International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Real Time Cyberbullying Detection

Shalni Prashar, Suman Bhakar

Abstract: ***Automated approaches for detecting cyberbullying on online platforms has remained a primary research concern over past years. Cyber bullying is defined as the use of electronic communication to bully a person, typically by sending messages of intimidating or threatening nature. The victims especially teenagers suffer from loss of confidence, depression, sleep disorder. The research on automated cyberbullying approach is mainly focused on data driven methods. Such methods work on a database of static texts, usually collected from online platforms and are not feasible for dynamic nature of a real-life social networking scenarios.***

The aim of our research is to develop a cyberbullying detection system using Fuzzy Logic. Three types of bullying emotions are considered in this research work namely aggression, abuse and threat. In the proposed approach chat between two users is continuously monitored and emotion present in each message is determined. Based on the emotion each user's behavior is categorized as decent or bullying. If the detected bullying nature is higher than a defined threshold value the account of user is ceased and reported automatically.

The proposed approach is tested with a chat application developed in Microsoft .Net Framework and approach can detect cyber bullying in good time. The proposed approach, if implemented with social networking platforms can serve as a useful aid for preventing online harassment. The developed algorithm can also be applied in surveillance and human behavioral analysis.

Keywords : Cyber bulling, Fuzzy Logic, k means clustering, Abusive language detection.

I. INTRODUCTION

The tremendous growth of social medial usage due to easy internet access has completely revolutionized our lifestyle. Social media activities such as using Instagram, YouTube, twitter is one of the most popular online activities. According to a survey about 2.65 billion people use social media [1]. The rate of penetration of social media is also very fast. The penetration rate of social media as on January 2019 is 45 percent. In future the aforementioned figures will increase for sure as it is anticipated that the digital markets which are presently less developed will soon catch a good development pace with consequent availability of cheap mobile devices and infrastructure.

The easy access and widespread usage of social medial platforms has also promoted its malicious uses. The positive usage of these platforms is often shadowed by the negative usage which includes abusive and offensive behaviors. The use of internet in order to bully someone.is called cyberbullying. The most traditional methods to combat cyberbullying includes the development of standards and guidelines that all users must adhere to, employment of human editors to manually check the bullying behavior, the use of profane word lists, and the use of regular expressions, etc[2].

The traditional methods are ineffective against the dynamic nature of present social media platforms . The cost of labor as well as maintenance involved in these approaches is also quite high. Furthermore, these methods cannot scale well. Thus, principal learning frameworks are required that can automatically detect bullying behaviors. Analyzing the problem from an all-round perspective it is clear that an automated and data-driven techniques can be a useful aid for analyzing and detecting bullying behaviors and can provide substantial aid to victims. Successful detection of cyberbullying would allow large-scale social media sites to identify harmful and threatening situations early and prevent their growth more efficiently.

Xu and etal [3] traced bully traces from texts extracted from Twitter. The tweets were reviewed to extract information about different types of emotions in them. The authors identified seven emotions namely embarrassment, fear, anger, empathy, pride, relief and sadness. Munezero, and etal[4] used emotion based features to determine emotional context of a post. The research work used two features namely the emotive words and SentiStrength. Serra and etal [5] proposed a rule-based framework is used for detecting cyberbullying. The proposed approach used user's age and his pattern of using mobile as the determining factors of cyberbullying. Based on these patterns a risk factor is assigned to the user. Chen and etal[6] evaluated an offensiveness score of two users. This offensiveness score is calculated by a Lexical Syntactic Framework (LSF). Mancilla and etal [7] developed a social computer game after carefully analyzing and studying the interactions of user in virtual environment. Based on this observation the detected bullying behavior in social gaming. R. Zhao and etal [8]. Developed semantic-enhanced marginalized denoising auto-encoder (smSDA) which is a novel representation algorithim to solve the problem .M. Yao and etal [9] introduced a new algorithm that drastically reduced the number of features used in classification. Y. J. Foong and etal[10].in this research an algorithim is developed for automatic identification and detection of cyber bullying . The dataset used was acquired from online forums and online communities.. R. Pawar and etal[11] proposed Multilingual Cyberbullying Detection System which has ability to detect Abusive behaviors in two different Indian languages-. Rosa and etal [12]. studied performance of Fuzzy Fingerprints, while detecting textual cyberbullying in social networks.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author
Shalni Prashar*, Pursuing Post-Graduation student, computer science program, Rajasthan College of Engineering for Women, Jaipur.
Suman Bhakar, Assistant Professor, Computer Science Department, Rajasthan College of Engineering for Women, Jaipur.

