Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 45 (2015) 485 – 492

Procedia Computer Science

International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)

Online Social Network Bullying Detection Using Intelligence Techniques

B.Sri Nandhini[a], J.I.Sheeba[b]

[a]*Department of Computer Science and Engineering, Pondicherry Engineering College, Pondicherry-605014*
[b]*Department of Computer Science and Engineering, Pondicherry Engineering College, Pondicherry-605014*

Abstract

Social networking sites (SNS) is being rapidly increased in recent years, which provides platform to connect people all over the world and share their interests. However, Social Networking Sites is providing opportunities for cyberbullying activities. Cyberbullying is harassing or insulting a person by sending messages of hurting or threatening nature using electronic communication. Cyberbullying poses significant threat to physical and mental health of the victims.

Detection of cyberbullying and the provision of subsequent preventive measures are the main courses of action to combat cyberbullying. The proposed method is an effective method to detect cyberbullying activities on social media. The detection method can identify the presence of cyberbullying terms and classify cyberbullying activities in social network such as Flaming, Harassment, Racism and Terrorism, using Fuzzy logic and Genetic algorithm. The effectiveness of the system is increased using Fuzzy rule set to retrieve relevant data for classification from the input. In the proposed method Genetic algorithm is also used, for optimizing the parameters and to obtain precise output.

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).

Keywords: Cyberbullying, Social Network, Fuzzy logic, Genetic Algorithm.

1. Introduction

With the proliferation of the Internet, cyber security is becoming an important concern. While Web 2.0 provides easy, interactive, anytime and anywhere access to the online communities, it also provides an avenue for cybercrimes like cyberbullying. Life annoying cyberbullying experiences among young people have been reported internationally, thus drawing attention to its negative impact. In the USA, traces of cyberbullying is highly increasing and it has officially been identified as a social threat. There is an urgent need to study cyberbullying in terms of its detection, prevention and mitigation. Traditional bullying is any activity by a person or a group aimed at a target group

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).
doi:10.1016/j.procs.2015.03.085

or individual involving repeated emotional, physical or verbal abuse. Bullying as a form of social turmoil has occurred in various forms over the years with the WWW and communication technologies being used to support deliberate, repeated and hostile behaviour by an individual or group, in order to harm others[1]. Cyberbullying is defined as an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time, against a victim who cannot easily defend him or herself[2].

Recent research has shown that most teenagers experience cyberbullying during their online activities including mobile phone usage[3], and also while involved in online gaming or social networking sites. As highlighted by the National Crime Prevention Council, approximately 50% of the youth in America are victimized by cyberbullying[4]. The implications of cyberbullying[5] become serious (suicidal attempts) when the victims fail to cope with emotional strain from abusive, threatening, humiliating and aggressive messages. The impact of cyberbullying is exasperated by the fact that children are reluctant to share their predicament with adults, driven by the fear of losing their mobile phone and/or Internet access privileges[6]. The challenges in fighting cyberbullying include: detecting online bullying when it occurs; reporting it to law enforcement agencies, Internet service providers and others and identifying predators and their victims.

Fuzzy rule-based system[7] is a mathematical tool for dealing with the uncertainty and the imprecision. The reasoning is based on compositional rule of fuzzy inference and the knowledge of specialists is important to determine the parameters.

The Genetic Algorithms[7] are direct, stochastic method for optimization. Since this algorithm uses populations with given results. Proportional to the fitness function selection the probability (P) of each individual to be selected is calculated as the proportion of its fitness function to the sum of the fitness functions of all individuals in the current generation. For the choice of parents Tournament selection is used. Parents are selected by Tournament and select the best parent among the population.

The growth of cyberbullying activities is increasing as equally as the growth of social networks. Cyberbullying activities poses a significant threat to mental and physical health of the victims. Study about effects of bullying is present but implementation for monitoring social network to detect cyberbullying activities is less. Hence, the proposed system focuses on detecting the presence of cyberbullying activity in social networks using fuzzy logic which helps government to take action before many users becoming a victim of cyberbullying. The system also uses genetic operators like crossover and mutation for optimizing the parameters and obtain precise type of cyberbullying activity.

