Towards the Detection of Cyberbullying Based on Social Network Mining Techniques

I-Hsien Ting
Department of Information Management
National University of Kaohsiung
Kaohisung, Taiwan
iting@nuk.edu.tw

Wun Sheng Liou
Department of Information Management
National University of Kaohsiung
Kaohisung, Taiwan
ericievonne@hotmail.com

Dario Liberona
Universidad Técnica Federico Santa María
Santiago, Chile
dario.liberona@usm.cl

Shyue-Liang Wang
Department of Information Management
National University of Kaohsiung
Kaohisung, Taiwan
slwang@nuk.edu.tw

Giovanny Mauricio Tarazona Bermudez
Universidad Distrital Francisco Jose de Caldas
Bogota, Colombia
gtarazona@gmail.com

***Abstract*—In recent years, users are widely intend to express and share their opinions over the Internet. However, due to the characters of social media, it appears negative use of social media. Cyberbullying is one of the abuse behavior in the Internet as well as a very serious social problem. Under this background and motivation, it can help to prevent the happen of cyberbullying if we can develop relevant techniques to discover cyberbullying in social media. Thus, in this paper we propose an approach based on social networks analysis and data mining for cyberbullying detection. In the approach, there are three main techniques for cyberbullying discovery will be studied, including keyword matching technique, opinion mining and social network analysis. In addition to the approach, we will also discuss the experimental design for the evaluation of the performance.**

Keywords—social network, social network analysis, social network mining, cyberbullying

I. INTRODUCTION

Recently, WWW has become the major media for users to communicate and express their opinion, from virtual community, forums, Blog to Social Network Services (SNS). Social Network Services are also called as social media, which has the character of free and anonymous and it allows user to interact with each other without boundary. However, abuse of social media is also becoming a very serious problem, such as cybercrime, cyber-terrorism, Internet fraud and cyberbullying, etc.

Cyberbullying is one of the most frequently happen Internet abuse and also a very serious social problem especially for teenager [1]. Therefore, more and more researchers are devoting on how to discover and prevent the happen of cyberbullying, especially in social media. Therefore, it would be very worth to develop an approach for cyberbullying detection.

Under this background and motivation, the main object of this paper is focusing on the development of cyberbullying detection approach. In the approach, three techniques will be used, including keyword technique, opinion mining and social networks analysis. These techniques will be combined as a experimental approach for cyberbullying detection.

II. CYBERBULLYING DETECTION

Nowadays, the major techniques for cyberbullying detection are based on keyword matching. The simplest one is based on dictionary which collect a set of keywords that related to cyberbullying. Some advanced techniques including intelligent tag based approach, machine learning and artificial intelligence approach, such as genetic algorithm or neural network[2][3]. However, these techniques are only focusing on keyword matching to detect cyberbullying and there are obviously two important factors are missing. First is the behavior model of cyberbullyer such as repeat and continuous attack, and second is tone and sentiment of their speak on Internet.

Social Networks Mining (SNM) is a new research area which is developed by combing Social Network Analysis and Data Mining. SNM is very helpful for analyzing the data in social media, related techniques including Web Mining, Social Networks Analysis, Text Mining, Natural Language Processing, Sentimental Analysis and Opinion Mining. These techniques are very common to be used for analyzing users' opinion in social networking websites [4][5]. Some of these techniques are considered very helpful to overcome the weakness part of current cyberbullying detection approach.

978-1-5386-2366-4/17/$31.00 ©2017 IEEE

III. Proposed Approach

In order to detect the cyberbullying behavior, we have proposed an approach which is based on techniques of SNM. The process of the cyberbullying approach is shown in figure 1. We have already collected data from four major social networking websites in Taiwan, which including Facebook, Twitter, Ptt (https://www.ptt.cc) and CK101 (https://ck101.com/). These four websites are very different and we intend to comprehensively study the happen of cyberbullying in different websites. Three features will be extracted from the data, which are keywords, SNA measurements and Sentiments. After features selection and pattern matching we then can detect posts in those websites where cyberbullying happen.

