How to Cite:
Das, B., Sengupta, A., & Das, A. (2022). Early detection of cyber-bullying from twitter data : An aspect based sentiment analysis approach. *International Journal of Health Sciences*, *6*(S5), 5989–5995. https://doi.org/10.53730/ijhs.v6nS5.10019

Early detection of cyber-bullying from twitter data – An aspect based sentiment analysis approach

Basabdatta Das
Techno College Hooghly, WB, India.
Email: basavdutta.das@gmail.com

Anirbit Sengupta
Dr. Sudhir Chandra Sur Institute of Technology and Sports Complex, WB, India
Email: anirbit87sengupta@gmail.com

Abhijit Das
RCC Institute of Information Technology, Kolkata, India.
Email: ayideep@yahoo.co.in

***Abstract*---**The vast volume of user created data in social media has created a wide-open path to data analytics in terms of extraction of features, analysis of polarity of information and classification of human emotional state. Sentiment analysis and data mining has reached at its peak in terms of its use in feature extraction and commercial analysis of products. Enormous approaches have been also seen in human behaviour analysis using ML (Machine Learning) algorithms and NLP (Natural Language Processing). To be very specific, it is not too hard to analyse data in granular level such as aspects. Social media play a distinct role in providing large dataset at no cost. And aspect-based sentiment analysis provides us the text analysis technique that categorisesuser provided data and assigns positive or negative value for the sentiment identified in it. In our approach we have tried to find out a solution of detection of cyber bullying in human interaction. This problem, we find a new threat in this digital era as commenting and replying to that comment play an important role in virtual communication, and just in this point, it takes no cost or hesitation factor to attack another person verbally or virtually. It is possible to identify these phrases or sentences and categorise them as influencer to cyber harassment, mental agony, breach to peace of mind which in the long run appears to tending depression, deterioration to mental trauma and perhaps suicidal tendency. This psychopathic crime has no means of detection directly so far. Neither any elaborative study has been performed towards early detection using aspect-based approach. We propose a solution in

Manuscript submitted: 27 Feb 2022, Manuscript revised: 9 April 2022, Accepted for publication: 18 June 2022

this problem having case studies from Indian background where the user uses colloquial languages and jargons outside of formal Dictionary.

***Keywords*---**Sentiment analysis, ML algorithm, NLP, Cyberbullying, aspect based.

1. Introduction

Cyberbullying is the new threat to social media in recent days. People are largely dependent over internet and they are keen to communicate over the social platform. This virtual world provides way fast medium to express our joy, sorrow and anger very easily. We never think twice to express our rage or grievance towards other and thus we ignore others feelings. Attacking verbally in group is very common too. Intentions of bullying others, thus becomes a new psychological threat which may even conclude in death. Numerous researches have been doneon detecting cyber bullying. Effective mechanism of this case can be seen in [1]. There are also methods of identifying good or bad words using NLP and Sentiment Analysis algorithms. But we feel that there is a gap in researches done so far and the actual problem of detection of cyber bullying at the earliest stage. More over jargons differ depending on geographical location. So, there is always a demand of detection of offensive local words. In our work, we propose a method of detecting cyber bullying in Indian context; by proposing a method of detection of profane words in Hindi and English mixed up languages.

2. Literature Review

Analysing product reviews using aspect based sentiment analysis is a very common work. One can gain the perspective of the method and application very well from the studies like [2,3, 4,5]. A study on the researches done in last three years result in quite a good number of study papers on this method. Aspect based approach is seen on language specific sentences in work done by Machado et al [4] on Portuguese language. Barros et al suggested some framework in their work[6] with words of Spanish and Portuguese languages. Nurul et al performed their work on Arabic languages [7]. Around thousand such references can be drawn on either of the two problems- aspect based categorisation of online data and language specific algorithm to apply with sentiment analysis using aspect-based methods. In [8], Dutta et al presented a review of different methods of sarcasm detection done so far. From which we can get a clear knowledge of methods invented so far for detecting good or bad words. Numerous studies have also been done to detect depression and cyber bullying so far [9, 10, 11, 12, 13, 14]. Other methods such as Deep Learning [15] and Grid search [16] and many more have also been adopted for the same problem field. Right at this point we found that study on detection of cyber bullying using aspect based sentiment analysis approach is hard to find. So we aim to design one such algorithm using Chi Square method [17] that can successfully identify the polarity of words obtained from the users of Indian subcontinent, especially in Hindi, English and a mix up of both the languages.

