



EVALUATING THE SENTIMENT SCORE WITH THE USE OF MATRIX REPRESENTATION IN SENTIMENT ANALYSIS

Nisha Yadav, Dr. Banita

Research Scholar, Associate Professor

Department of Computer Science and Engineering

Baba MastNath University, Haryana

Abstract:

Natural Language Processing and machine learning models has widely used in today's field. To analyze the various emotions and to categorize the various opinions of all the users, sentiment analysis has been used. In this study, Twitter database is used to calculate the opinions of all the users which are using Twitter. Twitter is basically the social platform which is used to communicate at faster mode. With the help of Twitter, we can easily communicate within a couple of seconds. We can share the data either in audio, video or textual format. Support Vector Machine has also been used in this study to categorize and to analyze the whole data set along with the performance which make the proposed work much accurate. The opinions are segregated into three parts, that is positive, negative and neutral aspect. The sentiment analysis basically divides the performance will be analyzed on 2 basic aspect that is semantic analysis and opinion mining. The basic approach for analyzing the data is firstly we use the various data set from the Twitter database. Then the data preprocessing will be done which will help to figure out and filter the various noise. After that Linguistic Data Processing with the help of Natural Language Processing is used. It basically mentioned the polarity of all the sentiments. And then we will apply in this study machine learning algorithm and it lasts after applying the machine learning algorithm, the result will be analyzed. the whole sentence into number of words and then words get segregated into tokens.

Keywords:

Twitter Analysis, Opinion based, Sentiment Score, Matrix Analysis

I. INTRODUCTION:

There are a number of machine learning algorithm which are used nowadays on the huge media. Twitter is the social platform from which we can easily communicate our messages. With each other. Such communication will be done within a couple of seconds. The Twitter social platform is much faster mode as compared to the others. The various machine learning algorithm which are enlisting in sentiment analysis are. Support vector machine maximum entropy. Naive Bayes and WEKA.

In this study, the machine learning algorithm which we use to analyze the data set, our support vector machine maximum entropy and little portion of newbies algorithm. Additional to this, various approaches of the sentiment analysis are also available. In this study. The work will be done on 2 aspect that is nature and working platform. The opinion mining is segregated into three categories that is positive, negative and neutral. A huge number of literature review has been taken to analyze the data set and to compare the proposed work with the earlier model. Support Vector Machine is basically used to categorize the performance on different aspects, that is positive, negative and neutral. All the positive words or the valuable words or answerable words are in the positive aspect. The negative aspect. Include the matter of fact that. No reply should be there or

ignorance of the sentences there. The graphical representation has also been made in this study, but it has not been shown in the current paper.

II. TWITTER ANALYSIS:

The Twitter sentiment analysis has been done with the use of Support Vector Machine classifier. The Support Vector Machine classifier is used to detect the accuracy of the data set on the three aspects that is positive, negative and neutral. That is more than 50%, then it rely in the positive aspect. If less than 50% then it rely on the negative aspect. And similarly for the neutral aspect that is from 50 to 60 ratio. We are again going to segregate the 50 to 60% ratio. That is 50 to 55%. Include the negative aspect and 55 to 60% include the positive aspect. For this analyzing the data. The support vector machine is basically work on the 50 to 60% of the data set on which the uncertainty is there. We basically do the this study of sentiment analysis to detect the accuracy of the data set which is available on 50 to 60%. This above said percent is calculated on the behalf of the probability of the sentence whatsoever is available. All these sentences have been taken from the Twitter data set. The Twitter data set as we are aware about the social platform, that number of the comments or reply or emojis or audio or video or textual format data has been shared publicly and there is no privacy on the Twitter. Just like Instagram, Instagram is also a social media platform. On which any user can easily check the videos or data available on another user. No matter whether that user is available on the list or contact list of the previous user or not because it is publicly available, the huge amount of the data has been shared on the Twitter data set. And for this reason, the cloud data set has been used or big data set has been used.

The huge data set contains the very critical opinion related information that has been used in the business purposes and other commercial and scientific industries. Hence the manual tracking and extraction of the data is useful. And it is not possible. Thus the sentiment analysis is used. Sentiment analysis is basically the procedure of extracting the various sentiments or opinions from the reviewers which has been expressed by the user or the customer which are available on the Twitter platform. Initially the text get converted into structured text and then the structured text gets extracted all the sentiments which are

available on the text and after the extraction of the sentiment the opinion has been made whether the opinion is positive, negative or neutral. Actually, the architecture of the proposed system also include the structured and unstructured test text. The structured text is directly display onto the screen as such, whereas the unstructured text get further segregated and operated with the three terms, that is sentiment extraction, polarity detection and opinions. Memorization. After going the data set from all the three aspects, the sentiment of text has been received. Natural language processing is the field of computer science where the artificial intelligence is much more concerned and in this the interaction between the computer and the human languages has been made. New vice classifier or algorithm has also been used. The theorem of newbies is strictly independent among all the predictors. There is one also data set or the procedure that is Wordnet which is available for evaluating the various texts in the English language. The word net is a lexical database used for the English languages. To find out the. Valuable context. The context gets undergo with the unsupervised learning. Unsupervised learning gets aggregated into two aspect that is sense, inventory and context. Then these Sense inventory and context will further classified with the help of Science Classifier. And in the Sense Classifier we can also add new valuable text which are available.

