A survey on Datamining in Cyber Bullying

K. Nalini
Research Scholar
Bharathiyar University
Coimbatore
immanuelsamen@rediffmail.com

Dr. L. Jaba Sheela
Professor, MCA Department
Panimalar Engineering College
Chennai
sujitha14@hotmail.com

Abstract—CRIME has always the very first thing that people want to avoid and police officer want to stop. A major challenge facing all law-enforcement and intelligence –gathering organizations is accurately and efficiently analyzing the growing volumes of crime data. Crime activity reports available from victims, governmental organizations, news press, and social networks play a significant role in public safety, including crime prevention, suppression and investigation, uniformed patrol and response. Cyber Crime is an area which covers crimes committed in the internet. Opportunities for connecting with classmates, friends and people with shared interests abound. Email, online chat, and social networking sites allow us to interact with people in the same town and people on the other side of the world. Unfortunately the opportunity for misuse comes with any new technology. There were sexual predators and bullies long before the advent of the internet and chat rooms. Cyber bullying and internet predation threaten minors, particular teens and teens who do not have adequate supervision when they use the computer. As the amount of criminal records growing every day, it is impossible to perform manual analysis on the dataset and extract useful information. Data mining has been studying for decades trying to get useful information out of large amount of data. Many efforts have used automated techniques to analyze different types of crimes, but without a unifying framework describing how to apply them. In particular, understanding the relationship between analysis capability and crime type characteristics can help investigators more effectively use those techniques to identify trends and patterns, address problem areas and even predict crimes.

Keywords—*Data Mining, Cyber Bullying, Sexual Predation, Machine Learning, Datasets*

*****

I. Introduction

"Cyber bullying is defined as an aggressive, intentional act carried out by a group or individual using electronic forms of contact (eg. Email and Chat rooms), repeatedly or overtime, against a victim who cannot easily defined herself". Cyber bullying consist in sending messages containing slanderous expressions, harmful for other people or verbally bullying other people in front of the rest of online community. According to recent studies almost 43%
of teens in the United States alone reported being victims of cyber bullying. In 2011, 70% of teens use social media sites on daily basis and nearly one in four teens hit their favorite Social-media sites 10 (or) more times a day. Scan safe's monthly "Global Threat Report" found that up to 80% of blogs contained offensive contents and 74% included porn in the format of image, video (or) offensive languages.

In particular the story of 13-year-old Megan Meir brought notoriety to the subject of Cyber bullying when she committed suicide after being harassed though a popular social networking site. (ABC NEWS, whereas 10% of the cases took place for less than 10 years. A 14-year-old Estonian teenager S.K[18] , who committed suicide in 2009, after being recurrently harassed by a Paedophile
through the internet. The Paedophile pretended to be a teenager girl in order to gain access to dozens victims. He could therefore internet with many children in a seemingly natural way. Sadly S.K could not bear the constant coercion from the Paedophile and this soon led to his suicide. The "MySpace Mom" case[19] is another tragic example:- L.D and Cyber bullies, pretended to be a teenager boy on MySpace and befriended a teenager girl (M.M). After several weeks exchanging messages abruptly ended their friendship, telling M.M that she was cruel. Some days later M.M committed suicide. Many other that occur on a yearly basis across the world, are highly indicative of how severe Cyber threats are.

Online security has been an important and urgent problem ever since the creation of the Internet. One of the burning online security problems are online slandering , bullying and particularly when the target subjects are under-age victims. Sexual predators have adapted their predatory strategies to these platforms and usually the target victims are kids. The number of children who are approached (or) solicited for sexual purpose through the Internet is staggering [1] and unfortunately online sexual predators always outnumber the law enforcement officers available in Police Cyber Crime Units [2].

