7th National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(1st and 2nd March 2019)
International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

Cyber-Bullying Detection Techniques- A Literature Survey

Vijay Banerjee[1], Pooja Gaikwad[2], Jui Telavane[3], Pallavi Vartak[4]

Department of Computer Engineering, Mumbai University, India

Abstract**—***As the computerized age gets increasingly progressed and innovation turns out to be all the more effectively open, new issues start to emerge with this expanded utilization of innovation.Cyberbullying is one of such issues. This paper presents the survey of main approaches for detection of cyberbullying across various social media platforms. Cyberbullying is the type of harassment that is done through electronic methods via web-based networking media stages, for example, twitter, Facebook, WhatsApp and so forth. This incorporates posting bits of gossip, dangers, an injured individual's close to home data, sexual comments or pejorative names, for example, despise, discourse and so on. Because of continued harassing the unfortunate casualty may have lower confidence, assortment of passionate reactions, for example, being terrified, bothered, outrage and sadness, tension, dejection, somatic symptoms and in worst cases the suicidal ideation etc. Thus, there is a need for the detection of cyberbullying across multiple social media platforms so that certain actions could be taken to reduce cyberbullying.*

Keywords**— *Cyberbullying, Participant Vocabulary Consistency, Deep Learning, Neural Network*.**

I. INTRODUCTION

With the quick development of innovation, the utilization of internet and web-based life increments. Alongside this progression there emerges the issue of Cyberbullying which is getting to be oppress. To take care of this issue cyberbullying discovery is essential. Because of the absence of mindfulness about the severity of Cyberbullying rates occurring on the online life stages and the impacts of cyberbullying on the victim

There are different procedures that can be utilized for the location of swear words in talks and remarks on the internet-based life. Profound learning, Machine Learning and Data Mining are a portion of the methodologies utilized for this reason. There is a requirement for a framework that can procedure more measure of information crosswise over social medias with equivalent exactness concerning less information. There additionally felts a requirement for understanding the intensions of the assailants behind the cyberbullying endeavor, for example, individual, social, bigot, Sexist and so forth.

The paper is a study of different cyberbullying location systems that exist today. It additionally centers around study of couple of neural system-based procedures. An unmistakable examination of the considerable number of techniques in done so as to locate the perfect methodology for the equivalent.

II. LITERATURE SURVEY

2.1 Cyber-bullying detection based on semantic-enhanced marginalized denoising auto-encoders [1].

In this paper, Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA) is created by means of semantic expansion of the well-known profound learning model stacked denoisingautoencoder. The semantic augmentation comprises of semantic dropout commotion and sparsity imperatives, where the semantic dropout clamor is structured dependent on space information and the word implanting system.

This paper learns the denoising and autoencoding along these lines ended up being helpful for the more proficient portrayal of the information.

2.2 Weakly supervised cyber -bullying detection with participant vocabulary consistency [2].

This paper proposes a pitifully regulated machine learning technique for at the same time deriving client jobs in harassment-based tormenting and new vocabulary markers of tormenting. The learning calculation considers social structure and deduces which clients will in general domineering jerk and which will in general be victimized. The model evaluations whether every

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

social communication is tormenting dependent on who takes an interest and dependent on wahat language is utilized, and it endeavors to amplify the understanding between these appraisals, i.e., participant-vocabulary consistency (PVC).

Through this paper the job of PVC is examined and the data identified with the recognition of the harassing jobs of client is found out.

2.3 Attention- based Bi-Directional Long Short-Term memory network for relation classification [3].

In this paper the first step is the tokenization and then word2vec tool is provided by Google. All these vectors are fed to CNN (Convolution Neural Network) with help of word vector embedding. Output of CNN is given to LSTM (Long Short-Term Memory) layers.

The system has introduced a novel model for deep sentiment representation by skillfully combining one-layer CNN and two-layer LSTMs.

2.4 Very Deep CNN for text classification [4].

This paper introduces another engineering for content preparing which works specifically at the character level what's more, utilizes just little convolutions and pooling activities. The paper demonstrates that the execution of this model increments with the profundity sing up to 29 convolutional layers, and report enhancements over the cutting edge ona few open content classification errands This paper comprehends the utilization of profound convolution systems for the content classification and its points of interest over RNN and LSTM

2.5 Comparative analysis of different word embedding models [5].

This paper presents a comparative analysis of different word embedding models namely continuous bag of words, Skip gram, Gloveand Hellinger PCA (Principal Component Analysis). The models are compared on different parameters.

Thus, this paper proves useful to understand the benefits of various word embedding models and the comparison between them so as to select the best model.

2.6 Cyberbullying Detection with Weakly Supervised Machine Learning [6].

This paper presents the participant-vocabulary consistent model a weakly supervised approach for simultaneously learning the roles of social media users in the harassment form of cyberbullying and the tendency of language indicators to be used in such cyberbullying. It evaluates PVC on all social Medias platform data sets with both quantitative and qualitative analysis.

