Automatic Averting From Cyberbullying Using Efficient Polarity Count Algorithm

Thenmozhi B, Prof. Rajaprabha M N

MCA Student, School of Information Technology and Engineering, VIT University, Vellore, India

Assistant Professor, School of Information and Technology, VIT University, Vellore, TamilNadu, India

ABSTRACT: Cyberbullying is the term which meant for the act of harassing or frightening an in individual by means of internet. There are many customs of cyberbullying, but its only intention is to threatening a person with negative words and harmful pictures. The advancement of technology also gives advantage to misuse it, cyberbullying is one such which is undetectable until and unless the victim reports it. To avoid this and to prevent the victims from bullies, automatic blocking bullies system with the help of data polarity identified by sentiment analysis. Using sentiment analysis, personality detection can be made through over the comments in the blogs, social media etc. By taking comments, messages and post for sentiment analysis, the rate of polarity is defined. Here, it is proposed for preventing an individual from harmful and hurting texts. By mining the texts alone, almost 70% of cyberbullying activities can be avoided.

KEYWORDS: cyberbullying; sentiment analysis; polarity; personality detection; mining.

I. INTRODUCTION

In recent years, the electronic communication and internet has revolutionized the entire mankind. Though information sharing and computer based communications is a great boon to us, it also holds some adverse effects which are quite inescapable. One of such bad misuse of technology is the cyberbullying. Cyberbullying has different kinds of characterizations and common description is that, it is the act of harassing a person through electronic or digital media so as to harm or threaten or abuse a person brutally for entertaining, rage or any other personal motives which usually carried out anonymously [1].

This cyberbullying has different forms of delivering based on the nature it as, like, flickering which meant for fighting verbally using some vulgar language, harassment by sending abusive texts to insult repeatedly, demeaning or denigration gossip, to pretend to be someone else, impersonation, sharing or outing someone's personal information or awkward secrets, trickery means to draw out and then expose personal information, exclusion that is to intentionally excludes someone from a group or site, Cyber stalking by repeated frightening threats [3][4].

Cyberbullying is very prominently found on social networking sites like Twitter, Google plus and Facebook. So that the statistics says that 70% of cyberbullying is targeted by texts only other than the texts cyberbullying is carried through abusing images or amended video to make an individual inconvenience [8]. So, in this paper proposing a method to block the bullies automatically from posting further by taking count of his or her posts, messages and comments polarity in any social media. Using sentiment analysis polarity of the text can be determined which enters into a proposing effective polarity count algorithm.

Sentiment analysis is usually used for measuring the opinions based on the text given. In paper [2], using sentiment valence personality detection is made automatically using the posts made by an individual in the social media and other blog activities. Text pre-processing like lemmatization, tokenization, sentence split, parts of speech, sentiment, regular expression etc. is carried through Natural Language Processing technique. Here Stanford Natural Language Processing website is used to find the accurate polarity level of the text [11]. The tree model used to determine the value and the statistics graph shows each value like negative, positive and neutral polarity level with its percentage separately.

This polarity rating helps us to determine the valence in the text made by the victim [5]. The algorithm is used to take the count of polarity in account to block the post or report the person when it reaches to a certain extent. This automatic blocking of abused post by a victim protect the internet from bullies and also reduces the harassment made through internet. The polarity level is categorized according to each sentiment analysis way like machine-learning, text

processing etc. as referred in [12]. As per the way of finding the polarity level also get differed, and also the corpus which is connected in analysing also contains the polarity level mapped with each of the text. This corpus also takes into account of the newly entered text.

II.   RELATED WORK

In Paper [1], NektariaPothan et.al, have proposed a work on cyberbullying detection using time series modelling. They have suggested a study in deducing the level of accuracy in cyberbullying attack with the help of classification methods. Also they have briefly discussed about categorizing potential patterns within usual linguistic style of individual analysts. They have utilized a set of real world data which consists of conversations of two individuals. The predator questions are annotated manually with the specified numeric label and these questions were formulated with the help of the singular value decomposition. The question set is subjugated as a signal and its magnitude will be based upon the bullying content's degree. These signals get analysed through neural networks and give the level of abuse of the questions. They have used dynamic time wrapping algorithm for getting the similarities between the signals and SVD for time series.