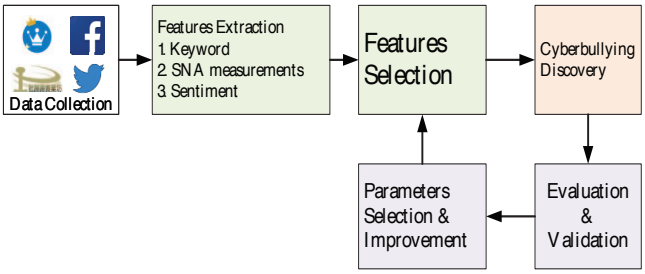

Fig. 1. Cyberbullying Detection Approach

About feature selection in the approach, which is considered as the most important step. In prior to do the features selection, we need to identify features first. These features can be used to identify the pattern of cyberbullying. Figure 2 shows the experiment to define important features to classify cyberbullying.

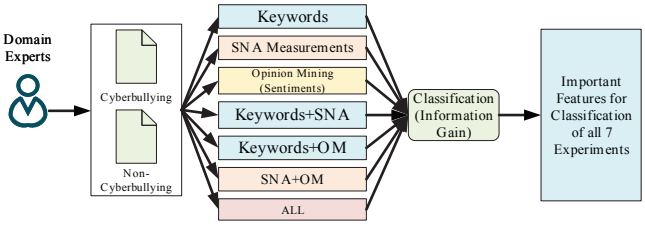

Fig. 2. Experimental Design and Claissifcation for Defining Important Features to Classify Cyberbullying Posts.

Currently, we collected 100 posts which are considered as cyberbullying happen for each website as well as another 100 without cyberbullying. Then, we extracted keywords, SNA measurements, and sentiments. Classification, which is one of the most well know data mining techniques will be used to analyze the information gain. This approach can help to identify important features from the seven experiments. Among all the features, sentimental is an important feature that can be used to understand the sentiment when user posts a message in social media. The extraction approach of sentimental features is shown in figure 3 and the approach is very common used in most of the sentimental extraction approach.

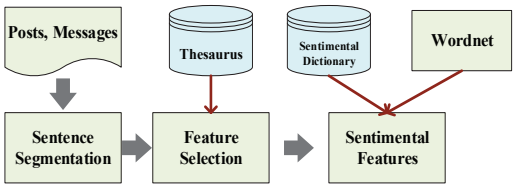

Fig. 3. Sentimental Features Extraction

About the performance evaluation, in figure 1, after cyberbullying discovery, we then evaluate the accuracy of the discovered cyberbullying posts. In our experiment, precission and recall are the measurments for us to validate the accuracy. The first evaluation result shows the precission accuracy is around **0.79** and the recall is **0.71**. It means more than **70%** of cyberbullying posts can be detected correctly by using our approach, which is better than what we exepcted, but maybe can be improved by considering different weights for different features.

IV. Conclusion and Future Works

In this paper, we have developed an approach towards the detection of cyberbullying behavior. The approach is designed based on technique of Social Networks Mining with three important features, including Keywords, SNA measurements and Sentiment. We also designed a series of experiments to implement the approach and evaluate the performance. The data are collected from four major social networking websites and the evaluation results show an acceptable accuracy for cyberbullying detection.

Due to this is a very first step of the research, we have some future directions to proceed the research. The first future research direction is to consider the weight for the three features. We intend to tune the weight and hopefully the accuracy can be improved. In the future, we would also like to study the situation when applying the approach to different countries and different culture. We believe some interesting results will be discovered, which maybe also very helpful for us to improve the performance of cyberbullying detection.

References

[1] J. Juvonen and E. F. G, "Extending the school grounds?-Bullying experiences in cyberspace," *J. Sch. Health*, vol. 78, pp. 496–505, 2008.

[2] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network."

[3] B. S. Nandhini and J. I. Sheeba, "Social Network Bullying Detection Using Intelligence Techniques," *Procedia - Procedia Comput. Sci.*, vol. 45, pp. 485–492, 2015.

[4] K. Jedrzejewski and M. Morzy, "Opinion Mining and Social Networks: A Promising Match," *2011 Int. Conf. Adv. Soc. Networks Anal. Min.*, pp. 599–604, Jul. 2011.

[5] A. Stavrianou, J. Velcin, and J.-H. Chauchat, "Definition and Measures of an Opinion Model for Mining Forums," *2009 Int. Conf. Adv. Soc. Netw. Anal. Min.*, pp. 188–193, Jul. 2009.