iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36613

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Implementation Analysis of Data Classification Approach for Sentiment Classification

Bhushan R. Chincholkar[1], J. M. Patil[2]
[1, 2]*Computer Engineering, Shree Sant Gajanan Maharaj College of Engineering, Shegaon, India*

Abstract: ***Sentiment analysis is one of the fastest growing fields with its demand and potential benefits that are increasing every day. Sentiment analysis aims to classify the polarity of a document through natural language processing, text analysis. With the help of internet and modern technology, there has bee***
n a tremendous growth in the amount of data. Each individual is in position to precise his/her own ideas freely on social media. All of this data can be analyzed and used in order to draw benefits and quality information. In this paper, the focus is on cyber-hate classification based on for public opinion or views, since the spread of hate speech using social media can have disruptive impacts on social sentiment analysis. In particular, here proposing a modified approach with two stage training for dealing with text ambiguity and classifying three type approach positive, negative and neutral sentiment, and compare its performance with those popular methods also as well as some existing fuzzy approaches. Afterword comparing the performance of proposed approach with commonly used sentiment classifiers which are known to perform well in this task. The experimental results indicate that our modified approach performs marginally better than the other algorithms.
Keywords: Ambiguity, cyber hate, fuzzy, Sentiment analysis.

I. INTRODUCTION

As internet is growing bigger, its horizons are becoming wider. Social Media and Micro blogging platforms like Facebook, Twitter, Tumblr dominate in spreading encapsulated news and trending topics across the globe at a rapid pace. A topic becomes trending if more and more users are contributing their opinion and judgments, thereby making it a valuable source of online perception [1]. These topics generally intended to spread awareness or to market public figures, political campaigns during elections, product endorsements and entertainment like product, award shows. Large organizations and firms cash in of people's feedback to enhance their products and services which further help in enhancing marketing strategies. It is intended to function an application to understand the attitudes, opinions and emotions expressed within an web mention. The intention is to realize an summary of the broader popular opinion behind certain topics. Precisely, it is a paradigm of categorizing conversations into positive, negative or neutral labels. Many people use social media sites for networking with other people and to stay up-to-date with news and current events. These sites (Twitter, Facebook, Instagram) offer a platform to people to voice their opinions. For example, people quickly post their reviews online as soon as they watch a product and then start a series of comments to discuss about the acting skills depicted in the product. This kind of information forms a basis for people to evaluate, rate about the performance of not only any product but about other products and to know about whether it will be a success or not. This type of vast information on these sites can used for marketing and social studies [2]. Therefore, sentiment analysis has wide applications and includes emotion mining, polarity, classification and influence analysis. Twitter is an online networking site driven by tweets which produces millions of tweet per day. Usually, these steams of raw twitter data comprises of irregular, noisy, poor structured sentence, ill formed, incomplete words, nondictionary terms, and irregular expressions in unstructured manner. Twitter sentiment analysis involves the use of natural language processing to extract, identify to characterize the sentiment content. However, doing the analysis of tweets expressed in not an easy job. A lot of challenges are involved in terms of tonality, polarity, lexicon and grammar of the tweets. They tend to be highly unstructured and non-grammatical. It gets difficult to interpret their meaning. Moreover, extensive usage of slang words, acronyms and word ambiguity are quite common while tweeting online. The categorization of such reviews into specific class polarity gets tough for natural processors involved. In this paper, focus is on the detection of online hate speech (cyberhate) in text posted to social media platforms on the basis of four types of online hate speech, namely, religion, race, disability, and sexual orientation, by proposing a completely unique approach, especially for handling with text ambiguity. The rest of this paper is structured as follows. In Section II, we detailed some related work of our project by highlighting important features. Next, Section III gives brief details about the proposed system. Section IV covers details of methodology & implementation of the project. The result and analysis part are specified in section V. Further, in Section VI, future work and conclusion is discussed.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

II. RELATED WORK

The paper "Sentiment analysis of twitter published in 2012 introduces a machine learning approach to implement sentiment analysis on the data. [3] They need performed sentiment classification of Twitter data where the classes are positive, negative and neutral. Two kinds of of models are utilized: Tree part and highlight based models and both these models beat the unigram pattern. For the element based approach they performed include examination, Which uncovers that the foremost critical components are people who that join the sooner extremity of words and their parts of discourse labels[4].

In "The Twitter Sentiment Analysis: The Bad the great and The OMG" paper, they need explored the utility of phonetic components for recognizing the assumption of twitter messages [5][6]. They have known the value of existing of lexical assets and additionlly includes that catch data about the casual and imaginative dialect utilized as a neighborhood of varied social sites. An approach has been introduced to unravel the issues.