Most adverse effects of cyberbullying are seen in adolescents, though this menace exists among all age groups. According to a report by Microsoft, India, ranks third in the world when it comes to cyberbullying of children and first in cyberbullying cases of adults. It includes all those activities that are meant to humiliate, disturb, defame, threaten or insult an individual. Photoshopping target's face over obscene pictures, posting and spreading defaming rumors and blackmailing the victim over his/her objectionable videos are common examples of cyber crime. Like other cases of bullying, children may not inform their parents of the cyber humiliation or threat they suffer out of fear of social stigma. Results are depression, anxiety, loss of self esteem, fear and isolation.

Cases of cyberbullying have been reported all over the world and India is no exception. Twenty one percent of the suicides among Indian adolescents are due to the trauma of cyber bullying suffered by the victims, as shown by studies. This is both due to easy availability of internet as

well as lack of rules against cyber defamation. A recent case that horrified me to the core was the one in which an Indian boy committed suicide because one of his friends posted his video online which had him making out with his cousin sister. Amanda's case gained much concern, as that tenth grade girl suicide after posting her own video recounting the cyber blackmail she was suffering.

Akash Ambani, the elder son of Mukesh and Nita Ambani, was one of the top trending tags on the micro-blogging site Twitter post his brother Anant's appearance at an IPL match. He became the subject of snide remarks targeting his weight. The most re-tweeted post was by @SirJadeja16h which read: "Red alert. = Expected earthquake in Kolkata later tonight coz Akash Ambani will be doing jhamping jhapang after MI win #IPLFinal".

What many people do not know about the younger Ambani scion is that the reason behind his obesity is the usage of steroids to treat his asthma. These tweets not only openly slander the victim they also breach the fine line between freedom of expression and potential cyber crime.

What makes cyber bullying so widespread in India is the fact that unlike many developed countries, India does not presently have laws to curb it. There are no serious punishments for cyber offenders. The recent most cyber crime laws which were made by the parliament in Feb 2013 include only financial matters like cases of fraud and phishing scams. Indian government must realize that India being a prominent IT hub, with a large section of society having access to internet, cyber laws are need of the hour.

In Mumbai, Two cases of cyber bullying, where profiles of two young women on social networking sites were replaced with obscene material, ended in acquittal after the prosecution failed to produce evidence. In both the cases, the court said the prosecution failed to submit electronic evidence. In one of the cases, the investigating officer was not available for examination in the court.

In the first case, in September 2006, a Thane resident prepared a fake profile of a college student, posted obscene comments about her and also provided her telephone number. Following this, the student started receiving vulgar messages and calls. On her complaint, the police registered a case, identified the cyber cafe from where the accused had posted the obscene material and arrested him. The accused had used his personal computer for the activity.

II. Literature Survey

Patchin and Hinduja[3] define Cyber bullying as "Willful and repeated harm inflicted through the medium of electronic text". In their recent studies found that students who experienced Cyber bullying (Both those who were victims and those who admitted to Cyber bullying others) perceived a poorer climate at their school than those who had not experienced Cyber bullying.

In a recent study on Cyber bullying detection Yin, et al [4] used a supervised learning approach for detecting harassment. They determined that the base line text mining system (using a bag of words approach) was significantly used content, sentiment and contextual features of documents to train a Support Vector Machine Classifier for Corpus of Online posts and only the contents of the posts were used to determine either a post is harassing (or) not and the characteristics of the author of the posts were not considered. They have used the combination of 3 features such as N-gram, TFIDF weighting and foul words frequency were used as the baselines. The results shows improvements over the baselines. Both C4.5 decision tree learner and an Instance-based learner were used to identify the true positive with 78% accuracy, by recording the percentage of curse and insult words within a post.

Dinakar et al.,[5] applied a range of binary and multiclass classifiers on a manually labeled Corpus of You tube Comments. Their findings showed that binary individual topic-sensitive classifiers can outperform the detection of textual Cyber Bullying compared to merge data sets or multiclass classifiers. They have illustrated the application of commonsense knowledge in the design of social network software for detection Cyber Bullying. The authors treated each comment on its own and did not consider other aspects to the problem as such the pragmatics of dialogue, conversation and the social networking graph. They concluded that taking into account such features will be more useful on social networking websites and can lead to a better modeling of the problem.