III. OPINION BASED APPROACH:

As far as the Internet is concerned, single user is connected with the social media and the Internet or various websites at one time. The factors of sentiment aspects are belief, feeling, emotions, opinion and attitude. All these factors result in the buying behavior and purchasing intention of any customer on electronic commerce platform. The various sampling technique which has been used are nonprobability sampling and probability sampling. The non probability sampling include the convenience sampling, judgmental sampling, quota sampling, and snowball sampling, whereas the probability sampling includes simple random, systematic. Rumbling, stratified sampling and cluster sampling. We can also prepare one question here on the future aspect to check whether a customer is regularly using the social media or any Internet websites or not. Basically 2 dimensions of the study, sentiment aspect and

behavioral aspect. The behavioral include the opinion or the reply or the data which has been posted by the user. To analyze any type of data, Support Vector Machine is well used to calculate the accuracy of the data set. Whereas to evaluate the language or English language analysis can be done with the help of sentiment analysis. We can easily. Conclude the meaning of any sentence on the positive, negative or neutral aspect with the use of lexicon platform.

IV. SENTIMENT SCORE:

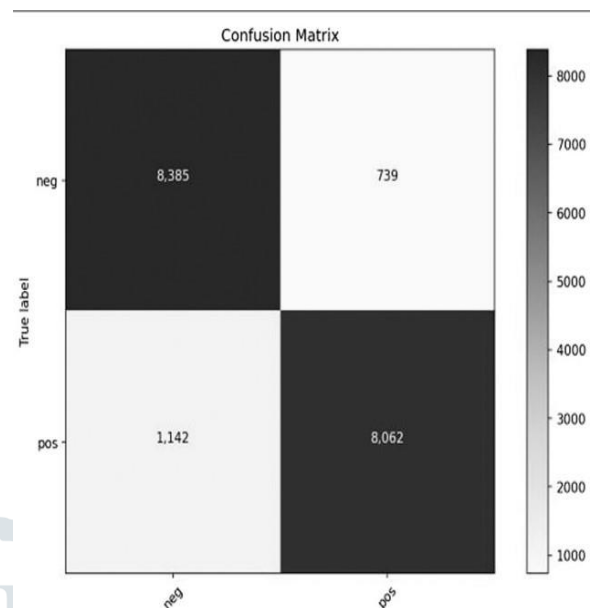
Demand analysis has widely used in number of the fields. Just like in the airline customers feedback, there is a huge amount of data set and for that deep learning technique has also been used as far as the literature survey is concerned. The sentiment analysis is basically the extracting information from any or entity and it will automatically identify the entity subjectivities. Similarly, sentiment analysis is a computational study of the people's sentiments, emotions, or opinions. And the social media or the reviewer site that is Twitter, blogs, microblogs has also used and all are merged. The user mode can be calculated using the tweets positive and negative term. Here the positive is represented by P and negative is represented by N. The sentiment score which is going to calculate is. $P - N / P + N + 2$. Sentiment score is represented by S.

Deep learning method is basically a subfield of the machine learning based models and it includes the composition of artificial neural networks. There are many algorithms which has also been implemented in the airline using the machine learning models, but the small amount of data and high accuracy. Is achieved. As far as the data set is concerned, huge amount of data set will lead to highest accuracy or defended accuracy. In deep learning algorithm for the proposed model, the classification has been done as follows. Initially the data set has been taken.

IV. MATRIX ANALYSIS:

The data set is undergone with the preprocessing and the text preprocessing. These text preprocessing include the stop word, lower casing, tokenization. And the huge amount of data set is segregated as a result in positive and negative aspect. The prediction is based on the test and trained model, whereas the

data split is depended on the train and test model. And the word embedding has also been done to define the classification. The evaluation is done in this study on different aspects like accuracy, F1 score and precision.



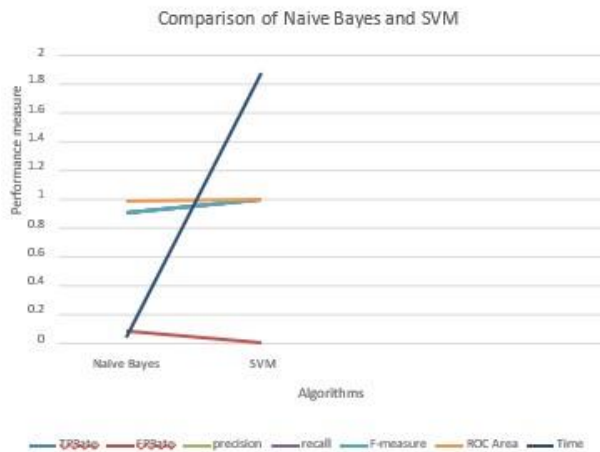
The data set includes the text speed processing, data splitting and feature extraction. The text preprocessing includes the removal of the words like Hew, She, It or any stop words. The recurrent neural network model has also been used in the sentiment analysis for deep learning models. And huge amount of data set has also been captured from the literature review purpose. The matrix which is received after the evaluation are segregated into two actual values, that is positive and negative. On the X axis and Y axis. The seaborn confusion matrix with labels has also been shown. Onto the below figure. Negative aspect with respect to the positive 1900 whereas +2 positive aspect are 5271. As the predicted values.

