Alert System for Controlling Cyberbullying Words Using Fuzzy Logic and Fuzzy Inference Engine

S. Senthil Kumar[1] and V.Kathiresan[2]

[1]Assistant Professor, Department of Commerce with Computer Applications,
[2]Head, Department of Computer Applications (PG),
Dr.SNS Rajalakshmi College of Arts and Science (Autonomous), Coimbatore, Tamil Nadu, India

***Abstract* - Cyberbullying is a socially aggressive and has powerful negative effects for individuals, specifically adolescents and youngsters. Cyberbullying allows the offender to mask his or her identity behind a computer. This anonymity makes it easier for the offender to strike blows against a victim without having to see the victim's physical response. The distancing effect that technological devices have on today's youth often leads them to say and do crueler things compared to what is typical in a traditional face-to-face bullying situation. In the recent times many methods for automatic thoughts of mining in the online data are becoming increasingly important, to increase the safety parameter of the people. This framework is proposed to extract Cyberbully polarity from the Forum using Fuzzy logic technique. At first, the given input is pre-processed and the useful content is gathered. Subsequently, the pre -processed data will be sent to the features extraction method. Probabilities of the words are calculated by using Fuzzy Decision Tree Method. Fuzzy rules can be applied in all these features to extract the certain set of cyberbully words like bad words, insulting words, threatening words and terrorism words from the given input, hence we use text mining here. Finally this method will return the reduced and accurate cyberbully words. This method is performed by human annotation using the existing methods like Mamdani Fuzzy System and Naive Bayes classifier. Extensive experiments are performed by using fuzzy logic on crime debate forum and the results show that this proposed approach is better than the traditional one. Aggressive text detection in social networks allows identifying offenses and misbehavior, and leverages tasks such as cyberbullying detection. Social media became a very useful platform to express ourselves. The expressions have adverse reactions as well. We intend to take data from these platforms and make use of it to improve on the safety parameter. For the development of the system we take the data available on Twitter and filter all the useful contents.**

***Keywords:* Cyberbully, text mining, Forum, fuzzy-logic, fuzzy decision tree, Naive Bayes classifier, feature extraction**

I. INTRODUCTION

Different types of cyberbullying are listed below:

Flaming: Online fights usually through emails, instant messaging or chat rooms where angry and rude comments are exchanged.

Denigration: Putting mean online messages through email, instant messaging, chat rooms, or websites set up to make fun of someone.

Exclusion: Intentionally leaving someone out of a group such as instant messaging, friend sites, or other online group activities.

Outing: Sharing secrets about someone online including private information, pictures, and videos.

Trickery: Tricking someone into revealing personal information then sharing it with others.

Impersonation: Pretending to be someone else when sending or posting mean or false messages online.

Harassment: Repeatedly sending malicious messages to someone online.

Cyberstalking: Continuously harassing and denigration including threats of physical harm.

II. WHAT CAUSES PEOPLE TO PARTICIPATE IN CYBERBULLYING

a. Children do it for revenge of other peers
b. Children do it simply for entertainment purposes
c. Some do it to boost their ego
d. Children do it to get a laugh from their peers
e. Children do it to help bolster or remind people of their own social standing

III. EFFECTS OF CYBERBULLYING

a. It can give people a bad reputation
b. Causes children to have low self-esteem
c. Causes children to become depress if they are constantly picked on by peers
d. Makes it hard for kids to communicate with others
e. Sometimes cause people to commit suicide due to embarrassment

IV. PROPOSED METHODOLOGY FOR FINDING AND ANALYZING CYBERBULLYING

A. Data collection

The first step in building this project was to gather data to compare the users. We had to analyse the moods of the people that are depicted through their tweets, whether angry words, foul language or jargons. Whether people responded positively or negatively to the occurrence of a particular event. Twitter's API provides a straightforward way to query for users and returns results in a JSON format which makes it easy to parse in a Python script, this makes our data reliable.

B. Pre-processing

Today, more than 80% of the data is unstructured – it is either present in data silos or scattered around the digital archives. In order to produce any meaningful actionable insight from data, it was important to know how to work with it in its unstructured form. One of the first steps in working with text data is to pre-process it. It is an essential step before the data is ready for analysis.

