A Hybrid Deep Learning Approach for Detecting Cyberbullying in the Twitter Social Media Platform

H Yugundhar¹
Dr R Maruthumuthu, Ph.D²

Master of Computer Applications, Madanapalle Institute of Technology and Science, India

ABSTRACT

Cyberbullying (CB) has become progressively common in web-based entertainment stages. It is essential to make social media platforms safer from cyberbullying given the widespread use of social media by people of all ages and their popularity. To detect CB on Twitter, a hybrid deep learning model known as DEA-RNN is presented in this paper. The proposed DEA-RNN model combines an optimized 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 ten thousand tweets, we thoroughly evaluated DEA-RNN and compared its performance to that of cutting-edge algorithms like Bi-LSTM, RNN, SVM, Multinomial Naive Bayes (MNB), and Random Forests (RF). In every scenario, the experimental results demonstrate that DEA-RNN was superior. In terms of detecting CB on the Twitter platform, it performed better than the considered existing methods. In scenario 3, DEA-RNN performed better, achieving an average accuracy of 90.45 percent, precision of 89.52 percent, recall of 88.98 percent, F1-score of 89.25 percent, and specificity of 90.94%..

1. INTRODUCTION

For people of all ages, social media platforms like Facebook, Twitter, Flickr, and Instagram have emerged as the preferred online platforms for interaction and socialization. While these platforms make it possible for people to communicate and interact in ways that were previously unimaginable, they have also contributed to harmful behaviors like cyberbullying. A form of psychological abuse known as cyberbullying has a significant impact on society. Events involving cyberbullying have been on the rise, particularly among young people who spend most of their time hopping between social media platforms. Due to their popularity and the anonymity afforded to abusers by the Internet, social media networks like Twitter and Facebook are particularly susceptible to CB. For instance, in India, 14% of all harassment takes place on Facebook and Twitter, with 37% of these incidents involving children [1]. In addition, cyberbullying may result in negative effects on mental health as well as serious mental health issues. Anxiety, depression, stress, and social and emotional difficulties brought on by cyberbullying events account for the majority of suicides [2][4]. This necessitates a method for spotting instances of cyberbullying in social media posts, tweets, and comments.
This article focuses primarily on the issue of Twitter's detection of cyberbullying. The primary tasks in combating cyber bullying threats are the detection of cyber bullying events from tweets and the provision of preventative measures, as cyber bullying is becoming a prevalent issue on Twitter [5]. As a result, there is a greater need to expand research on cyberbullying that is based on social networks in order to gain more insight and assist in the creation of efficient tools and strategies to effectively combat the issue [6]. It is virtually impossible to manually monitor and control cyberbullying on the Twitter platform [7]. Additionally, mining social media messages for the purpose of detecting cyberbullying is quite challenging. Twitter messages, for instance, are frequently brief, full of slang, and may include emojis and gifs. As a result, it is impossible to ascertain the intentions and meanings of individuals solely from social media messages. In addition, bullying can be difficult to spot if the bully conceals it with tactics like sarcasm or passive aggression. Cyberbullying detection on social media is an open and active research topic, despite the difficulties presented by messages on social media. The majority of efforts to detect cyberbullying on the Twitter platform have focused on tweet classification, with some use of topic modeling techniques. Text classification that is based on supervised machine learning (ML) models is frequently used to divide tweets into those that are about bullying and those that are not [8][17]. Classifiers based on deep learning (DL) have also been used to divide tweets into bullying and non-bullying tweets [7, 18] and [22]. If the class labels cannot be changed and are not relevant to the new events, supervised classifiers perform poorly [23]. Topic modeling approaches have long been used as a means of extracting the most important topics from a set of data in order to form the patterns or classes in the entire dataset. In addition, it may only be suitable for a predetermined collection of events, but it is unable to successfully handle tweets that change on the _y. Although the idea is similar, general unsupervised topic models are not effective for short texts. As a result, specialized unsupervised topic models for short texts were used [24]. The trending topics in tweets are effectively identified and extracted by these models for further processing. Using these models, meaningful topics can be extracted by utilizing bidirectional processing. However, in order to acquire sufficient prior knowledge, these unsupervised models require extensive training, which is not always sufficient [25]. Taking into account these limits, an effective tweet characterization approach should be created to overcome any barrier between the classifier and the subject model with the goal that the flexibility is fundamentally capable.

2. LITERATURE SURVEY

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.


