2018 Third International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT) 14-15, December 2018

Identification of potential cyber bullying tweets using hybrid approach in sentiment analysis

Akankshi Mody
Student, Computer Department,
D. J. Sanghvi College of Engineering Mumbai.
akankshimody@gmail.com

Shreni Shah
Student, Computer Department,
D. J. Sanghvi College of Engineering Mumbai.
shrenishah96@gmail.com

Reeya Pimple
Student, Computer Department,
D. J. Sanghvi College of Engineering Mumbai.
reeya26@gmail.com

Narendra Shekokar
Computer Department,
D. J. Sanghvi College of Engineering Mumbai.
Narendra.shekokar@djsce.ac.in

***Abstract*— With the rise of the Internet, cyber bullying is becoming more and more widespread. Cyber bullying has resulted in such disastrous consequences that there is a pressing need to detect it. The aim of this study is to do the same by using sentiment analysis. We perform cyber bullying detection using a novice approach on Tweets using Natural Language Processing and Machine Learning techniques. After processing a tweet, it can be flagged down if the tweet is a potential cyber bullying threat.**

Keywords—NLP(Natural Language Processing), Machine Learning.

I. INTRODUCTION

On an average, Twitter has 330 million active users worldwide. It serves as a platform where people express their opinions, expressions and lifestyle. But, sometimes Tweets are targeted negatively at specific individuals causing serious consequences such as anxiety, lowered self-esteem, depression, and even suicide. This phenomenon is essentially cyberbulling and needs to be addressed. A recent study indicated that one in ten parents online claim that their child has been a cyberbullying victim. We suggest the use of Sentiment Analysis to detect such Tweets. Sentiment Analysis refers to the use of Natural Language Processing, computational linguistics, and text analysis to study subjective information.

A tweet typically consists of 140 characters or less. It includes text, emoticons, symbols and URLs. To improve performance, data cleaning and preprocessing steps are executed. These steps include removal of symbols, URLs and stop words, acronym expansion and lemmatization. Emoticons are used widely to express opinions and feelings and cannot be neglected while analyzing the sentiment. Emoticons are typically made up of symbols such as ')',':','='. These emoticons help reinforce the sentiment of the text or indicate irony.

Many methods have been suggested for sentiment analysis such as bag-of-words approach and supervised learning approach using various machine learning classifiers. Since freedom of speech is an inherent right of every individual, it is very important that only the tweets that are extremely negative and are targeted at a specific person are flagged down for removal. Hence, in this paper, we propose a hybrid approach where the results from knowledge based approach are verified with those from machine learning algorithms.

The paper is structured as follows. Section 2 introduces the methods currently used for sentiment analysis and assigning polarity to text. Section 3 explains our proposed solution. Section 4 illustrates our expected outcomes.

II. LITERATURE REVIEW

A. Data Pre-processing

(Agarwal et al.,2011)[3]explain that appropriate text pre-processing methods can hugely affect the performance of classifier. (Saif et al.,2014)[8] checked whether stopword removal affects the performance of the classifiers. They concluded that pre-compiled lists of stopwords can negatively impact the performance of classification. (Zhao et al.,2015)[5] studied how six different preprocessing methods affect the sentiment polarity classification on Twitter. They suggest that removing negation and acronym expansion can improve the classification accuracy.

B. Sentimental Analysis on Emoticons

(Shi Yuan go et al.,2016)[7] They suggest that we should match each emoji to the sentiment each emoticon expresses. As for a micro blog, they count the occurrence of emoticons and check whether an emoticon is positive or negative. If the occurrence number of positive emoticons is more than negative emoticons, the micro blog would be labeled as positive and vice versa. However, if a sentence does not of emojis or there

are equal number of emojis, then they will be sent to a Naive Bayes Classifier. (Alexander Hogenboom et al.) [10] does extensive work on sentiment analysis using emoticons. They build a comprehensive emoticon lexicon for the same.

C. Sentimental Analysis of Text

All of the existing cyberbullying detection approaches use large text bodies wherein Twitter has a maximum limit of only 140 characters. That, along with the fact that Twitter uses its own unique language structure makes it different to mine opinions from. (Go et al., 2009)[2] performed sentiment analysis on twitter. They identified the tweet polarity using emoticons as noisy labels.

D. Classification

(Anastasia Giachanou Go et al.,2016)[6] They analyzed that Lexicon-based methods do not require training data and also they provide as advantage with identifying polarity of a sentence.. Lexicon-based approaches have been largely applied on conventional text but hardly on Twitter based text as it is difficult to process. Twitter consists of data that is not only composed of varied versions of acronyms like 'lol' but also symbols used in emojis and hashtags. This makes Twitter data difficult to be analyzed and is hence less explored for analysis. (Shi Yuan go et al.,2016)[7] They provided with the output as microblogs with four sentiments: angry, sad, disgusted and anxious. For every negative micro-blog, they count the occurrence of the words belonging to each of the four emotions. After that, the negative micro-blog was labeled by that most frequently appearing emotion.

