European Journal of Molecular & Clinical Medicine

ISSN 2515-8260 Volume 09, Issue 03, 2022

DETECTION OF CYBER BULLYING ON SOCIAL MEDIA

T.Jaya Sri
Asst. Professor, Department of computer science and engineering
QIS COLLEGE OF ENGINEERING AND TECNOLOGY

A. Venkata Sai, J.Kamakshi Venkata Nagasaisiribhumika, S.Keerthi, G.Anusha5, SK.Charishma Banu6
B.Tech., Scholars, Department of Computer science and engineering
QIS COLLEGE OF ENGINEERING AND TECNOLOGY

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.

Introduction

The Internet has created never before seen opportunities for human interaction and socialization. In the past decade, social media, in particular, has had a popularity explosion. From MySpace to Face book, Twitter, Flickr, and Instagram, people are connecting and interacting in a way that was previously impossible. The widespread usage of social media across people from all ages created a vast amount of data for several research topics, including recommender systems [1], link predictions [2], visualization, and analysis of social networks [3].

While the growth of social media has created an excellent platform for communications and information sharing, it has also created a new platform for malicious activities such as spamming [4], trolling [5], and cyber bullying [6]. According to the Cyber bullying Research Center (CRC) [7], cyber bullying occurs when someone uses the technology to send messages to harass, mistreat or threaten a person or a group. Unlike traditional bullying

where aggression is a short and temporary face to- face occurrence, cyber bullying contains hurtful messages which are present online for a long time. These messages can be accessed worldwide, and are often irrevocable. Laws about cyber bullying and how it is handled differ from one place to another. For example, in the United States, the majority of the states incorporate cyberbullying into their bullying laws, and cyber bullying is considered a criminal offense in most of them [8]. Popular social media platforms such as Face book and Twitter are very vulnerable to cyber bullying due to the popularity of these social media sites and the anonymity that the internet offers to the perpetrators. Although strict laws exist to punish cyber bullying, there are very less tools available to effectively combat cyber bullying. Social media platforms provide users with the option to self-report abusive behavior and content in addition to providing tools to deal with bullying. For example, Twitter has features that include locking accounts for a brief period of time or banning the accounts when the behavior becomes unacceptable. The body of work produced by the research community with regards to cyber bullying in social networks also needs to be expanded to get better insights and help develop effective tools and techniques to tackle the issue.

To identify cyber bullies in social media, we first need to understand how social media can be modeled. The common way of modeling relationship in social psychology [9] is to represent it as a signed graph with positive edge corresponds to the good intent and negative edge corresponds to malicious intent between people.

Mining social media networks to determine cyber bullies imposes several challenges and concerns. First, it is typically hard to accurately interpret user's intentions and meanings in social media based merely on their messages (e.g. posts, tweets, comments) which are typically short, use slang languages, or may include multimedia contents such as pictures and videos. For example, Twitter limits its users' messages to 140 characters, which could be a mix of text, slangs, emojis, and gifs. As a result, it is hard to determine the opinion expressed by a message correctly. For this we utilize sentiment analysis [11], [12] to determine whether the user's attitude towards other users is positive, negative, or neutral. Second, bullying could be hard to detect if the bully chooses to disguise it through techniques such as sarcasm or passive-aggression. In this situation, a single text (message) cannot determine the user's intention. So, we collect the entire conversation between two or more users to identify the context in which the user attitude exists. Third, the large size and dynamic and complex structure of social media networks makes it challenging to identify cyber bullies. For example, on Twitter, hundreds of millions of tweets are sent every day on the social network platform. In this case we construct the social network as a graph and assign value based on the maliciousness of the user. Because the network analysis reduces the complex relationship between the users to the simple existence of nodes and edges [10] There are several works in the literature concerning detecting malicious users from unsigned networks with positive edge weights, including community detection [13],

node classification [14] and link prediction [2]. On the other hand, methods that analyze signed social networks are scarce [15].

In this paper, we study the problem of cyber bullying in social media in an attempt to answer the following research question: Can tweet contexts (conversations) help improve the detection of cyber bullying 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. We call 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 KB.

