Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, September : 2023

A HYBRID DEEP LEARNING APPROACH FOR DETECTING CYBER BULLYING IN THE TWITTER SOCIAL MEDIA PLATFORM

#1SANNAPURI GAYATHRIRANI,
#2Dr.V.BAPUJI, ***Associate Professor& HOD,***
Department of Master of Computer Applications,
VAAGESWARI COLLEGE OF ENGINEERING, KARIMNAGAR, TELANGANA

ABSTRACT: Cyberbullying (CB) is becoming more common in online entertainment situations. Given the popularity of social media and its widespread use by people of all ages, it is critical to keep the platforms safe from cyberbullying. DEA-RNN, a hybrid deep learning model for CB identification on Twitter, is introduced in this study. The proposed DEA-RNN model combines an improved Dolphin Echolocation Algorithm (DEA) with Elman-type recurrent neural networks (RNNs) to reduce training time and fine-tune the Elman RNNs' parameters. Using a dataset of 10,000 tweets, we fully evaluated DEA-RNN and compared its performance to that of cutting-edge algorithms such as Bi-LSTM, RNN, SVM, Multinomial Naive Bayes (MNB), and Random Forests (RF). The results of the experiments demonstrate that DEA-RNN was superior in every situation. In terms of detecting CB on the Twitter site, it outperformed previously considered strategies. In scenario 3, DEA-RNN fared better, with an average accuracy of 90.45%, precision of 89.52, recall of 88.98, F1-score of 89.25, and specificity of 90.94%.

Index terms: cyber bullying, social media, Recurrent Neural Network, Deep Learning.

1. INTRODUCTION

The most common locations for people of all ages to interact online are social media platforms like Facebook, Twitter, Flickr, and Instagram. In addition to facilitating hitherto impossible types of communication and connection, these platforms have facilitated negative phenomena like stalking. Cyberbullying is a sort of psychological abuse with far-reaching implications for our culture. Young individuals who spend a lot of time switching between different social media sites are particularly vulnerable to cyberbullying. Because of the widespread usage of social media sites like Twitter and Facebook and the anonymity they provide, these platforms are particularly susceptible to CB. Facebook and Twitter account for 14% of all abuse in India, with 37% of that coming from teenagers [1]. When it comes to your mental health, cyberbullying may be harmful and could lead to more severe issues. Anxiety, sadness, stress, and social and emotional problems associated with cyberbullying are major contributors to suicide. Therefore, it is important to have a system in place for identifying instances of cyberbullying in online content such as posts, tweets, and comments.

How Twitter identifies abusive content is the primary topic of this piece. Finding instances of cyberbullying in tweets and taking preventative measures are crucial given the growing prevalence of the issue on Twitter [5]. As a result, there's an increasing demand for research on cyberbullying on social networks to better understand the issue and provide solutions [6]. Handling trolling on Twitter is a full-time job in and of itself [7]. Furthermore, it is laborious to go through social media posts in search of cyberbullying. For instance, tweets tend to be succinct, rife with slang, and peppered with emojis and gifs. Therefore, it is not possible to infer someone's motivations or values from their social media activity alone. Covert forms of bullying, such as passive aggression or sarcasm, can sometimes be difficult to identify. Despite the fact that cyberbullying can be difficult to detect

because of the nature of social media messaging, research into the topic is open and flourishing. While some have employed topic modeling methods, the vast majority of research looking for abuse on Twitter have concentrated on classifying tweets. To distinguish abusive tweets from others, supervised machine learning (ML) text classification models are frequently employed [8]_[17]. Algorithms based on deep learning (DL) have also been used to distinguish between bullying and non-bullying tweets [7, 18, 22, and 22]. Unchangeable class names that don't adapt to new occurrences are problematic for supervised classifiers [23]. Extracting the most salient topics from a dataset in order to discover underlying patterns or classes has long been a goal of topic modeling techniques. It may also be limited in the quantity of events it can process if tweets that are updated in real time pose problems. Even if the principles are identical, general, unsupervised topic models struggle with short texts. Unsupervised topic models for short texts [24] were developed to circumvent this issue. These models extract recurring ideas from tweets for further processing. These models can be used to extract relevant information through two-way processing. These unsupervised models require extensive training, which is not necessarily sufficient to extract substantial historical knowledge [25]. With these restrictions in mind, it's important to develop a robust method for classifying tweets in order to avoid any issues between the classifier and the topic model and guarantee generally good adaptability.