April Kontostathis et al.,[6] used a Language-Based method of detecting Cyber Bullying. They collected the data from the website Formspring.me, a question and answer formatted website that contains a high percentage of bullying content. The data was labeled using a web service, Amazon's mechanical turk. They used the labeled data in conjunction with machine learning techniques provided by the Weka Tool Kit, to train the computer to recognize bullying content. Both C4.5 decision tree learner and an Instance-based Learner were used to identify the true positive with 78.5% accuracy, by recording the percentage of curse and insult words within a post.

Maral Dadvar et al., [7] have investigated the Gender-Based Approach for Cyber Bullying detection in Myspace. They have used the content of the text written by the users but not the user's information. They approached Support Vector Machine model to train a gender-specific text classifier using WEKA. They have utilized the Myspace posts as dataset which was provided by Fundacion Barcelona Media. The dataset consists of more than 3,81,000 posts in about 16,000 threats. Overall 34%of posts are written by female and 64% by male authors. The Gender Specific Approach improved the Baseline by 39% in precision, 6% in recall, 15% in F-measure.

Ying Chen et al.,[8] investigated existing text mining methods in detecting offensive contents for protecting adolescent online safety. Specifically, they proposed the Lexical Syntactical Feature(LSF) approach to identify offensive contents in social media and further predict a user's potentiality to send out offensive contents. Their research has several contributions. First they practically conceptualize the notion of online offensive contents and further distinguish the contribution of pejoratives/profanities and obscenities in determining offensive contents, and introduce hand authoring syntactic rules in identifying name-calling harassment. Second, they improved traditional

Machine-Learning methods by not only using lexical features to detect offensive language, but also incorporating style feature, structure features and Content-specific features to better predict a user's potentiality to send out offensive content in social media. Experiment result shows that the LSF Sentence offensiveness prediction and user offensiveness estimate algorithm outperform, traditional learning-based approaches in turns of precision, recall and F-score. The LSF tolerates informal and misspelling contents and it can easily adapt to any formats of English writing styles.

Chou et al.,[9] applied two term weighting method to detect internet abuse in the workplace of software programmers. They have used six classification methods such as Naïve Bayes, Multinomial Naïve Bayes, Back propagation neural network, K-nearest neighbor, C4.5 decision tree, Support Vector Machine in Online news websites such as New York Times online with several sections such as general news, sports, entertainment, business and technology. They approached text categorization to detect internet abuse in the workplace.

Androutsopouloes et al.,[10] used Naïve Bayesian Classification, memory-based classification and total cost ratio that allows that performance of a filter to be compared easily to that of the Baseline inorder to filter unsolicited bulk email. They applied both the methods to achieve very high classification accuracy and clearly outperformed the anti-spam keyword patterns of a widely used e-mail reader. Their findings suggest that it is entirely feasible to construct learning-based anti-spam filters when spam messages are simply to be flagged or when addition mechanism are available to inform the senders of block messages.

Javier Paraper et al.,[11] have presented automatic methods for detecting sexual predation in Chat rooms. They have successfully shown that a learning based method is a feasible way to approach this problem and have proposed innovative sets of features to derive the classification of chat participants as predators or non-predators. They demonstrated that the set of features utilized and the relative weighting of the misclassification costs in the SVMs are two main factors that should be taken into account to optimize performance. They carefully analyzed the relation importance of the classifier's features as a preliminary effort to understand the psycho-linguistic, contextual and behavioural characteristics of several predators in the internet. Their approach is promising for intelligence gathering and prioritizations of investigative resources to assist police Cyber Crime units in their hunt for sexual predators in the Internet.

In a recent study on cyberbullying detection [12], Electronic aggression, or cyber bullying, is a relatively new phenomenon. In a series of two studies, exploratory and confirmatory factor analyses (EFAs and CFAs respectively) were used to examine

whether electronic aggression can be measured using items similar to that used for measuring traditional bullying, and whether adolescents respond to questions about electronic aggression in the same way they do for tradition al bullying. EFA and CFA results revealed that adolescents did not differentiate between bullies, victims, and witnesses; rather, they made distinctions among the methods used for the aggressive In general, it appears that adolescents differentiated themselves as individuals who participated in specific mode of online aggression, rather than as individuals who played a particular role in online aggression.

