International Conference on Innovation in Engineering and Technology (ICIET) 27-29 December, 2018

An Approach to Detect Abusive Bangla Text

Md Gulzar Hussain*, Tamim Al Mahmud (Member, IEEE)†, Waheda Akthar‡
Department of Computer Science and Engineering*
Green University of Bangladesh, Dhaka-1207, Bangladesh
Email: gulzar.ace@gmail.com*, tamim@cse.green.edu.bd†, wahedaakthar@gmail.com‡

***Abstract*—Maximum research work on abusive text detection is in the English language and some of these are to detect offensive text. But in Bangla language, a few work is found. Detecting abusive text in Bangla language will be useful for preventing cybercrime such as online harassment, blackmailing and cyber bullying which are becoming the main concern in Bangladesh nowadays. Our goal is to detect abusive Bangla comments which are collected from various social sites where people share their sentiment, opinions, views etc. in this paper. We proposed a root level algorithm to detect abusive text and also proposed unigram string features to get a better result.**

Keywords: Abusive Text, Bangla Language, Natural Language Processing (NLP), unigram.

I. Introduction

Human languages are natural language, but computer languages e.g. C++ and C, are not. For example, Bangla, English, Chinese and French are natural languages. In computer science, it is probably a most challenging problem to make a computer to understand natural languages. Natural language processing is one of the fields of computer science, where we study that how computer and human can communicate with each other. Research works are increasing day by day in this field. Natural Language Processing contributing in almost every sector such as Customer Service, Automotive and Healthcare etc. Acording to Tractica(2017) report, market profit would be around $22.3 billion by 2025 on NLP hardware, software, and different services. It forecasts that this software service solutions utilizing Artificial Intelligence will increase to a market value to $5.4 (billion) by 2025 from $136 (million) in 2016 [1].

Bangla language is spoken by 205 million native speakers in 2010 and is the seventh most spoken native language by population in the world[2]. This total count covers 3.05 percent of the total world population. And there are approximately 250-300 million total speakers worldwide in 2017[3]. In Bangladesh total internet users are 81.7 million, among them, 30.0 million are active social media users and 28.0 million of them uses mobile phones to access social media in January 2018[4]. Facebook is the most popular social networking sites in the world with 2.23 billion monthly active users, Youtube is second with 1.9 billion monthly active users and Instagram is third with 1 billion monthly active users in August 2018[5].

978-1-5386-5229-9/18/$31.00

Social networking has nowadays become a part of human life. People share their information, feelings, and emotions using these social sites such as Facebook, Twitter, Instagram etc. As social networking increasing day by day, cyber oppression, online nuisance, blackmailing using these social sites are also increasing rapidly. These kinds of incidents are happening through negative images sharing, abusive comments and messages. There are approximately 25-30 million Facebook users in Bangladesh and 72 percent of them are male and 28 percent are female According to BTRCs November 2017s report[6]. In January 2018, among all the cities in the world Dhaka ranked second with the largest number of active Facebook users[4]. In Bangladesh, 49 percent of students are subject to cyberbullying[7] and 73 percent of women face cyber crime[8] via online or offline. In paper [9], the negative effects on children of cyberbullying are explained and another one [10] explained that the victims of cyberbullying are more intense to commit suicide than non victims. In Bangladesh cybercrime and cyberbullying are increasing[11]. A 16 years old girl from a small town in Sylhet committed suicide because of sexual harassment which is one kind of cyberbullying by a man from Dhaka through social networking[12].

42 million Facebook users which are 1.9 percent of total users of Facebook use Bangla language to communicate with each other[4]. Bangla language is also used by users in other social media sites. And the use of Bangla language in all social sites is increasing day by day. A lot of research has been done in the field of abusive text detection in English Language using social networks. Though we found some works on sentiment analysis on Bangla language but recently very few research has been done to detect abusive Bangla text using social network sites. So, in this field, there are lots of research scope for us. In this article, we are proposing a new technique to detect abusive Bangla text. We are just classifying a Bangla text as if it is abusive or not.

The rest of the paper is organized as follows. **Section II** discusses about the related works. **Section III** discusses the methodology and illustrate some sample comments following our methodology. **Section IV** shows details of our result analysis. **Section V** demonstrates the discussion and finally **Section VI** refers the conclusion.