In This Paper "The Twitter Sentiment Classification using Discrete Supervision" published in 2009 introduces a novel approach for naturally grouping the sensation of varied twitter message [5]. These messages are either classified because the the positive or negative with reference to the information. The paper describes the preprocessing of varied steps so as to realize very high accuracy. The principle commitment of this paper is to utilizing tweets with emoticons for much off regulated learning Diverse machine learning classifiers and highlight extractors have been utilized alongside the utilization of unigrams, bigrams, unigrams and bigrams, and parts of discourse as components.

III. PROPOSED SYSTEM

A. Process Cyber Hat and DSA

In proposed system there are basically two modules are available. First module indicate dual sentiment analysis (DSA) framework intimately. And second module indicate the prediction user review rating on user previously review data.

Fig. 3.1 illustrates the method of a dual sentiment analysis. In this fig, the rectangle crammed with slash denotes the original data and the rectangle filled with backslash denotes the reversed data. DSA contains two main stages: dual training (DT) and dual prediction (DP).

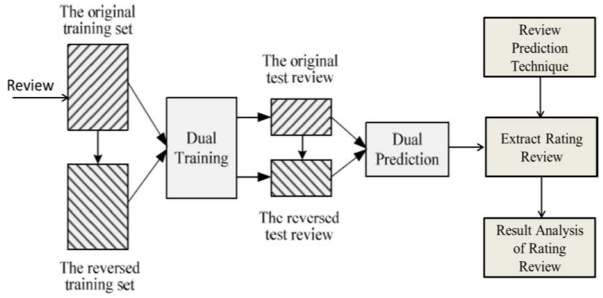


Fig. 3.1 Block Diagram for process of dual sentiment analysis (DSA).

1) *Dual Training (DT):* The original training specimens are reversed to their opposites. Informs to them as "original training set" and "reversed training set". In our data expansion technique, there's a one-to-one correspondence among the original and reversed reviews. The classifier is trained by maximizing a mixture of the likelihoods of the original and reversed training samples. This process is named dual training.
2) *Dual Prediction (DP):* Dual prediction works in addressing the polarity shift problem. This point think" I don't like this book. It's boring" is an ingenious test review, and "I like this book. It is interesting" is the reversed test review.
3) *Review Prediction Technique:* In review prediction technique, users predict the reviews like positive and negative. This collection of reviews used for various opinion analysis tasks. Users give the reviews on product online. The admin then checks reviews of uploaded product, and displays a web rating for each product
4) *Extract Rating Review:* After gives the user online reviews, system extract the reviews like positive, negative and neutral .Then it displays a web rating reviews and extracting from the positive, negative and neutral reviews.
5) *Result Analysis of Rating Review:* In the result analysis, all the reviews are gathers for a specific product and calculate the amount of count for positive-negative-neutral. This Positive-negative neutral count is generated the graphically view of positive-negative- neutral analysis. These systems also sorts and display the analysis and calculate the result. This provides an automatic product rating review system supported on sentiment analysis.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

IV. METHODOLOGY AND IMPLEMENTATION

Usually encounter product rating websites where users are allowed to review on products online. These rating are provided as input to the web site admin. The admin then checks reviews and displays an web rating for every product. Here propose an online system that automatically allows users to post reviews and store them. The system now analyzes this data to check for user sentiments associated with each review. Our system consists of a sentiment library. The system breaks user review to see for sentiment keywords. Once the keywords are found it associates the review with a sentiment rank. The system now gathers all review for a particular product and then calculates the positive-negative-neutral count. This provides an automatic product rating system supported on sentiment analysis.

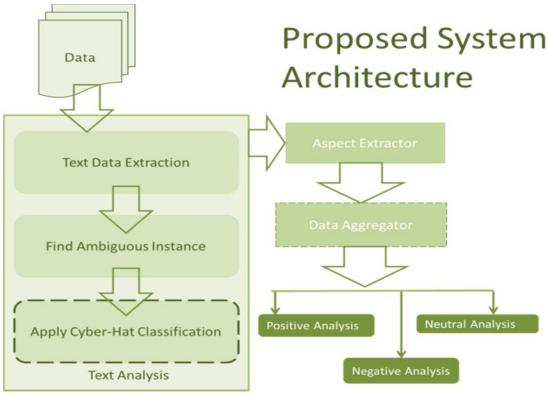


Fig. 4.1 Architecture Diagram

A. *Features*

1) *User Registration/ Login:* Users may login to system and register if not already registered.
2) *Reviews:* Users may reviews on a product provided within the system.
3) *Review Storage:* The system stores each review for further processing.
4) *Sentiment Analysis:* The system process reviews for finding out sentiments and their weightage.
5) *Single review Count:* One user can review just one occasion for a specific product.
6) *Review Rating and Analysis:* The system further processes sentiment ratings and then analysis for a particular product.
7) *Graphical Analysis:* The system displays graphically positive-negative-neutral product reviews.