The Experimental results revealed that the abovementioned technique outperforms the traditional classifiers when tested in an imitation of real life situations.

From the literature it is revealed that the most widely used approach for detecting cyberbullying behavior is training and evaluating classifiers on the static data. The published results though demonstrate efficiency but cannot ensure similar performance of these algorithms in real life scenario, such as instant messaging on different social networking applications.

The proposed research considers three types of cyber bullying expressions namely aggression, abuse and threat.In the proposed approach chat between two users is continuously monitored and emotion present in each message is determined. Based on the emotion each user's behavior is categorized as decent or bullying. If the detected bullying nature is higher than a defined threshold value the account of user is ceased and reported automatically. The proposed approach can detect cyber bullying in good time and thus prevent victim from undergoing mental or physical stress.

II. PROPOSED METHODOLOGY

A.Proposed Algorithim

. In this research sentimental data analysis is used for detecting abusive language in social media texts. To imitate social media platform, a chat application is developed in C#. In the frontend the application allows user to send messages to other person, Abusive language is detected in backend processing with fuzzy logic and k means clustering algorithm. The user behavior while use the app can be categorized as calm , happy , neutral , aggression , abuse or threat by our proposed algorithm. If the user behavior is threatening then a warning is generated and the user's account is ceased for a given time period. The proposed algorithm is explained in detail in the following section.

B. Dataset Collection

The first step is dataset preparation. In this research 'labMT ' dictionary by Mechanical Turk is used . The dictionary is open source and consists of 5000 words. These 5000 words are gathered from four sources namely Twitter, the New York Times, Google Books, and music lyrics as most frequently used words. Each word in dictionary has a pre-defined happiness score linked to itself

C.Dataset Classification

The downloaded dictionary is in raw form and to make it useable in the proposed research we need to classify the words present in dictionary as calm , happy , neutral , aggression , abuse or threat. For this purpose we constructed individual dictionary related to these categories. CL,HP,NE,AR,AB, TH denotes the constructed dictionary for calm, happy , neutral, aggression , abuse and threat respectively. The set D denotes the collection of these dictionaries. The set L denotes labMT dictioary.

$$D=\{CL,HP,NE,AR,AB, TH\} \quad \ldots(1)$$

K means clustering algorithm is applied to classify the LabMT dictionary . The pseudo code is given below. Here W_(i) denotes the i^th word in labMT dictionary, H_(W_(i))is the happiness average of the i^th word, H_(CL) denotes the array which consists of happiness average of CL, H_(HP) denotes the array which consists of happiness average of HP, H_(NE) denotes the array which consists of happiness average of NE, H_(AR) denotes the array which consists of happiness average of AR, H_(AG) denotes the array which consists of happiness average of AG, H_(TH) denotes the array which consists of happiness average of TH. The result of the classification is shown in fig 1.