3. Proposed Method

Twitter is a widely used social platform for virtual communication in recent era. So we choose this platform to extract our data set and build the database. As we are intending to work on Indian user, we prefer to use HINGLISH language, which is a mixof Hindi and English language. The basic problem of working with this type of languageis we have to create the grammar at our own. Also we need to clearly specifythe list of all possible words that can be treated as abusive or offending. We divideour wholeprocedure of identifying cases of bullying into four subparts. Figure 1 shows the diagrammatic classification and functions or procedures included in eachsubparts.

3.1 Collection of Data Set

We first collect the data from a set of users. For our study we have taken 4000 user profiles and collected the data on a span of one week. Then these data are needed to be manipulated and pre-process. As we have to build our own grammar, we first discarded all common words like “is”, “the”, “to”, “ke”, “liye”, “magar” and formulated our own bag of words.

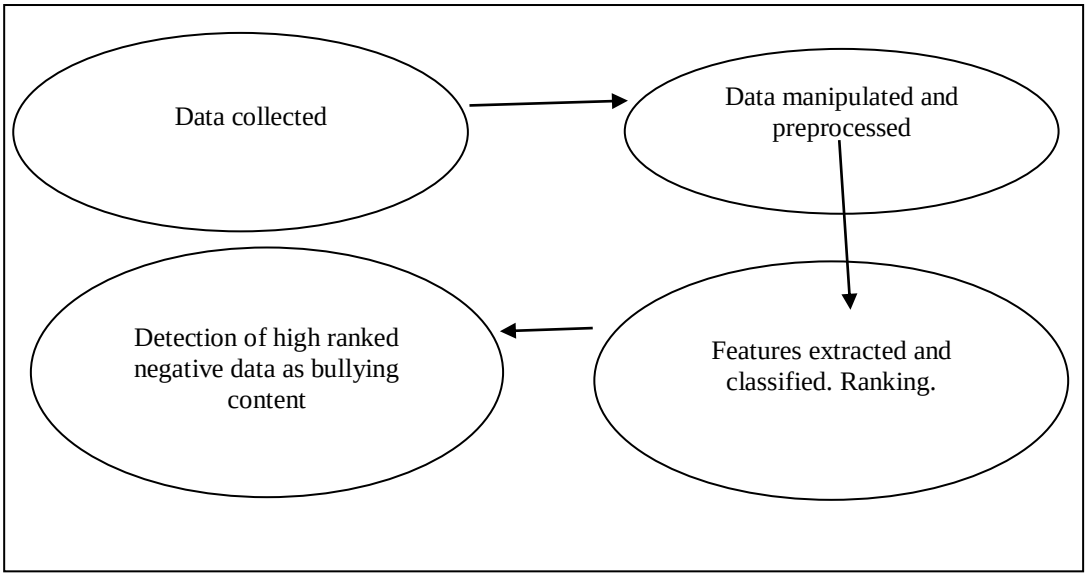

Figure 1: Block Diagram of Proposed Method

3.2 Data Manipulation and Pre-Processing

Then we have to perform feature extraction and polarise them. Work done so far gives different set of words. We prepared three types, bullying, praising and neutral words. After collecting data from Twitter, we try to formulate an algorithm that creates alist of negative words that has been used in a twitter. This process normalizes the BagOf Words (BOW) after eliminating articles and prepositions and extracting meaningfulwords in Hindi language written in English alphabet.

