DEEP SENTIMENT ANALYSIS ON SOCIAL THREADS FOR PUBLIC OPINION ON INDIAN GOVERNMENT DECISION

Er. Sukhdeep Singh[1], Er.Sumreet Kaur[2]
Assistant Professor, Institute of Engineering and Technology –Bhaddal, Ropar India

***Abstract:** The social networks are adding more number of users every year and rising their popularity. There are several political opinion, product review, service review or other threads from where a lot of information can be collected using the opinion mining or emotion mining. Sentiment analysis is a branch of text mining which deals with the emotion analysis under the sub-category of emotion mining. The sentiment analysis can give a public opinion or review about any political news, product or several other similar things. In the paper, we are proposing the use of deep sentiment analysis on the social network threads in order to know the public opinion. The public opinion will be calculated on the basis of multi-keyword phrases to improve the accuracy of the sentiment analytical system. Also the proposed model will be made capable of calculating the emotion from the smiley's written in the text or attached in the form of the images. The image based (for smiley images) emotion extraction method will developed to achieve the goal of emotion calculation from smiley images. The results would be measured using the elapsed time, accuracy, precision, recall and false predictive ratio.*
***Keywords:** opinion mining, text mining, sentiment, sentiment analysis, social sentiment.*

I. INTRODUCTION

Data extracted from social networking websites is unstructured and fuzzy in nature. In everyday life conversations as noticed on social networking websites, people do not care about the spellings and accurate grammatical construction of a sentence that leads to different types of uncertainties, such as lexical, syntactic, and semantic. So, analyzing and extracting information patterns from such data sets are more complex. A significant amount of research has already been carried out to categorize data/sentence into various categories of emotion. Most of the emotions that have been worked upon are either positive or negative or finding the polarity i.e. the level of emotion expressed by the author. India being the largest democracy, gives its people the right to express their emotions and sometimes some ill willed people take due advantage of such facts. As the social media is becoming a need of life today, so is the interaction/discussions of people through social media has also increased. A great impact of all this can be seen on the Indian Army Fans page on Facebook [9]. The page that is followed by around 1.6 million people in the country has many thought leaders that can easily influence the thought process of several others. And it can also be said that it can be really helpful to control various difficult situations in the country if we identify those thought leaders who mainly

circulate negative (anger/fear) emotions among people. This will mainly help to control situations like riots, anger among people due to false information.

2.secondly the concept of smiley which people mostly put in comments instead of text in which we will calculate that the person is happy ,sad ,anger etc .Below are some examples of similes used in facebook. The proposed model will be made capable of understanding the emotion created by the smileys. The images of the smiley will undergo the emotion analysis on the images, which is a method to extract the emotion from the emotions. The emotion calculated from the smiley images will be added to the total emotion of the message.

So as to achieve this, an accurate understanding of how emotions are represented both in the human mind and in the computer environment is essential in the study of affect detection. The relationship between emotion and text is also important when mapping textual information to emotion space. In general, the study of emotions in written text is conducted from two opposite points of view. 1. The first is the viewpoint of a writer. This is concerned with how emotions influence a writer of a text in choosing certain words and/or other linguistic elements. 2. The second point of view is concerned with how a reader interprets the emotion in a text, and what linguistic clues are used to infer the emotion of the writer. In this work, the second point of view is taken into consideration because we are interested in the way people infer emotions. When an event occurs, each individual has his own perceptions and his own thought process that leads to a reaction regarding that event. As each one of us react in a different way so is the way to express our emotions is also different either verbal or textual. Now if we study only the textual expression of a group of people regarding an event, we shall have a variety of text, with different languages and ways to express. This becomes quite difficult to mine the relevant information from such variety of text. So as to make this task much simpler and efficient we need to work onto certain areas if we are working with the dictionary based techniques. We are able to prepare bag

of words for anger and fear especially effective for data obtained from Facebook which has posts in English. But still it needs more improvement with regard to different types of data. One limitation of the model is that as the data sets contain the text that has posts/comments in different languages like Hindi, Urdu, Tamil and Punjabi. So the model is unable to consider comments in such languages. Also we need a much rich bag of words that can provide us more reliable and precise results. While study many other researches and the results delivered by many online tools available for the categorization of data into various emotions are not effective. Most of them provide only 60% of the correct result, which are much less than the required amount. A lot of scope is there to work on the specialized tools for various emotions that contain most commonly used language specific bag of words so as to achieve more precise results. Also we need to be extra careful while categorizing a word into a particular category of emotions, because there are various words that represent different emotions in different context. So the most of the emphasis must be on obtaining a perfect bag of words, where we live in a "nothing is perfect" era.

