

A Survey On Different Approaches of Detecting Cyber Bullying Messages

¹M. Devi, ²M. Chitra Devi

¹Research Scholar, Department of Computer Science Sakthi Arts and Science College For Women, Oddanchatram, Tamil Nadu, India

²PG Head & Associate Professor, Department of Computer Science, Sakthi Arts and Science College For Women, Oddanchatram, Tamil Nadu, India

ABSTRACT

Opinion mining or sentiment analysis is considered as an important application of NLP (Natural Language Processing). Opinion mining is extracting the views that people express online. Those Websites which permits social interaction and collaboration can be considered as social media site, including networking sites such as Facebook, MySpace, and Twitter. Such sites offer today's youth a platform for amusement, entertainment, thrill, correspondence and communication with friends and furthermore have developed radically and exponentially as of late. This is the reason, there are various side effects, as cyberbullying has emerged as a serious issue afflicting children, adolescents and young adults. Machine learning techniques have conceivable ability to make automatic detection of bullying messages in social media, and this could develop a healthy and comparatively safe social media environment. Social media getting more and more popular in our day today life. By the popularity of the social media affects the people who involving into it. This makes the technology to work or to feel smarter and makes us lazier. On resulting to this robust and discriminative numerical representation learning of text messages is a critical issue. Hence here we propose a learning method to tackle this issue which is named as Semantic Enhanced Marginalized Denoising Auto Encoder (smsda).

Keywords: Social Media, Twitter, Machine Learning, Cyberbullying Detection.

I. INTRODUCTION

Opinion extraction, sentiment mining, affect analysis, analysis of emotion, review analyzing or mining, etc. are numerous comparative names with somewhat unique errands and with slightly different tasks. However, they are presently considered within the shades of opinion mining or sentiment analysis. This can be considered as an important application of NLP (Natural Language Processing). Some of social media sites, includes social networking sites such as Facebook, MySpace, and Twitter; video sites such as YouTube; photo sharing such as Flickr, Photobucket, or Picasa; gaming sites and virtual worlds such as

Kaneva, Club Penguin, Second Life, and the Sims; live casting such as Ustream or Twitch; instant messaging like Google talk, yahoo messenger or skype and blogs.

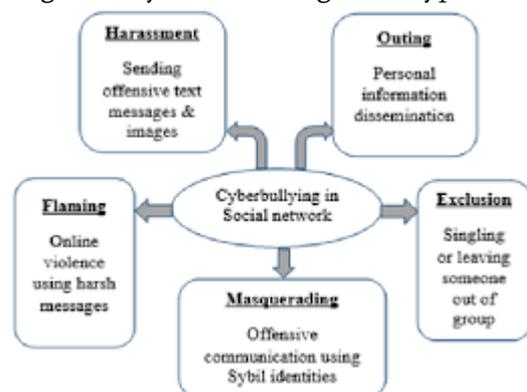


Figure-1: Cyberbullying in social network

II. SENTIMENT ANALYSIS APPROACHES

Sentiment Analysis or Opinion Mining intends to adjudge the viewpoint stance of a person with respect to any subject. There are two fundamental approaches used for opinion mining, those are:

- Machine Learning
- Semantic Orientation.

The machine learning approach is said to be belonging to supervised classification approach. This approach is more precise as each of the classifiers is initially trained on a collection of representative data which is known as corpus. In this way, it is called “supervised learning”. In a machine (supervised) learning based classification, two sorts of documents are required: training set as well as test set. A training set is utilized to prepare and learn the classifier and a test set is utilized to test the performance on execution of the automatic classifier. A huge number of machine learning techniques are available to access which classifies the opinions. Machine learning techniques like Naïve Bayes, Maximum Entropy (ME) and Support Vector Machines (SVM) have achieved extremely significant success in text categorization.

Figure-2 shows the procedure followed:

1. Tweets for various cases acts as input
2. The tweets are collected from the twitter using OAuth API.
3. For each of the tweets the stop words are removed and clean tweets are also obtained.