Abstract:
The use of new technologies along with the popularity of social networks has given the power of anonymity to the users. The ability to create an alter-ego with no relation to the actual user, creates a situation in which no one can certify the match between a profile and a real person. This problem generates situations, repeated daily, in which users with fake accounts, or at least not related to their real identity, publish news, reviews or multimedia material trying to discredit or attack other people who may or may not be aware of the attack. These acts can have great impact on the affected victims’ environment generating situations in which virtual attacks escalate into fatal consequences in real life. In this paper, we present a methodology to detect and associate fake profiles on Twitter social network which are employed for defamatory activities to a real profile within the same network by analysing the content of comments generated by both profiles. Accompanying this approach we also present a successful real life use case in which this methodology was applied to detect and stop a cyberbullying situation in a real elementary school.


As the size of Twitter© data is increasing, so are undesirable behaviors of its users. One of such undesirable behavior is cyberbullying, which may even lead to catastrophic consequences. Hence, it is critical to efficiently detect cyberbullying behavior by analyzing tweets, if possible in realtime. Prevalent approaches to identify cyberbullying are mainly stand-alone and thus, are time-consuming. This research improves detection task using the principles of collaborative computing. Different collaborative paradigms are suggested and discussed in this paper.
Preliminary results indicate an improvement in time and accuracy of the detection mechanism over the stand-alone paradigm.

3. PROPOSED SYSTEM

In this research, we present DEA-RNN, a hybrid deep learning-based approach for automatically detecting bullying in tweets. The DEA-RNN approach combines Elman type Recurrent Neural Networks (RNN) with an upgraded Dolphin Echolocation Algorithm (DEA) for fine tweaking the parameters of the Elman RNN. DEA-RNN can deal with the dynamic nature of brief texts and topic models for effective extraction of trending topics. In all circumstances and with multiple assessment metrics, DEA-RNN beat the examined existing techniques in detecting cyberbullying on the Twitter network. This article's contributions can be summarised as follows:

Create an improved DEA optimisation model for automatically tuning RNN parameters to improve performance; Propose DEA-RNN by combining the Elman type RNN and the improved DEA for optimal tweet classification; A new Twitter dataset based on cyberbullying keywords is collected to evaluate the performance of DEA-RNN and the existing methods; The efficiency of DEA-RNN in recognising and classifying cyberbullying tweets is evaluated using According to the extensive experimental data, DEA-RNN beats other competing models in terms of recall, precision, accuracy, F1 score, and specificity.

3.1 IMPLEMENTATION

Service Provider
In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train & Test Tweet Data Sets, View Tweet Datasets Trained and Tested Accuracy in Bar Chart, View Tweet Datasets Trained and Tested Accuracy Results, View Predicted Cyber bullying Detection Type, Find Cyber bullying Detection Type Ratio, Download Predicted Data Sets, View Cyber bullying Detection Ratio Results, View All Remote Users.

View and Authorize Users
In this module, the admin can view the list of users who all registered. In this, the admin can view the user’s details such as, user name, email, address and admin authorizes the users.

Remote User
In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT CYBERBULLYING TYPE, VIEW YOUR PROFILE.
5. CONCLUSION

To improve the efficiency of topic models for the detection of cyber-bullying events, this paper developed an effective tweet classification model. DEA RNN was created by consolidating both the DEA enhancement and the Elman type RNN for proficient boundary tuning. Besides, it was tried in correlation with the current Bi-LSTM, RNN, SVM, RF, and MNB strategies on a recently made Twitter dataset, which was removed utilizing CB catchphrases. In all scenarios, the experimental analysis revealed that the DEA-RNN outperformed the other currently available approaches in terms of accuracy, recall, precision, and specificity. This implies the effect of DEA on the presentation of RNN. The feature compatibility of the DEA-RNN model decreases when the input data are increased beyond the initial input, despite the fact that the hybrid proposed model achieved higher performance rates than the other models that were taken into consideration. The Twitter dataset was the sole focus of the current study; other Social Media Platforms (SMP), such as Facebook, Instagram, Flickr, and YouTube, should be looked into to find the pattern of cyberbullying. The use of data from multiple sources for cyberbullying detection will then be looked into in the future. In addition, we restricted our analysis to tweets' contents; We were unable to conduct the users' behavior-based analysis. Future works will include this. While other types of media, such as images, videos, and audio, remain an open research area and potential future research directions, the proposed model works to detect cyber bullying using tweet text. In addition, our objective is to classify and locate CB tweets in a live stream.

6. REFERENCES