3.3 Feature Extraction and Classification, Ranking

To test the goodness of fit of bullying terms to our method, we adopt CHI-SQUARE [17] filtering approach that measures independenciesbetween two events and try to measure the degree of polarity of differentusers. If they provide a high rank of negative polarity, we can detect bullying contentthere. To compute expected frequency of terms extracted from tweets, we use the formula

$$E_{ij} = [(n_i) * (n_j)]/N;$$

Where E_{ij}= expected outcome
n_i= number of time i^{th} event occurs
n_j= number of time j^{th} event occurs
N= number of total outcomes

Now we compute the test statistics to find degree of freedom as

$$\chi^2 = \sum\sum \frac{(f_o - f_e)^2}{f_e}$$

Wheref_o=observed frequency,
f_e= expected frequency .

Now the Null hypothesis is

H_0 : μ_0=the number of tweets that are of negative polarity are below the threshold level

And the alternative hypothesis is

H_A: μ_1 = the number of tweets that are of negative polarity are above the threshold level

We then proceed to apply Bayesian Logistic Regression Method(BLR) toclassify and detect polarity of data. The Naïve Bayes method is used to apply MultinomialNB classifier to classify each user depending upon their words. We also try tofind out basic minimized risk using SVM classification algorithm. Thus we rank eachsentence of a user in terms of negative or positive polarity. The weight given to eachsentence by CH2 and Weight_Preproc help to determine rank of each user.We take each user as a node of the graph G(V,E). Each user is connected toother node via the ranked sentences they use in social platform. The cumulative sumcalculated from the sentences add up to formulate the degree of polarity of the userindividuals.

3.4 Detection of Cyber-Bullying

The last part of our method is of detecting most negatively polarized user. This canbe calculated from the weighted graph. Three major polarities are detected such as bullying words to be negative polarity, praising words to be positive polarity and neutral words. Most negative node is then kept under surveillance and if needed, an alert can be raised. The result of calculations over the dataset is in figure 2.

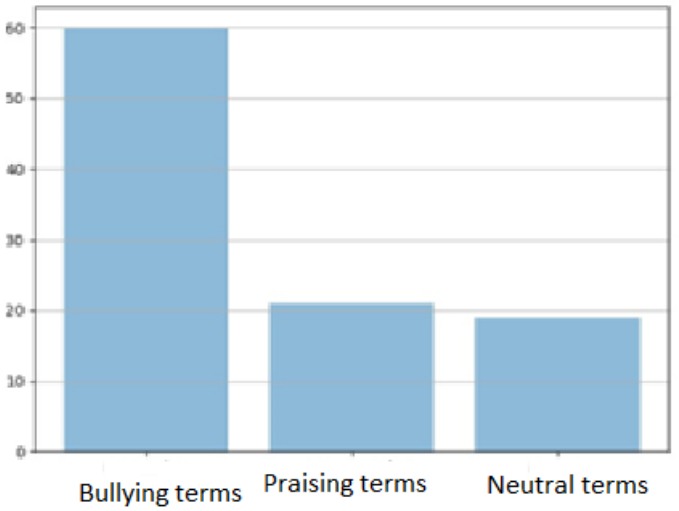

Figure 2: Plotted Graph of the Chi Square Method

The result shows that 60 percent of the data is of negative polarity, 20.5 percent words are of positive polarity and 18 percent words are considered as neutral polarity. If we discard the proportion of neutral dataset, the result clearly demonstrates the hatred nature of input data used by 4000 users over the specified temporal platform.

4. Conclusions

There are huge scope of modification and invention of new paths relating to sentimentanalysis and detection of cyber bullying. Further improvement can be done byconstructing an application to automatically detect the adverse cases. We can alsoapply this method for different languages like Bengali. Finding can be done for sarcasm and ironical words. We further aim to study the same problem domain under various other analytical methods such as Regression Analysis, SVM, Decision Tree and other Opinion Mining and Deep learning methods.

