International Journal of Information and Computation Technology.
ISSN 0974-2239 Volume 3, Number 6 (2013), pp. 569-576

http://www. irphouse.com /ijict.htm

Improved Textual Cyberbullying Detection Using Data Mining

Paridhi Singhal and Ashish Bansal

Department of Information Technology, Shri Vaishnav Institute of Technology and Science, Baroli, Sanwer Road, Indore, India.

Abstract

The most common definition of bullying is based on Olweus's (1991, 1993) definition, which states that ''A person is being bullied when he or she is exposed, repeatedly and over time, to negative actions on the part of one or more other persons''. Internet and its applications have become more popular in this age. The new generation spends a lot of their time on social networking websites and shares their personal information with friends and sometimes makes them public. This information can be of advantage for the bullies to bull innocent people just for their fun, thus required an advance and appropriate data analysis system by which we can prevent users to get bulled. As electronic aggression is basically using internet for the intention of harming others. Mostly cyberbullying is done through social networks, text messaging, chat rooms, emails etc .and the common topics around which cyberbullying revolves are physical appearance, race and ethnicity ,sexuality and sexual identity ,intelligence , social acceptance and rejection etc. In this paper we are giving a survey on cyberbullying then according to that, we will design and implement a social networking web site by which we can simulate bullying, and prevent the users to get bulled. Moreover we will provide a case study and the performance study of our system.

Keywords: Bullying, Social networks, Cyber harassment, Text-mining.

1. Introduction

The word "bully" can be traced back as far as the 1530s (Harper, 2008). In its most basic sense bullying involves two people, a bully or intimidator and a victim. The bully abuses the victim through physical, verbal, or other means in order to gain a sense of superiority and power. These actions may be direct (i.e. hitting, verbally assaulting face-to-face, etc.) or indirect (i.e. rumors, gossip, etc.). When bullying behavior is carried out through the use of information and communication technologies such as e-mail, mobile phones, instant messaging (IM), social networking websites, apps, and other online technologies it becomes increasingly difficult to deal with and goes beyond the traditional boundaries of the school environment .According to a survey conducted by research firm Ipsos puts India on top when it comes to cyberbullying of adults. Also, according to a recent study by Microsoft Corporation, India ranks third overall, after China and Singapore, in cyberbullying of children. Cyberbullying includes motives and actions that seek to humiliate, threaten, control, insult or slander the victims. Most common cyberbullying cases include photoshopping the target's face on nude bodies, spreading false rumors through anonymous or public profiles about someone, posting defamatory messages about or to someone, or using filmed footage of potential victims in provocative situations to blackmail them.

Parents, being less enlightened about modern technology than young adults, are often in the dark about the horrific experiences of their children. In many cases the children themselves do not tell their parents either under threat from the bully or out of fear of social stigma. Since children often take the pressure alone on their young shoulders, a feeling of depression and self-isolation sets in. In extreme cases it leads one to commit suicide. The biggest problem regarding cyberbullying is that the age group of the offenders ranges from as young as eight to the legal adult age of eighteen and beyond. In schools, children are bullied on the basis of their physical features, race, age or even their level of knowledge. Even if no actual long-lasting harm may be intended, the victims are often left permanently c are not brought to justice and bullying in itself is left unchecked, there is a very high risk of the problem evolving into something more harmful and, maybe, uncontrollable.

India, being one of the prominent IT hubs in the world, ought to have good enough laws effective immediately that curb cyber crime and punish its offenders. With rapid globalization, availability of cheap mobile phones and laptops, easy to use technology, and inexpensive network charges, India is a potential sitting duck for all forms of cyber crimes including bullying. In the meantime, people must try to become more aware of cyberbullying.

2. Cyberbullying Defined

Patchin and Hinduja define cyberbullying as "willful and repeated harm inflicted through the medium of electronic text [3]." Willful harm excludes sarcasm between friends comments meant to criticize or disagree with an opinion but not meant to attack the individual.Nine different types of cyberbullying were identified [1][2][4]:

Flooding consists of the bully monopolizing the media so that the victim cannot post a message [2].

