iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.38511

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 XI Nov 2021- Available at www.ijraset.com

Feature Set Selection for Sentiment Analysis

Ganesh K. Shinde[1], Vaibhav N. Lokhande[2], Vikas B. Gore[3], Rasika T. Kalyane[4], Umesh M. Raut[5]
[1, 2,3,4,5] *Department of CSIT Dr BAMU Aurangabad Maharashtra India*

Abstract: With proliferation of online blogging web sites, hundreds of thousands of text posts are generated. Using this rich information facilitate educated purchasing of objects, discovering and public developments involving more than a few merchandises in the market, discovering political inclination of societies previous to a country wide election, and many others. Considering that the final decade, Sentiment evaluation has received increased attention from many researchers as a procedure for addressing subject matters, such as the a fore mentioned ones. This paper specializes in Sentiment evaluation and use of sentiment Features. In this paper we have created the feature set and given input to svm and result verified for sentiment.
Keywords: Sentiment analysis, support vector machine, maximum entropy, with features, without features, artificial intelligence.

I. INTRODUCTION

In recent years, we have noticed that opinion postings in social media (e.g. stories, discussion board discussions, blogs, Twitter, feedback, and) have helped reshape corporations, and read public sentiments and emotions, which have impacted on our social and political methods [1][2]. The dataset used in this paper is a tweets. Those tweets shall be extracted and processed so it will produce information such as sentiment that consists in tweets. Sentiment analysis on tweets is used to find out whether a tweet consists of positive or negative sentiment. Two kinds of learning that usually used in the sentiment analysis, which is supervised learning and unsupervised learning [3]. The machine learning system is belong to supervised studying, this procedure quite often want so much if training information that have been labelled manually. Without labelled training data, supervised learning won't capable to be processed [4]. Whereas, the lexicon- cantered method is belong to unsupervised finding out, which does not want already trained information and most effective depend on the dictionary [3]. As mentioned above ways have special characteristics, however it may complementary if both methods are mixed. The combination of each methods can be finished by way of utilizing lexicon-based system to create labelled tweets which can be utilized as training information in Support Vector Machine process so there shall be no training approach on this blend ways [5] .Micro blogging web sites such as Twitter have gained increased repute. Every day a big amount of texts are made available on line through this online media. The knowledge captured from these texts, could be active for scientific surveys from a amusing or political perspective [6]. On one hand, companies and product owners who aim to alleviate their products/services may strongly account from the affluent acknowledgment [7]. On the opposite hand, customers would also gain knowledge of about positivity or negativity of customers respecting. Sentiment analysis is the system of extracting the polarity of individuals' opinions from normal language texts [9]. Twitter messages are most cases informal. On account that of the anomalistic nature of informal textual content, evaluation or processing of this sort of text is more often than not more difficult when compared to formal textual content. Opinion mining is a type of natural language processing for tracking the mood of the public about a particular product. Opinion mining, which is also called sentiment analysis, involves system to collect and categorize opinions about a product. Automated opinion mining often uses machine learning, a type of artificial intelligence to mine text for sentiment. There are several challenges in opinion mining. The first is that a word that is considered to be positive in one situation may be considered negative in another situation. A second challenge is that people don't always express opinions the same way. Finally, most reviews are both positive and negative Comments, which is somewhat manageable by analysing sentences one at a time. In general, sentiment analysis has been investigated mainly at three levels:

1) *Document Level:* The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment.
2) *Sentence Level:* This level task goes to the sentences and determines whether or not each and every sentence expressed a positive, negative, or neutral opinion.
3) *Entity and Aspect Level:* Both the document level and the sentence level analyses do not discover what exactly people liked and did not like. Aspect level performs finer-grained analysis. Aspect level was earlier called feature level. Aspect level directly looks at the opinion itself.
4) *Twitter:* Twitter blogging website online that permit person to ship their tweet with the maximum characters used are one hundred forty characters.