References

1. Basabdatta Das, Barshan Das, AvikChatterjee, Abhijit Das, “Chapter 13 - Designing of Latent Dirichlet Allocation Based Prediction Model to Detect Midlife Crisis of Losing Jobs due to Prolonged Lockdown for COVID-19”,Cyber-Physical Systems, Academic Press,2022,Pages 219-230,ISBN 9780128245576.
2. Soni, Chetan Kumar, and Atul D. Newase. "TEXT TO TEXT TRANSFER LEARNING BASED SENTIMENT ANALYSIS FROM THE REVIEWS AND FEEDBACKS OF E-COMMERCE PORTAL." *Harbin GongyeDaxueXuebao/Journal of Harbin Institute of Technology* 54, no. 6 (2022): 173-180.
3. Mohan, I., and M. Moorthi. "Retraction Note to: Topic flexible aspect based sentiment analysis using minimum spanning tree with Cuckoo search." *Journal of Ambient Intelligence and Humanized Computing* (2022): 1-1.
4. Machado, MateusTarcinalli, and ThiagoAlexandreSalgueiroPardo. "Evaluating Methods for Extraction of Aspect Terms in Opinion Texts in Portuguese–the Challenges of Implicit Aspects."
5. Vassilikopoulou, Aikaterini, Irene Kamenidou, and Constantinos-VasiliosPriporas. "Negative Airbnb reviews: an aspect-based sentiment analysis approach." *EuroMed Journal of Business* ahead-of-print (2022).
6. Barros, Meléndez, and Jose de Jesus. "A deep learning approach for aspect sentiment triplet extraction in portuguese and spanish." PhD diss., Universidade de São Paulo,2022.
7. NurulAtiraBinti Musa, Children Delinquency In Social Media: Legal And Shariah Perspectives, https://oarep.usim.edu.my/jspui/handle/123456789/16731, 2022.
8. Dutta, Poulami, and Chandan Kumar Bhattacharyya. "Multi-Modal Sarcasm Detection in Social Networks: A Comparative Review." In *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 207-214. IEEE, 2022.
9. Salas-Zárate, Rafael, GinerAlor-Hernández, María del Pilar Salas-Zárate, Mario Andrés Paredes-Valverde, Maritza Bustos-López, and José Luis Sánchez-Cervantes. "Detecting depression signs on social media: a systematic literature review." In *Healthcare*, vol. 10, no. 2, p. 291. MDPI, 2022.
10. Buford, Cayla, “The Dark Triad and Dark Behaviors: An Analysis of the Relationship Between Social Networking Sites, Deviant Behaviors, and the Dark Triad” Barry University ProQuest Dissertations Publishing, 2022.
11. Karthika, C. "Cyberbullying on celebrities: A case study on actress Parvathy." AIP Conference Proceedings. Vol. 2463. No. 1. AIP Publishing LLC, 2022.
12. MohdFadhli, SitiAisyah, et al. "Finding the Link between Cyberbullying and Suicidal Behaviour among Adolescents in Peninsular Malaysia." Healthcare. Vol. 10. No. 5. MDPI, 2022.
13. Sayfullaevich, P. S. (2021). Clinical and pathogenetic approaches to early rehabilitation of ischaemic stroke patients. International Journal of Health & Medical Sciences, 4(4), 373-380. https://doi.org/10.21744/ijhms.v4n4.1788
14. 13.Giumetti, Gary W., and Robin M. Kowalski. "Cyberbullying via social media and well-being." Current Opinion in Psychology (2022): 101314.

15. Byrne, Elizabeth, Judith A. Vessey, and Lauren Pfeifer. "Cyberbullying and social media: Information and interventions for school nurses working with victims, students, and families." The Journal of School Nursing 34.1 (2018): 38-50.
16. Ahmed, Nova. "Deep Learning Approach for Classifying the Aggressive Comments on Social Media: Machine Translated Data Vs Real Life Data." (2022).
17. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2022). Post-pandemic health and its sustainability: Educational situation. International Journal of Health Sciences, 6(1), i-v. https://doi.org/10.53730/ijhs.v6n1.5949
18. DenizKılınç,FatmaBozyiğit, “Application of Grid Search Parameter OptimizedBayesian Logistic Regression Algorithm to Detect Cyberbullying in Turkish MicroblogData”, Academic Platform Journal of Engineering and Science 7-3, 355-361, 2019.
19. Phillip G.Sokolove, WayneN.Bushell, “The chi square periodogram: Its utility for analysis of circadian rhythms”.