II. LITERATURE SURVEY

Poria, Soujanya et. al. [1] has proposed the use of Sentic patterns for the purpose of the sentiment analysis from the social data. The authors proposed the use of dependency-based rules for concept-level sentiment analysis. In this work, the authors have introduced a novel paradigm to concept-level sentiment analysis that merges linguistics, common-sense computing, and machine learning for improving the accuracy of tasks such as polarity detection. Yassine, Mohamed et. al. [2] has worked on the development of a framework for emotion mining from text in online social networks. This paper presents a new perspective for studying friendship relations and emotions' expression in online social networks where it deals with the nature of these sites and the nature of the language used. It considers Lebanese Facebook users as a case study. The technique adopted is unsupervised; it mainly uses the k-means clustering algorithm. Narayanan, Vivek et. al. [3] has worked on a fast and accurate sentiment classification using an enhanced Naive Bayes model. The authors have explored different methods of improving the accuracy of a Naive Bayes classifier for sentiment analysis. They have also observed that a combination of methods like effective negation handling, word n-grams and feature selection by mutual information results in a significant improvement in accuracy.

III. RESEARCH GAPS

The existing scheme is entirely based upon the use of sentic pattern analysis of the purpose of emotion detection from the dataset collected from the social media data. The existing scheme evaluates various emotions like joy, serenity, sadness, grief, surprise, interest, rage, anger, fear, terror, trust, admiration, disgusting, etc. The existing model is aimed at fetching the words containing the emotions and then calculates the message score on the basis of the points pre-

stored in the word database. The existing model is completely based upon the knowledge-based dataset. The existing model does evaluate the points on the basis of the concept-level dependency based rules on the combination of the words. The existing model is not capable of evaluating the three word combination, e.g. "pretty big fear", "extreme anger burst", etc.

IV. SOLUTION DESIGN PROPOSAL

The proposed work will be based upon the evaluation of the message based on the one, two or three word combinations. The proposed work will be using the large knowledge based database for the purpose of tokenization, which is the process to fetch the emoticon words out of the message. The emoticon words the processed for the score calculation based upon the single words and word combinations in the knowledge database. The proposed model will be able to learn the new word combination based upon the knowledge database for learning. The proposed model will also offer the user feedback for the higher accuracy and a human-software intelligent interaction, which can help the machines to learn the new rules for the higher accuracy than the existing system.
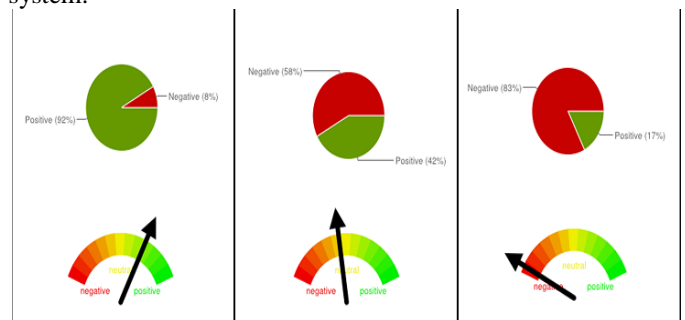

Fig 1.1

In fig 1.1 shows the use of a rich set of sentiment analysis features like positive, negative, fear, anger, etc. The proposed feature selection method can improve opinion classification performance. The proposed Feature Relation Network is rule-based sentiment classification method that finds the text features like anger and fear also. The proposed algorithm consisted of four basic components: Post/Thread Acquisition, Tokenization, and Polarization & Negative Emotion Analysis.

STEP 1:- Post acquisition will be used to read the post saved in the excel file.

STEP 2:- **Tokenization** module, the program reads the

known words from the user comment based on the word list prepared on the basis of most common words. This word is list is obtained by matching the words in the user comments along with a valid word list file already saved as a text file. The text file is loaded into the memory and passed to the tokenization process for the further computations. The process of tokenization then extract all of the words from the user comment and filter them on the basis of a list of common words containing no emotion.

STEP 3:- filtered list is obtained after deleting those matching words in the user comment. These common words are not given in the word weight file containing the rank/weight of each word being used in the common English language, which contains a neutral, positive or negative emotion. Then the user comments are polarized in three major categories under this step. The three major categories are positive, negative and neutral. The tokenized comments are compared with a list of words. The file contains the ranking for each of the word listed on the list. The rank or weight or strength of the words has been listed in the document, which ranges between -5 to +5. The words are classified on the basis of their use and its impact in the natural English language spoken in our daily lives. All of the comments marked as negative undergoes the negativity analysis, which checks the comments for the different negative emotions. The negative emotions analyzed under this step are anger, fear and mixed emotion (of Anger and Fear). The user comment is compared with two different files, out of which one is containing the words representing anger and other for fear. The comment is marked on the basis of higher weight. For example if anger is far more than fear, then the comment will be marked as it is spreading the anger. If anger and fear will be equal, almost equal and non-zero, then it will be marked as both anger and fear under the type mixed emotion. All of the comments marked as positive undergoes the positivity analysis, which checks the comments for the different positive emotions. The positive emotions being analyzed under this study will be joy, hurray, normal, highly satisfied, etc. The point calculation will be similar to the negative emotion technique.