A wide range of learning strategies have been adopted for sexual predation classification in the literature. Villatoro-Tello et al.[25] applied two-stage approach with an initial conversation-level classification that tries to filter out conversation with no sexual predation, and a subsequent predator-victim classification. The two-stage method designed in was highly effective but the main reason behind such high performance was a pre-processing step that removed 90% of the conversations : a) conversations that had only on participant were removed, b) conversations that had less than six interventions per-user were removed, and c) conversations that had long sequences of unrecognized characters (apparently images) were removed. Such heuristic pruning was favourable for a particular experimental setting but can most likely not be used with other datasets.

There are also some software products available for fighting against cyberbullying e.g., [13], [14],[15], [16], [17]. However, filters generally work with a simple key word search and are unable to understand the semantic meaning of the text. While some filters block the webpage containing the keyword, some shred the actual offensive words themselves. Other software products exhibit a blank page on detection of the keywords. However, removal of the offensive word from the sentence can totally distort the meaning and sense of the sentence. Moreover, Internet programmers can easily dodge filters. It can be argued that filters are not an effective anti-cyberbullying solution as there are many ways to express inappropriate, illegal and offensive information. Another limitation is that filtering methods have to be set up and manually.

III. Proposed work

Challenges in detecting Cyber bullying

Bullying is a social problem that is too large and old in nature, there are various peoples are involved directly or indirectly. This problem is unsolved in real life and it is just like a hard problem. In this age of technology there are too few places to spend time thus youngsters are involve in internet net based social networking web sites or different kinds of web based applications where they are able to search new friends, persons and may be able to share personal data over internet. Secondly the social networking websites provide the provision for their privacy and content management scheme, but most of the users are now much aware about these privacy policies. Required to improve the user interaction with these privacy policies by which the surf internet in secure environment. The problem exist in the real world system of bullying and their effects, we find that it is a problematic situation of internet surfing.

 ISSN: 2321-8169
1865 – 1869

For a number of issues related to cyber bullying detection, research has been done based on the text mining paradigm such as online sexual predator recognition [20] and spam detection [21]. Nevertheless, very little study has been done on technical solutions, for which is why there is insufficient proper training datasets. Moreover, privacy issues and ambiguities can be the reasons in describing cyber bullying.

Although bullying messages are posted everyday comparing to hundreds of thousands of messages posted every second, they are very sparse. Collecting enough training data is a big challenge, since random sampling will lead to few bully messages. One possibility is using hash tags(#bully etc) [22] or using a set of commonly used terms of abuse[23], however it leads to a very biased training dataset.

In order to face these types of challenges we need to design an effective framework that incorporates word-level features and user based features to detect and prevent offensive content IRC logs. We should also design the effective strategy to detect and evaluate the level of offensiveness of a user and word level offensiveness in a message and we need to check whether this proposed framework is efficient and effective enough to be deployed on real time. In this proposed solution we provide the primary way by which we identify the bullying, additionally using the text and data mining technique we analyse text content in the posts and provide the conclusion is there any kind of bullying exist or not.