SaravananSagadevan et.al, in [2] have proposed a paper on 'sentiment valences for automatic personality detection of online social networks users using three factor model' using PEN model. Here the Three Factor Personality (PEN) Model acts as a personality framework helps in depicting the characteristics of a user. Personality is individual differences of perceptions, emotions, cognitions, motivations, and actions when those individuals interacts with various environmental situations. PEN model determine three main traits to represent the characteristics of humans: extraversion, neuroticism and psychoticism. The two poles of the personality are extraversion and neuroticism. Extraversion trait reflects positive characteristic such as sociable, loving and enthusiastic while neuroticism trait reflects negative characteristics such as sadness and hate. Psychoticism reflects aggression and anti-social behaviours. The linguistic patterns used by the users in the Facebook and social networks to communicate helps in identifying their individual personalities. This study has conducted over thousands of Malaysian users to identify user personality. The user status messages extracted to study the majority perceptions. In general the study revealed the perceptions are homogenous.

Ana Kovaevi has proposed a work on cyberbullying detection using web content mining [6]. The author discuss about the existing possible concepts on making the internet to be a safer component to the environment by detecting cyberbullying through mining the web content. The author presented as a survey of characteristics and classifiers of data sets from youtube comments, twitter, chat on the MySpace and some web sites for identifying whether it is cyberbullying or not. Such problem of cyberbullying has to be solved or atleast minimized to some extent for victim user to be safer.

MarliesRybnicek et al., have proposed a work on 'Facebook Watchdog: A Research Agenda For Detecting Online Grooming and Bullying Activities' where they have used text mining, social media analytics and image analysis techniques to detect bullying activities to alert the public about the enduring attacks [7]. Here they have aimed to create an app which can be integrated with Facebook, if any threats have been detected in the timeline then the app intimates the user and the trusted friend of the user by sending a warning message. In text mining the statistical approach has been used so bag-of-words approach gives the threatening words list, whereas in social media analytics with the help of clustering and pattern recognition the profile activity and the mood of the user is predicted and in image analysis face-recognition, texture analysis and skin detection algorithm plays major role in identifying sexual/harassing images. As a future development they are working in collaboration with psychological experts.

In paper [8], WalaaMedhat et.al, has took a survey on Sentiment analysis algorithms and applications to find how the mining differs in each field with different algorithm. Sentiment analysis and Opinion mining recently become prime research area in mining field. Here they deal in mining opinions, attitudes and emotions from the product reviews or feedbacks provided by the people. Primarily sentiment analysis identifies whether the opinion is positive or negative, then it finds and extracts subjective information from raw data using different algorithm and techniques. Lexicon based or statistical method is used as a feature selection method depending upon the type of resource. SA techniques are used in identifying the differences or polarity between the reviews and find the mood of the reviewer. Sentimental analysis treaties with three level of granularities: Document level, Sentence level and Aspect level. The fields such as transfer learning, emotion detection, and building resources are related to sentiment analysis.

In paper [9], it is described ontology in information science defines the interrelationships of the entities that exist in a particular domain. It helps in limit complexities, organize information and problem solving. Ontology-based Sentiment Analysis Process for Social Media content acts as a medium to extract and analyses the abundance of information on social media sites. This OSAPS process is presented with negative sentiments in this paper. The tweets related to negative sentiments such as delay in delivery, lost package or inappropriate customer service. Then extract information like nouns or verbs from the content using GATE software. After using GATE software Portege software is used to build ontology model and to query the model while ontology model consists of class, subclass, objects and object properties. SentiStrength tool is then used to differentiate positive and negative tweets. Finally SPARQL is used to query the ontology model and retrieve the information.

III.   DETAILED DESCRIPTION

Here with the help of "Sentiment tree bank model" from Stanford University we have processed sentiment analysis over the text to find the polarity at the end [13]. On common note, text processing is a method that consumes a lot of time. It deals with a lot of supervised training, assessing the resources and effectual module for customized results. In order to tackle these scenarios Sentiment detection tree bank model is being used which could perform the following processes like True case, Parse, Regular Expression NER, Conference graph and Sentiment , Named-Entity-Recognition (NER), Tokenization, Sentence split, Parts Of Speech (POS), Lemmatization, Gender, [10][11].

A.   Tokenization

The whole sentences are broke into bits and pieces (.i.e. words & phrases) which are termed as tokens. These tokens acts as the source for the upcoming phases like text mining or parsing. These tokenizers are classified further into different types like Stanford tokenize, customer parser, OpenNLPtokenizer, Hypothetical Tokenizer, white space parser and PTBTokenizer which is most used among others [15]. The notable types of the above mentioned tokenizers are white space parser which can perform the worse part of the tokenization and Hypothetical tokenizer provides much better tokenization, which can segregate the sentences by punctuators and white spaces, which may or may not be available in the token list & this can able to handle noisy text as well. Also termed as Word segmentation.