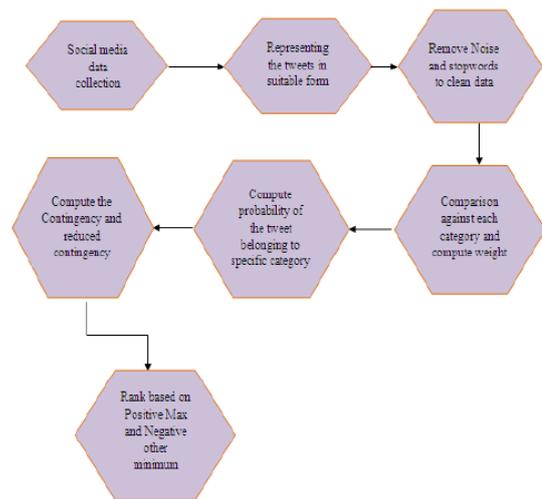


Figure-2: Process to detect bullying messages

Internet has become very popular and used around the world in our day to day life. By the growing of internet the cyber security is becoming the most important factor. Currently web 2.0 allows us to access the online related services and some users have been affected by the cybercrimes like cyber bullying experiences internationally. By these kind of issues the growth of social media gets the negative impacts from the various users. We propose an effective predator and victim identification with semantic enhanced marginalized denoising auto-encoder approach to detect cyber-bullying message from social media through the weighing scheme of feature of selection. We present Model to extract the cyber bullying network, which is used to identify the most active cyber bullying predators and victims to ranking algorithms the existing filters generally work with the simple key word search and are unable to understand the Semantic meaning of the text. So we propose Semantic Enhanced Marginalized Denoising Auto-Encoder.

III. LITERATURE SURVEY

Y. Bengio, A. Courville, and P. Vincent Represented Learning: A Review and New Perspectives: The success of machine learning algorithms generally depends on data representation, and we hypothesize that this is because different representations can entangle and hide more or less the different

explanatory factors of variation behind the data. Although specific domain knowledge can be used to help design representations, learning with generic priors can also be used, and the quest for AI is motivating the design of more powerful representation-learning algorithms implementing such priors. This paper reviews recent work in the area of unsupervised feature learning and deep learning, covering advances in probabilistic models, auto-encoders, manifold learning, and deep networks. This motivates longer-term unanswered questions about the appropriate objectives for learning good representations, for computing representations (i.e., inference), and the geometrical connections between representation learning, density estimation and manifold learning.

A. M. Kaplan and M. Haenlein Users of the world, unite! The challenges and opportunities of Social Media The concept of Social Media is top of the agenda for many business executives today. Decision makers, as well as consultants, try to identify ways in which firms can make profitable use of applications such as Wikipedia, YouTube, Facebook, Second Life, and Twitter. Yet despite this interest, there seems to be very limited understanding of what the term "Social Media" exactly means; this article intends to provide some clarification. We begin by describing the concept of Social Media, and discuss how it differs from related concepts such as Web 2.0 and User Generated Content. Based on this definition, we then provide a classification of Social Media which groups applications currently subsumed under the generalized term into more specific categories by characteristic: collaborative projects, blogs, content communities, social networking sites, virtual game worlds, and virtual social worlds. Finally, we present 10 pieces of advice for companies which decide to utilize Social Media.

R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner : Bullying in the digital age: a critical review and meta-analysis of cyber bullying

research among youth. Although the Internet has transformed the way our world operates, it has also served as a venue for cyber bullying, a serious form of misbehavior among youth. With many of today's youth experiencing acts of cyber bullying, a growing body of literature has begun to document the prevalence, predictors, and outcomes of this behavior, but the literature is highly fragmented and lacks theoretical focus. Therefore, our purpose in the present article is to provide a critical review of the existing cyber bullying research. The general aggression model is proposed as a useful theoretical framework from which to understand this phenomenon. Additionally, results from a meta-analytic review are presented to highlight the size of the relationships between cyber bullying and traditional bullying, as well as relationships between cyber bullying and other meaningful behavioral and psychological variables. Mixed effects meta-analysis results indicate that among the strongest associations with cyberbullying perpetration were normative beliefs about aggression and moral disengagement, and the strongest associations with cyberbullying victimization were stress and suicidal ideation. Several methodological and sample characteristics served as moderators of these relationships. Limitations of the meta-analysis include issues dealing with causality or directionality of these associations as well as generalizability for those meta-analytic estimates that are based on smaller sets of studies ($k < 5$). Finally, the present results uncover important areas for future research. We provide a relevant agenda, including the need for understanding the incremental impact of cyberbullying (over and above traditional bullying) on key behavioral and psychological outcomes B. K. Biggs, J. M. Nelson, and M. L. Sample: Peer relations in the anxiety-depression link: test of a mediation model. We employed a five-month longitudinal study to test a model in which the association between anxiety and depression symptoms is mediated by peer relations difficulties among a sample of 91 adolescents ages 14-17 ($M=15.5$, $SD=.61$) years. Adolescents completed measures of anxiety

symptoms, depression symptoms, peer group experiences (i.e., peer acceptance and victimization from peers), and friendship quality (i.e., positive qualities and conflict). As hypothesized, Time 1 anxiety symptoms predicted Time 2 (T2) depression symptoms, and this association was mediated by T2 low perceived peer acceptance and T2 victimization from peers, both of which emerged as unique mediators when they were considered simultaneously in the model. Contrary to expectations, qualities of adolescents' best friendships at T2 did not emerge as mediators and were largely unrelated to symptoms of anxiety and depression. Implications of the findings include the importance of addressing peer relations difficulties, especially peer acceptance and victimization, in the treatment of anxiety and the prevention of depression among anxious youth. K. Dinakar, R. Reichart, and H. Lieberman says about Modeling the Detection of

Textual Cyberbullying

The source of cyberbullying has assumed alarming proportions with an ever-increasing number of adolescents admitting to having dealt with it either as a victim or as a bystander. Anonymity and the lack of meaningful supervision in the electronic medium are two factors that have exacerbated this social menace. Comments or posts involving sensitive topics that are personal to an individual are more likely to be internalized by a victim, often resulting in tragic outcomes. We decompose the overall detection problem into detection of sensitive topics, lending itself into text classification sub-problems. We experiment with a corpus of 4500 YouTube comments, applying a range of binary and multiclass classifiers. We find that binary classifiers for individual labels outperform multiclass classifiers. Our findings show that the detection of textual cyberbullying can be tackled by building individual topic-sensitive classifiers.

