A survey and proposal on social media cyber hate detection using text post analysis

Monika Soni,

Mtech(0816CS14MT13)

Dr.Kalpana Rai

Professor

Department of Computer Science, Sagar Institute of Research Technology and Science Bhopal Madhya Pradesh India

Abstract: The data mining approaches are used for analyzing the raw data and identifying the application targeted patterns. In this context it is applied for both structured and unstructured format of data, for performing prediction, classification, categorization and finding relationship among data attributes. Thus various nature of data mining algorithms can be used for automated data analysis i.e. classifiers, clustering, soft computing and association rule mining. In this presented work the soft computing techniques and supervised learning techniques are applied on unstructured text data for classifying the cyber hate speech in social media platforms. That becomes a serious issue in today's social networking. This paper provides the overview of the proposed work and their functional aspects. Additionally the implementation strategy of the work is also demonstrated.

Keywords: survey, text mining, social media text analysis, cyber hate detection, data mining model.

I. INTRODUCTION

The data mining techniques are the mathematical model for evaluation and analysis of data. Therefore different kinds of algorithms are available for computing the required features and providing the target patterns, which are able to satisfying the required application. The data which are needed to be analyzed can be structured or unstructured [1]. The structured data is obtained in a specific organization or format. On the other hand the unstructured data is not available in a specific format. The unstructured data may contain the irregular length of feature and data, additionally the internal structure of data is also not followed by similar features. Therefore the classification and categorization of unstructured data is complex process as compared to structured data. Thus processing of text data is a complex nature for classification [2].

The proposed work introduces a new technique for classifying the text according to the sentiments hidden on the expressed text by some author. The technique involves the NLP (natural language processing) techniques for converting text into the POS (part of speech) features [3]. Here for optimizing text post analysis

the POS features the available fuzzy logic is involved. That optimizes the created NLP based features according to their pattern similarity. Additionally during this process it generates temporary class labels for each data patterns [4]. Finally, the SVM (support vector machine) which is a supervised learning classifier involved. That algorithm classifies the complex nature of text.

This section provides the overview of the proposed work. Additionally next section provides the recent contributions for processing text according to the hidden sentiments over the text. In further the basic overview of the used concepts is included and then a model is provided that is used for the proposed system design. Finally future extension of the proposed work is also listed.

II. LITERATURE SURVEY

The proposed work is a soft computing approach for classifying the complex text nature of text. In this context the following essential contributions are providing guidelines to design and develop the required system.

Vimala Balakrishnan et al [5] presents a digital tormenting recognition model dependent on client character. The model intends to perceive harassing designs in Twitter, in light of connections among character and digital tormenting. Arbitrary Forest, calculation was utilized for digital tormenting arrangement, connected related to a benchmark calculation incorporating seven Twitter highlights. Discoveries show that calculating client's character significantly improves digital tormenting recognition. Observed to be critical in distinguishing menaces, accomplishing up to 96% (accuracy) and 95% (review).

Opinion examination is an extremely prevalent application territory of content mining and AI. Be that as it may, techniques for the most part have a place with discriminative learning. With regards to content grouping, occurrences are normally fluffy and in this manner are not viewed as obvious, particularly given the way that mark alloted. Conventional fluffy strategies ordinarily include age of a solitary fluffy classifier and utilize a fixed guideline of defuzzification yielding the class. This procedure is probably going to get the classifier experiencing the content uncertainty circumstance on feeling information. *Han Liu et al [6]* center around digital abhor arrangement, since the spread of detest discourse through online networking. Programmed identification of cyberhate has in this manner turned into a need inquire about region. Specifically, creators propose an altered fluffy methodology with two phase preparing for managing content uncertainty and grouping four sorts of despise discourse, to be specific, religion, race, incapacity, and sexual direction. The exploratory outcomes demonstrate that the proposed fluffy strategy beats different techniques much of the time.

A noteworthy piece of their work concentrated on assessment examination and feeling mining. Multi-class slant investigation, addresses the recognizable proof of the careful supposition

passed on by the client as opposed to in general assessment extremity. *Mondher Bouazizi et al [7]* present an errand not the same as the customary multi-class arrangement. That keeps running on an informational collection gathered from Twitter. Creators allude to this errand as "measurement." By mean the distinguishing proof of every single existing assumption inside an online post as opposed to ascribing a solitary notion name to it. They propose a methodology that consequently credits various scores to every assumption in a tweet, and chooses the notions with the most noteworthy scores. To arrive at this objective, creators added to recently presented device SENTA the important parts to run and perform such an assignment. Creators study the plausibility of measurement, and propose a way to deal with perform it on an informational index made of tweets. The examinations demonstrate the possibility of this assignment and arrive at a F1 score equivalent to 45.9%.