We can also count the various factors with respect to the sentiment on three aspects, positive, negative and neutral aspect.

V. RESULT:

The implementation is done in the study and it has been verified that 85% of accuracy has been achieved with the help of Support Vector Machine classifier. That one score and precision analysis has also worked on three aspects, that is accuracy, macro average and weighted average. The overall negative and positive

sentiment tweets has also been analyzed. And it lost the neighbors algorithm and the support vector machine. Performance has been graphically represented. Such representation will clearly detect the performance.



References:

- [1] Augustyniak, Łukasz, et al. "Comprehensive Study on Lexicon-based Ensemble Classification Sentiment Analysis." *Entropy* 18.1 (2015): 4.
- [2] Bandgar, B. M., and Binod Kumar. "Real time extraction and processing of social tweets." *International Journal of Computer Science and Engineering, E- ISSN 2347-2693* (2015): 1-6.
- [3] Bhattacharyya, Pushpak, et al. "Facilitating multi-lingual sense annotation: Human mediated lemmatizer." *Global WordNet Conference*. 2014.
- [4] Bird, Steven. "NLTK: the natural language toolkit." *Proceedings of the COLING/ACL on Interactive presentation sessions*. Association for Computational Linguistics, 2006.
- [5] Bird, Steven, Ewan Klein, and Edward Loper. *Natural language processing with Python*. "O'Reilly Media, Inc.", 2009.
- [6] Cambria, Erik. "Affective computing and sentiment analysis." *IEEE Intelligent Systems* 31.2 (2016): 102-107.
- [7] Carrillo de Albornoz, Jorge, Laura Plaza, and Pablo Gervás. "A hybrid approach to emotional sentence polarity and intensity classification." *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2010.
- [8] Carvalho, Jonnathan, Adriana Prado, and Alexandre Plastino. "A Statistical and Evolutionary Approach to Sentiment Analysis." *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02*. IEEE Computer Society, 2014.
- [9] Chowdhury, Gobinda G. "Natural language processing." *Annual review of information science and technology* 37.1 (2003): 51-89.
- [10] "Company | About." *Twitter*. Twitter, 30 June 2016. Web. 04 Mar. 2017.
- [11] Esuli, Andrea, and Fabrizio Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining." *Proceedings of LREC*. Vol. 6. 2006.
- [12] Fernández-Gavilanes, Milagros, et al. "Unsupervised method for sentiment analysis in online texts." *Expert Systems with Applications* 58 (2016): 57-75.
- [13] Firmino Alves, André Luiz, et al. "A Comparison of SVM versus naive-bayes techniques for sentiment analysis in tweets: a case study with the 2013 FIFA confederations cup." *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web*. ACM, 2014.
- [14] Gastelum, Zoe N., and Kevin M. Whattam. "State-of-the-Art of Social Media Analytics Research." *Pacific Northwest National Laboratory* (2013).
- [15] Gonçalo Oliveira, Hugo, António Paulo Santos, and Paulo Gomes. "Assigning Polarity Automatically to the Synsets of a Wordnet-like Resource." *OASIS-OpenAccess Series in Informatics*. Vol. 38. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.
- [16] González, Cristóbal Barba, et al. "A Fine Grain Sentiment Analysis with Semantics in Tweets." *International Journal of Interactive Multimedia and Artificial*

- Intelligence 3.Special Issue on Big Data and AI (2016).
- [18] Hemalatha, I., Dr GP Saradhi Varma, and A. Govardhan. "Case Study on Online Reviews Sentiment Analysis Using Machine Learning Algorithms." *International Journal of Innovative Research in Computer and Communication Engineering* 2.2 (2014):3182-3188.
- [19] Hemalatha, I., Dr GP Saradhi Varma, and A. Govardhan. "Preprocessing the informal text for efficient sentiment analysis." *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* 1.2 (2012): 58-61.
- [20] Hemalatha, I., Dr GP Saradhi Varma, and A. Govardhan. "Sentiment analysis tool using machine learning algorithms." *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* 2.2 (2013): 105-109.
- [21] Hull, David A. "Stemming algorithms: A case study for detailed evaluation." *JASIS* 47.1 (1996): 70-84.
- [22] Isah, Haruna, Paul Trundle, and Daniel Neagu. "Social media analysis for product safety using text mining and sentiment analysis." *Computational Intelligence (UKCI), 2014 14th UK Workshop on.* IEEE, 2014.