E. Polarity Assignment

(A. Cernian et al., 2015) [1] performs an exhaustive search on product reviews that reveals the meaning of the sentence (by context). They compute the sentence score based on the score of individual words and length of words. (M. Lailiyah et al., 2017) [11] use part-of-speech (POS) to get sentiment score of the sentence.

III. PROPOSED ARCHITECTURE

We propose a hybrid approached for sentiment analysis of the tweets, which will help in identifying the highly negative tweets and their subject, which will aid us in flagging down the cyber bullying tweets. Our framework consists of three main steps, which includes knowledge based sentiment analysis, whose result is then reinforced with machine learning based analysis. The resulting polarity is then aggregated with the polarity obtained from the emoticons, which segregates the tweets into varying levels of positive and negative text. First, we capture the live data from Twitter for building our training dataset (step 1). The data is segmented into emoticons and text, which are then processed separately. While splitting, we consider the emoticons (":)", ":-(","¡3") along with the emojis included in the Unicode list.

Once we have compiled the emoticons lexicon, we assign individual polarity to each emoticon (step 2). Then the overall polarity of all the emoticons in one tweet is calculated through aggregation and is fed back into the network.

The textual data in each tweet is processed separately using a two pronged strategy. The first strategy is the knowledge based approach (step 3). We utilize SentiWordNet for this. It is a lexicon for opinion mining which assigns three sentiment scores: positivity, negativity, objectivity to each word. We consider all three scores to get the unbiased sentiment of each.

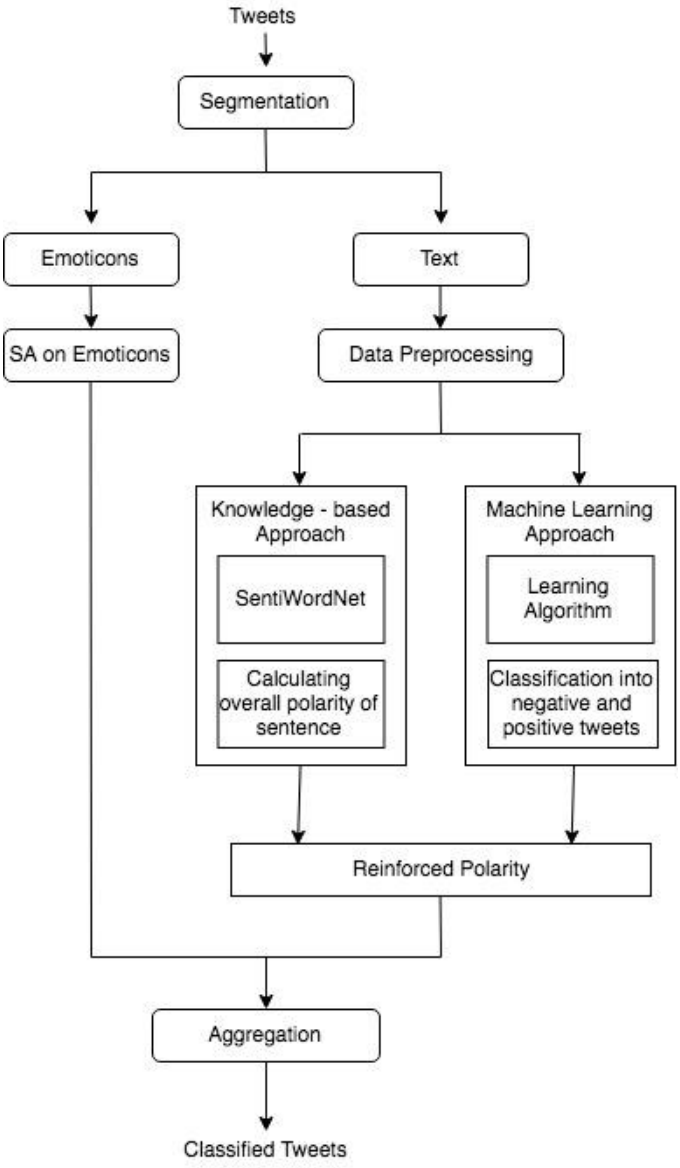

Fig. 1 Architecture Diagram

The same textual data is processed through another algorithm which uses the Machine Learning approach. In this approach,

we make use of Naive Bayes and Linear Support Vector Classifier (SVC) to get the most accurate polarity of the text, in a way that minimizes the false positives. We obtain the dual polarity of the text using two separate strategies (step 4). This helps us reinforce the polarity obtained by one method. The next stage (step 5) is Polarity Aggregation. In this step, we aggregate the polarities obtained using Emoticon Analysis and Sentiment Analysis of Text using a two pronged strategy.

The main aim of this proposed solution is to identify the Cyber Bullying tweets. The highly negative tweets are obtained after Polarity Aggregation. For the final stage (step 5), we use POS Tagging to identify the subject of the tweets. Once we ensure that the highly negative tweets are aimed at a person, the tweets are flagged down as Cyber Bullying Tweets.

IV. RESULTS

We receive three polarities for each sentence: (1) Sentiment from Emoticons Analysis (2) Sentiment from Knowledge Based Approach (3) Sentiment from Machine Learning Approach. This hybrid approach helps us reinforce out results such that we do not violate an individual's freedom of speech.