Masquerade involves the bully logging in to a website, chat room, or program using another user's screen name to either bully a victim directly or damage the victim's reputation [4].

Flaming, or **bashing**, involves two or more users attacking each other on a personal level. The conversation consists of a heated, short lived argument, and there is bullying language in all of the users' posts [4].

Trolling, also known as **baiting**, involves intentionally posting comments that disagree with other posts in an emotionally charged thread for the purpose of provoking a fight, even if the comments don't necessarily reflect the poster's actual opinion [1].

Harassment most closely mirrors traditional bullying with the stereotypical bully-victim relationship. This type of cyberbullying involves repeatedly sending offensive messages to the victim over an extended period of time [4].

Cyberstalking and **cyberthreats** involve sending messages that include threats of harm, are intimidating or very offensive, or involve extortion [4].

Denigration involves gossiping about someone online. Writing vulgar, mean, or untrue rumors about someone to another user or posting them to a public community or chat room or website falls under denigration [4].

3. Literate Survey

In a recent study on cyberbullying detection [5]**,** 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. In another study [7], gender specific features were used and users were categorized into male and female groups. In other study [8], NUM and NORM features were devised by assigning a severity level to the badwords list (nosewaring.com). NUM is a count and NORM is a normalization of the badwords respectively. The dataset consisted of 3,915 posted messages crawled from the Web Site, Formspring.me. They employed replication of positive examples up to ten times and accuracy on the range of classifiers was reported. Their findings showed that the C4.5 decision tree and an instancebased learner were able to identify the true positives with 78.5% accuracy. Dinakar et al. in [9] considered detecting cyberbullying in the form of sexuality, race, intelligence and profanity label-specific comments.On

4,500 manually labeled YouTube comments, the accuracy was reported for binary and multiclass classifiers such as, naive Bayes, JRIP, J48 and SMO. Their results indicated that binary label-specific classifiers outperformed multiclass classifiers with 66.7% accuracy.

Other interesting works [6]in this area performed harassment detection from forum and chat room 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. Contextual features are based on the similarity measure between posts, with the intuition that the posts which are dramatically different from their neighbors are more likely to be harassing posts. Research on online sexual predators detection [10], [11] associate the theory of communication and text-mining methods to differentiate between predator an victim conversations, as applied to one-to-one communication such as in a chat-log dataset. The formal method is based on the keywords search while the latter uses a rule-based approach. There are also some software products available for fighting against cyberbullying e.g., [12], [13], [14], [15], [16]. 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.

4. Facts on Cyber Bullying

Acoording to a survey, figure 1 shows analysis of the occurrence of cyberbullying by various social networking sites and email , and it also shows the the percentage of kids who faced online bullying and the youth who have committed suicide because of cyber bullying. It also shows the data gender wise.According to this surve maximum percentage of bullying is done through Facebook ie around 70% and minimum online bullying is done through emails ie around 25%.Around 48 % of kids have faced cyber harassment among them 45% were girls and 30% were boys and youth who have committed suicide due to cyberbullying are around 55 %.

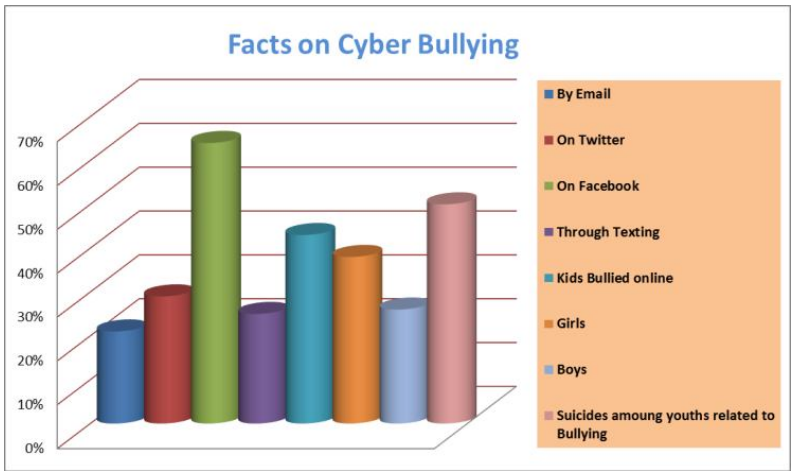

Fig. 1: A graph showing cyberbullying analysis.