Literature Work

There are many attitudes that proposes systems which can detect cyberbullying automatically with high exact. First one is author Nandhini et al [1]. have proposed a model that uses Naïve Bayes machine learning approach and by their work they achieved 91% accuracy and got their dataset from MySpace.com, and then they proposed another model [2]. Naive Bayes classifier and genetic operations (FuzGen) and they achieved 87% accuracy. Another approach by Romsaiyudet al [3]. they enhanced the Naïve Bayes classifier for extracting the words and examining loaded pattern clustering and by this approach they achieved 95.79% accuracy on datasets from Slashdot, Kongregate, and MySpace. However, they have a problem that the cluster processes do not work in parallel. Moreover, in the accurate proposed by Bunchanan et al[4].they used War of Tanks game chat to get their dataset and manually classified them and then differentiate them to simple Naïve classification that uses sentiment analysis as a feature, their results were poor when compared to the manually classified Results. Furthermore, Isa et al[5]. proposed an approach after getting their dataset from kaggle they used one classifier Naïve Bayes. The Naïve Bayes classifier yielded average accuracy of 92.81% while SVM with poly kernel yielded accuracy of 97.11%, but they did not mention their training or testing size of the dataset, so the results may not be acceptable. Another Approach by Dinakar et al[6]. that aimed to detect explicit bullying language pertaining to Sexuality, Race & Culture and intelligence, they acquired their dataset from YouTube comment section. After applying Naïve Bayes classifiers, SVM yielded accuracy of 66% and Naïve Bayes 63%. Moving on to Di Capua et al[7]. they proposed a new way for cyberbullying detection by adopting an unsupervised approach, they used the classifiers inconsistently over their dataset, applying SVM on Form Spring and achieving 67% on recall, applying GHSOM on YouTube and achieving 60%

precision, 69% accuracy and 94% recall, applying Naïve Bayes on Twitter and achieving 67% accuracy. Additionally, Haidar et al[8]. proposed a model to detect cyberbullying but using Arabic language they used Naïve Bayes and achieved 90.85% precision and SVM achieved 94.1% as precision but they have extraordinary of false positive also they are work on Arabic language. Another type of proposal using machine learning. One of the proposed methods is Zhang et al. [9] in their paper uses novel pronunciation-based convolution neural network (PCNN), thereby lighten the problem of noise and bullying data scarcity to overcome class imbalance. 1313 messages from twitter, 13,000 messages from spring. Accuracy of the twitter dataset was not calculated thanks to it being imbalanced. While Achieving 56% on precision, 78% recall and 96% accuracy, while achieving high accuracy their dataset was unbalanced, so that gives false results and that reflects in precision score which is 56%. The authors Nobata et al. [10] showed that using abusive language has enlarged recently, they used a framework called Vowpal wabbit for classification, and they also developed a supervised classification methodology with NLP features that execute machine learning approach, The F-Score reached 0.817 using dataset collected from comments posted on Yahoo News and Finance. Zhao et al. [11] proposed framework particular for cyberbullying detection, they used word embedding that creates an inventory of pre-defined insulting words and allocate weights to get bullying features, they used SVM as their main classifier and got recall of 79.4%. Then another approach was proposed by Parime et al. [12] they got their dataset from MySpace and manually marked them, and they used the SVM Classifier for the classification. Moreover, Chen et al. [13] proposed a replacement feature extraction method called Lexical Syntactic Feature and SVM as their classifier and that they reached 77.9% precision and 77.8% recall. Furthermore, Ting et al. [14] proposed a strategy based on SNM, they collected their data from social media and then used SNA calculation and sentiments as features. Seven experiments were made, and they reached around 97% precision and 71% as recall. Furthermore, Harsh Dani et al. [15] introduced a new framework called SICD, they used KNN for classification. Finally, they achieved 0.6105 F1 score and 0.7539 AUC score. SVM classifier was one of the proposals used in the research papers.

ExistingSystem

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.

- Built conversation.
- Constructed Bullying Signed Network.
- 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.

Implementation

- **Registration based Social Authentication Module**
- **Security Module**
- **Attribute-based encryption module.**
- **Multi-authority module.**

Registration -Based Social Authentication Module:

The system prepares trustees for a user Alice in this phase. Specifically, Alice is first authenticated with her main authenticator (i.e., password),and then a few(e.g., 5) friends, who also have accounts in the system, are selected by either Alice herself or the service provider from Alice's friend list and are appointed as Alice's Registration.