Step 1 for i=1, i<=5000, i++

Step 2 if W_(i) ∈CL

H_(CL) (i)= H_(W_(i))

Step 3 if W_(i) ∈HP

H_(HP) (i)= H_(W_(i))

Step 4 if W_(i) ∈NE

H_(NE) (i)= H_(W_(i))

Step 5 if W_(i) ∈AR

H_(AR) (i)= H_(W_(i))

Step 6 if W_(i) ∈AB

H_(AB) (i)= H_(W_(i))

Step 6 if W_(i) ∈TH

H_(TH) (i)= H_(W_(i))

Step 7 End

Step 7 Number of clusters K = 6

Step 8 for i=1, i<=5000, i++

Step 9 if ⟦min(H_(CL))≤H⟧_(W_(i))≤max(H_(CL))

Step 10 W_(i) ∈CL

Step11 if ⟦min(H_(CL))≤H⟧_(W_(i))≤max(H_(CL))

Step 12 W_(i) ∈HP

Step 13if ⟦min(H_(HP))≤H⟧_(W_(i))≤max(H_(HP))

Step 14 W_(i) ∈NE

Step 15 if ⟦min(H_(NE))≤H⟧_(W_(i))≤max(H_(NE))

Step 16 W_(i) ∈AR

Step 17 if ⟦min(H_(AR))≤H⟧_(W_(i) ≤max(H_(AR))

Step 18 W_(i) ∈AB

Step 19 if ⟦min(H_(AB))≤H⟧_(W_(i))≤max(H_(AB))

Step 18 W_(i) ∈TH

Step 19 if ⟦min(H_(TH))≤H⟧_(W_(i))≤max(H_(TH))

D. Chat Application Development

In this step chat application is developed using the principles of asynchronous programming in C#. When the user sends a message, it is stored in the string named, 'UserString'.

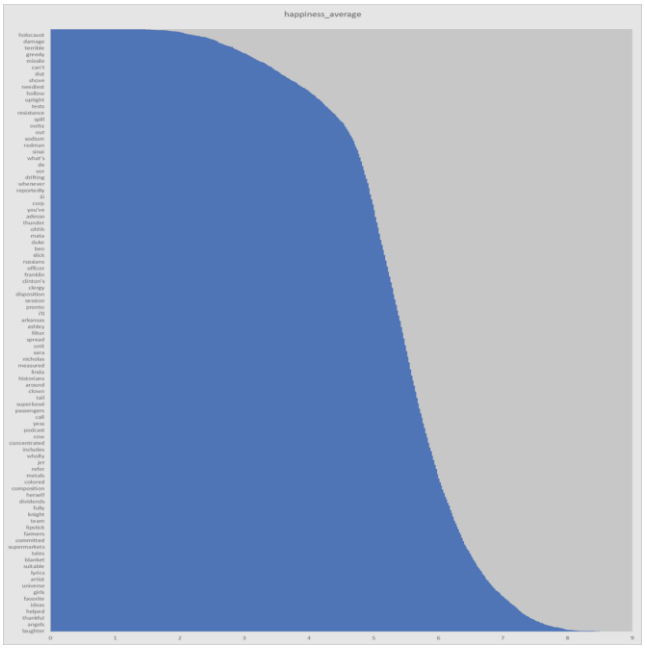

Fig 1 : Happiness Average value of different words.

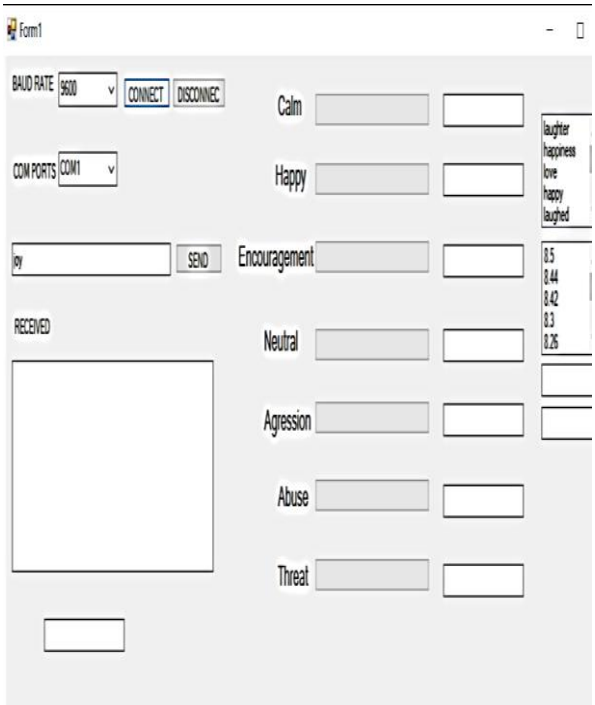

Fig 2 : Developed GUI in C#.