II. Related Works

Detecting abusive text in social sites is a challenging work due to the changing nature and the variation in the language used. Researchers tried to develop many approaches to detect abusive or offensive text to get a better result. In paper [13], they developed a machine learning based method to detect hate speech on online user comments and categorized the sentences into Hate Speech, Derogatory and Profanity categorizes. They used Vowpal Wabbits regression model to measure different aspects of the user comment and used N-grams, Linguistic, Syntactic features. Using multiclass classifier, [14] categorized tweets into hate speech, offensive and neither of these two and differ hate speech from offensive language. The authors of [15] proposed a Lexical Syntactic Feature (LSF) architecture for detecting offensive content and identify potential offensive users in social media. In [16], the authors proposed a statistical topic modeling to detect profanity-related offensive content in Twitter.

When it comes to work with Bangla language it becomes more difficult to detect abusive text. The authors of [17] calculated the gross positiveness and negativeness of an article or a sentence with regards to the entire sense of the sentence. In this procedure of regaining information of an article, they applied Tf.Idf (term frequencyinverse document frequency) for a improved result. In this experiment, they wanted to figure out some patterns to understand positive and negative sentences so that they can categorize them. In paper [18], they tried to extract the negative or positive opinion or sentiment of a full text from Bangla Microblog posts. For classification, they used Support Vector Machine (SVM) and Maximum Entropy (MaxEnt) and used a semi-supervised bootstrapping approach for the development of the training corpus. By combining the results of word2vec word co-occurrence score with the sentiment polarity score of the words [19], the authors tried to do sentiment classification of Bengali comments and found the accuracy of 75.5%.

The only paper we found worked to detect abusive Bangla text is [20], used Random Forest (RF), Support Vector Machine (SVM), Multinomial Nave Bayes (MNB) and compared with unigram, bigram and trigram based Count-Vectorizer and Tf.idf-Vectorizer characteristics to detect Bengali abusive text. They found that Support Vector Machine Linear kernel trigram with Tf.idf-Vectorizer features accomplishes the best.

III. Methodology

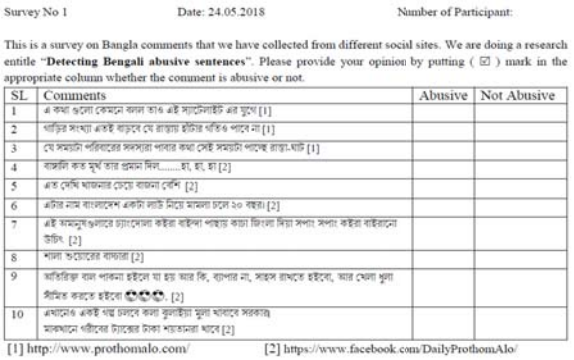
For conducting the experiment we proposed an algorithm to detect abusive Bangla text.

A. Dataset Collection

For the experiment, we collected various comments on posts from Facebook pages, Prothom-Alo news, and YouTube channels e.g. Prothom Alo [21], Mashrafe Bin Mortaza [22], Shakib Al Hasan [23], SalmoN TheBrownFish [24], Naila Nayem [25] and Prothom Alo News Portal [26]. Only public comments are collected without the commenters information due to protect privacy. In total, we collected 300 comments as we had to do the whole experiment manually. We used 250 comments to train our algorithm and 50 comments to test our algorithm.

B. Survey

To label every comment we had to run survey on the comments. We created a survey form like Figure 1 and operate this survey in Green University of Bangladesh. For every comment, we take opinion from at least 50 persons to take a decision if the comment is abusive or not and found results like Figure 2. Then we labeled every comment using the result.



Survey No 1 Date: 24.05.2018 Number of Participant:

This is a survey on Bangla comments that we have collected from different social sites. We are doing a research entitle "**Detecting Bengali abusive sentences**". Please provide your opinion by putting (☑) mark in the appropriate column whether the comment is abusive or not.