Security Module:

Authentication is essential for securing your account and preventing spoofed messages from damaging your online reputation. Imagine a phishing email being sent from your mail because someone had forged your information. Angry recipients and spam complaints resulting from it become your mess to clean up, in order to repair your reputation. trustee-based social authentication systems ask users to select their own trustees without any constraint. In our experiments (i.e., Section VII), we show that the service provider can constrain trustee selections via imposing that no users are selected as trustees by too many other users, which can achieve better security guarantees

ISSN 2515-8260 Volume 09, Issue 03, 2022

Attribute-based encryption module.

Attribute-based encryption moduleis using foreach and every node encrypt data store. After encrypted data and again the re-encrypted the same data is using for fine-grain concept using user data uploaded. the attribute-based encryption have been proposed to secure the cloud storage. Attribute-Based Encryption (ABE). In such encryption scheme, an identity is viewed as a set of descriptive attributes, and decryption is possible if a decrypter's identity has some overlaps with the one specified in the ciphertext.

Multi-authority module.

A multi-authority system is presented in which each user has an id and they can interact with each key generator (authority) using different pseudonyms. Our goal is to achieve a multi-authority CP-ABE which achieves the security defined above; guarantees the confidentiality of Data Consumers' identity information; and tolerates compromise attacks on the authorities or the collusion attacks by the authorities. This is the first implementation of a multi-authority attribute-based encryption scheme.

CONCLUSION

Although the digital revolution and the rise of social media enabled great advances in communication platforms and social interactions, a wider proliferation of harmful behavior known as bullying has also emerged. This paper presents a novel framework of Bully Net to identify bully users from the Twitter social network. We performed extensive research on mining signed networks for better understanding of the relationships between users in social media, to build a signed network (SN) based on bullying tendencies. We observed that by constructing conversations based on the context as well as content, we could effectively identify the emotions and the behavior behind bullying. In our experimental study, the evaluation of our proposed centrality measures to detect bullies from signed network, we achieved around 80% accuracy with 81% precision in identifying bullies for various cases. There are still several open questions deserving further investigation. First, our approach focuses on extracting emotions and behavior from texts and emojis in tweets. However, it would be interesting to investigate images and videos, given that many users use them to bully others. Second, it does not distinguish between bully and aggressive users. Devising new algorithms or techniques to distinguish bullies from aggressors would prove critical in better identification of cyber bullies. Another topic of interest would be to study the relationship between conversation graph dynamics and geographic location and how these dynamics are affected by the geographic dispersion of the users? Are the proximity increase the bullying behaviour?

References

1] J. Tang, C. Aggarwal, and H. Liu, “Recommendations in signed socialnetworks,” in Proceedings of the International Conference on WWW,2016, pp. 31–40.

[2] D. Liben-Nowell and J. Kleinberg, “The link-prediction problem forsocial networks,” Proceedings of the ASIS&T, vol. 58, no. 7, pp. 1019–1031, 2007.

[3] U. Brandes and D. Wagner, “Analysis and visualization of socialnetworks,” in Graph drawing software, 2004, pp. 321–340.

[4] X. Hu, J. Tang, H. Gao, and H. Liu, “Social spammer detection withsentiment information,” In Proceedings of IEEE ICDM, pp. 180—189,2014.

[5] E. E. Buckels, P. D. Trapnell, and D. L. Paulhus, Trolls just want tohave fun, 2014, pp. 67:97–102.

[6] S. Kumar, F. Spezzano, and V. Subrahmanian, “Accurately detectingtrolls in slashdot zoo via decluttering,” in Proceedings of IEEE/ACMASONAM, 2014, pp. 188–195.

[7] J. W. Patchin and S. Hinduja, “2016 cyberbullying data,” 2017.

[8] C. R. Center, “https://cyberbullying.org/bullying-laws.”

[9] D. Cartwright and F. Harary, “Structural balance: a generalization ofheider’s theory.” Psychological review, vol. 63, no. 5, p. 277, 1956.

[10] J. Leskovec, D. Huttenlocher, and J. Kleinberg, “Signed networks insocial media,” in Proceedings of the SIGCHI CHI, 2010, pp. 1361–1370.

[11] R. Plutchik, “A general psychoevolutionary theory of emotion,” inTheories of emotion, 1980, pp. 3–33.

[12] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithmsand applications: A survey,” Proceedings of the Ain Shams engineeringjournal, vol. 5, no. 4, pp. 1093–1113, 2014.