E. Chat Application Development

The UserString is separated into individual words, by using the following pseudo code.

Step 1 Len = length(UserString)

Step 2 Count=1

Step 3 Do

Step 4 i=count

Step 5 For i, UserString(i) != '', i++

Step 6 W_{kuser} (i) = UserString(i)

Step 7 Count =i+2;

Step 8 While(UserString(i) != '\n')

F. Chat Application Development

Once the sentence is separated into words then the value of each sentiment is calculated. The value of a particular sentiment class is denoted by V and is calculated by using following equation

$$V_S = \frac{number\ of\ words\ beloning\ to\ sentiment\ S}{total\ number\ of\ words} \qquad (2)$$

G. Developing Fuzzy logic system for abusive language detection

The input to fuzzy logic is the value of each sentiment as obtained in previous step. The input to fuzzy logic framework can be denoted by equation 4.3.

I={V_CL,V_HP,V_NE,V_AR,V_AB,V_TH} (3)

Gaussian membership function is used to represent input fuzzy variables. Triangular membership functions are used for output fuzzy variable. (Fig 4.2-4.4).

G(x;c,σ)=e^(-1/2 ⟦((x-c)/σ)⟧ ^2) (4)

Based on fuzzy rules the output is calculated which gives a cumulative result of the user behavior in form of the six sentiments. If the detected user behavior is calm, happy or neutral then no action is taken. If the user behavior falls in category of aggression, abuse or threat then the necessary action is performed.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed algorithm of 'Detecting cyberbullying by fuzzy logic and K means clustering is implemented in Microsoft Visual studio, .net frame work. Fig 2 shows the developed graphical user interface of the chat system. Since the chat system is designed to works between two applications the settings of Baud rate and COM ports are given. The user needs to connect to the application by clicking connect button.

The user then types the message to be sent and click on send button. When the system receives a message, it analyses it and detects the relevant emotion. The detected emotion is highlighted. If detected emotion is threat then system blocks the account for some time.

Fig 3 shows the detected emotion of 'threat' when the input string is 'I have thoughts about suicide'. Performance analysis of the proposed algorithm is discussed in later section.

Fig 3 : Detected emotion.

Table- I: Name of the Table that justify the values

PARAMETER	VALUE
True Positive Rate = Trials in which positive emotion is detected correctly/Total number of positive trials.	0.81
False Positive Rate = Trials in which positive emotion is detected incorrectly/Total number of positive trials.	0.19
True Negative Rate = Trials in which negative emotion is detected correctly/Total number of negative trials.	0.85
False Negative Rate = Trials in which negative emotion is detected incorrectly/Total number of negative trials	0.15

The performance of the proposed approach is evaluated with performance matrix which consists of True Positive Rate, False Positive Rate, True Negative Rate and False negative rate. True Positive Rate refers to samples that are correctly classified as positive. False positive Rate is the samples that are wrongly classified as positive. True negative Rate refers to the samples that are correctly identified as Negative. False Negative Rate refers to samples that are wrongly identified as negative. A detailed explanation about these rates can be found in [13]. The performance matrix is shown in Table I.

IV. CONCLUSION

The proposed algorithm for automatic detection of cyberbullying by fuzzy logic and k means clustering is successfully implemented. Using this approach six emotions namely calm, happy, neutral, aggression, abuse and threat can be detected from the user's text messages. The proposed approach can identify bullying behaviors with an accuracy of 85 percent, which is much higher than previous approaches, such as [14]. The proposed algorithm can prove to be highly useful in real time detection of cyberbullying and prevent emotional stress on victims.