2. Related Work

In a recent study on cyberbullying detection[8], gender specific features were used and users are categorized into male and female groups. It is limited only to gender feature. In other study[9], NUM and NORM features were devised by assigning a severity level to the bad words list (nosewaring.com). NUM is a count and NORM is a normalization of the bad words respectively. The dataset consisted of 3,915 posted messages crawled from the Web Site, Formspring.me.It showed only 58.5% accuracy, which is very less accuracy.

Proposed a system allowing OSN users to have a direct control on the messages posted on their walls[4]. This is done by using flexible rule-based system, this system allows users to customize the filtering criteria to be applied to their walls, and a Machine Learning based classifier will automatically label messages using content-based filtering. This approach is incapable of capturing more complex relationships at a deeper semantic level.

In[9] a research work by Massachusetts Institute of Technology a system to detect cyberbullying through textual context in YouTube video comments is developed. The system classifies the comment in a range of sensitive topics such as sexuality, culture, intelligence, and physical attributes and determining what topic it is. The system shows less precise classification outcome and increased false positives.

In[12] using a bag-of-words approach examined a baseline text mining system and improved by including sentiment and contextual features. Even with those models, a vector machine learner produce a recall level of 61.9%.

In[11] bullying traces is identified using a variety of natural language processing techniques. Online and offline instances of bullying are traced. To identify the bullying they use sentiment analysis system and Latent Dirichlet Analysis to identify topics. In this method, the instances of bullying is not accurately detected.

Other interesting works[12] in this area performed harassment detection from comments and chat datasets provided by a content analysis workshop (CAW). Various features were generated including: TFIDF as local features; sentiment feature, which includes second person and all other pronouns like 'you', 'yourself', 'him', 'himself' and foul words; and contextual features. Increased false positive is its limitation. Research on online sexual predator's detection[12] associate the theory of communication and text-mining methods to analyse difference between predator and victim conversations, as applied to one-to-one communication such as in a chat-log dataset. The as usual methods are based on the keywords. It involves high semantic and contextual work.

Generally most existing systems are focusing on effects after cyberbullying incident and there is no system for online cyberbullying detection. Intelligence techniques are also not used in cyberbully detection. The proposed system is to detect the cyberbullying activities and classify them as Flaming, Harassment, Racism and Terrorism, which helps to prevent the cyberbullying victims from facing effects of cyberbullying and take necessary actions like blocking, law enforcement or taking corresponding legal actions accordingly.

3. Proposed Architecture

In the proposed architecture the process of detecting cyberbully activities begins with input dataset from social network. Input is text conversation collected from formspring.me. Input is given to data pre-processing which is applied to improve the quality of the research data and subsequent analytical steps, this includes removing stop words, extra characters and hyperlinks. After performing pre-processing on the input data, it is given to Feature Extraction. Feature Extraction is done to obtain features like Noun, Adjective and Pronoun from the text and statistics on occurrence of word (frequency) in the text. The extracted features are given to Learning Algorithm. The Learning algorithm unit is the central element of the architecture and is composed of a genetic algorithm for modeling adaptive and exploratory behaviour. Knowledge is given as Fuzzy rule set. The main functionality is to adjust the representation of the information needed for classification and yet retains the essential knowledge from the past. This knowledge is kept in a population of chromosomes, which is processed by the genetic algorithm. All the chromosomes in the population are competing to predict the classification of cyberbully activities. The output from learning unit is given to Classifier technique classifies the cyberbully activities using the fitness value of chromosome. The ability of a chromosome to classify the activity is called the fitness of the chromosome. The chromosome with higher fitness value gives the classified output. The output is classified bullying words present in the conversation.

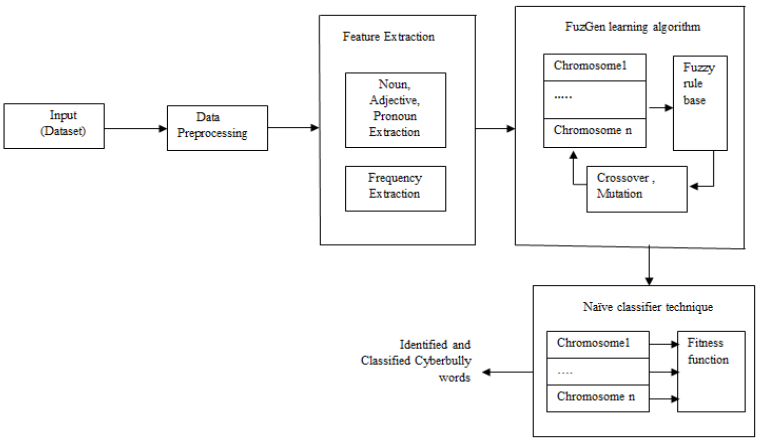

Fig. 1. Proposed cyberbully identification and classification system

In the proposed framework for detecting cyberbully activities, following steps have been included