This model requires no extra space beyond the storage of vectors and raw data. Thus, this paper helps in studying the PVC technique.

2.7 CNN comparator: comparative analytics of CNN [7].

This paper uses a visualization technique that has 4 linked view that's helps to analyze learning parameters. These study uses AlexNet as the neural network architecture. All the 4-view combined help to understand the insight of CNN clearly so that we can improve the training process

As training process of CNN leads to large number of parameters over time, this results in decreased performance. This paper helps to view the learning process of the CNN.

2.8 Fast deep neural network based on intelligent dropout and layer skipping [8].

This paper is based onrapid way to compute the feature using fast beta wavelet transform.The intelligent dropout methodis based on a unit is efficiency and not randomly selectedit is possible to classify the image using efficient unit of earlier layer and skip the all hidden layer from the output layer.

This paper proves that the FWT are the best or it is extracting the feature of input image.

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

2.9 Detection of cyberbullying on social media using data mining techniques [9].

In this paper the research was carried out using data mining technique here there are several stages as data collection, preprocessing, TF-IDF, weighting, data validation and classification using naive Bayes classifier.

This paper proves that TF-IDF weighted and validation data using for the cross validation and then do classification.

2.10 Detecting cyberbullying and aggression in social commentary using NLP and machine learning [10]

In this paper the user experimenting with different work process a robust methodology for extracting text is done.User work in certain ways to identify and classify bullying in the text by analyzing and network-based attributes feature distinguish them from regular user. This paper shows the training in machine learning model using supervised learning.

2.11 Cyber-Bullying Revelation in Twitter Data using Naive Bayes classifier algorithm [11].

This paper use naïve Bayes as the classifier for the content classification in email application it deals with theclassification of spam words when massage is received and it is processed using feature set extraction method inwhich feature probabilities are found using NB and SVM are compared for precision factor. This paper just classifies the massage into cyberbullying. A strong representation and learning of text massage are crucial for consistent detection system. The main methodfor the data extraction is web base mining technology.

2.12 Semi-supervised sequence learning [12].

This paper is based on supervised sequence learning model using CNN and LSTM. The paper recommends the use of LSTM-RNN to be more useful than CNN and RNN for the purpose of data training using the proposed approach. This paper tests the semi-supervised method on five benchmarks to check the results using LSTM as the training method.

This paper proves that CNN-LSTM is the better method than conventional CNN and gives better results than the previous methods for training unlabeled data.

2.13 Multi-Category classification by SoftMax combination of binary classifiers [13].

This paper explains how to efficiently extend binary classification method for multi-category classification. The paper also explains that most common approach to multi-category classification are binary-classifier based methods such as “one-versus-all” and “one-versus-one” that solves the multicategory classification problem. posteriori probabilities are obtained from the combination are used to do multicategory classification.

This paper helps understand the advantages of multicategory classification and also methods to implement that using two methods in detail.

2.14 Tweet sentiment analysis by incorporating sentiment specific word embedding and weighted text features [14].

The method here is proved to be better than the SSWE and NRC techniques. In this the sentiment of tweet is polarized into 3 types. The paper suggests SSWE word embedding algorithm for data representation as it also do sentiment classification. The WTFM has 2 features i.e. negation feature and the tf.idf word weighing scheme.

In the model here (SSWE + WTFM) the four features of WTFM concatenated with SSWE. The SSWE captures the semantic and syntactic feature and using original n-gram polarity of tweets it predicts 2-dimentional vector (f0, f1).

III. ANALYSIS TABLE

Analysis Tableis the summary of the studied research papers on the various methods used for cyberbullying detection. There are also research papers based on the technologies, word embedding techniques, neural network models and some analytical papers.