System Architecture

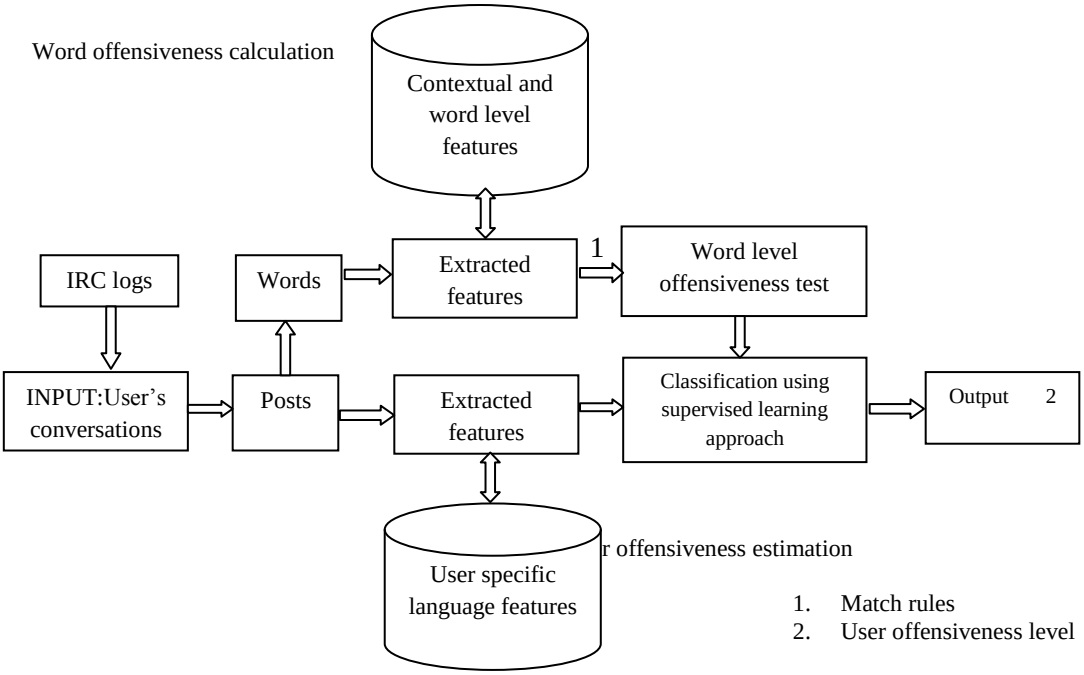

Figure 1. User offensiveness classification

We propose a contextual and word level features based framework to detect offensive content and identify offensive users in IRC logs. We would like to include 2 phases of offensiveness detection. Phase 1 aims to detect the offensiveness on the word level and phase 2 derives offensiveness of user level. In phase 1 we need to apply the natural language processing techniques such as word level features and contextual level features. In phase 2 we incorporate user-level features by using style ,structure and cyber bullying features. The framework is illustrated in fig.1.

IV. Datasets

Data collection is the first step in any research project in text mining. Data collection for the study of cybercrime needs to focus primarily on capturing data from and social networking sites; however, there are both legal and technical issues that must be overcome. There is very little reliable labeled data about predator communication; much of the work that has appeared in both computer science and communication studies forums is focused on anecdotal evidence and chat log transcripts from Perverted Justice [26]. Perverted-Justice .com began as a grass-roots effort to

identify cyber predators. PJ volunteers pose as youth in chat rooms and respond when approached by an adult seeking to begin a sexual relationship with minor. When these activities result in an arrest and conviction, the char log transcripts are posted online. New chat logs continue to be added to the web site. There were 325 transcripts, representing arrests and convictions, on the site as of July 2009. Using these datasets, we would like to compare the performance of different classification algorithms included in WEKA. 1. Random Forest 2. J48 (WEKA'S C4.5 implementation) 3 Sequential Minimal Optimization.

to be more proactive in addressing the role that newer technologies, particularly cell phones are peer-to peer devices, play in new incarnations of Cyber Crime.

Conclusion

Cyber Bullying is a growing problem in the social web and it is becoming major threat to teenagers and adolescents. In this paper we represented a survey on the current scenario of cyberbullying and various methods available for the detection and prevention of cyber harassment. Our concept depends upon the text analysis, the data which is uploaded or text written by any user is first analyzed and after that, we estimate the roles of user, is it a bully? or a victim? As more researchers enter this field of research should attempt

