowledging the repetitive nature of cyberbullying on social media i.e. a sequence of aggressive messages sent from bully to a victim with the intent of harm.

[3] uses sequential hypothesis testing formulation to drastically reduce the number of features used in classification, while still maintaining high accuracy. This approach focuses High accuracy, Timeliness, and scalability. Models are trained using semi-supervised ML algorithms, using an Instagram dataset collected using snowball sampling, labeled manually(to a small extent) by a group of experts. The limitation of this approach was the use of a single data set that was only valid for Instagram, with no way to check the validity of labels, and the time overhead due to difficulty in capturing comment based labels.

[4] extracted the corpus and cleaned it from the Reddit database, followed by training a word embedding model based on word2vec skip-gram model. Then, the features of this model were used to train a random forest classifier for classifying cyberbully comments, This new word embedding model made using domain knowledge performed better than 4 pre-trained word embedding models, as well as handcrafted feature extraction methods.

[5] proposed a model for cyberbullying identification that uses research based on psychology; it describes the design for an app referred as BullyBlocker, which aims to intimate the parents of the user if cyberbullying is detected. It uses traditional methods to analyze social media data of the user by going through their messages and comments and rank them as warning signs or give them a bullying rank. It is specifically made for adolescents and uses old methods for detection in Facebook, but it has the potential to grow by acting as a data-collecting app over which ML classification can be run.

III.PROPOSED ALGORITHM

A. *Design Considerations:*

- Initial battery energy (IBE) is 50Jules for each node.
- Nodes are able to calculate its residual battery energy (RBE).
- Keeping track of previously used paths.
- Considered all possible paths at beginning.
- Receiving energy is not considered.
- The time when no path is available to transmit the packet is considered as the network lifetime.
-

B. *Description of the Proposed Algorithm:*

Aim of the proposed algorithm is to maximize the network life by minimizing the total transmission energy using energy efficient routes to transmit the packet. The proposed algorithm is consists of three main steps.

Step 1: Calculating Transmission Energy:

The transmission energy (TE_{node})of each node relative to its distance with another node is calculated by using eq.(1)[8].

$$TE_{node} \;\alpha\; d^n$$

$$TE_{node} = k\, d^n \qquad \text{eq. (1)}$$

where k is constant and n is path loss factor which is generally between (2-4) [8].

Step 2: Selection Criteria:

Node should have more residual battery energy (RBE) than the required transmission energy (TE_{node}) to transmit the packet to the next node in the route. All the nodes in the route will be checked with this condition even if one node of a route is not satisfying this condition then that route will not be considered as a feasible solution. All the other routes having all the nodes with sufficient amount of energy are considered as the feasible solution. And those nodes having equal RBE than (TE_{node}) are made to go into sleep mode. This selecting criterion helped to prolong the network life by avoiding the link breakage. We tried to avoid the repeated use of the path. But at one stage we have to compromise with energy efficiency when we have a route with less energy consumption but it is already being used and a rout with maximum consumption of energy which is not used. So till this point we avoided repeated use of the paths and tried to increase the network life. Transmission energy of a node to node in a rout is calculated according to the distance and the total transmission energy (TTE_R) for that rout is calculated using eq. (2).

$$TTE_R = \sum_{i=1}^{m} TE \qquad \text{eq. (2)}$$

where m is the number of hops in the route, TE = TE_{node} is the transmission energy between the nodes. The route having minimum total transmission energy i.e. min (TTE_R)will be selected as energy efficient route.

Step 3: Calculating Residual Battery Energy (RBE):

After transmitting the packet, residual battery energy for each node of the route is calculated using eq. (3) with parameters initial battery energy (IBE) and TE_{node}.

$$RBE = IBE - TE_{node} \qquad \text{eq. (3)}$$

IV. PSEUDO CODE

```
Step 1: Generate all the possible routes.
Step 2:  Calculate the $TE_{node}$ for each node of each route using eq. (1).
Step 3:  Check the below condition for each route till no route is available to transmit the packet.
if ($RBE <= TE_{node}$)
Make the node into sleep mode.
else
        Select all the routes which have active nodes
end
Step 4:  Calculate the total transmission energy for all the selected routes using eq. (2).
Step 5: Select the energy efficient route on the basis of minimum total transmission energy of the route.
Step 6:  Calculate the RBE for each node of the selected route using eq.  (3).
Step 7: go to step 3.
Step 8: End.
```