For example, looking into this incoming text, "You look awkward", these words will be considered as a single semantic group for next round of processing. Now the sentence will be tokenized as "you", "look", "awkward". The analytical rule for tokenization are considered to eradicate the empty spaces, brackets, slashes and quotation marks. The next comes the stemming process which normalizes these tokens which in turn deals with parts of speech. Stem word is nothing but filtering out the prefixes and suffixes of a particular word. Usually performed by Lovin's stemmer, Porter stemmer, Paice stemmer, Dawson stemmer and so on. History suggests that porter stemmer is the most effective and produces the least error rates comparing to the others. The instructions of porter stemmer is classified as context-stemmer and it has delivered 93.3% of meaningful words.

B.   Sentence split

A continuance of tokens are segregated into sentences.

C.   Parts of speech,

Every token is appended with a POS like verb, subject, adjective, noun etc. For example, "You look awkward" is represented as you (S) look (V) awkward (A).

D.   Lemmatization

The lemmas for all the tokens are delivered in bulk. It is the way of processing and organizing varies types of words, so that they can be assess as a single component. This process is used to create the word lemma for the entire token in the dictionary form. It will find the proper lemma for the given input which in turn finds the single item which is assessed by the group of words. The lemmatization process is done for entire text processing based on the POS and segregation between these words which have unlike meanings. In order to find lemmas, lemmatization has to run before parsing phase.

E.   Named Entity Recognition (NER)

The entities like Organization, Location, Persons and number entities like Number, Money and Date/Time can be assessed here, where number entities are evaluated using rule based system and Named entities are considered using CERF sequence taggers. The entity detection is usually done by picking the subset of tokens. One of the chunkier need

to find an operational and then adjectives and noun in order. Considering the given input, chunk parser will be created and executed. The results derived will be in the form of tree. And then comes the RegerNEX process, which can able to map between regular expression and NE classes [15]. RegerNEX consists of two rules and for each further divided into two fileds which is differentiated by one tab. The regular expression involves with one or more expressions differentiated by non-tab white spaces.

F. True case

One can identify the true cases of tokens in the text where the information could not be found, e.g., Text in caps. The true case text is achieved by re adjusting the token text. The dependency and the constituent representation is used to provide full syntactic analysis.

G. Regular Expression NER

Java regular expression is used here to perform a low level rule based NER. The main agenda of this is to provide a user free framework to induce named Entity labels. For examples, he following words VARITIES, GODS, OCEANS, will be identified as variety, God, Ocean by the regular expressions that is present in the file already.

H. Co-reference graph

Here nominal and pronominal co-reference resolution algorithms are used. The equivalent annotations is saved from the finished co-reference graph with all nodes stated within it.
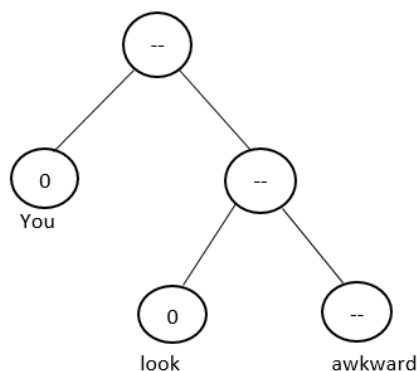
I. Sentiment

Fig. 1. Sentiment Tree

Here we are using socker et al's sentiment model. The sentence core level map will be in the form of binary tree for that particular sentence. The sentiments are further segregated into Very Positive (++), Positive (+), Neutral (0), Negative (-) and Very Negative (- -). The sentences will be differentiated based on tokenization and other concepts mentioned above. The final sentiment of the incoming text will be present in the root node of the tree. The root node will contains the output of the leaf node and the word will be present in the leaf node. For example sentiment tree is shown in Fig. 1 [14].

IV.   PROPOSED METHODOLOGY

To avoid cyberbullying getting aggressive and to prevent the victim from getting bullied, I have proposed an idea of automatic blocking or reporting of the negative comments or posts made by an individual. Every comment or post must go through the below architecture before getting posted in the social media. So that the post can be checked for whether the content is bullying someone or not and how extent it is, based on the extent of bullying comment it is decided whether to post or to enquiry the victim whether to be posted or not.