V. RESULT ANALYSIS

In this system, user gives the reviews like positive, negative and neutral. This collection of reviews used for various opinion analysis tasks. Users give the reviews on products online. The admin then checks reviews, and displays an online rating for every product. Usually encounter product rating websites where users are allowed to review on products online. These rating are provided as input to the web site admin. The admin then checks reviews and displays an online rating for each product. Online system that automatically allows users to post reviews and store them. In the result analysis, first calculate the amount of positive-negative-neutral count then on the basis of counting, it represents the graphically. The system now analyzes this data to check for user sentiments associated with each review. The system now gathers all review for a specific product and then calculates the amount of count for the positive-negative-neutral. This provides an automatic product rating system based on sentiment analysis.

In result analysis, first show the Review analysis table and then graphical analysis with the assistance of particular product.

1) *Product 1:* Real Me TV

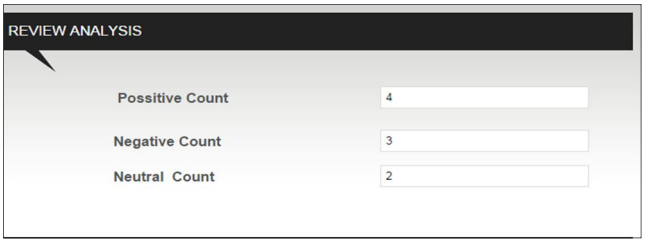
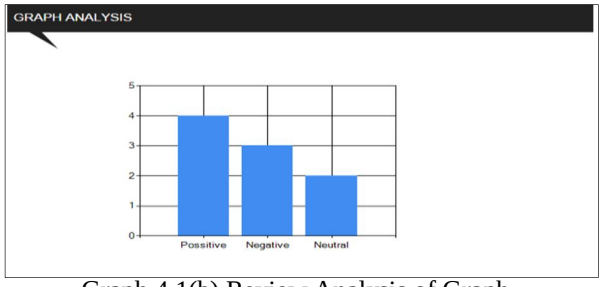


Table 4.1(a) Review Analysis

©IJRASET: All Rights are Reserved



Graph 4.1(b) Review Analysis of Graph

VI. CONCLUSION & FUTURE SCOPE

A. Conclusion

Sentiment Analysis is extremely important research because Sentiment Analysis helps in summarizing opinion and reviews of public. They consider as research filed. However, Sentiment Analysis still needs to improve and progress. Moreover, there are many challenges just like the polarity during a sentence.

In this dissertation, specialize in creating reversed reviews to assist supervised sentiment classification. DSA is highlighted by the technique of one-to-one correspondence data expansion and therefore the manner of using a pair of samples in training (dual training) and prediction (dual prediction). Then highlights the essential ideas about Sentiment Analysis and then explains in details the Sentiment Classification and Technique.