Jufeng Yang et al [8] explores the issue of visual slant examination, which includes an abnormal state reflection in the acknowledgment procedure. Creators propose a structure to use full of feeling locales, where first utilize an off-the-rack objectness device to produce the applicants, and utilize a competitor choice technique. A convolutional neural system (CNN) is associated with every contender to process the estimation scores, and the full of feeling districts, taking the objectness score just as the slant score. At last, the CNN yields from neighborhood districts are amassed with the entire pictures to create the last forecasts. Broad trials demonstrate that the proposed calculation beats the cutting edge draws near.

Companies have constantly wanted brief client experience criticism about their items. A positive client experience can be made by examining client feelings. They have advanced as an asset for removing assessments for applications in different fields. Supposition investigation can be utilized to get the general client experience of an enormous client base on a constant. In *Sandeep Ranjan et al [9]* inquire about, an aggregate of 153,651 unmistakable tweets for Twitter handle of 5 well known telecom marks in India were removed for five months to build up a forecast model utilizing the notion score. The outcomes were approved factually utilizing connection investigation. Positive client opinions about the brand which they incline toward are reflected by higher development pace of new supporters.

III. BACKGROUND

This section provides the key terms that are used for understanding the proposed concept, some essential key terms are reported as:

A. **Data mining:** the data mining is a framework which supports the automated data analysis. Additionally recovery of patterns that are best fit on processing the data. Here various applications which are based on data can be implemented using the data mining techniques such as decision making, pattern recognition, classification, prediction and many more applications.

B. **Text mining:** the text mining is a sub-domain of data mining techniques and algorithm. The text mining is completely used for classification, categorization and information retrieval task. The similar kinds of algorithm and techniques are used as data mining algorithms for performing required text processing task. But in initial phases for computing essential

features the methods can be different as compared to traditional data mining approaches.

C. **Classification:** that is the supervised approach of pattern learning. In this technique the classifier or learning algorithm learn on labelled data samples and the identify the similar patterns where no class labels are present. Those techniques are also used for prediction and other intelligence task.

D. **Supervised learning:** the supervised learning techniques are also known as learn by example. Therefore in order to learn with the data, some predefined samples are required by using feedback from the predefined samples the system learn by examples and after successful learning the similar patterns of data can be recognized.

E. **Categorization:** categorization is also called as the clustering of data. That is the unsupervised approach of data mining. In this technique the data is evaluated internally and based on their internal similarity or differences the groups are created.

F. **NLP:** NLP is known as the natural language processing. That is the domain of artificial intelligence where the sentiments and emotions are recovered on the basis of available data. That data may have in the format of text, audio or visual.

G. **POS tagging:** the POS tagging is also termed as part of speech tagging of the text. The sentences are built with the combination of part of speech information and that information help us to understand the syntax and formation

IV. PROPOSED WORK

The proposed work is motivated from the research article given in [6]. That technique works on fuzzy approach to classify the data and obtaining the negative patterns in social text for identifying the text that contains the hate. In this context the authors introduced as multi-class classifier to distinguish the negative patterns from other text patterns. But the given technique is a rule based classification technique which having a complexity of O(n) for classifying the single text sentence. Therefore a large data classification needs a significant amount of time to be process. Therefore the proposed work is concentrated to reduce the computational complexity with the cost of similar accuracy.

The proposed approach is described using the figure 1. Additionally the required components are also reported in this section which is used for processing the data.
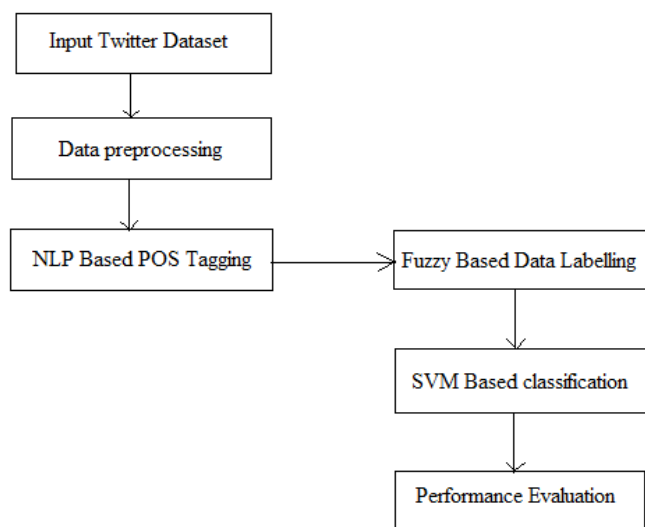
Figure 1 proposed system