V. METHODOLOGY

At very first step, a detailed literature survey on sentiment analysis and emotion mining algorithms will be conducted to find the pros and cons of the existing algorithms. Also a literature study will be conducted on the other mining algorithms based on the knowledge based or learning based solutions. The best sentiment analysis algorithm out of all of the surveyed ones would be used for the purposed of the new sentiment analysis solution. A series basic social datasets will be prepared which will be used to conduct all of the experiments on the proposed model implementation. Afterwards, the simple sentiment analysis system would be developed, which will work with single word fetching and evaluation to produce the emotion mining results. Then the two word and three word combination evaluation models would be developed using the latter as the base of the sentiment analysis solution. Then the model would undergo debugging process and the results would be obtained and analyzed as the last step of the research.

VI. CONCLUSION & FUTURE WORK

The proposed system will be applicable on wide variety of the social network threads to calculate the user sentiment in order to examine the public opinion on any product, political news or service provided to them. The public opinions are quite important to reform the strategies for the governments, companies or other organizations in order to improve their decisions or services. The deep sentiment analysis gives a wide angle view about the public opinion about their decision, which may help them to improve it in better order than the simple positive-negative based sentiment analysis. Also, the emotion calculation from the images attached by users such as smiley's or other similar emoticons has been analyzed for the emotional weight. The smiley based emotion calculations emphasize the emotion created by the user in their messages, hence it becomes important to calculate such emotion in order to increase the accuracy of the system. In the research, a semantic sentiment analysis approach would be described and developed. The higher accuracy and low count of false positives is the major target of this development.

REFERENCES

[1] Poria, Soujanya, Erik Cambria, Grégoire Winterstein, and Guang-Bin Huang. "Sentic patterns: Dependency-based rules for concept-level sentiment analysis." Knowledge-Based Systems 69 (2014): 45-63.

[2] Yassine, Mohamed, and Hazem Hajj. "A framework for emotion mining from text in online social networks." In Data Mining Workshops (ICDMW), 2010 IEEE International Conference on, pp. 1136-1142. IEEE, 2010.

[3] Narayanan, Vivek, Ishan Arora, and Arjun Bhatia. "Fast and accurate sentiment classification using an enhanced Naive Bayes model." In Intelligent Data Engineering and Automated Learning–IDEAL 2013, pp. 194-201. Springer Berlin Heidelberg, 2013.

[4] Cambria, Erik, Yangqiu Song, Haixun Wang, and Newton Howard. "Semantic multidimensional scaling for open-domain sentiment analysis." Intelligent Systems, IEEE 29, no. 2 (2014): 44-51.

[5] GunWooPark, SooJin Lee, SangHoon Lee, "To Enhance Web Search Based on Topic Sensitive_Social Relationships Ranking Algorithm in Social Networks", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology – Workshops, 2009.

[6] Ronen Feldman, 'Techniques and Applications for Sentiment Analysis', Communications of the ACM, Vol. 56 No. 4, Pages 82-89, 2013.

[7] Federico Neri, Paolo Geraci&FurioCamillo, 'Monitoring the Web Sentiment, the Italian Prime

Minister's case', International Conference on Advances in Social Networks Analysis and Mining, 2010 IEEE.

[8] Baumer E. P. S., Sinclair J. & Tomlinson B.,'America is like Metamucil: Fostering critical and creative thinkingabout metaphor in political blogs', In Proceedings of 28th International Conference on Human Factor in Computing Systems ACM, Atlanta, GA, USA,pp. 34-45, 2010

[9] Sorensen L.,'User managed trust in social networking comparing facebook, myspace and linkdin', In Proceedings of 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory andAerospace & Electronic System Technology, (Wireless VITAE 09), Denmark, pp. 427-431, 2009.

[10] Evans B. M., Kairam S. &Pirolli P., 'Do your friends make you smarter: An analysis of social strategies inonline information seeking',Information Processing and Management, 46 (6), pp. 679-692, 2010.

[11] Li J., Li Q., Khan S. U. &Ghani N.,'Community-based cloud for emergency management', In Proceedingsof the 6th IEEE International Conference on System of Systems Engineering (SoSE), Albuquerque, USA, pp. 55-60, 2011.

[12] Li J. & Khan S. U.,'On how to construct a social network from a mobile ad hoc network', Technical Report,NDSU-CS-TR-09-009, 2009.

[13] Lionel Martin, 'User Behavior Under the Influence of Groups in Social Media', EDIC Research Proposal, 2009.