RESEARCH ARTICLE OPEN ACCESS

Predicting Cyberbullying on Social Media

Durga Bhavani[1],Nithya Konda[2], BhanuSree[3] ,Gowthami[4]
1 Department of Information Technology,
Malla Reddy Engineering College for Women(UGC-Autonomous),
Hyderabad, India
*Email:*bhavani.bsr4@gmail.com

2 Department of Information Technology,
Malla Reddy Engineering College for Women(UGC- Autonomous),
Hyderabad, India
Email : nithyakonda5902@gmail.com

3 Department of Information Technology,
Malla Reddy Engineering College for Women(UGC- Autonomous),
Hyderabad, India
Email : bhanusreelingoju@gmail.com
3 Department of Information Technology,
Malla Reddy Engineering College for Women(UGC- Autonomous),
Hyderabad, India
Email : lingutlagowthami259 @gmail.com

Abstract:

Prior to the innovation of information communication technologies (ICT), social interactions evolved within small cultural boundaries such as geo spatial locations. The recent developments of communication technologies have considerably transcended the temporal and spatial limitations of traditional communications. These social technologies have created a revolution in user-generated information, online human networks, and rich human behaviour-related data. However, the misuse of social technologies such as social media (SM) platforms, has introduced a new form of aggression and violence that occurs exclusively online. A new means of demonstrating aggressive behaviour in SM websites are highlighted in this paper. The motivations for the construction of prediction models to fight aggressive behaviour in SM are also outlined. We comprehensively review cyberbullying prediction models and identify the main issues related to the construction of cyberbullying prediction models in SM. This paper provides insights on the overall process for cyberbullying detection and most importantly overviews the methodology. Though data collection and feature engineering process has been elaborated, yet most of the emphasis is on feature selection algorithms and then using various machine learning algorithms for prediction of cyberbullying behaviours. Finally, the issues and challenges have been highlighted as well, which present new research directions for researchers to explore.

I. INTRODUCTION

Machine or deep learning algorithms help researchers understand big data [1]. Abundant information on humans and their societies can be obtained in this big data era, but this acquisition was previously impossible [2]. One of the main sources of human-related data is social media (SM). By applying machine learning algorithms to SM data, we can exploit historical data to predict the future of a wide range of applications. Machine learning algorithms provide an opportunity to effectively predict and detect negative forms of human behaviour, such as cyberbullying [3]. Big data analysis can uncover hidden knowledge through deep learning from raw data [1]. Big data analytics has improved several applications,
and forecasting the future has even become possible through the combination of big data and machine learning algorithms [4].

An insightful analysis of data on human behaviour and interaction to detect and restrain aggressive behaviour involves multifaceted angles and aspects and the merging of theorems and techniques from multidisciplinary and interdisciplinary fields.

The accessibility of large-scale data produces new research questions, novel computational methods, interdisciplinary approaches, and outstanding opportunities to discover several vital inquiries quantitatively. However, using traditional methods (statistical methods) in this context is challenging in terms of scale and accuracy. These methods are commonly based on organized data on human behavior and small-scale human networks (traditional social networks). Applying these methods to large online social networks (OSNs) in terms of scale and extent causes several issues. On the one hand, the explosive growth of OSNs enhances and disseminates aggressive forms of behaviour by providing platforms and networks to commit and propagate such behaviour. On the other hand, OSNs offer important data for exploring human behaviour and interaction at a large scale, and these data can be used by researchers to develop effective methods of detecting and restraining misbehaviour and/or aggressive behaviour. OSNs provide criminals with tools to perform aggressive actions and networks to commit misconduct. Therefore, methods that address both aspects (content and network) should be optimized to detect and restrain aggressive behaviour in complex systems.

II. EXISTING SYSTEM

State-of-the-art research has developed features to improve the performance of cyberbullying prediction. For example, a lexical syntactic feature has been proposed to deal with the prediction of offensive language; this method is better than traditional learning-based approaches in terms of precision. Dadvar *et al.* examined gender information from profile information and developed a gender-based approach for cyberbullying prediction by using datasets from Myspace as a basis. The gender feature was selected to improve the discrimination capability of a classifier. Age and gender were included as features in other studies, but these features are limited to the information provided by users in their online profiles.

