BullyNet Unmasking Cyberbullies on Social Networks

ABSTRACT

One of the most harmful consequences of social media is the rise of cyberbullying, which tends to be more sinister than traditional bullying given that online records typically live on the internet for quite a long time and are hard to control. In this paper, we present a three-phase algorithm, called BullyNet, for detecting cyberbullies on Twitter social network. We exploit bullying tendencies by proposing a robust method for constructing a cyberbullying signed network. We analyze tweets to determine their relation to cyberbullying, while considering the context in which the tweets exist in order to optimize their bullying score. We also propose a centrality measure to detect cyberbullies from a cyberbullying signed network, and we show that it outperforms other existing measures. We experiment on a dataset of 5.6 million tweets and our results shows that the proposed approach can detect cyberbullies with high accuracy, while being scalable with respect to the number of tweets.

**EXISTING SYSTEM**

The first method of determining bullying messages was done using a combination of text-based analytics and a mix of text and user features. Zhao et al. [18] proposed a text based Embeddings-Enhanced Bag-of-Words (EBoW) model that utilizes a concatenation of bullying features, bag-of-words, and latent semantic features to obtain a final representation, which is then passed through a classifier to identify cyberbullies.

Xu et al. [21] used textual information to identify emotions in bullying traces, as opposed to determining whether or not a message was bullying. Singh et al. [19] proposed a probabilistic socio-textual information fusion for cyberbullying detection. This fusion uses social network features derived from a 1.5 ego network and textual features, such as density of bad words and part-of-speech-tags. Hosseinmardi et al. [20] used images and text to detect cyberbullying incidents. The text and image features were gathered from media sessions containing images and the corresponding comments, which was then fed into various classifiers. Chen [25] proposed an novel method in identifying cyberbullies within a multi-modal context. To understand cyberbullying Kao et al. [26] proposed a framework by studying social role detection. By using words and comments, temporal characteristics, and social information of a session as well as peer influence Cheng et al. [27], [28] proposed frameworks for detecting cyberbullies.

The second method was aimed at identifying the person behind the cyberbullying incidents. Squicciarini et al. [22] used MySpace data to create a graph, which integrated user, textual, and network features. This graph was used to detect cyberbullies and predict the spreading of bullying behavior through node classification. Gal´an-Garc´ıa et al. [23] used supervised machine learning to detect the real users behind troll profiles on Twitter, and demonstrated the technique in a

real case of cyberbullying. In a recent paper on aggression and bullying in Twitter, Chatzakou et al. [24] found cyberbullies and aggressors using user, text, and network-based features.

Disadvantages

* The system is less effective due to lack of Constructing bullying signed network.
* The system doesn’t effective due to lack of training large scale datasets.

**PROPOSED SYSTEM**

In the proposed system, the system studies the problem of cyberbullying in social media in an attempt to answer the following research question: Can tweet contexts (conversations) help improve the detection of cyberbullying in Twitter?. Our intuition is that each tweet should be evaluated not only based on its contents, but also based on the context in which it exists. The system calls such a context a conversation, which is a set of tweets between two or more people exchanging information about a certain subject. Thus, our solution consists of three parts. First, for each conversation, a conversation graph is generated based on the sentiment and bullying words in the tweets. Second, we compute the bullying score for each pair of users in a conversation graph, and then combine all graphs to create an SSN called bullying signed network (B). The inclusion of negative links can bring out information that would otherwise be missed with only positive links [16]. Finally, we propose a centrality measure called attitude & merit (A&M) to detect bullying users from the signed network B.

Our main contributions are organized as follows:

1) Collected, preprocessed and labelled the Twitter dataset.

2) Proposed a novel efficient algorithm for detecting cyberbullies on Twitter.

a) Built conversation.

b) Constructed Bullying Signed Network.

c) Proposed Attitude and Merit Centrality.

3) Experimented on 5.6 million tweets collected over 6 months. The results show that our approach can detect cyberbullies with high accuracy, while being scalable with

respect to the number of tweets.

**Advantages**

* The system is more effective due to presence of Conversation Graph Generation Algorithm, Bullying Signed Network Generation Algorithm, and Bully Finding Algorithm.
* The system is more effective due to the techniques to analyze large number of datasets.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP
* Coding Language - Java/J2EE(JSP,Servlet)
* Front End - J2EE
* Back End - MySQL