[13] L. Tang and H. Liu, “Community detection and mining in social media,”Synthesis lectures on data mining and knowledge discovery, vol. 2, no. 1,pp. 1–137, 2010.

[14] S. Bhagat, G. Cormode, and S. Muthukrishnan, “Node classification insocial networks,” in Social network data analytics, 2011, pp. 115–148.

[15] J. Tang, Y. Chang, C. Aggarwal, and H. Liu, “A survey of signed networkmining in social media,” In Proceedings of the ACM Comput. Surv.,no. 3, pp. 42:1–42:37, 2016.

[16] J. Kunegis, J. Preusse, and F. Schwagereit, “What is the added valueof negative links in online social networks?” in Proceedings of theInternational Conference on WWW, 2013, pp. 727–736.

[17] Z. Wu, C. C. Aggarwal, and J. Sun, “The troll-trust model for ranking insigned networks,” in Proceedings of the ACM International Conferenceon WSDM, 2016, pp. 447–456.

[18] R. Zhao, A. Zhou, and K. Mao, “Automatic detection of cyberbullyingon social networks based on bullying features,” in Proceedings of theICDCN, 2016.

[19] V. K. Singh, Q. Huang, and P. K. Atrey, “Cyberbullying detection usingprobabilistic socio-textual information fusion,” In Proceedings of theIEEE/ACM ASONAM, pp. 884—-887, 2016.

[20] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, andS. Mishra, “Detection of cyberbullying incidents on the instagram socialnetwork,” In Proceedings of the CoRR, 2015.

[21] J.-M. Xu, X. Zhu, and A. Bellmore, “Fast learning for sentiment analysison bullying,” in Proceedings of the First International WISDOM, 2012,pp. 10:1–10:6.

[22] A. C. Squicciarini, S. M. Rajtmajer, Y. Liu, and C. H. Griffin, “Identificationand characterization of cyberbullying dynamics in an onlinesocial network,” in Proceedings of the IEEE/ACM ASONAM, 2015, pp.280–285.

[23] P. Galan-Garcia, J. De La Puerta, C. G´omez, I. Santos, and P. Bringas,“Supervised machine learning for the detection of troll profiles in twittersocial network: Application to a real case of cyberbullying,” vol. 24, pp.42–53, 2014.

[24] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini,and A. Vakali, “Mean birds: Detecting aggression and bullying ontwitter,” in Proceedings of the ACM on WebSci, 2017, pp. 13–22.

[25] L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, “Xbully:Cyberbullying detection within a multi-modal context,” in Proceedingsof the Twelfth ACM International Conference on Web Search and DataMining, 2019,pp.339-347.

ISSN 2515-8260 Volume 09, Issue 03, 2022

Authors Profile

Mrs.T.JayaSriworking as Assistant professor of Computer Science & Engineering in Qis college ofEngineering and Technology (Autonomous&NAAC‘A’Grade), Ponduru Road, Vengamukkalapalem, Ongole,Prakasam Dist. Affiliated to Jawaharlal Nehru Technological University, Kakinada.

A.Venkata sai pursuing B.Tech in the department of Computer Science & Engineering from Qis college ofEngineering and Technology (Autonomous &NAAC‘A’Grade), Ponduru Road, Vengamukkalapalem, Ongole,Prakasam Dist. Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2018-2022 respectively.

J.Kamakshi Venkata NagasaiSiribhumika pursuing B.Tech in the department of Computer Science & Engineering from Qis college ofEngineering and Technology (Autonomous &NAAC‘A’Grade), Ponduru Road, Vengamukkalapalem, Ongole,Prakasam Dist. Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2018-2022 respectively.

S.Keerthi pursuing B.Tech in the department of Computer Science & Engineering from Qis college ofEngineering and Technology (Autonomous &NAAC‘A’Grade), Ponduru Road, Vengamukkalapalem, Ongole,Prakasam Dist. Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2018-2022 respectively.

G.Anushapursuing B.Tech in the department of Computer Science & Engineering from Qis college of Engineeringand Technology (Autonomous &NAAC‘A’Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist.Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2018-2022 respectively.

SK.Charishma Banupursuing B.Tech in the department of Computer Science & Engineering from Qis college of Engineeringand Technology (Autonomous &NAAC‘A’Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist.Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2018-2022 respectively.