II. LITERATURE REVIEW

Level classification is most promising subject in sentiment evaluation document in Sentiment classification. Reference [11] showed that there is a correlation between sentiment measures computed utilizing phrase frequencies in tweets and both client self-assurance polls and political polls. Accordingly, they illustrated that inclination of public towards special entities might be examined through analysis of tweets. Reference [12] measured presidential efficiency over a exact time interval by way of extracting general public sentiment from Twitter. For this motive they used the Senti Strength lexicon [13]. Many researchers have developed distinctive methods for sentimental analysis. The researcher Seyed-Ali Bahrainian et al. [9] presented a approach to Sentiment analysis of quick informal texts with a primary focus point on Twitter posts referred to as "tweets". Additionally the process proposed by means of Noriaki Kawamaet al. In [15] the place "the hierarchical technique to sentiment analysis, identifies each an item and its score by means of dividing topics, which is mainly handled as one entity. [16] Developed sentiment ontology to conduct context-sensitive sentiment evaluation of on-line opinion posts in stock markets. ZHU Nanli et.Al. [5] Introduced a survey on the cutting-edge progress in sentiment evaluation, and makes an in-depth introduction of its research and application in industry and Blogsphere. [14] Adopts a suite of sentiment aspects as well as some non-sentiment aspects to procedure and analyse a manually annotated information set of tweets. [15] Measured presidential performance over a special time period by using extracting general public sentiment from Twitter. For this purpose they used the SentiStrength lexicon [11].

III.HYBRID POLARITY DETECTION

This section describes target-oriented hybrid sentiment analysis system. It consists of three major modules.

A. Pre-processing Module

@username is replaced with "ATUSER". URLs are removed.
"#word" is replaced with "word".
Slangs are replaced with their actual phrase equivalences.
The target of sentiment word is replaced by "TARGET".

B. Sentiment Feature Generator Module

This module starts with replacing slangs with their equivalences using a slang dictionary. To build this slang dictionary, we use SentiStrength lexicon [13]. In the second step this module uses the SentiStrength lexicon [13] to tag all words present in dictionary for each document with their corresponding sentiment scores. Likewise, according to a list of emoticons resent in dictionary, it tags happy emoticons with a sentiment score of "+1" and sad ones with a score of "-1". Also it further, tags all intensifiers (e.g. finally) and diminishes (e.g. may) with their corresponding scores. Also, it tags negation words with "NEGATE". Finally, if a word did not belong to any of the mentioned categories in the dictionary, it tagged that with the score "0". Having all words in a document tagged by their score now, we handle occurrence of intensifiers, diminishes, and negations. Firstly, we intensify the strength of a word that appears after an intensifier words, by the score of that intensifier word. Similarly, in the case of diminishers, we weaken the strength of a word that appears after a diminisher word by the strength of that diminisher. Finally, for negations, we flip the polarity of the score of a word that appears after a negation word. Then we lessen the flipped sentiment score by 1. That is, if the flipped score is positive, we subtract it by 1 and if it negative we sum it by 1. Our aim in features extracting is to capture sequence of sentiment relevant words that show a document sentiment change. Additional, we define some features that present the neighborhood of the target of sentiment which we defined as iPhone. Table below shows this feature set.

C. Machine Learning Classifier

The machine learning module is a linear support vector machine that takes as input the feature set described in the previous table and according to that classifies the tweets to separate classes. We explain the heuristic regarding "if statements" in detail. In the following example sentence is given "If I don't get an iPhone for Diwali, I would be sad." In the above statement, it shown that the actual polarity of the given sentence regarding target i.e. iPhone is positive, whereas, the sentence only contains words and patterns that usually occur in a negative context. "Don't get an iPhone" is a negative sentiment regarding iPhone and it sets f10 to "1". Likewise, the word "sad" is a negative word and its occurrence sets f3 and f14 to "1". These features are often more likely to be present in negative sentences. However, because of the presenting patterns, the inverse sentiment feature (f7) is also set to "1". The heuristics search to find out if both the "if clause" and the "main clause" in the example sentence, include a negative sentiment pattern and if so it detects this pattern as an inverse sentiment

TABLE I

F1	sentiment score
f2	Number of positive words
f3	Number of negative words
f4	Number of negation words
f5	Number of negation words followed by a positive word
f6	Number of negation words followed by a negative word
f7	inverse sentiment
f8	Number of positive words followed by target
f9	Number of negative words followed by target
f10	Number of negation words followed by target
f11	Number of positive words followed by a negative word
f12	Number of negative words followed by a positive word
f13	Number of target words followed by a positive word
f14	Number of target words followed by a negative word

IV. EVALUATION

A. Dataset

Dataset consists of 15,000 tweets. We have no annotated data and hence we download tweets from twitter and classified with svm and MaxEntropy classifier. Tweets are classified as shown in following graph. The data classification with Max Entropy gives better result as compared to svm.