The proposed system contains an input module that accepts the social media text dataset. The obtained data or text from a third party data source can contains the noise components. Thus the proposed system first preprocess the data and reduces the noise contains in the initial dataset. After preprocessing of data NLP parser is applied on data for POS (part of speech tagging). The POS tagging convert the unstructured data into a structured data format as a 2D vector. The vector of data is further optimized and pre-labeled using the fuzzy categorization technique. That process may works iteratively for sub-dividing the entire input data into two parts. First part of data is labeled as the normal text and second part of data contains the hate based data. After pre-labeling the data is trained using a supervised learning algorithm namely SVM (support vector machine). The SVM algorithm is first trained on pre-labeled data and produces the class labels as a final outcome of the classification system. After that the performance of system is computed and summarized according to need.

V. CONCLUSION & FUTURE WORK

This paper is motivated to work with the social media text data and recover the hate patterns from the text. In this context a survey on existing techniques are performed and some essential contributions are surveyed that are providing guidelines for developing the required system. Finally a data model is proposed that is being used with the proposed application domain. The proposed system is not only identifying the cyber hate based text it is also suitable for classifying the other kinds of sentiments accurately in efficient manner. Some essential applications are reported here:

a. Classifying sentiments in multi-class manner
b. Classifying the student's educational pressure
c. Classifying social media sentiments over a given topic
d. Classifying the reviews and sentiments on any product and services for online platform

After successful implementation of the proposed sentiment analysis technique the following fruitful outcomes are expected.

a. Obtaining a rich literature of text processing using NLP based techniques
b. Offering the new classification technique for optimizing the performance of accuracy of exiting classification
c. Comparing the performance of classifiers of exiting technique and the proposed classification technique.

Thus in near future the proposed system is implemented and their performance is reported using suitable classification parameters.

REFERENCES

[1] V. Gupta, Prof. Devanand, *"A survey on Data Mining: Tools, Techniques, Applications, Trends and Issues"*, International Journal of Scientific & Engineering Research Volume 4, Issue3, March-2013

[2] V. Borkar, K. Deshmukh, S. Sarawagi, *"Automatic segmentation of text into structured records"*, SIGMOD '01 Santa Barbara, California, Copyright 2001 ACM 0-89791-88-6/97/05

[3] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. A. Hawalah, A. Gelbukh, Q. Zhou, *"Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques"*, Cogn Comput (2016) 8:757–771, DOI 10.1007/s12559-016-9415-7

[4] D. Gupta, Er. J. Dangra, *"A new approach for clustering of text data based on fuzzy logic"*, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 4 Issue 8, August 2015

[5] V. Balakrishnan, S. Khan, T. Fernandez, H. R. Arabnia, *"Cyberbullying detection on twitter using Big Five and Dark Triad features"*, Personality and Individual Differences 141 (2019) 252–257

[6] H. Liu, P. Burnap, W. Alorainy, and M. L. Williams, *"A Fuzzy Approach to Text Classification With Two-Stage Training for Ambiguous Instances"*, IEEE Transactions On Computational Social Systems, VOL. 6, NO. 2, APRIL 2019

[7] M. Bouazizi and T. Ohtsuki, *"Multi-Class Sentiment Analysis in Twitter: What If Classification Is Not the Answer"*, Special Section on Emerging Trends, Issues and Challenges for Array Signal Processing and Its Applications in Smart City, Vol 6, 2018, 2169-3536 2018 IEEE

[8] J. Yang, D. She, M. Sun, M. M. Cheng, P. L. Rosin and L. Wang, *"Visual Sentiment Prediction based on Automatic Discovery of Affective Regions"*, IEEE Transactions On Multimedia,

[9] S. Ranjan, S. Sood, V. Verma, *"Twitter Sentiment Analysis of Real-time Customer Experience Feedback for Predicting Growth of Indian Telecom Companies"*, 2018 4th International Conference on Computing Sciences, 978-1-5386-8025-4/18/$31.00 ©2018 IEEE