SL	Comments	Abusive	Not Abusive
1	এ কথা গুলো কেমনে বলা তাও এই সাইটেলাইট এর যুগে [1]		
2	গাড়ির সংখ্যা একই রাস্তায় যে রাস্তায় হাঁটার গতিও পাবে না [1]		
3	যে সমস্যটা পরিবারের সদস্যরা পাবার কথা সেই সমস্যটা পাচ্ছে রাস্তা-ঘাট [1]		
4	বাঙ্গালি কত মূর্খ তার প্রমান দিল.....হা, হা, হা [2]		
5	এত বেশি খাচ্ছনায় চেয়ে বাচ্চনা বেশি [2]		
6	একটার নাম বাংলাদেশ একটা লাট দিয়ে মানলা চলে ২০ বছর। [2]		
7	এই মামলুক গুলারে ট্যাংদেলা কইরা বাইন্ধা পাছায় কাচা ডিংলা দিয়া সপাং সপাং কইরা বাইরানো উচিৎ. [2]		
8	শালা স্টুয়োডের বাচ্চারা [2]		
9	অতিরিক্ত বাল পাকনা হইলে যা হয় আর কি, ব্যাপার না, সাহস রাখতে হইবো, আর খেলা ধুলা শীমিত করতে হইবো 😂😂😂. [2]		
10	এখানেও একই ধান্দা চলবে কলা বুলাইয়া মুলা খাবারে সরকার। মাঝখানে গরীবের ট্যাক্সের টাকা শয়তানরা খাবে [2]		

[1] http://www.prothomalo.com/ [2] https://www.facebook.com/DailyProthomAlo/

Fig. 1. A sample survey form

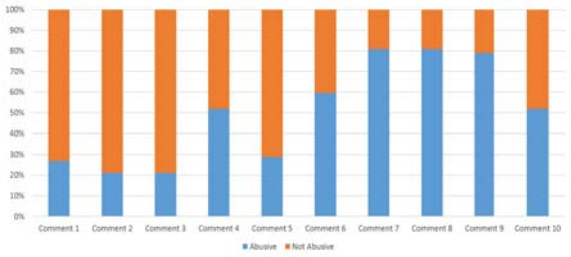


Fig. 2. A sample survey result

C. Preprocessing

The raw comments contain special characters e.g. @, #, - etc., punctuations, and emotions. At the time of preprocessing, we removed these special characters, punctuations and Unicode emotions manually. Conjunctions are also removed as they are unnecessary to abusive text detection. For this paper, we consider only Bangla Unicode characters.

D. Proposed algorithm

The proposed algorithm has two parts. One is for training and another is for testing.

To train the data, following algorithm is proposed-

- Step 1: Start

- Step 2: For every comments of the dataset following steps a and b should be taken
- Step 3: Initial

$$nWc = NumberOfWordsInTheComment$$

- Step 4:
  - Step a. If the comment is labeled as abusive then for every words in the comment
    - Step i. If the word is not the list of weighted words then

$$Weight_{Abusive} = \frac{1}{nWc}$$

    - Step ii. Else

$$Weight_{Abusive} = (old)Weight_{Abusive} + \frac{1}{nWc}$$

  - Step b. Else for every words in the comment
    - Step i. If the word is not the list of weighted words then

$$Weight_{NotAbusive} = \frac{1}{nWc}$$

    - Step ii. Else

$$Weight_{NotAbusive} = (old)Weight_{NotAbusive} + \frac{1}{nWc}$$

- Step 5: End

To classify the data, following algorithm is proposed-

- Step 1: Start
- Step 2: For every comments of the dataset following steps a and b should be taken
  - Step 3:
  - Step a. for every words in the comment
    - Step i. Word is not in the Term Weight List than

$$TotalWeight_{Abusive} = 0$$

      and

$$TotalWeight_{NotAbusive} = 0$$

      and Skip the next Step ii and iii.
    - Step ii.

$$TotalWeight_{Abusive} = Weight_{Abusive} + TotalWeight_{abusive}$$

    - Step iii.

$$TotalWeight_{NotAbusive} = Weight_{NotAbusive} + TotalWeight_{NotAbusive}$$

  - Step b. If

$$TotalWeight_{Abusive} > TotalWeight_{NotAbusive}$$

    then set the label of that comment as abusive that is 1 else set as not abusive that is 0
- Step 4: End

E. Feature Extraction

To find out, with what kind of feature our proposed algorithm performs better we can conduct the experiment with three types of string property; unigram, bigram, and trigram. The word relevancy in single sentence are not considered in the unigram features. But which words are more abusive, it can be found using this feature. The relevancy between two successive words in a sentence are considered in the bigram structures. In trigram, the correlation between three successive words is considered in a sentence.

