International Journal of Education and Science Research Review
E-ISSN 2348-6457
Volume-9, Issue-5 Sep - Oct – 2022
P-ISSN 2349-1817
www.ijesrr.org
Email- editor@ijesrr.org

STUDY ON FACEBOOK POSTS USING HYBRID METHOD RELATED TO SENTIMENT

Patil Yogendra Vana
Research Scholar
School of Science & Technology
Glocal University Saharanpur (U.P)

Dr. Praveen Kumar
Research Supervisor
School of Science & Technology
Glocal University Saharanpur (U.P)

ABSTRACT:

The analysis is further improved by incorporating language discourse features to detect intensity of sentiment and the prominent emotions expressed through these posts. Social Media is a popular medium of communication amongst youngsters to remain connected with their friends. Facebook is one of the most preferred Social Media Sites which store the gigantic amount of data which can be explored for Sentiment Analysis. In this study, we have applied hybrid analysis approach which combines the best features of a lexical analysis and SVM machine learning classification algorithm on Facebook Posts.

KEY WORD: *Emotion lexicons, Hybrid Analysis, Sentiment Analysis, Social Networking Sites, Support Vector Machine.*

INTRODUCTION

Nowadays, enormous information is available on Social Networking Sites due to the penetration of the Internet amongst 4.2 billion users worldwide1. People use Social Networking Sites as the most popular and fast communication medium to remain in touch with their friends, family, and peer. There are 3.03 billion active Social Networking Sites users1, most of them range in the age group of 18-49 years who share their emotions, photos, daily life activities, chats, opinions about products, politics, social issues, movies and many more. These sites play a vital role in spreading mass opinion thus can be used to build a positive public opinion about different societal issues [1]. Sentiments spread through Social Networking Sites are contagious which can be used as a tool for the well-being of mankind [2], [3], [4]. Sentiment analysis is the systematic process of collecting and analyzing emotions from the enormous volume of unstructured online data in real time. People, through the medium of the Internet, use various Social Networking Sites, blogs forums etc., to express their feelings and thoughts. Their sentiments can include a situation, event or object [5].

Sentiment analysis generally uses a lexical method or machine learning method for detecting sentiment polarity of data. The lexical analysis uses lexicons to identify the semantic orientation of the textual data while machine learning classifier requires a labeled dataset for classification. In our study, we have tried to identify the basic emotions underlying Facebook Posts of youngsters. In a nutshell, the progression of our research begins by collecting user-generated posts from Facebook. The extracted posts are cleaned, transformed, and accordingly classified into positive or negative sentiments. We recommend a hybrid method for sentiment analysis which combines features of both methods. Each sentence is evaluated, and the overall score is combined to predict the sentiment polarity, the degree of sentiment and the basic emotions exhibited by him. In the end, the inclination of a youngster towards negativity is identified through the level of negativity found in his posts. Section II reviews the related work carried out by the researchers in this area. Section III describes the methodology used to detect emotions using a hybrid approach. Section IV gives information about experimental setup. Section V highlights the results and comparative analysis of different approaches used. The conclusion is discussed in section VI.

II. RELATED WORK

Sentiment analysis incorporates several tasks to generate contextual knowledge from a huge textual data starting with systematically gathering data and concluding with the concrete results which can be useful in the design of opinion mining system [7]. Self-articulation is an imperative use of social media, in the form of sharing comments, activities, and happenings of daily life, sharing the opinion about different things, etc. Social media is penetrated in our lives in such a way that it has started dominating face to face interaction with virtual communication. The reason behind this changing dynamic is the widespread use of Social Media by young users. Facebook is immensely helpful in keeping touch with family and friends that live at distant. With its unique features of providing posts, photos, and profile information, it makes you aware of day to day happenings of your circle of friends and family. On Facebook, you can have thousands of friends as it offers you with a facility to add friends without any complicated technicalities. Using machine learning techniques, Facebook posts can also be used to find the personality traits of a user [11]. Some authors used Naïve Bayes classifier to identify how people feel about certain topics [12] while others used Support Vector Machine and Naïve Bayes classifiers, to classify Facebook posts of Tunisian users for studying their behavior and state of minds during Arabic Spring Era [13]. In very few studies, the hybrid analysis approach which combines both lexical and machine-based techniques are used to detect the emotions of the user from the contents posted on Facebook. The authors have applied hybrid analysis to create an application called SentBuk and used it for

effective e-learning [14]. In another study, the authors have used regular expression-based rules and statistical text mining techniques to predict sentiments expressed in suicide notes [15].