B. Evaluation Metrics

We evaluate the methods presented in this paper using accuracy on overall accuracy as presented in the following:
Overall accuracy = (TP+TN)/ (TP+FP+TN+FN)
Where TP, FP, TN, and FN are the number of true positives, false positives, true negatives and false negatives. Furthermore, for testing any of the supervised classifiers aswell as our hybrid method we use 10-fold cross validation.

C. Experimental Results

Firstly, we tested the data using two classifiers. Classifying data using SVM and MaxEnt for 15000 tweets it is having accuracy 88% and 87% with features respectively. For 5000 tweets it is having accuracy 86% and 84% with features respectively. SVM give good result as compared to MaxEnt.

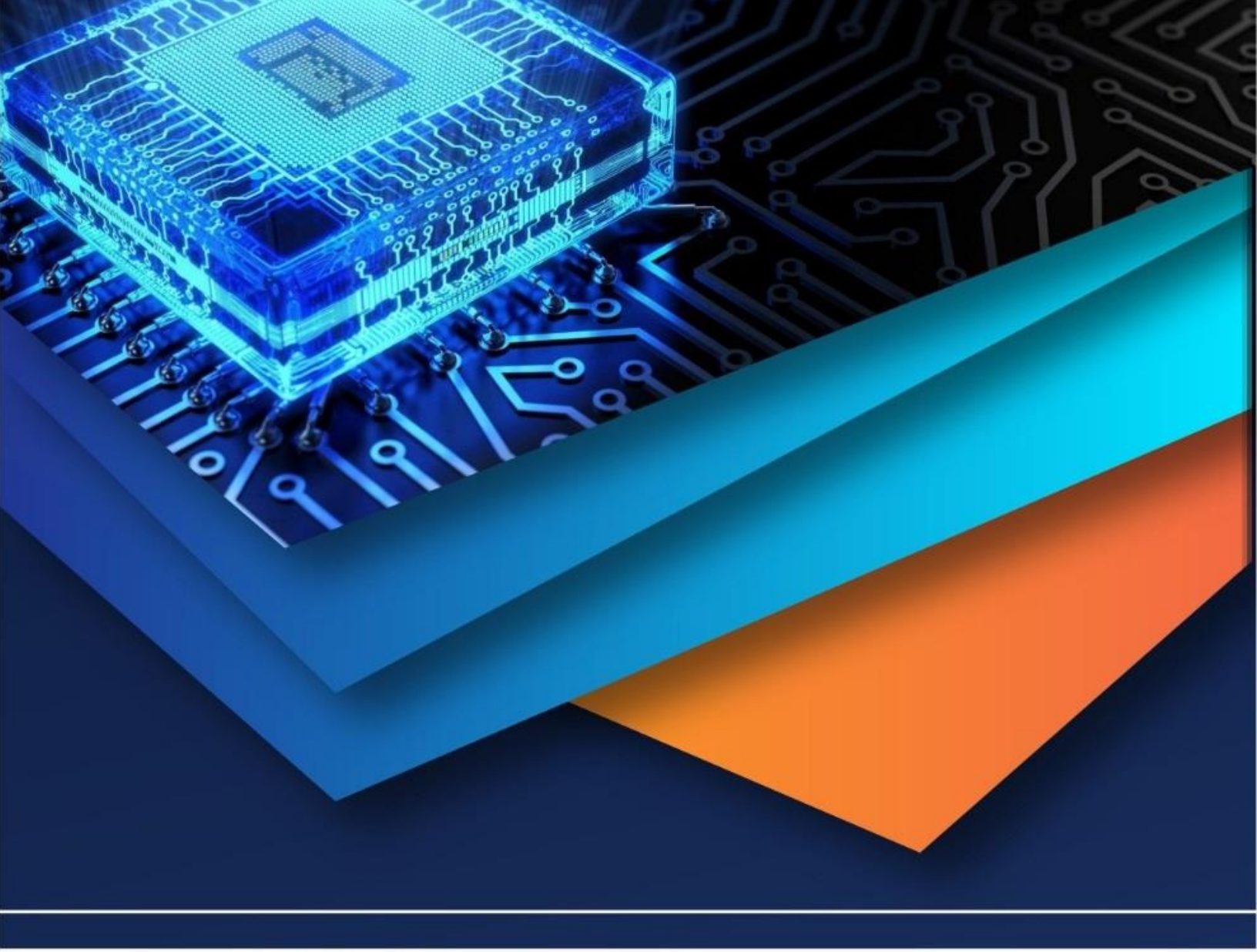
V. CONCLUSIONS

We introduced Hybrid method in which we combines Sentiment Lexicon with machine learning classifier for polarity detection of sentiment tweets. We conclude that sentiment features method also be solution to sentiment analysis.

REFERENCES

[1] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai and A. Kappas, Sentiment strength detection in short informal text, Journal of the American Society for Information Science and Technology (2010), 2544–2558S.
[2] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers
[3] Tan,S., Wang, Y., & Cheng,X.(2008). "Combining learnbased and lexicon-based techniques for sentiment detection without using labeled examples", In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in informationretrieval, July 20-24, 2008, Singapore, Singapore
[4] Pang, B., Lee, L., & Vithyanathan, S. (2002). "Thumbs Up? SentimentClassification Using Machine Learning Techniques." Proceedings of The ACL-02 conference on Empirical methods in natural language processing (pp. 79-86). Stroudsburg: Association for Computational Linguistic
[5] Ley, Z., Riddhiman, G., Mohamed, D., Meichun, H., & Bing, L. (2011). "Combining lexicon-based and learningbased methods for twitter sentiment analysis". HP Laboratories, Technical Report HPL-2011, 89.