F. Challenges

It is hard to process Bangla text. Also, Bangla language is more complex than the English language due to its structure and way of using a word. Sometimes it becomes hard to differentiate between hate speech, funny speech and abusive speech.

G. Representation with Test Comments

A step-wise representations of our proposed algorithm is given below using some sample comment. Consider unlabeled comments of figure 3,

Serial	Comments
01	শালা শুয়োরের বাচ্চা।
02	শালা বাবুকে আরও খেতে দাও।
03	বাচ্চা দেরকে খেতে দাও।

Fig. 3. Unlabeled sample comments

Here think the first and second comments as training data. After running the survey suppose we found the following labeled comments in figure 4-

Serial	Comments	Label
01	শালা শুয়োরের বাচ্চা।	1
02	শালা বাবুকে আরও খেতে দাও।	0
03	বাচ্চা দেরকে খেতে দাও।	

Fig. 4. Labeled sample comments after survey

Now the part of preprocessing the comments and use of training algorithm comes into play. So, the following table found given in figure 5-

For the third comment, if we run the test algorithm then the following table of figure 6 calculated-

Figure 6 shows that for the third comment sum of

$$TotalWeight_{abusive} < TotalWeight_{notabusive}$$

So that the third comment is not abusive.

Comment	Weight $_{abusive}$	Weight$_{notabusive}$
শালা	0.33	0.2
শুয়োরের	0.33	0.0
বাচ্চা	0.33	0.0
বাবুকে	0.0	0.2
আরও	0.0	0.2
খেতে	0.0	0.2
দাও	0.0	0.2

Fig. 5. After preprocessing and running the training algorithm

Comment	বাচ্চা	দেরকে	খেতে	দাও	Sum
TotalWeight$_{abusive}$	0.33	0	0	0	0.33
TotalWeight$_{notabusive}$	0.0	0	0.2	0.2	0.40

Fig. 6. After running the classifying algorithm

IV. Experimental Result

We divided our data set into three sets with 100 comments, 200 comments, and 300 comments. We provide the experimental results of Correct Abusive, Wrong Abusive, Correct not-Abusive, Wrong not-Abusive for the binary classification task using our proposed algorithm with unigram feature for 20% test data in TABLE I and Figure 7. From TABLE I and Figure 7 we can see that the number of accurateness is increasing with the increase of number of comments.

Number of Comments	Correct Abusive	Wrong Abusive	Correct not Abusive	Wrong not Abusive
100 comments	10	0	4	6
200 comments	16	4	10	10
300 comments	25	5	18	12

TABLE I
CORRECT AND WRONG RESULT FOR ABUSIVE AND NOT ABUSIVE CLASS FOR 20% TEST DATA OF THREE SETS OF COMMENTS

V. Discussion

A. Future Work

We are proposing some ideas for our future works.

- We will try to implement the whole idea to make it faster and automated.
- Our algorithm can be integrated with various Machine learning algorithms like Nave Bayes, Random Forest, and Support Vector Machine etc. to observe if the result become more accurate than the previous methods.

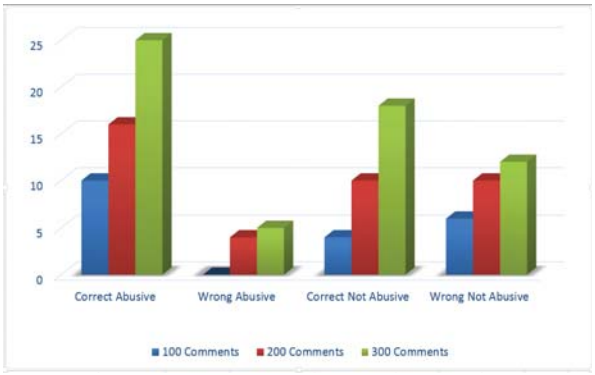


Fig. 7. Correct and wrong result for abusive and not abusive class for three sets of comments

- New features can be integrated to get more accurate results.
- The proposed algorithm can be modified to differentiate funny sentences, hate speeches, angry sentences, and abusive sentences.
- An application can be developed to detect abusive texts when people browse various social sites using browsers or mobile app.
- The number of comments in the dataset has to be increased to get a more accurate result.