III. HYBRID SENTIMENT ANALYSIS

The hybrid analysis combines both lexical and machine learning analysis techniques and compensates shortcomings of each technique to generate better results [16]. In our study, the analysis is based on the combination of the algorithm used to calculate the aggregate sentiment score of lexicons expressed by the post and Support Vector Machine classifier. The score works as a new feature for the SVM's training dataset. The proposed approach has the advantage of having the improved score of the lexicons obtained by applying the aggregate sentiment scoring algorithm which takes into consideration the language discourse features and the flexibility of the SVM. Fig. 1 shows the proposed method's flow.

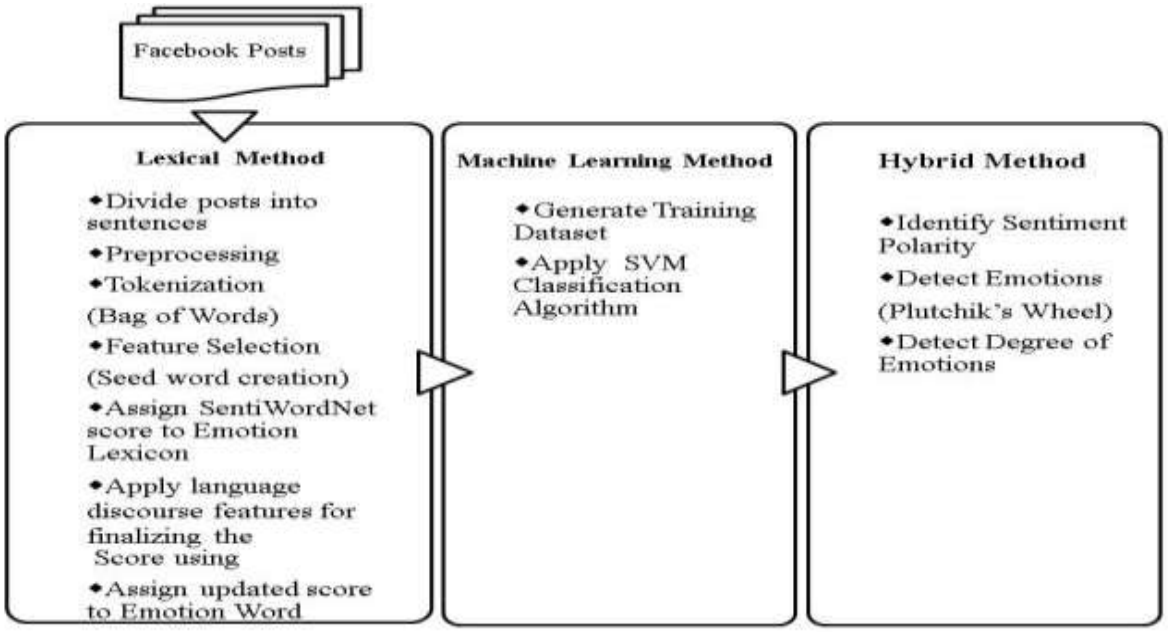

Fig. 1. Proposed architecture of Hybrid Analysis

A. *Phase I using Lexical Analysis method*

The lexical analysis begins with the creation of seed words i.e. emotion lexicons. The seed words are precompiled, context-oriented emotion lexicons and incorporate emotion lexicons related to the domain of analysis. Each time a new post is incorporated, it is tokenized, and each token is matched against the seed words. The user post is analyzed using seed words and the sentiment of data is decided based on the combined polarity of each emotion lexicon found in the post [17] [18].

a. Preprocessing

To get more accurate results, the unstructured data is subjected to preprocessing before being used for finding sentiment polarity. It involves the removal of unwanted special symbols except '!' or '?', removal of repeated characters, unnecessary spaces, spelling corrections, conversion of data to lowercase, conversion of abbreviations to full forms, changing Internet slangs to their respective full forms, stemming etc. The entire post which is in the form of a paragraph is segregated into sentences. Each sentence is then tokenized into its constituent tokens. Hence each token can be analyzed individually.

b. Emotion Lexicon creation

The emotion lexicons are the words which represent strong semantic orientation. They are context specific words used to express human feelings. The research work centers on eight basic emotions given by Plutchik's wheel of emotions [19]. Hence, we have stored the emotion lexicons (Seed Words) depicting emotions like joy, trust, surprise, anticipation fear, anger, disgust, and sadness in a database. Sentiment polarity and weight associated with each emotion word are assigned using SentiWordNet2 and Hindi SentiWordnet3.

c. Feature Selection

In sentiment analysis, feature selection is considered as one of the most important tasks. Once the dataset is tokenized and brought to a stage where one can identify individual token, features are selected by matching the tokens with the emotion lexicons stored in the database exclusively created for storing lexicons expressing emotions. The features can be selected based on their presence or frequency.