- Data Pre-processing
- Feature Extraction
- FuzGen learning algorithm
- Naïve classifier technique

3.1. Data Pre-Processing

The data pre-processing (Kontostathis 2009) is an important phase in representing data in feature space to the classifiers. Social network data are noisy, thus pre-processing has been applied to improve the quality of the research data and subsequent analytical steps, and this includes removing stop words, unwanted characters, etc.

3.2. Feature Extraction

This module is used for extracting the data required from the processed data. The part of speech for every word in the conversation (Kontostathis, 2009) is obtained using natural language processing technique and then features like Noun, Adjective and Pronoun are extracted from the tagged output and statistics on occurrence of word in the text are also extracted.

3.3. FuzGen Learning Technique

The learning module incorporates the adaptive component of the system by means of a GA with fuzzy set genes. GAs are adaptive search and optimization algorithms that work by mimicking the principles of natural genetics (Deb, 1996). In the proposed system, the function to be optimized is a hypothetical representation of cyberbully terms in the Social Network.

In the following, the elements of the GA model, namely: the fuzzy gene types and the GA operators are presented.

3.3.1. The Fuzzy Set Genes

A gene G is, G = (t, g, and c), where
t is frequency of the term,
g identifies the gene type and
c is a non-negative real number .

When *G(t=c),* gene type represents the occurrences of a cyberbully term.

When *G(t<c).* This gene type is completely satisfied by dataset that have no occurrences of the cyberbully term *t*.

When *G(t≥c).* Genes of this type are satisfied completely by dataset with at least *c* occurrences of the cyberbully term *t*.

3.3.2. The GA Operators

Selection, crossover, and mutation are the genetic operators of evolutionary process. Choice of chromosomes from population to reproduce is done by selection. Using crossover an offspring chromosome is produced by taking sequences of genes from each of two parent chromosomes selected and combining them. The mutation is the random alteration of a gene in the chromosome selected.

3.4. Classifier Technique

It assigns a fitness value to the terms by using the present information in the population of chromosomes. The chromosome with higher fitness value will be obtained as classified output. Fitness function is the difference in the term t identified as a type cyberbully term and the actual type of the
Term t.
Fitness=$\sum^{?}$ (? ??? ? ??? ? ?(?) − ? ? ?(?))

Where, F is the fitness function,

P is the number of terms in training dataset

T is a term in the dataset.

4. Proposed Algorithm

The proposed algorithm takes the processed social network conversation dataset as input from the pre-processing and feature extraction unit. It performs evolutionary process using Genetic algorithm and the evaluation of the chromosome is done using Fuzzy rule set. The fitness value of the chromosome is calculated for every individual in the population and the cyberbully terms in the input dataset will be given as output.

Input: Conversation dataset from Social Network.

//Initializing initial population and evaluating Fitness value

Step1: Current population is assigned to the initial population.

Step2: Evaluate the current population with the fuzzy rule set given as knowledge base.

Step3: The fitness value of the current population is calculated using the function EvalPop ().

Step4: The current population is considered as best population since it is the initial population.

Step5: The fitness value of the current population is assigned as the best fitness value.

Step6: The size of the term set retrieved from input is assigned as null.

// For Parent selection

Step7: The size of the current term set is compared with the size of evolved term set, Ne, if the size of N is less than Ne, then the following steps takes place.

Step8: The offspring population is initialized as null

Step9: If the size of offspring population is less than current population then following steps will be executed

Step10: Parents are selected by using the tournament selection mechanism and children are created by using mutation and cross over mechanism, where Tournament selection is a method of selecting an individual from a population of individuals in a genetic algorithm.

Step11: Once the offspring population is created, it is joined to current population.

Step12: End of while loop.

Step13: Evaluate the Fuzzy rule set for the offspring population.

Step14: Once the offspring population is created, it is joined to current population.