When a person enters a comment or post or some test in social media it goes into text pre-processing for analysing sentiment in later step. Using sentiment analysis the polarity percentage of the text can be determined. With the help of Polarity percentage, it can be easily defined that whether the text is bullying or not. So after finding the

polarity percentage, it goes through the polarity count algorithm which is efficient and effective in finding whether the text can be posted or it must be blocked or it must be sent for enquiry to the victim whether it is hurting or not.
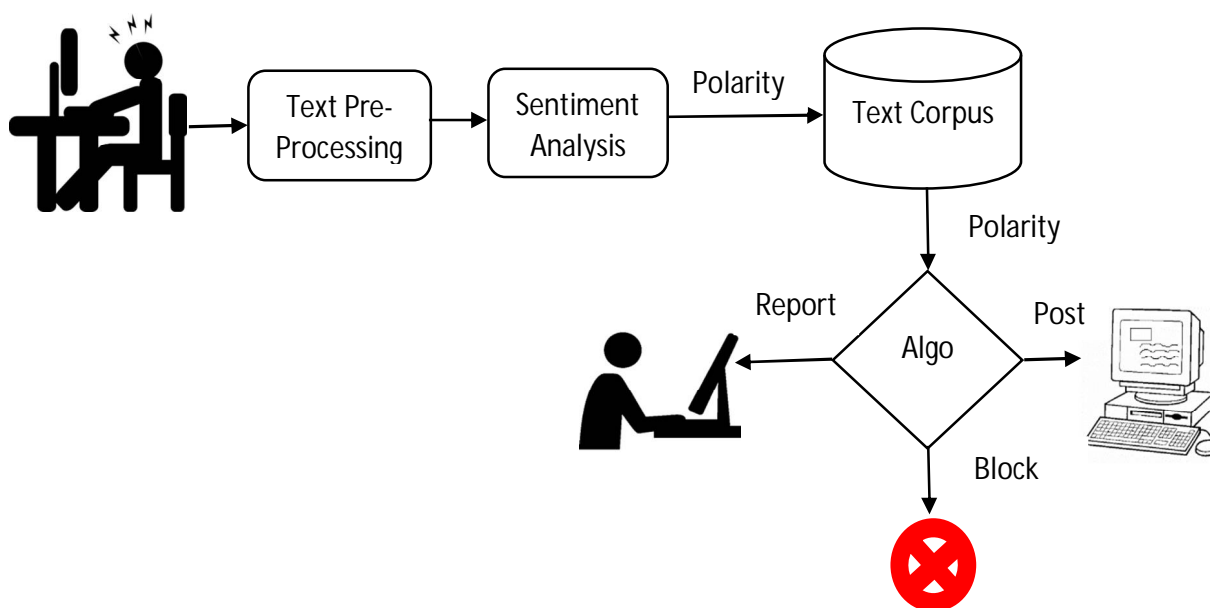

Fig. 2. Proposed Architecture

The polarity of the word is determined by the value mapped with each word in the corpus and also if new words comes it takes into account with its polarity value for later use. The polarity from sentiment analysis is distinguished in five ways as Very-Positive (VP), Positive (P), Neutral (NU), Negative (N) and Very Negative (VN). The polarity refined from the corpus consists of both the sentiment analysis value and also cyber bullying content's value is added to the polarity.

V.   ALGORITHM

Aim of the proposed Polarity count algorithm is to effectively count the positive and negative polarity of the text and efficiently result the system whether the text is bullied or not. This algorithm acts as the main part of this proposed methodology where the result is decided by this process. The algorithm with input and output is mentioned below as stepwise

Algo 1: Polarity Count Algorithm
Input: Polarity values from corpus {VP, P, NU, N and VN}
Output: Identifies whether to block or report or to post the text.
Step 1: Neg := VN + N;
Step 2: Pos := VP + P;
Step 3: If (Neg>= 80)
Step 4: Block;
Step 5: Else If (Neg>= 50)
Step 6: Report;
Step 7: Else
Step 8: Post;
Step 9: End;

Let the output be as mentioned as the results in the upcoming section. This algorithm helps in differentiating the text from the normal and bullying text, which in turn helps the methodology in protecting an individual from cyberbullying. As said, this algorithm plays a major role in categorizing the text.

VI. RESULTS AND DISCUSSION

This polarity value goes into the polarity count algorithm, according to the algorithm the result of the process is defined. The result can be either of the three followings

Negative Polarity	Action to be taken
< 50	Post the text as defined by the person.
>= 50 &&<= 80	Report to the person, whether the text can be posted or it cannot be. As per the persons detection either of the first or second result is carried forward.
>= 80	Block the comment from posting, to avoid bullying comment.

Table. 1. Action table

After reporting to the person, the person has their own choice to post it or delete which in turn also helps the victim in understanding what the person wants to convey and whether he/she is really hurting through the text. This also warn the victim by showing the real face of the person, so that the victim can also block him/her in all the ways as a prior step from getting hurt again. Since the polarity level is less the text has only 50% of harmful comment which is least than other.