In figure 2, a table shows the statistics of the parent awareness on the cyberbullying .According to the table, India is ranked top among all the countries.

	My child has experienced cyber bullying	A child in my community has experienced cyber bullying	Net Parent awareness of cyber bullying in country
TOTAL	12%	26%	38%
India	32%	45%	77%
Indonesia	14%	53%	47%
Sweden	14%	51%	65%
Canada	18%	31%	49%
Australia	13%	35%	48%
Brazil	20%	25%	45%
Saudia Arabia	14%	25%	41%
United States	15%	26%	41%
South Africa	10%	30%	41%
Turkey	5%	35%	40%
Mexico	8%	28%	36%
Argentina	9%	27%	36%
China	11%	25%	36%
Great Britain	11%	25%	36%
South Korea	8%	27%	35%
Poland	12%	20%	32%
Belgium	15%	13%	25%
Russia	5%	15%	20%
Germany	7%	12%	19%
Japan	7%	12%	19%
Hungary	7%	11%	18%
Italy	3%	15%	18%
Spain	5%	11%	16%
France	5%	10%	15%

Fig. 2: Table showing cyberbullying statistics according to different countries.

5. Conclusion

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? and then provide help as required by the user using data mining techniques. Also we will be using a User Identity for registration on our site ie one will have to provide an identity proof for registering on our site else they will not be able to make an account.With this feature we will be able to check the problem of fake accounts and also cyberbullying will be controlled to a limit as user accounts will be directly linked to their original identity.This mechanism will be very helpful for our society and the victims.

References

[1] Glossary of cyberbullying terms. (2008, January). Retrieved from http://www.adl.org/education/curriculum_connections/cyber bullying/glossary.pdf

[2] Maher, D. (2008). Cyberbullying: an ethnographic case study of one australian upper primary school class. Youth Studies Australia, 27(4), 50-57.

[3] Patchin, J., & Hinduja, S. “Bullies move beyond the schoolyard; a preliminary look at cyberbullying.” Youth violence and juvenile justice. 4:2 (2006). 148-169.

[4] Willard, Nancy E. Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, and Distress. Champaign, IL: Research, 2007.

[5] The changing face of bullying: An empirical comparison between traditional and internet bullying and victimization, Daniell M. Law a , Jennifer D. Shapka a, Shelley Hymel a, Brent F. Olson a, Terry Waterhouse b

[6] 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

[7] 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.

[8] K. Reynolds, A. Kontostathis, and L. Edwards, "Using Machine Learning to Detect Cyberbullying," In Proceedings of the 2011 10th International Conference on Machine Learning and Applications Workshops (ICMLA 2011), vol. 2, pp. 241-244, December 2011.

[9] 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.
[10] 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).
[11] I. Mcghee, J. Bayzick, A. Kontostathis, L. Edwards, A. Mcbride, and E. Jakubowski, "Learning to Identify Internet Sexual Predation," Inte rnational Journal on Electronic Commerce 2011, vol. 15, pp. 103-122, 2011.
[12] Bsecure. Available: http://www.safesearchkids.com/BSecure.html
[13] Cyber Patrol. Available: http://www.cyberpatrol.com/cpparentalcontrols.asp
[14] eBlaster. Available: http://www.eblaster.com/
[15] IamBigBrother. Available: http://www.iambigbrother.com/
[16] Kidswatch. Available: http://www.kidswatch.com/
[17] Bullying and Cyberbullying:History, Statistics, Law, Prevention and Analysis by Richard Donegan*Strategic Communication Elon University
[18] Detecting the Presence of Cyberbullying Using Computer Software by Jennifer Bayzick ,April Kontostathis, Lynne Edwards
[19] The problem of cyber bullying amongst school students in India: The loopholes in IT Act by Debarati Halder* and K. Jaishankar**
[20] Cyberbullying Detection: A Step Toward a Safer Internet Yard by Maral Dadvar and Franciska de Jong.