Step15: Token competition is carried out to obtain the best individuals from the joint population.

Step16: The Joint population is assigned to the current population.

Step17: The Fitness value of the joint population is calculated using the function EvalCurPop ().

// Updating the Best fitness value and Best population for obtaining classified output.

Step18: Fitness values of the current population is checked with the best fitness value. If the current fitness value is greater than following steps occur

Step19: The best fitness value is updated with the current fitness value.

Step20: The best population is updated with the current population.

Step21: End of if loop.

Step22: The size of the current term set is incremented.

Step23: End of while loop.

Output: Identified Cyberbully terms and their type from the input dataset.

In the proposed algorithm, step 1 to step 6 is for initializing the parameters and evaluating. Consider the current population as the initial population. Evaluate the current population with the fuzzy rule set given as knowledge base. After evaluating population with rule set, the fitness value of the current population is calculated using the function EvalPop ().The current population is considered as best population since it is the initial population. The fitness value of the current population is assigned as the best fitness value. The size of the term set retrieved from input, n is assigned as null. Step 7 to step 11 are followed for parent selection. The size of the current term set is compared with the size of evolved term set, Ne, if the size of N is less than Ne, then the following steps takes place. The offspring population is initialized as null. If the size of offspring population is less than current population then

following steps will be executed. Parents are selected by using the tournament selection mechanism and children are created by using mutation and cross over mechanism, where Tournament selection is a method of selecting a best individual from a population of individuals in a genetic algorithm. Tournament selection involves running several "tournaments" among a few individuals chosen at random from the population. Once the offspring population is created, it is joined to current population. Evaluate the Fuzzy rule set for the offspring population. Once the offspring population is created, it is joined to current population. Token competition is carried out to obtain the best individuals from the joint population. The Joint population is assigned to the current population. The Fitness value of the joint population is calculated using the function EvalCurPop (). Step 18 to step 23 is for updating the best population. Fitness values of the current population is checked with the best fitness value. If the current fitness value is greater than following steps occur, the best fitness value is updated with the current fitness value. The best population is updated with the current population.

Experimental Results

A. Dataset

For this work, we considered the datasets described below for the experiment on cyberbullying detection, which are available from the workshop on Content Analysis for the Web 2.0 [10]. The dataset contains data collected from two different social networks: Formspring.me and MySpace. Formspring.me, is a discussion-based site, users broadcast their message. MySpace is a popular social networking website. Datasets were provided in the form of text document for each conversation set between two users.

B. Evaluation parameters

Precision: The total number of correctly identified true bullying posts out of retrieved bullying posts.

Recall: Number of correctly identified bullying cases from total number of true bullying cases.

F-1 measure is the harmonic mean of precision and recall.

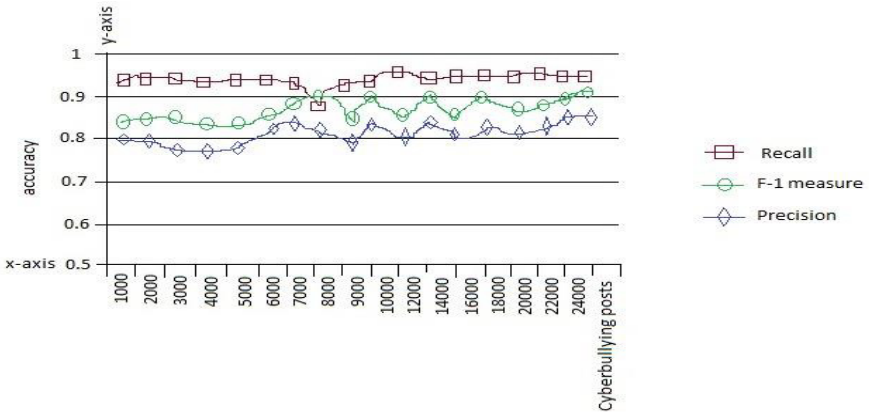

Fig. 2. Precision, Recall and F-1 measure on combined dataset

Table 1. Comparison of proposed approach with fuzzy classification rule

	Accuracy		F – Measure		Recall	
Dataset	Fuzzy classification rule	Proposed rule	Fuzzy classification rule	Proposed rule	Fuzzy classification rule	Proposed rule
Myspace	.35	.87	.44	.91	.60	.98
Formspring.me	.42	.86	.31	.92	.58	.87

Conclusion

The proposed system focuses on detecting the presence of cyberbullying activity in social networks using fuzzy logic which helps government to take action before many users becoming a victim of cyberbullying. The system also uses genetic operators like crossover and mutation for optimizing the parameters and obtain precise type of cyberbullying activity which helps government or other social welfare organization to identify the cyberbullying activities in social network and to classify it as Flaming, Harassment, Racism or Terrorism and take necessary actions to prevent the users of the social network from becoming victims.