2. LITERATURE SURVEY

Proceedings of the 4th International Conference on Behavioral, Economic, and SocioCultural Computing (BESC 2017), 2018-January, "Towards the detection of cyberbullying using social network mining techniques," doi: 10.1109/BESC.2017.825640.

People who use the Internet have developed a greater yearning to express themselves and connect with others in recent years. It would appear, however, that social media users are being targeted for exploitation. One of the worst things you can do on the internet is engage in cyberbullying, which is a serious social issue. Keeping this in mind and as motivation, developing methods to identify abuse on social media can contribute to putting a stop to it. To investigate cyberbullying, we employ techniques from data mining and social network analysis. Phrase matching, opinion mining, and social network analysis are the three key components of this strategy that will be examined. The experimental strategy for measuring results will also be discussed with the proposal.

P. Galán-Garca, J. G. de la Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, "Supervised machine learning for the detection of troll profiles in the Twitter social network: Application to a real case of cyberbullying," 2014.

Because of the proliferation of new resources and the rise in prominence of social networks, users are now able to maintain their anonymity online. It is impossible to verify the authenticity of a page since a false identity might be created that has no relation to the actual user. As a workaround, some users create aliases or alter their profiles so that they no longer reflect who they actually are. They then publish articles, critiques, or media with the intent of attacking or slandering others, who may or may not be aware of the content. Real-world deaths could result from virtual attacks because of the potential impact on the victims' environments. In this study, we demonstrate how to identify and connect bot accounts on Twitter that distribute misinformation with their legitimate counterparts. We do this by considering what each character has to say about the other. We also provide an example of how this strategy was implemented successfully to address cyberbullying in a real elementary school.

A. Mangaonkar, A. Hayrapetian, and R. Raje, "Collaborative detection of cyberbullying behavior in Twitter data," 2015, doi:

10.1109/EIT.2015.7293405,
http://dx.doi.org/10.11109/EIT.2015.7293405.
Twitter users are getting into more and more trouble as a result of having access to more and more information. Cyberbullying is one of these actions that can have devastating consequences. This highlights the significance of monitoring Twitter, ideally in real time, for instances of cyberbullying. The most prevalent approaches to detecting cyberbullying require considerable individual effort and time. In order to enhance detection in this study, we use concepts from collaborative computing. Several methods of cooperative effort are demonstrated and discussed in this research.

Early findings indicate that the detection method outperforms the standalone model in terms of speed and accuracy.

3. PROPOSED SYSTEM

In this paper, we discuss DEA-RNN, a technique that uses a combination of deep learning and neural networks to detect abusive language in tweets automatically. The DEA-RNN technique allows for the Elman RNN's parameters to be fine-tuned by combining that network with a refined version of the Dolphin Echolocation Algorithm (DEA). Short texts and subject models evolve over time, but DEA-RNN can swiftly identify emerging topics. When compared to other methods for detecting cyberbullying on Twitter, DEA-RNN performed the best. This held true over a wide range of criteria. A brief rundown of the page's contents follows.

Create a more robust DEA optimization model for automatically adjusting RNN parameters to maximize performance. The DEA-RNN technique, which combines the Elman-type RNN and the modernized DEA, provides the highest accuracy when classifying tweets. So that DEA-RNN and other approaches may be compared, a new Twitter dataset is created using cyberbullying-related keywords. Extensive experimental data demonstrates that the DEA-RNN beats competing models in recognizing and categorizing tweets, including cyberbullying tweets, in terms of recall, precision, accuracy, F1 score, and specificity.