FUTURE SCOPE

In future author suggests that the developed prototype can be implemented on Wide Area Network with combination of different dictionaries. Cloud Frameworks can be used for this purpose. The fuzzy rules can be further modified to detect non textual forms of cyberbullying such as abuse through obstructed texts, images and videos.

REFERENCES

1. https://ourworldindata.org/rise-of-social-media.
2. H.Dani , J.Li and H. Liu , " Sentiment Informed Cyberbullying Detection in Social Media", Machine Learning and Knowledge Discovery in Databases, Springer pp 52-67,2017.
3. Xu, J.M., Jun, K.S., Zhu, X. and Bellmore, A. (2012a). Learning from Bullying Traces in Social Media. IN: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Montreal. June 3 8, 2012.
4. Munezero, M., Montero, C.S., Kakkonen, T., Sutinen, E., Mozgovoy, M. and Klyuev, V. (2014). Automatic Detection of Antisocial Behaviour in Texts. Informatica. Special Issue: Advances in Semantic Information Retrieval, 38(1), p.3 – 10.
5. Serra, S.M. and Venter, H.S. (2011). Mobile Cyber-Bullying: A Proposal for a Pre-Emptive Approach to Risk Mitigation by Employing Digital Forensic Readiness. Information Security South Africa (ISSA), p.1-5..
6. J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.
7. C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
8. Chen, Y., Zhou, Y., Zhu, S., and Xu, H. (2012). Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. IN: International Conference on Privacy, Security, Risk and Trust (PASSAT) and Social Computing (SocialCom). Amsterdam, September 3-5, 2012. New York: IEEE.M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.
9. Mancilla-Caceres, J., Espelage, D. and Amir, E. (2015). A Computer Game-Based Method for Studying Bullying and Cyberbullying. Journal of School Violence, 14(1), 66-86
10. M. Yao, C. Chelmis and D. Zois, "Cyberbullying Detection on Instagram with Optimal Online Feature Selection," 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, 2018, pp. 401-408.
11. Y. J. Foong and M. Oussalah, "Cyberbullying System Detection and Analysis," 2017 European Intelligence and Security Informatics Conference (EISIC), Athens, 2017, pp. 40-46. inct forums and achieves reasonable detection performances
12. R. Pawar and R. R. Raje, "Multilingual Cyberbullying Detection System," 2019 IEEE International Conference on Electro Information Technology (EIT), Brookings, SD, USA, 2019, pp. 040-044.
13. S. Mane, J. Srivastava, San-Yin Hwang and J. Vayghan, "Estimation of false negatives in classification," Fourth IEEE International Conference on Data Mining (ICDM'04), Brighton, UK, 2004, pp. 475-478.
14. D.Zois , A. Kapodistria ,M. Yao and C. Chelmis,"Optimal Online Cyberbullying Detection", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),pp :2017-2021,2018.

AUTHORS PROFILE

Shalni Prashar, is a pursuing post-graduation student within the computer science program from Rajasthan College Of Engineering For Women, in Jaipur. She has completed her graduation in computer science from Rajasthan College Of Engineering For Women. Currently she is working with the designation of web designer with web aggregator since last 2 year and has good knowledge of cyber security which is basically her field of interest and has done multiple task in the mentioned domain. She had participated in multiple field in her graduation period like international conference and multiple workshop which were conducted within different engineering college. She also has interest in research field of database, data structure, artificial intelligence..

Suman Bhakar, is Assistant Professor in Cumputer Science Department in Rajasthan College of Engineering for Women Jaipur. She received her Ph.D. degrees in computer science from the Manipal University Jaipur Rajasthan, India in 2019.She received her Master degree in computer science department from the university of chennai,Tamil Naidu. She had completed her Bachelor degree from chennai.Her Research interests include Augmented reality, image processing,Security.She has knowledge in java.She has paricipant on several seminar on security.she has good experience in artificial intelligence.She had made several project on image processing.