References

[1] B. Belsey. (6th June 2013). Cyberbullying.org. Available: http://www.cyberbullying.org/
[2] P. K. Smith, J. Mahdavi, M. Carvalho, S. Fisher, S. Russell, and N. Tippett, "Cyberbullying: Its nature and impact in secondary school pupils," Journal of Child Psychology & Psychiatry, vol. 49, pp. 376-385, 2012.
[3] M. A. Campbell, "Cyber bullying: An old problem in a new guise?" Australian Journal of Guidance and Counselling, vol. 15, pp. 68-76, 2011.
[4] Sara Bastiaensens, Heidi Vandebosch , KarolienPoels, Katrien Van Cleemput , Ann DeSmet, Ilse De Bourdeaudhuij," Cyberbullying on social network sites. An experimental study into bystanders' behavioural intentions to help the victim or reinforce the bully," Elsevier, Computers in Human Behaviour 31 (2014) 259–271.
[5] Christina F. Brown, Michelle Kilpatrick Demaray, Stephanie M. Secord, "Cyber victimization in middle school and relations to social emotional outcomes", Elsevier, Computers in Human Behaviour 35 (2014) 12–21.
[6] Jin-Liang Wanga, Linda A. Jackson , James Gaskin , Hai-Zhen Wang, "The effects of Social Networking Site (SNS) use on college student's friendship and well-being", Elsevier, Computers in Human Behaviour 37 (2014) 229–236.
[7] Victoria Lopez, Alberto Fernandez, Maria José del Jesus , Francisco Herrera, "A hierarchical genetic fuzzy system based on genetic programming for addressing classification with highly imbalanced and borderline data-sets", Elsevier, Knowledge-Based Systems 38 (2013) 85–104.
[8] M. Dadvar, F. d. Jong, R. Ordelman, and D. Trieschnigg, "Improved cyberbullying detection using gender information," In Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012), pp. 23-25, February 2012.
[9] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the Detection of Textual Cyberbullying," in Proc. IEEE International Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 2011.
[10] K. Reynolds, A. Kontostathis, and L. Edwards, "Using Machine Learning to Detect Cyberbullying," In Proceedings of the 2011 10thConference on Machine Learning and Applications Workshops, vol. 2, pp. 241-244, December 2011.
[11] I. McGhee, J. Bayzick, A. Kontostathis, L. Edwards, A. Mcbride, and E. Jakubowski, "Learning to Identify Internet Sexual Predation, "International Journal on Electronic Commerce 2011, vol.15, pp. 103-122, 2011.
[12] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, "Detection of Harassment on Web 2.0," in Proc. Content Analysis of Web 2.0 Workshop, Madrid, Spain, 2009.
[13] A. Kontostathis, L. Edwards, and A. Leatherman, "Chat Coder: Toward the Tracking and Categorization of Internet Predators," In Proceedings of Text Mining Workshop 2009 held in conjunction with the Ninth SIAM International Conference on Data Mining, 2009.
[14] Bsecure. Available: http://www.safesearchkids.com/BSecure.html
[15] Cyber Patrol. Available: http://www.cyberpatrol.com/cpparentalcontrols.asp
[16] eBlaster. Available: http://www.eblaster.com/
[17] IamBigBrother. Available: http://www.iambigbrother.com/
[18] Kid swatch. Available: http://www.kidswatch.com/
[19] Xu, Jun-Ming; Kwang-Sung Jun; Xiaojin Zhu; and Amy Bellmore. Learning from bullying traces in social media. In Proceedings of the 2012 Conference of North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Montreal, Canada, 2012, pp.656-666.
[20] Patchin, J. and S. Hinduja. (2006). "Bullies move beyond the schoolyard; a preliminary look at cyberbullying," Youth violence and juvenile justice. 4:2, pp.148-16