[6] B. O'Connor, R. Balasubramanyan, B. Routledge and N. Smith, From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series, in: International AAAI Conference on Weblogs and Social Media, North America, May 2010.
[7] M. Gamon, A. Aue, S. Corston-Oliver and E. Ringger, Pulse: mining customer opinions from free text, in: Proc. of the 6th International Conference on Advances inIntelligent Data Analysis, Madrid, Spain, September 8– 10, 2005, pp. 121–132.

[8] H. Tang, S. Tan and X.A. Cheng, Survey on sentiment detection of reviews, Expert Systems with Applications: An International Journal 36(7) (2009), 10760–10773.
[9] S.-A. Bahrainian and A. Dengel, in: 2013 IEEE/WIC/ACM International Joint Conferences on Sentiment Analysis Using Sentiment Features, Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Vol. 3, 17–20 Nov.2013, pp. 26–29.
[10] S.A. Bahrainian, S.M. Bahrainian, M. Salarinasab and A. Dengel, Implementation of an Intelligent Product Recommender System in an e-Store, in: Proc. of the 6thInternational Conference on Active Media Technology (AMT'10), Toronto, Canada, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 174–182.
[11] O'CONNOR, B.; BALASUBRAMANYAN, R.; ROUTLEDGE, B.; SMITH, N.. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. InternationalAAAI Conference on Weblogs and Social Media, North America, may. 2010.
[12] Lai, P., Extracting Strong Sentiment Trends from Twitter,2011.
[13] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas. A., Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, pages 2544- 2558, 2010.
[14] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R., Sentiment analysis of Twitter data. In Proceedings of the Workshop on Languages in Social Media (LSM '11). Association for Computational Linguistics, Stroudsburg, PA, USA, 30-38,2011.
[15] Xiaohui Yu,Yang Liu, Aijun An" An Adaptive Model for Probabilistic Sentiment Analysis", IEEE Computer Society ,Volume, Issue No. : 4191-4/10, pp-661-667, November 2010.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY