ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

A Survey on Predator and Victim Identification with Semantic-Enhanced Marginalized Denoising Auto-Encoder Approach for Cyberbullying Detection

Yogita Jagdale[1,] Prof. Gayatri M Bhandari [2]

M.E Student, Dept. of Computer, JSPM's, BSIOTR, Wagholi, Pune[1]

Asst. Professor, Dept. of Computer, JSPM's, BSIOTR, Wagholi, Pune [2]

ABSTRACT: The rapid growth of social networking is supplementing the progression of cyber-bullying activities. Most of the individuals involved in these activities belong to the younger generations, especially teenagers who are the worst scenario are at more risk of suicidal attempts 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 a graph model to extract the cyberbullying network, which is used to identify the most active cyberbullying 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. (smSDA) is developed via semantic extension of popular deep learning model stack denoising auto-encoder. The semantic extension consists of semantic dropout noise and sparsity constraints, where the semantic dropout noise is designed based on domain knowledge and the word embedding technique. The experiment show effective of our approach.

KEYWORDS: Cyberbullying detection, Text Mining, Representation learning, Stacked Denoising Auto encoders, Word Embedding.

I. 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 [1] a number of life threatening cyberbullying experiences among young people have been reported internationally, thus drawing attention to its negative impact. With the rapid growth of social media, users especially adolescents are spending significant amount of time on various social networking sites to connect with others, to share information, and to pursue common interests. In 2011, 70% of teens use social media sites on a daily basis and nearly one in four teens hit their favourite social-media sites 10 or more times a day[2]. While adolescents benefit from their use of social media by interacting with and learning from others, social media may have some side effects such as Cyber-bullying, which may have negative impacts on the life of people, especially children and teenagers.."**Cyber-bullying**" is when a child, preteen or teen is tormented, threatened, harassed, humiliated, embarrassed or otherwise targeted by another child, preteen or teen using the Internet, interactive and digital technologies or mobile phones[6]. As a side effect of increasingly popular social media, Cyber-bullying has emerged as a serious problem afflicting children, adolescents and young adults. For victims, they are easily exposed to harassment since all of us, especially youth, are constantly connected to Internet or social media [5]. As reported in, Cyber-bullying victimization rate ranges from 10% to 40%. In the United States, approximately 43% of teenagers were ever bullied on social media. The same as traditional bullying, Cyber-bullying has negative, insidious and sweeping impacts on children. The outcomes for victims under Cyber-bullying may even be tragic such as the occurrence of self-injurious behaviour suicides. One way

DOI: 10.15680/IJIRCCE.2016. 0412072

to address the Cyber-bullying problem is to automatically detect and promptly report bullying messages so that proper measures can be taken to prevent possible tragedies.

II. Related work

Text Mining and Cybercrime

Cyberbullying and internet predation frequently occur over an extended period of time and across several technological platforms [4]. This chapter describes the current state of research in the areas of cyberbullying and internet predation. It discusses the several commercial products which claim to provide chat and social networking site monitoring for home use. The chapter provides a summary of research into internet predation and cyberbullying. It reviews the technology that is available for capturing internet messages (IM) and internet relay chat (IRC). The chapter discusses the datasets that are currently available for research in the area. It surveys several research articles for both internet predation and cyberbullying detection, as well as provide a summary of the literature as it relates to legal issues.

Detection of harassment on Web 2.0

Web 2.0 has led to the development and evolution of web-based communities and applications. These communities provide places for information sharing and collaboration [5]. They also open the door for inappropriate online activities, such as harassment, in which some users post messages in a virtual community that are intention-ally offensive to other members of the community. It is a new and challenging task to detect online harassment; currently few systems attempt to solve this problem. In this paper, we use a supervised learning approach for detecting harassment. Our technique employs content features, sentiment features, and contextual features of documents. The experimental results described herein show that our method achieves significant improvements over several baselines, including Term Frequency-Inverse Document Frequency (TFIDF) approaches.

Modelling the Detection of Textual Cyberbullying

The scourge 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-problem[6]. 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.

Improved Cyberbullying Detection Using Gender Information

As a result of the invention of social networks, friendships, relationships and social communication are all undergoing changes and new definitions seem to be applicable[7]. One may have hundreds of "friends" without even seeing their faces. Meanwhile, alongside this transition there is increasing evidence that online social applications are used by children and adolescents for bullying. State-of-the-art studies in cyberbullying detection have mainly focused on the content of the conversations while largely ignoring the characteristics of the actors involved in cyberbullying. Social studies on cyberbullying reveal that the written language used by a harasser varies with the author’s features including gender. In this study we used a support vector machine model to train a gender-specific text classifier.

Improving Cyberbullying Detection with User Context

The negative consequences of cyberbullying are becoming more alarming every day and technical solutions that allow for taking appropriate action by means of automated detection are still very limited[8]. Up until now, studies on cyberbullying detection have focused on individual comments only, disregarding context such as users’ characteristics and profile information.

Stacked Denoising Auto encoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion

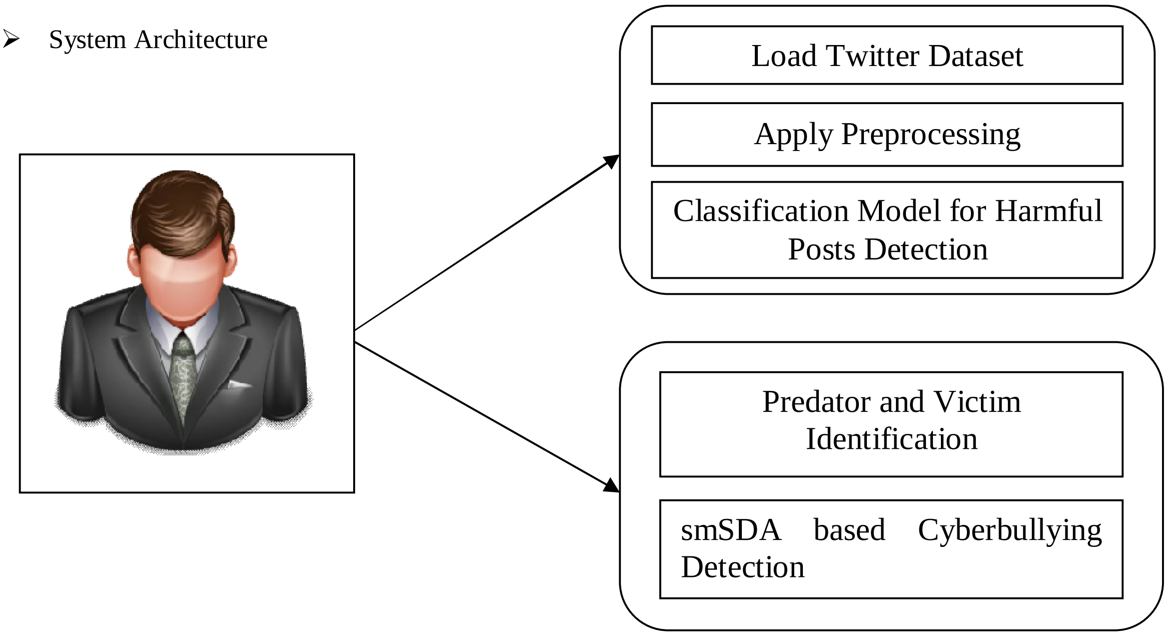
We explore an original strategy for building deep networks, based on stacking layers of denoising auto encoders which are trained locally to denoise corrupted versions of their inputs [9]. The resulting algorithm is a straightforward variation on the stacking of ordinary auto encoders. It is however shown on a benchmark of classification problems to yield significantly lower classification error, thus bridging the performance gap with deep belief networks (DBN), and in several cases surpassing it. Higher level representations learnt in this purely unsupervised fashion also help boost the performance of subsequent SVM classifiers. Qualitative experiments show that, contrary to ordinary auto encoders, denoising auto encoders are able to learn Gabor-like edge detectors from natural image patches and larger stroke detectors from digit images.

Auto encoders, Unsupervised Learning, and Deep Architectures

Auto encoders play a fundamental role in unsupervised learning and in deep architectures for transfer learning and other tasks. In spite of their fundamental role, only linear auto encoders over the real numbers have been solved analytically [10]. Here we present a general mathematical framework for the study of both linear and non-linear auto encoders. The framework allows one to derive an analytical treatment for the most non-linear auto encoder, the Boolean auto encoder. Learning in the Boolean auto encoder is equivalent to a clustering problem that can be solved in polynomial time when the number of clusters is small and becomes NP complete when the number of clusters is large. The framework sheds light on the different kinds of auto encoders, their learning complexity, their horizontal and vertical composability in deep architectures, their critical points, and their fundamental connections to clustering, Hebbian learning, and information theory.

III SYSTEM ARCHITECTURE AND ALGORITHM

- System Architecture


Load Twitter Dataset
Apply Preprocessing
Classification Model for Harmful Posts Detection
Predator and Victim Identification
smSDA based Cyberbullying Detection

- Algorithm

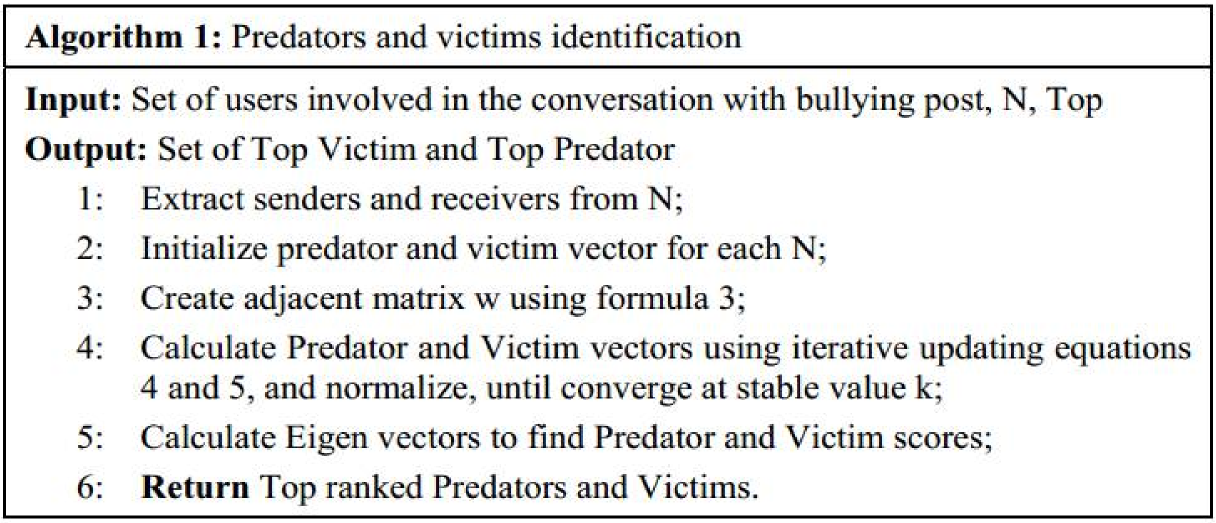
$$p(u) \leftarrow \sum_{u \to y} v(y) \tag{1}$$

$$v(u) \leftarrow \sum_{y \to u} p(y) \tag{2}$$

$$w_{ij} = \begin{cases} n \textit{ if there exist } n \textit{ bullying posts from } u_i \textit{ to } u_j \\ 0 \textit{ otherwise} \end{cases} \tag{3}$$

$$p(u_i) = w_{i1}v_1 + w_{i2}v_2 + \ldots + w_{iN}v_N \tag{4}$$

$$v(u_i) = w_{i1}p_1 + w_{i2}p_2 + \ldots + w_{iN}p_N \tag{5}$$



```
Algorithm 1: Predators and victims identification
Input: Set of users involved in the conversation with bullying post, N, Top
Output: Set of Top Victim and Top Predator
    1: Extract senders and receivers from N;
    2: Initialize predator and victim vector for each N;
    3: Create adjacent matrix w using formula 3;
    4: Calculate Predator and Victim vectors using iterative updating equations
       4 and 5, and normalize, until converge at stable value k;
    5: Calculate Eigen vectors to find Predator and Victim scores;
    6: Return Top ranked Predators and Victims.
```