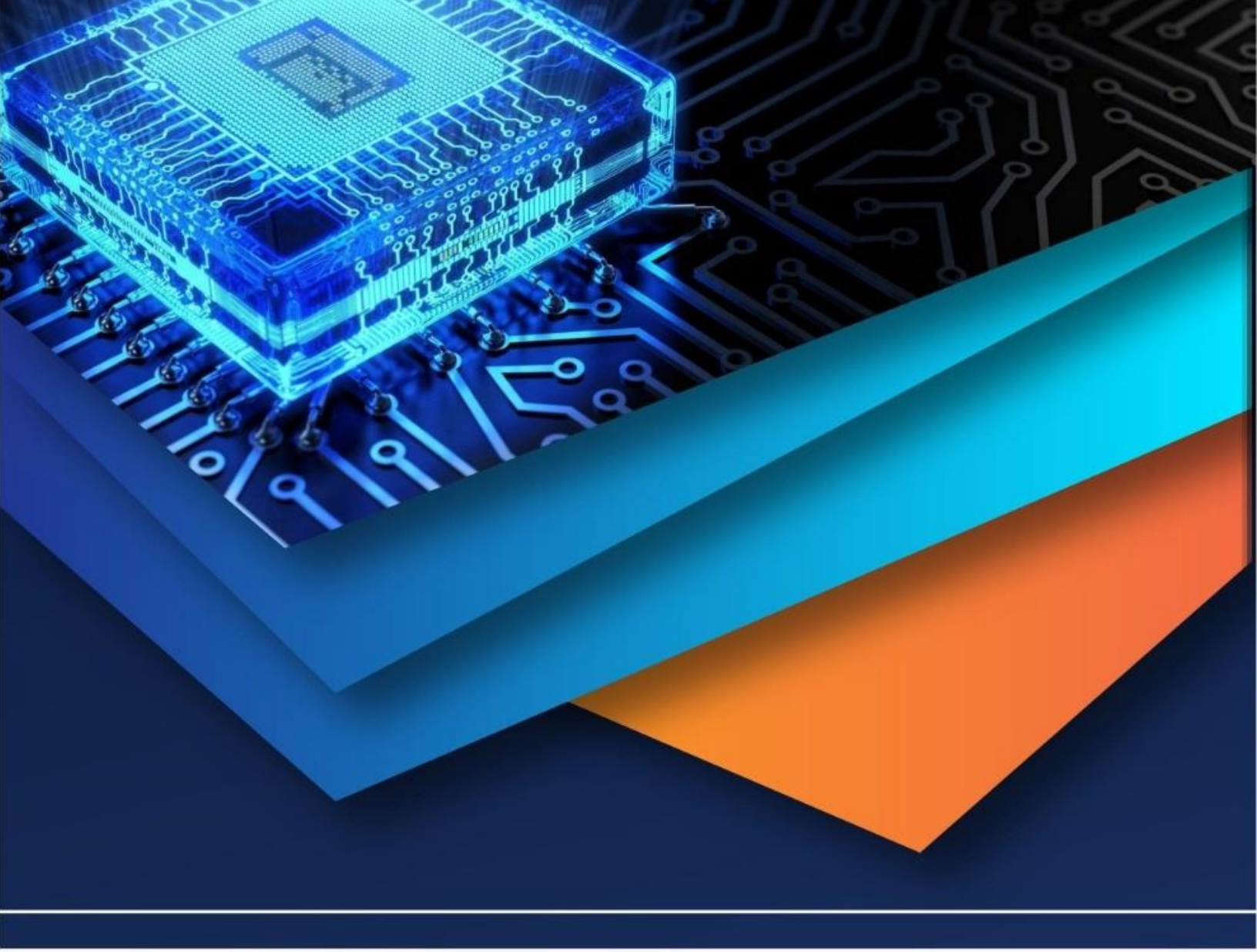
B. Future Scope

In the future scope, can expand the DSA and cyber hat algorithm to a bigger range of sentiment analysis tasks. And consider also an idea for more complex polarity shift patterns such as transitional, subjunctive and sentiment-inconsistent sentences in creating reversed reviews

REFERENCES

[1] M. Desai and M. Mehta, "Techniques for sentiment analysis of Twitter data: A comprehensive survey", 2016 International Conference on Computing, Communication and Automation (ICCCA), pp.12-18, 2016.
[2] Dr. Khalid N. Alhayyan & Dr. Imran Ahmad “Discovering and Analyzing Important Real-Time Trends in Noisy Twitter Stream”, SYSTECS, CYBERNETICS AND INFORMATICS, VOLUME 15 - NUMBER 2 - YEAR 2017, pp. 25-31.
[3] The David Haussler, D.C. Uthus and D. W. Aha. “Plans toward automated chat summarization”. In Meeting of the Association for Computational Linguistics, pp.1–7,2011.
[4] M Hu and B Liu. 2004,” The Mining of data and summarizing the customer reviews based on an issue”. KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168–17, August 2004..
[5] S M Kim and E Hovy, ”Determining the sentimental analysis of various opinions”, Proceedings of the COLING conference, Geneva, pp. 1-8,2004.
[6] Erumit, A.K.Nabiyev, V. Cebi, A.Karadeniz Tech. University, Trabzon, Turkey, “The Design of motion problems based on The graphical theory in counting the number of words in math”, Turkey ,pp.68-81,2006.
[7] Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, “Sentiment analysis of Twitter data,” in Proc. Workshop Lang. Social Media, Stroudsburg, PA, USA, pp. 30–38,Jun 2016.
[8] N. Pendar, “Toward spotting the pedophile telling victim from predator in text chats,” in Proc. International Conference Semantic Computer., Irvine, CA, USA, pp. 235–241, Sep.2007.
[9] K. Thiel and M. Berthold, “The KNIME text processing feature: An introduction,” KNIME, Zürich, Switzerland, Tech. Rep. 120403F, 2, pp.1-27, 2012.
[10] M.-C. De Marneffe, B. MacCartney, and C. D. Manning, “Generating typed dependency parses from phrase structure parses,” in Proc. 5th International. Conference. Language Resource Eval., Genoa, Italy, pp. 71–80,May 2006.
[11] V. Tursi and R. Silipo , “From Words to Wisdom- An Introduction to Text Mining With KNIME”. Zurich, Switzerland: KNIME Press, 2018.
[12] N. Djuric, J. Zhou, R. Morris, M. Grbovic, V. Radosavljevic, and N. Bhamidipati, “Hate speech detection with comment embeddings,” in Proc. 24th Int. Conf. World Wide Web, pp. 29– 30,May 2015.
[13] Nemes, “Regulating hate speech in cyberspace: Issues of desirability and efficacy,” J. Inf. Communication Technology. Law, vol. 11, no. 3, pp. 193–220,2002.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)