4. IMPLEMENTTAION

Service Provider

The Service Provider must supply a valid username and password in order to access this section. After login in, he will be able to access the following features: data sets for training and testing tweets; expected cyberbullying detection type; cyberbullying detection type ratio; forecasted data sets; results for cyberbullying detection ratio; and a list of all remote users. A bar chart depicting the Trained and Tested Accuracy of Tweet Datasets.

View and Authorize Users

An administrator will have access to a list of participants in this module. Here, the administrator can check details such a user's name, email, and physical location, as well as grant access to the site.

Remote User

This module contains n participants. The first step in doing anything involves registering. Information provided by users during registration is stored in a database. He can then enter his username and password to access his account. Once logged in, users have the option to view their profiles, do cyberbullying type predictions, or register and login.

5. RESULTS AND DISCUSION

6. CONCLUSION

In order to improve topic models' ability to detect cyberbullying, this study aimed to develop a reliable method for categorizing tweets. To improve the effectiveness of boundary adjustment, the DEA RNN was developed by combining the DEA enhancement and the Elman-type RNN. The Bi-LSTM, RNN, SVM, RF, and MNB techniques were also used to a newly created Twitter dataset that had been cleaned up using CB catchphrases. In terms of accuracy, recall, precision, and specificity, the DEA-RNN outperformed all other approaches in experimental studies. This demonstrates that DEA influences the way RNN is displayed. However, the DEA-RNN model's compatibility with additional data decreases after the initial input, which makes the hybrid recommended model less desirable. This study only analyzed data from Twitter, thus future research into cyberbullying should expand to include additional SMPs such as Facebook, Instagram, Flickr, and YouTube. The use of many data sets to detect cyberbullying is an area that will be investigated in the future. We also ignored Twitter users' habits and instead focused on what they really said in their tweets. This will be reflected in future creations. The proposed approach detects cyberbullying by analyzing tweets' text, but other media types, such as images, videos, and audio, are currently under investigation and may be investigated in the future. Additionally, a real-time feed of CB texts that can be searched and sorted would be ideal.

REFRENCES

[1] F. Mishna, M. Khoury-Kassabri, T. Gadalla, and J. Daciuk, ``Risk factors for involvement incyber bullying: Victims, bullies and bully_victims," *Children Youth Services Rev.*, vol. 34, no. 1,pp. 63_70,Jan. 2012, doi: 10.1016/j.childyouth.2011.08.032.

[2] K. Miller, ``Cyberbullying and its consequences: How cyberbullying is contorting the mindsofvictimsandbulliesalike,andthelaw'slimited availableredress,"*SouthernCaliforniaInterdiscipl.Law J.*, vol. 26, no. 2,p. 379, 2016.

[3] A. M. Vivolo-Kantor, B. N. Martell, K. M. Holland, and R. Westby, ``A systematic reviewand content analysis of bullying and cyber-bullying measurement strategies," *Aggression ViolentBehav.*,vol. 19, no. 4,pp. 423_434, Jul. 2014, doi:10.1016/j.avb.2014.06.008.

[4] H. Sampasa-Kanyinga, P. Roumeliotis, and H. Xu, ``Associations between cyberbullying andschoolbullyingvictimizationandsuicidalideation,plansandattemptsamongCanadianschoolchildren,"*PLoSONE*,vol.9,no.7,Jul.2014,Art.no.e102145,doi:10.1371/journal.pone.0102145.

[5] M.Dadvar,D.Trieschnigg,R.Ordelman,andF.deJong,``Improvingcyberbullyingdetectionwithusercontext," in*Proc.Eur.Conf.Inf.Retr.*,inLecture NotesinComputerScience:IncludingSubseriesLectureNotesinArti_cialIntelligenceandLectureNotesinBioinformatics,vol. 7814,2013, pp. 693_696.

[6] A. S. Srinath, H. Johnson, G. G. Dagher, and M. Long, ``BullyNet: Unmasking cyberbullieson social networks," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 2, pp. 332_344, Apr. 2021,doi:10.1109/TCSS.2021.3049232.

[7] A.Agarwal,A.S.Chivukula,M.H.Bhuyan,T.Jan,B.Narayan,andM.Prasad,