IV. RESULTS

In the above system architecture we firstly load the twitter dataset. twitter is the real time information network that connect you to the latest stories, ideas opinions and news about what you find interesting and after that user can register user can read the post tweets which are defined as the message posted on twitter thus preprocessing has applied to improve the quality of research. predators and victim are linked to each other by the means of post identified by their username. smSDA provide robust and discriminative representation to achive good performance on testing document.

V. CONCLUSION AND FURURE WORK

In this thesis we propose an approach for cyberbullying detection and the identification of the most active predators and victims. To improve the classification performance we employ a weighted TFIDF function, in which bullying-like features are scaled by a factor of two. The overall results using weighted TFIDF outperformed other methods. This captures our idea to scale-up inductive words within the harmful posts. However, bullying-like feature sets are limited to a static set of keywords. Therefore, dynamic strategies are required to be implemented to find emerging harmful and abusive words from the streaming text. To improve classifier's training in the absence of a sufficient number of positive examples, oversampling of positive posts is used. Also, throughout our experiments, we note that comparatively better performance was observed for false negative compared to false positive cases in individual and combined datasets. This is because of the fewer positive cases available for classifier's training. Therefore advance methods, which are capable of dealing with a few training sets in automatic cyberbullying detection, and to reduce false positive and false negative cases need to be developed, In addition, we proposed a cyberbullying graph model to rank the most active users (predators or victims) in a network. The proposed graph model can be used to answer various queries regarding the bullying activity of a user. It can also be used to detect the level of cyberbullying victimization for decision making in further investigations. Our future research in cyberbullying detection will continue to reduce false cases and train classifiers with fewer positive examples. We also plan to continue the in-depth analysis of cyberbullying victimization and its emerging patterns in stream text, to help the detection and mitigation of the cyberbullying. As a next step we are planning to further improve the robustness of the learned representation by considering word order in messages.

ACKNOWLEDGMENT

I thank my project guide Prof**. Gayatri M Bhandari** and ME. Coordinator **Prof. Archana Lomte** for the guidelines in completion of this paper. I also wish to record my thanks to our Head of Department **Prof. G. M. Bhandari** for consistent encouragement and ideas.

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] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proceedings of the National academy of Sciences of the United States of America, vol. 101, no. Suppl 1, pp. 5228–5235, 2004.

[3] 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.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," the Journal of machine Learning research, vol. 3, pp. 993–1022, 2003.

[5] 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.

[6] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 35, no. 8, pp. 1798–1828, 2013.

[7] G. Gini and T. Pozzoli, "Association between bullying and psychosomatic problems: A meta-analysis," Pediatrics, vol. 123, no. 3, pp. 1059–1065, 2009.

[8] J. Juvonen and E. F. Gross, "Extending the school grounds? bullying experiences in cyberspace," Journal of School health, vol. 78, no. 9, pp. 496–505, 2008.

[9] 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.

[10] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," Unsupervised and Transfer Learning Challenges in Machine Learning, Volume 7, p. 43, 2012.

BIOGRAPHY

Yogita M Jagdale is the ME student of the computer engineering in JSPM'S Bhivrabai Sawant Institute Of Technology research and has received BE degree in 2013 from Amravati her research interest in network security.

Dr. Gayatri Mahendra Bhandari Awarded Ph.D in Computer engineering in 2016, completed M.Tech in 2008 from COEP, completed BE in 2001. Academic experience about 15 yrs. Area of Specialization are N/w security, audio processing, cloud computing etc.