VII. CONCLUSION AND FUTURE WORK

From the proposed idea, it is clearly known 60-70% of text cyberbullying can be avoided from posting and also reporting the person helps us to improve that people does not miss something they want to know. This helps in preventing a person from getting bullied and also protect the internet from cyberbullying crimes. Using sentiment analysis polarity of the text has been defined and also analysing the text from corpus helps in identifying the most used cyberbullying text.

Cyberbullying not only occurs through text but also by various other ways like picture, video etc. To make as free cyberbullying internet it is very difficult but to prevent people from getting bullied is little easier where we have to put some more efforts in making secure social media. As a future work we can concentrate on image processing to avoid cyberbullying through pictures and videos. And also we can keep the count of each person's bullying and Non-bullying messages into account and report the site about the person when it reaches to an extent. This may also help in punishing the person legally by reporting to cyber-crime corps.

REFERENCES

1. NektariaPotha and ManolisMaragoudakis, "Cyberbullying Detection Using Time Series Modeling", IEEE International Conference on Data Mining Workshop, pp.373-382, 2014.
2. SaravananSagadevan, NurulHashimah A. H. M. and MohdHeikalHusin, "Sentiment Valences for Automatic Personality Detection of Online Social Networks Users using Three Factor Model", The Third Information Systems International Conference, Procedia Computer Science 72, pp. 201-208, 2015.
3. Margaret Anne Carter, "Protecting Oneself from Cyber Bullying on Social Media Sites – a Study of Undergraduate Students", 3rd World Conference on Learning, Teaching and Educational Leadership, Procedia - Social and Behavioral Sciences 93, pp. 1229-1235, 2013.
4. This emotional Life – Suzanne Phillips and PsyD, http://www.pbs.org/thisemotionallife/blogs/coping-cyberbullying-use-technology-terrify.
5. Farhan Hassan Khan, Usman Qamar and Saba Bashir, "SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection", Science direct journal of applied soft computing, Vol. 39, pp. 149-153, 2016.
6. Ana Kovacevic, "Cyberbullying detection using web content mining", 22nd Telecommunications forum Telfor (TELFOR), published by IEEE, pp. 939-942, 2014.
7. MarliesRybnicek, Rainer Poisel and Simon Tjoa, "Facebook Watchdog: A Resarch Agenda For Detecting Online Grooming and Bullying Activities", 2013 IEEE International Conference on Systems, Man, and Cybernetics, Vol.10, pp. 2854-2859, 2013.

8. WalaaMedhat, Ahmed Hassan and HodaKorashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, Production and hosting by Elsevier, Vol. 5, pp. 1093–1113.
9. Pratik Thakora andSreelaSasib, "Ontology-based Sentiment Analysis Process for Social Media Content", INNS Conference on Big Data, Procedia Computer Science, Vol. 53, pp 199–207,2015.
10. Chee Kian Leong, Yew Haur Lee and WaiKeongMak, "Mining sentiments in SMS texts for teaching evaluation", Expert Systems with Applications, Vol. 39, pp. 2584–2589, 2012.
11. Sentiment Analysis and its process, http://nlp.stanford.edu/sentiment/index.html.
12. Kelly Reynolds, April Kontostathis and Lynne Edwards, "Using Machine Learning to Detect Cyberbullying", 10th International Conference on Machine Learning and Applications (ICMLA), IEEE conference publication , Vol. 2, pp. 241-244, 2011.
13. ChetashriBhadane,HardiDalal and HeenalDoshi, "Sentiment Analysis: Measuring Opinions", International Conference on Advanced Computing Technologies and Applications (ICACTA), Procedia Computer Science 45, pp. 808-814, 2015.
14. Sentiment Analysis Tree bank model, http://nlp.stanford.edu/sentiment/treebank.html.
15. Deqing Wang, Hui Zhang,RuiLiu ,Weifeng v. and Datao Wang, "T-Test feature selection approach based on term frequency for text Categorization", Science Direct journal Pattern Recognition Letters,Vol. 45, pp. 1–10, 2014.

BIOGRAPHY

Thenmozhi B is a student currently pursuing final year in Master of Computer Application at VIT University. She have completed her Bachelors in Computer Application from the same University with 94.8% as a gold medalist. Since her core is Information Technology and Computer Application she ismore excel in these areas. As her course is research based learning she have also published some papers on Big Data Analysis, Data Mining, Cloud Computing, Internet security etc. Her interest towards programming languages also made her to do some projects in .Net, Java, HTML and C++.

Rajaprabha M N is a Faculty in the School of Information Technology and Engineering Department, VIT University, Vellore, TamilNadu, India. She received her M.Tech(CSE), B.E(CSE) from Amrita Institute of Technology and Science. Her research interests are Cloud Computing and Data Mining.