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

TABLE 1
ANALYSIS TABLE

Title	Techniques/ Methods	Features useful for project	Accuracy
Cyber-bullying detection based on semantic-enhanced marginalized denoising auto-encoders	Word-Embedding, mSDA, smSDA, LSA, LDA, BWM, BoW	Process of word Embedding, word semantics, Nature of bullying.	84.1(Training dataset) {the accuracy is different with respect to different Word-embedding methods.}
Weakly supervised cyber - bullying detection with participant vocabulary consistency.	PVC, Weakly supervised network, BoW.	Gives prospective about how to train data in real scenarios.	0.783 i.e. 78.3% Dropout is not used and overfitting is an issue
Attention- based Bi-Directional Long Short-Term memory network for relation classification	Attention- based Bi-Directional Long Short-Term memory network for relation classification	Attention- based Bi-Directional Long Short-Term memory network for relation classification	83.7% (Max)
Very Deep CNN for text classification	NLP, CNN, LSTM, N-Gram	How CNN works in text-based classification environment.	---
Comparative analysis of different word embedding models	Continuous BoW, Skip-Gram, GloVE, Hellinger PCA.	GloVE	GloVE(max Accuracy)
Cyberbullying Detection with Weakly Supervised Machine Learning	PVC, Weakly supervised Machine Learning	Optimization of objective function	---
CNN comparator: comparative analytics of CNN	CNN, TFlearn framework, Visualization technique	Training process of CNN, Learning patterns	72.79% Accurate
Fast deep neural network based on intelligent dropout and layer skipping.	Convolution neural network, fast wavelet transforms.	This prevent overfitting and reduce the classification time.	85% accurate
detection of cyberbullying on social media using data mining techniques.	Data mining, Naive Bayes Mode	Detection on cyber-bullying on twitter can be done.	75% Accurate
Detecting cyberbullying and aggression in social commentary using NLP and machine learning.	Natural language processing and machine learning.	In cyberbullying involves training a machine learning model using supervised learning.	75%-90% Accurate
Cyber-Bullying Revelation in Twitter Data using Naive Bayes classifier algorithm	Naive-bayes, SVM, stemming, Lemmatization	Precision Factor, Feature extraction	This proves that Naïve Bayes and other traditional machine learning algorithms are insufficient to provide perfect solution for an issue.
Semi-supervised sequence Learning	CNN, LM-LSTM, SA-LSTM	Efficient Data training	75% Accurate
Multi-Category classification by softmax combination of binary classifiers	SoftMax combination	SoftMax Function.	60.73% Accurate
Tweet sentiment analysis by incorporating sentiment specific word embedding and weighted text features	CNN and LSTM word2vec	---	87.2% Accurate

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

IV. CONCLUSION

A proper research and study of various research papers and case studies has been done. Through these researches we can conclude that detection of the cyberbullying incidences can be done in a more ideal manner through the use of deep neural network. Through the regressive analysis of the existing systems for the detection of cyberbullying it can be concluded that machine learning or data mining approaches have many drawbacks such as week classification, Inefficient word embedding, week training ability etc. It limits the correctness of the detection and works only on some limited classified features of cyberbullying. The paper concludes that the neural network approach can give better results for the detection of cyberbullying incidences over machine learning of deep learning approach.

REFERENCES

[1] RuiZhao,Kezhi Mao "CyberBullying Detection based on Semantic – Enhanced Marginalized Denoising Auto-encoders" IEEE Transaction on Affective Computing, 2015.

[2] ElahehRaisi,Bert Huang "Weakly Supervised Cyberbullying Detection with ParticipantVocabulary Consistency" Social Network Analysis and Mining, May 24,2018.

[3] PengZhou,WeiShi,JunTian,ZhenyuQi,BingchenLi,HoungWei,Hao,BoXu "Attention- based Bi-directional Long Short Term Memory Network for Relation Classification" proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,pages 207-212,August 12,2016.

[4] Alexis Conneau,HolgerSchwenk,Yann Le cun "Very Deep CNN for Text Classification" Association for Computational Linguistics, Volume1, pages 1107-1116,7 April 2017.

[5] MS.SnehalBhoir,TusharGhorpade,Vanita Mane "Comparative Analysis of Different Word Embedding Models" IEEE,2017.

[6] ElahehRaisis,Bert Huang "Cyberbullying Detection with Weakly Supervised Machine Learning" International Conference on Advances in Social Networks Analysis and Mining IEEE/ACM,2017.

[7] HaipengZeng,HammadHaleem,XavierPlantaz,NanCao and HuaminQu "CNN Comparator: Comparative Analytics of CNN" arXiv,15 Oct,2017.

[8] VandanaNandaKumar,BinsuC,Kovoor,Sreeja M.U "Cyber-Bullying Revelation in Twitter Data using Naive-Bayes Classifier Algorithm" International Journal of Advanced Research in Computer Science. Volume 9, No. Jan-Feb 2018.

[9] Andrew M.Dal,QuocV.Le "Semi-Supervised Sequence Learning " arXiv,4 Nov 2015.

[10] Kaiob Duan, S.Sathiya Keerthi,Wei Chu,Shirish Krishnaj Shevade and Anu Neow Poo "Multi-Category Classification by Softmax Combination of Binary Classifiers" Department of Computer Science and Automation,Bangalore.

[11] QuanzhiLi,SameenaShah,RuiFang,ArminehNourbakhsh,XiaomoLiu"Tweet Sentiment Analysis by Incorporating Sentiment Specific Word Embedding and Weighted Text Features" International Conference on Web Intelligence IEEE/WIC/ACM,2016.

[12] AsmaEIAdel,RidhaEjbali,MouradZaied,Chokri Ben Amar "Fast Deep Neural Network based on Intelligent Dropout and Layer Skipping" IEEE,2017.

[13] Hariani,Imam Raid "Detection of Cyberbullying on Social Media using Data Mining Techniques" International Journal of Computer Science and Information Security, Vol.15, No.3, March 2017.

[14] KahitizSahay,Harsimran Singh Khaira,PrinceKukreja,Nishchay Shukla "Detecting Cyberbullying and Aggression in Social Commentary using NLP and Machine Learning" International Journal of Engineering Technology Science and Research, ISSN-2394-3386, Volume5, Issue1, January 2018.