C. Cleaning & structuring

One approach is to directly remove useless html characters by the use of specific regular expressions. Another approach is to use appropriate packages and modules (for example htmlparser of Python), which can convert these entities to standard html tags. It is necessary to keep the complete data in standard encoding format. UTF-8 encoding is widely accepted and is recommended to use.

D. Removal of Stop-words

When data analysis needs to be data driven at the word level, the commonly occurring words (stop-words) should be removed. One can either create a long list of stop-words or one can use predefined language specific libraries.

E. Removal of Punctuations

All the punctuation marks according to the priorities should be dealt with. For example: ".", ",","?" are important punctuations that should be retained while others need to be removed.

F. Removal of Expressions

Textual data (usually speech transcripts) may contain human expressions like [laughing], [Crying], [Audience paused]. These expressions are usually non relevant to content of the speech and hence need to be removed. Simple regular expression can be useful in this case.

G. Split Attached Words

We humans in the social forums generate text data, which is completely informal in nature. Most of the tweets are accompanied with multiple attached words like Rainy Day, Playing In The Cold etc. These entities can be split into their normal forms using simple rules and regex.

H. Slangs lookup

Social media comprises of a majority of slang words. These words should be transformed into standard words to make free text. The words like luv will be converted to love, Heloto Hello. These words are extremely useful for our system and hence they will be collected for further analysis.

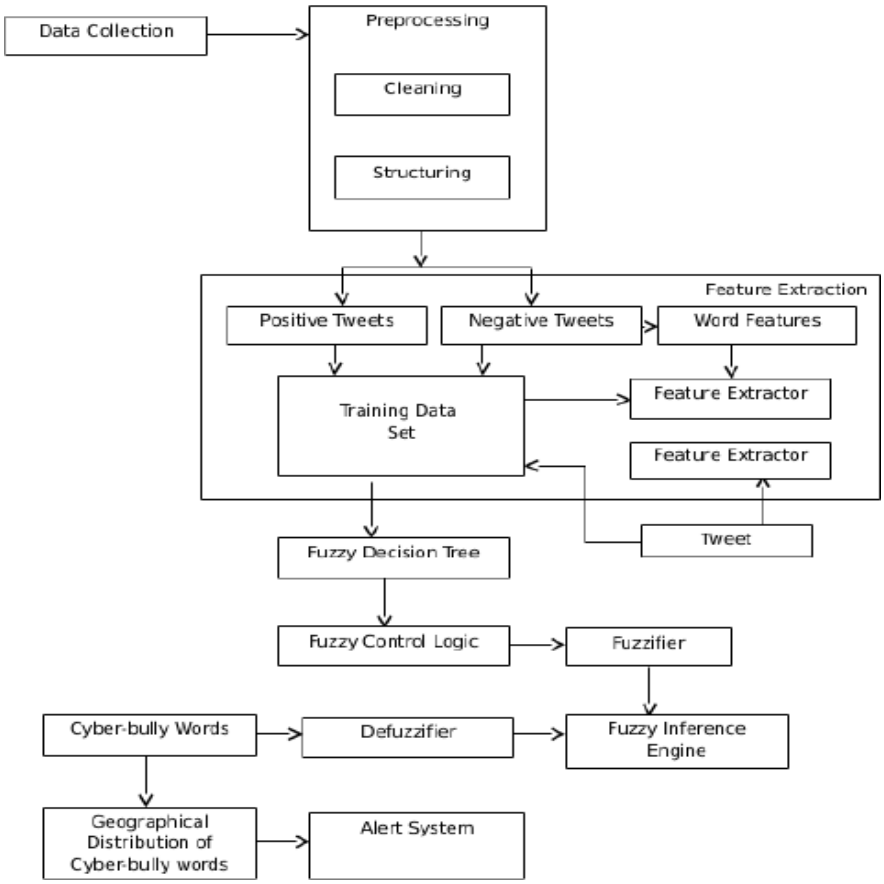

Fig.1Cyberbullying System Architecture using fuzzy logic

I. Feature Extraction

The feature extraction techniques are used to obtain the important words in the text. After pre-processing, each word from the document was represented by a vector of features. Using the training data set we are designing our system.