IV. DATA SETS

We used two social media datasets, namely Twitter and MySpace for the problem we study. Both datasets contain labeled social media post. Twitter is a micro blogging website which allows users to post 140 characters messages called "Tweets". The retweets are removed from the dataset. The posts in this dataset have been manually labeled as bully or normal. MySpace is a social networking website which allows a registered users to view pictures, read chat and check other users profile information. The MySpace dataset used in the experiments is crawled from MySpace's groups feature. Each post in the dataset is manually labeled as normal or bully.

	Twitter	Myspace
No. of posts	7321	3245
No. of Features	3709	4236
No. of Positive Posts	2102	950
No. of Negative Posts	5219	2295
No. of users	7043	1053
Average posts per user	1.04	2.98

Table-1: Verifying Sentimental score Distribution

V. CONCLUSION

The paper addresses the text-based cyber bullying detection problem, where we have developed semantic enhanced marginalized denoising auto encoder as a specialised illustration learning model for cyber bullying detection. In addition, word embeddings have been wont to automatically expand and refine bullying word lists that's initialized by domain information. The performance of our

approaches has been experimentally verified through cyber bullying methods. As a next step we are unit coming up with to additional improve the strength of the learned illustration by considering ordination in messages.

VI. REFERENCES

- [1]. A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," *Business horizons*, vol. 53, no. 1, pp. 59-68, 2010.
- [2]. R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner, "Bullying in the digital age: A critical review and metaanalysis of cyberbullying research among youth." 2014.
- [3]. M. Ybarra, "Trends in technology-based sexual and non-sexual aggression over time and linkages to nontechnology aggression," National Summit on Interpersonal Violence and Abuse Across the Lifespan: Forging a Shared Agenda, 2010.
- [4]. B. K. Biggs, J. M. Nelson, and M. L. Sampilo, "Peer relations in the anxiety-depression link: Test of a mediation model," *Anxiety, Stress, & Coping*, vol. 23, no. 4, pp. 431-447, 2010.
- [5]. S. R. Jimerson, S. M. Swearer, and D. L. Espelage, *Handbook of bullying in schools: An international perspective*. Routledge/Taylor & Francis Group, 2010.
- [6]. G. Gini and T. Pozzoli, "Association between bullying and psychosomatic problems: A meta-analysis," *Pediatrics*, vol. 123, no. 3, pp. 1059-1065, 2009.
- [7]. A. Kontostathis, L. Edwards, and A. Leatherman, "Text mining and cybercrime," *Text Mining: Applications and Theory*. John Wiley & Sons, Ltd, Chichester, UK, 2010.
- [8]. J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore, "Learning from bullying traces in social media," in *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies*. Association for Computational Linguistics, 2012, pp. 656-666.
- [9]. Q. Huang, V. K. Singh, and P. K. Atrey, "Cyber bullying detection using social and textual analysis," in *Proceedings of the 3rd International Workshop on Socially-Aware Multimedia*. ACM, 2014, pp.3-6.
- [10]. D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, "Detection of harassment on web 2.0," *Proceedings of the Content Analysis in the WEB*, vol. 2, pp. 1-7, 2009.
- [11]. K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detection of textual cyberbullying." in *The Social Mobile Web*, 2011.
- [12]. V. Nahar, X. Li, and C. Pang, "An effective approach for cyberbullying detection," *Communications in Information Science and Management Engineering*, 2012.
- [13]. M. Dadvar, F. de Jong, R. Ordelman, and R. Trieschnigg, "Improved cyberbullying detection using gender information," in *Proceedings of the 12th -Dutch-Belgian Information Retrieval Workshop (DIR2012)*. Ghent, Belgium: ACM, 2012.
- [14]. M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, "Improving cyberbullying detection with user context," in *Advances in Information Retrieval*. Springer, 2013, pp. 693-696.
- [15]. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *The Journal of Machine Learning Research*, vol. 11, pp. 3371-3408, 2010.
- [16]. P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," *Unsupervised and Transfer Learning Challenges in Machine Learning*, Volume 7, p. 43, 2012.
- [17]. M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," *arXiv preprint arXiv: 1206.4683*, 2012.
- [18]. T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse processes*, vol. 25, no. 2-3, pp.259-284, 1998.