V. 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 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.

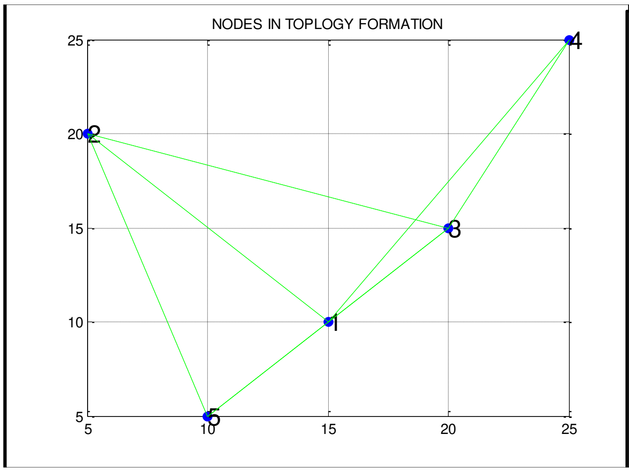


Fig.1. Ad Hoc Network of 5 Nodes

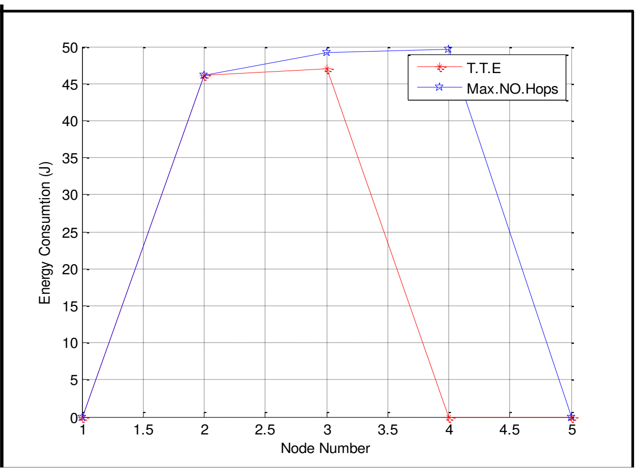


Fig. 2. Energy Consumption by Each Node

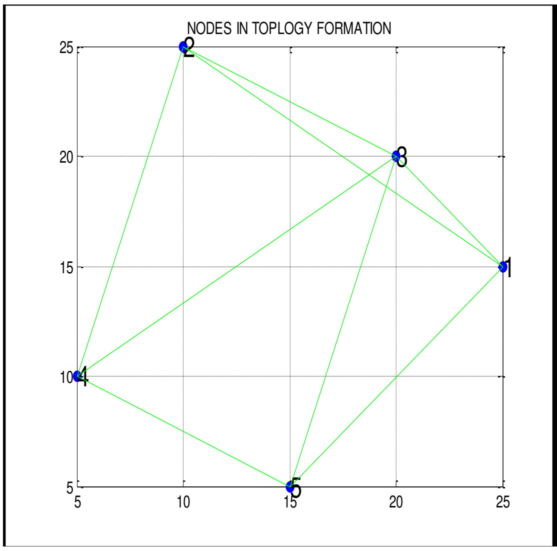


Fig. 3. Ad Hoc Network of 5 Nodes

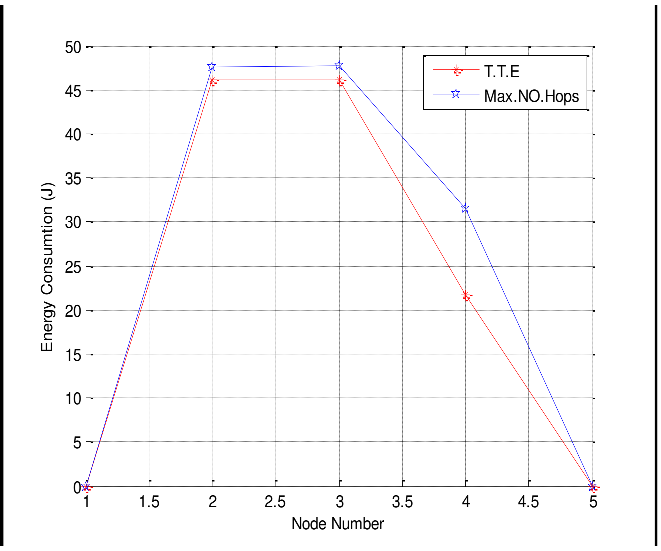


Fig 4. Energy Consumption by Each Node

VI. 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.

REFERENCES

[1] DataTurks. (2018, July 12). Tweets Dataset for Detection of Cyber-Trolls. Retrieved November 07, 2020, from https://www.kaggle.com/dataturks/dataset-for-detection-ofcybertrolls?select=Dataset+for+Detection+of+Cyber-Trolls.json

[2] Samghabadi, Niloofar Safi, et al. ”Detecting nastiness in social media.” Proceedings of the First Workshop on Abusive Language Online. 2017.

[3] Yao, Mengfan, Charalampos Chelmis, and Daphney? Stavroula Zois. ”Cyberbullying ends here: Towards robust detection of cyberbullying in social media.” The World Wide Web Conference. 2019.

[4] Huang, Qianjia, Vivek Kumar Singh, and Pradeep Kumar Atrey. ”Cyberbullying detection using social and textual analysis.” Proceedings of the 3rd International Workshop on Socially-Aware Multimedia. 2014.
[5] T. Bin Abdur Rakib, L. K. Soon, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Springer Verlag, 2018), vol. 10751 LNAI, pp. 180–189.
[6] Y. N. Silva, D. L. Hall, C. Rich, BullyBlocker: toward an interdisciplinary approach to identify cyberbullying. Social Network Analysis and Mining. 8 (2018), doi:10.1007/s13278-018-0496-z.
[7] E. Raisi, B. Huang, Weakly supervised cyberbullying detection with participant-vocabulary consistency. Social Network Analysis and Mining. 8 (2018), doi:10.1007/s13278-018-0517-y.
[8] Homa Hosseinmardi, Sabrina Arredondo Mattson, Rahat Ibn Rafiq, Richard Han, Qin Lv, Shivakant Mishra. (2015). Detection of Cyberbullying Incidents on the Instagram Social Network. ”
[9] Dadvar, Maral Eckert, Kai. (2018). Cyberbullying Detection in Social Networks Using Deep Learning Based Models; A Reproducibility Study. 10.13140/RG.2.2.16187.87846.
[10] Nandhini, B. Sri, and J. I. Sheeba. ”Cyberbullying detection and classification using information retrieval algorithm.” Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering Technology (ICARCSET 2015). 2015.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462 ✆ 6381 907 438 ✉ ijircce@gmail.com

Scan to save the contact details

www.ijircce.com