J. Fuzzy decision tree

Fuzzy logic for solving the introduced uncertainties were called Fuzzy Decision Trees.(FDTs)In this a method, selecting a word split at every new node of the tree and a rule for determining when a node should be considered terminal. The of frequency calculation words were given input to this method. It will find a probability of all the words, based on probability values the words were assigned into three ranges like Low, Medium, and High. Finally results were obtained and it was concluded whether the particular forums contains cyberbully words or not and assigning ranges were categorized as Low or Medium or High.

K. Fuzzy control logic

Fuzzy control logic was forming a knowledge-based consisting of IF-THEN fuzzy rules. These rules were obtained from human experts based on their respective domain of knowledge and observations. Fuzzy control logic contains 3 steps like Evaluation, Fuzzification and Defuzzification. Fuzzification transforms the real-valued input variables into a fuzzy set and using membership functions graphically to describe a situation. Rule Evaluation would evaluate the fuzzy rules. Defuzzification transforms back these fuzzy sets into real-valued variable outputs. The output of Feature extraction and Fuzzy decision tree system were given input to this method and assign fuzzy rules. Based on the fuzzy rules, the output of cyberbully words were categorized as Low or Medium or High.

L. Fuzzifier

Fuzzifer can be defined as crisp inputs which were translated into linguistic values by using a membership function. In this system, three word features were extracted where each word was associated with the vector of three features. These three features were operated as input to the fuzzifier separately and to manage the vector set, each feature is categorized into three sets like Low, Medium and High.

M. Fuzzy Inference Engine

After fuzzification, the inference engine was activated .Fuzzy rules were implemented to this process in order to assign linguistic values to all the three features. The inference engine was linked with all the category values and they were later converted into a single group that was eventually categorized as Low, Medium and High. The vital role of the Fuzzy inference engine was to assign the fuzzy rules. The important sentences were extracted from these rules according to the criteria of the features. In fuzzy rules, the set of fuzzy propositions associated with the if-condition rule was known as the premise or the antecedent. In the rule, if x is Low and y is High then z is Small .The premise consists of the two fuzzy propositions x is Low and y id High connected by the and (&&)operator.

N. Defuzzifier

The output of the linguistic variables from the inference engine will be converted into the final crisp values by the defuzzifier using membership function for representing the final sentence score. In this step, Defuzzification utilizes the output membership function which can be classified as output (Low, Medium, and High) that converts the fuzzy results from the inference engine into a crisp output to derive the final evaluation of each sentence. The suitable words were represented as High and it must be considered as the principal cyberbully words.

O. Cyberbully words

Slang term used to describe online harassment, which can be in the form of flames,comments made in chat rooms. We classify the words we get, as positive and negative then study their frequency. We discard the neutral words.

P. Geographical distribution of cyberbully words

The maps library for R language is a powerful tool for creating maps of countries and regions of the world. The coordinate system of the graph is latitude and longitude, so it's easy to overlay other spatial data on this map. We can plot all the tweets on the map and study their density.

Q. Alert system

The alert system includes all the tweets on the graph of India. The positive responses will be depicted in blue & negative in red. Based on the density of their occurrence we will warn the user if there is a chance of a tragedy. We can also depict statistical analysis with the help of bar graphs, etc.

V. CONCLUSION

Fuzzy logic approach had been known as powerful tool for risk assessment due to the fact that most approaches from classical statistics assume that they deal with exact measurements. But in most, if not all real scenarios, there is no precise measurement. Based on that fact, in this paper new method for assessing cyber security risk is introduced.

REFERENCES

[1] Sonia, A. Singhal, H. Banati, ―Fuzzy Logic Approach for Threat Prioritization in Agile Security Framework using DREAD model,‖. In IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No. 1, July 2011, Mauritius.

[2] Sodiya, A.S., Longe, H.O.D. and Fasan, O.M., ―Software Security Risk Analysis using Fuzzy Expert System,‖ In Journal of INFOCOMP: Journal of Computer Science, Brazil, Vol. 7, No. 3, 70—77, 2007.

[3] Ming-Chang Lee, ―Information Security Risk Analysis Methods and Research Trends: AHP and Fuzzy Comprehensive Method‖, International Journal of Computer Science & Information Technology (IJCSIT) Vol 6, No1, pages 29-45, February 2014

[4] National Institute of Standards and Technology NIST (Feb. 2012), Framework for Improving Critical Infrastructure Cyber security, Version 1.0