References

[1] Mcghee, Jennifer Bayzick, April Kontostathis, Lynne Edwards, Alexandra Mcbride, and Emma Jakubowski. Learning to identify internet sexual predation.
[2] *Int*. J. Electron. Commerce, 15(3):103–122, April 2011.
[3] Nick Pendar. Toward spotting the pedophile: Telling victim from predator in text chats. In Proc. First IEEE International Conference on Semantic Computing, pages 235–241, 2007.
[4] Patchin, J., & Hinduja, S. "Bullies move beyond the schoolyard; a preliminary look at cyberbullying." Youth violence and juvenile justice. 4:2 (2006). 148-169.
[5] D. Yin, B. D. Davison, Z. Xue, L. Hong, A. Kontostathis, and L. Edwards,"Detection of Harassment on Web 2.0," In Proceedings of the Content Analysis In The Web 2.0 (CAW2.0) Workshop at WWW2009, 2009.
[6] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the Detection of Textual Cyberbullying," International Conference on Weblog and Social Media - Social Mobile Web Workshop, Barcelona, Spain 2011, 2011.
[7] A. Kontostathis, L. Edwards, and A. Leatherman, "ChatCoder: 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 (SDM 2009).
[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] Ying Chen, Sencunzhu, Yiluzhou, Heng Xu, "Detecting Offensive Language in Social Media to protect Adolescent online safety", International conference on social computing, p71-80,sep-2012.
[10] C.H. Chou, Atish P. Sinha, Huimin Zhaor, "A Text mining Approach to Internet Abuse Detection" , Proc. Of the 5th Workshop on e-business (WeB), Milwarkee, WI 2006.
[11] Ion Androutsopoulos, John Koutsias, Konstantinos V. Chandrinos, and Constantine D. Spyropoulos. An experimental comparison of naive bayesian and keyword-based anti-spam filtering with personal e-mail messages. In Proceedings of the 23rd annual international
[12] ACM SIGIR conference on Research and development in information retrieval, SIGIR '00, pages 160–167, New York, NY, USA, 2000. ACM..
[13] Javier Parapar, David E. Losada, A´ lvaro Barreiro, Combining Psycho-linguistic, Content-based and Chat-based Features to Detect Predation in Chatrooms, Journal of Universal Computer Science, vol. 20, no. 2 (2014), 213-239.
[14] Daniell M. Law A , Jennifer D. Shapkaa, Shelley Hymel A, Brent F. Olson A, Terry Waterhouse B. The changing face of bullying: An empirical comparison between traditional and internet bullying and victimization, Computers In Human Behavior, Vol 28,Issue-1,Jan 2012, Pages 226-232.
[15] Bsecure. Available: http://www.safesearchkids.com/BSecure.html
[16] Cyber Patrol. Available: http://www.cyberpatrol.com/cpparentalcontrols.asp
[17] eBlaster. Available: http://www.eblaster.com/
[18] IamBigBrother. Available: http://www.iambigbrother.com/
[19] Kidswatch. Available: http://www.kidswatch.com/
[20] http://www.publico.es/espana/263683/retrato-de-una-cibervictim.
[21] http://www.foxnews.com/story/2007/11/16/mom-myspace-hoax-led-todaughter-suicide/
[22] A. Kontostathis, "ChatCoder: Toward the tracking and categorization ofinternet predators." In:
[23] Proceedings of SDM 2009, Sparks, NV, May 22009.[8]
[24] P.N. Tan,F. Chen, A. Jain, "Information assurance: Detection of webspam attacks in social media." Proceedings of Army Science Conference, Orland, Florida. 2010.
[25] J-M.Xu, K-S.Jun, X.Zhu and A.Bellmore, Cyber bullying detection in twitter, Vol-7, Issue-16, Pages 10-14, Reporter 2014.
[26] H. Sanchez and S. Kumar, Twitter bullying detection, Data mining Course Report, 2011.
[27] www.noswearing .com/dictionary.
[28] Esau villatoro-Tello, Antonio Juarez-Gonzalez, Hugo Jair Escalante, Manuel Montes-y Gomez, and Luis Villasenor Pineda. A Two-step Approach for Effective Detection of Misbehaving Users in Chats – Notebook for PAN at CLEF 2012.
[29] www.Perverted-justice.com