VI. Conclusion

We tried to detect Bangla abusive text using our proposed algorithm. Though this type of root level algorithm is not appropriate for nowadays, We hope our future plan will give a better result than already developed ideas. There are still many opportunities to improve our experimental methodology. Natural language processing using various machine learning algorithms and techniques, but work in Bangla language is not increasing as expected due to limited resources and mentoring.

References

[1] R. Madhavan. (2018) Natural language processing current applications and future possibilities. [Online]. Available: https://www.techemergence.com/nlp-current-applications-and-future-possibilities/
[2] Wikipedia. (2010) List of languages by number of native speakers. [Online]. Available: https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers
[3] wikipedia. (2017) Bengali language. [Online]. Available: https://en.wikipedia.org/wiki/Bengali_language
[4] W. A. Social and Hootsuite. (2018) 2018 digital yearbook. [Online]. Available: https://digitalreport.wearesocial.com/
[5] P. KALLAS. (2018) Top 15 most popular social networking sites and apps [august 2018]. [Online]. Available: https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/
[6] I. Tarik. (2018) Demographics of facebook population in bangladesh, april 2018. [Online]. Available: http://digiology.xyz/demographics-facebook-population-bangladesh-april-2018/
[7] D. Unb. (2016) 49% bangladeshi school pupils face cyberbullying. [Online]. Available: https://www.thedailystar.net/bytes/%E2%80%9849-bangladeshi-school-pupils-face-cyberbullying%E2%80%99-287209

[8] b. Senior Correspondent. (2017) 73 percent women subject to cyber-crime in bangladesh. [Online]. Available: https://bdnews24.com/bangladesh/2017/03/09/73-percent-women-subject-to-cyber-crime-in-bangladesh
[9] C. L. Nixon, "Current perspectives: the impact of cyberbullying on adolescent health," *Adolescent health, medicine and therapeutics*, vol. 5, p. 143, 2014.
[10] S. Hinduja and J. W. Patchin, "Bullying, cyberbullying, and suicide," *Archives of suicide research*, vol. 14, no. 3, pp. 206–221, 2010.
[11] Asaduzzaman. (2016) Cybercrime cases on the rise. [Online]. Available: https://en.prothomalo.com/bangladesh/news/122235/Cybercrimes-on-the-rise-due-to-section-57
[12] S. S. Preetha. (2015) Digital sexual harassment in digital bangladesh. [Online]. Available: https://www.thedailystar.net/in-focus/digital-sexual-harassment-digital-bangladesh-82480
[13] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang, "Abusive language detection in online user content," in *Proceedings of the 25th International Conference on World Wide Web*, ser. WWW '16. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2016, pp. 145–153. [Online]. Available: https://doi.org/10.1145/2872427.2883062
[14] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," *arXiv preprint arXiv:1703.04009*, 2017.
[15] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting offensive language in social media to protect adolescent online safety," 09 2012, pp. 71–80.
[16] G. Xiang, B. Fan, L. Wang, J. Hong, and C. Rose, "Detecting offensive tweets via topical feature discovery over a large scale twitter corpus," in *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2012, pp. 1980–1984.
[17] M. M. Nabi, "Detecting sentiment from bangla text using machine learning technique and feature analysis," 2016.
[18] S. Chowdhury and W. Chowdhury, "Performing sentiment analysis in bangla microblog posts," in *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*. IEEE, 2014, pp. 1–6.
[19] M. Al-Amin, M. S. Islam, and S. Das Uzzal, "Sentiment analysis of bengali comments with word2vec and sentiment information of words," 04 2017.
[20] S. C. Eshan and M. S. Hasan, "An application of machine learning to detect abusive bengali text," in *2017 20th International Conference of Computer and Information Technology (ICCIT)*, Dec 2017, pp. 1–6.
[21] Prothom alo - facebook home. [Online]. Available: https://www.facebook.com/DailyProthomAlo/
[22] Mashrafe bin mortaza - facebook home. [Online]. Available: https://www.facebook.com/Official.Mashrafe/
[23] Shakib al hasan - facebook home. [Online]. Available: https://www.facebook.com/Shakib.Al.Hasan/
[24] Salmon thebrownfish. [Online]. Available: https://www.youtube.com/user/salmanmuqtadir
[25] Naila nayem - facebook home. [Online]. Available: https://www.facebook.com/artist.nailanayem/
[26] Prothom alo - online news portal. [Online]. Available: https://www.prothomalo.com/