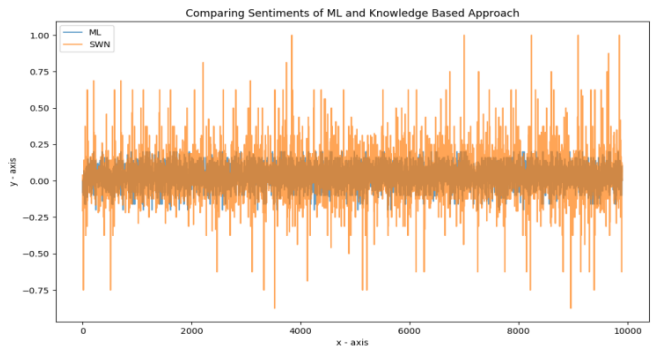

Fig. 2 Comparing Polarities of the Two Approaches

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

In Fig. 2, we can see the polarities obtained by testing the data using two different approaches. Using this data, we can reinforce the polarity to obtain the Confusion Matrix as seen in Table 1. We calculate the accuracy of classification to be 70.3%.

TABLE I CONFUSION MATRIX (WITH NOUNS)

		PREDICTION	
		Positive	***Negative***
ACTUAL	**Positive**	30.3%	14.5%
	Negative	15.1%	40.0%

TABLE II CONFUSION MATRIX (WITHOUT NOUNS)

		PREDICTION	
		Positive	***Negative***
ACTUAL	Positive	21.4%	9.3%
	Negative	24.1%	45.2%

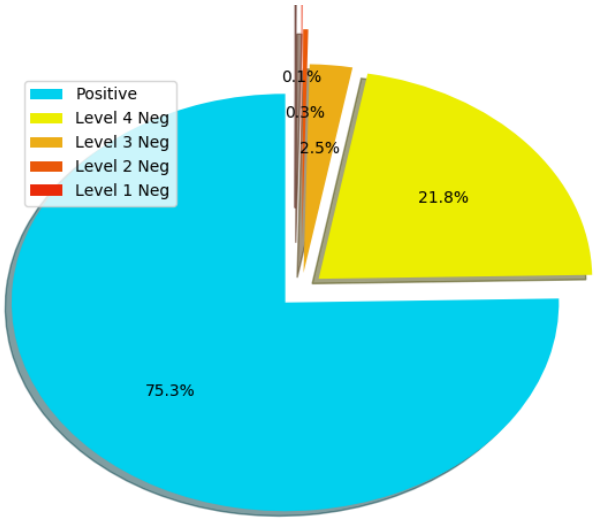

Fig. 3 Final Classification

We expect the overall polarity to help us flag down negative Tweets that are specifically targeted at a person and that are essentially contributing to cyberbullying.

V. FUTURE SCOPE

Social media has been on the rise since the past few years and its growth will only be augmented by internet penetration in all regions of the world. The use of this algorithm can be extended to other social media platforms like Facebook and Instagram, which do not conform to a fixed number of characters, unlike Twitter, which is a microblog. Vernacular languages can also be included once a comprehensive dictionary of their meaning and polarity has been made.

REFERENCES

[1] A. Cernian, V. Sgarciu and B. Martin, "Sentiment analysis from product reviews using SentiWordNet as lexical resource," *2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, Bucharest, 2015, pp. WE-15-WE-18.

[2] Go, A., Bhayani, R. and Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision. In CS224N Project Report, Stanford University.

[3] Agarwal, A., Xie, B., Vovsha, I., Rambow, O. and Passonneau, R. (2011). Sentiment analysis of Twitter data. In Proceedings of the Workshop on Languages in Social Media LSM '11.

[4] Akshat Bakliwal, Piyush Arora, Senthil Madhappan Nikhil Kapre, Mukesh Singh and Vasudeva Varma , Mining Sentiments from Tweets

[5] Zhao Jianqiang, "Pre-processing Boosting Twitter Sentiment Analysis?", , vol. 00, no. , pp. 748-753, 2015, doi:10.1109/SmartCity.2015.158

[6] Anastasia Giachanou and Fabio Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods"

[7] Shi Yuan, Junjie Wu, Lihong Wang, Qing Wang "Hybrid Method for Multi-class Sentiment Analysis of Microblogs",2016

[8] Saif H, Fernandez M, He Y, and Alani H, "On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter," Proc. 9th Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland ,2014,pp.80-817.

[9] Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A. Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. In Proceedings of Conference of the North American Chapter of the Association of Computational Linguistics on Human Language Technologies (NAACL-HLT 2013). The Association for Computational Linguistics, 380– 390.

[10] Alexander Hogenboom et al., "Exploiting Emoticons in Sentiment Analysis",2013

[11] M. Lailiyah, S. Sumpeno and I. K. E. Purnama, "Sentiment analysis of public complaints using lexical resources between Indonesian sentiment lexicon and Sentiwordnet," *2017 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Surabaya, 2017, pp. 307-312.