Several studies focused on cyberbullying prediction based on profane words as a feature. Similarly, a lexicon of profane words was constructed to indicate bullying, and these words were used as features for input to machine learning algorithms. Using profane words as

features demonstrates a significant improvement in model performance. For example, the number of ``bad'' words and the density of ``bad'' words were proposed as features for input to machine learning in a previous work.The study concluded that the percentage of ``bad'' words in a text is indicative of cyberbullying. Another research expanded a list of pre-defined profane words and allocated different weights to create bullying features. These features were concatenated with bag-of-words and latent semantic features and used as a feature input for a machine learning algorithm.

III . PROPOSED SYSTEM

- The proposed system is constructing cyberbullying prediction models is to use a text classification approach that involves the construction of machine learning classifiers from labeled text instances. Another means is to use a lexicon-based model that involves computing orientation for a document from the semantic orientation of words or phrases in the document. Generally, the lexicon in lexicon-based models can be constructed manually or automatically by using seed words to expand the list of words. However, cyberbullying prediction using the lexicon-based approach is rare in literature.
- The primary reason is that the texts on SM websites are written in an unstructured manner, thus making it difficult for the lexicon-based approach to detect cyberbullying based only on lexicons. However, lexicons are used to extract features, which are often utilized as inputs to machine learning algorithms. For example, lexicon based approaches, such as using a profane-based dictionary to detect the number of profane words in a post, are adopted as profane features to machine learning models. The key to effective cyberbullying prediction is to have a set of features that are extracted and engineered.

IV.CONCLUSION

This study reviewed existing literature to detect aggressive behaviour on SM websites by using machine learning approaches. We specifically reviewed four aspects of detecting cyberbullying messages by using machine learning approaches, namely, data collection, feature engineering, construction of cyberbullying detection model, and evaluation of constructed cyberbullying detection models. Several types of discriminative features that were used to detect cyberbullying in online social networking sites were also summarized. In addition, the most effective supervised machine learning classifiers for classifying cyberbullying messages in online social networking sites were identified. One of the main contributions of current paper is the definition of evaluation metrics to successfully identify the significant parameter so the various machine learning algorithms can be evaluated against each other. Most importantly we summarized and identified the important factors for detecting cyberbullying through machine learning techniques specially supervised learning. For this purpose, we have used accuracy, precision recall and f-measure which gives us the area under the curve function for modelling the behaviours in cyberbullying. Finally, the main issues and open research challenges were described and discussed.

VI. REFERENCES

[1] V. Subrahmanian and S. Kumar, ''Predicting human behavior: The next frontiers,'' Science, vol. 355, no. 6324, p. 489, 2017.

[2] H. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, ''Homophily in the digital world: A LiveJournal case study,'' IEEE Internet Comput., vol. 14, no. 2, pp. 15–23, Mar./Apr. 2010.

[3] M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana, ''Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network,'' Comput. Hum. Behav., vol. 63, pp. 433–443, Oct. 2016.

[4] L. Phillips, C. Dowling, K. Shaffer, N. Hodas, and S. Volkova, ''Using social media to predict the

future: A systematic literature review,'' 2017, arXiv:1706.06134. [Online]. Available: https://arxiv.org/abs/1706.06134

[5] H. Quan, J. Wu, and Y. Shi, ''Online social networks & social network services: A technical survey,'' in Pervasive Communication Handbook. Boca Raton, FL, USA: CRC Press, 2011, p. 4. [6] J. K. Peterson and J. Densley, ''Is social media a gang? Toward a selection, facilitation, or enhancement explanation of cyber violence,'' Aggression Violent Behav., 2016. [7] BBC. (2012). Huge Rise in Social Media. [Online]. Available: http://www.bbc.com/news/uk-20851797 [8] P. A. Watters and N. Phair, ''Detecting illicit drugs on social media using automated social media intelligence analysis (ASMIA),'' in Cyberspace Safety and Security. Berlin, Germany: Springer, 2012, pp. 66–76. [9] M. Fire, R. Goldschmidt, and Y. Elovici, ''Online social networks: Threats and solutions,'' IEEE Commun. Surveys Tuts., vol. 16, no. 4, pp. 2019–2036, 4th Quart., 2014. [10] N. M. Shekokar and K. B. Kansara, ''Security against sybil attack in social network,'' in Proc. Int. Conf. InSf. Commun. Embedded Syst. (ICICES), 2016, pp. 1–5.