The features selected are assigned binary values (0 or 1) when selected based on presence or absence of sentiment in the feature vector or it can be stored as an integer or decimal value used to depict the intensity of sentiment in the data. The core of accurate sentiment polarity depends on the extraction of emotion-oriented words called features from this informal data. Better the selection of emotion words accurate would be the result of the sentiment polarity.

d. Prediction of Sentiment Polarity and Sentiment Degree

Each emotion word is assigned the corresponding value of its emotion score using SentiWordNet 4.0. When an emotion word is detected in the post corresponding score of its emotion value is assigned to the word. The emotion score of all lexicons present in the sentence is added to calculate the total score of the sentence. Based

on the total score of the emotion lexicons, the polarity of the sentence is identified as Negative, Positive or Neutral. The degree of sentiment expressed through the sentence is calculated by summing the individual score of emotion lexicon along with the other details of emotion score as mentioned below.

1. Identifying domain specific emotion keywords: The accuracy of the classifier is greatly influenced by the context in which words are used in the sentence. Emotion words change with respect to a domain. Hence taking this into consideration only domain-specific words are selected.

2. Negation: When negation occurs in a sentence it mainly influences the original meaning of positive or negative emotion words by inverting their polarities [20], [21], [22], [23].

3. Double negation: It is observed that, if negation is used more than once in a sentence then it invalidates the effect of negation on emotion words. The emotion words in such sentences are generally adjectives or adverbs [24], [25].

4. Effects of conjunctions: Conjunctions are used to connect words, clauses or sentences. They provide meaningful information about the sentence.

5. The presence of conjunction in a sentence makes the calculation of polarity difficult. When it appears in a sentence, we need to find which part of the sentence contributes more to the final emotional polarity of the sentence [26].

6. Intensifiers and Diminishers: They increase or decrease the polarities of negative or positive emotion words. They do not have their own sentiment orientation, but their presence strongly conveys the sentiments which they are associated with. They never invert polarities of the emotion words [27].

7. Punctuation marks: The punctuations like an exclamation mark and question mark are used to further increase or decrease the strength of the emotion expressed. An exclamation mark used in a sentence conveys strong emotions such as surprise, astonishment and any other such emotions. It adds additional emphasis to the emotion expressed. In contrast, the question mark indicates confusion.

B. *Phase II using Machine learning method*

Lexical analysis generates better results when used for small datasets. If the emotion lexicon is not found in the Seed word database, then the dataset cannot be evaluated properly. In contrast to this, Machine Learning analysis works on many labeled training datasets and produces better results [28], [29], [30]. There are various popular classification algorithms which outperform lexical analysis. In Machine learning sentiment analysis, the bag-of-words feature selection method is used predominantly. The entire dataset is treated as a bag (group) of words where the sequence of words in a sentence is retained even after the removal of stop words and

stemming. Support Vector Machine (SVM) is a dominant linear classification algorithm. It treats the dataset as the points plotted in space. They are expected to be separated by enough space. It calculates a maximum margin hyperplane which divides the data points into two classes. In text classification, Support Vector Machine is considered the best classification algorithm [31], [32]. SVM algorithm generally divides the training dataset into minimum of two classes. These classes are separated from each other by the maximum possible distance drawn by the hyperplane. The sum of the distances of the closest points of the two classes from the hyperplane defines the margin of the classes.

The linear equation is given as under.

$$Y=BX+A \quad (1)$$

Where the point (X, Y) has two-dimensional values X, Y and A is a constant value.

A point with value X will be classified in which class is given by equation 2.

$$W=\sum_{j=0}^{n} \alpha j\, yj\, xj \quad (2)$$

Among the advantages of SVM is that it achieves excellent results in high-dimensional data with very few samples. It is robust to outliers and noise. It can learn both simple linear and very complex nonlinear functions by using kernel function. We are using multiclass SVM to classify multiple sentiment levels and different emotions exhibited by the posts. SVM applies a technique of one-versus-all to select the class which classifies the test data with optimal possible distance.

IV. EXPERIMENTAL SETUP

The methodology of this study is divided into two phases. The first phase begins with the lexical analysis method which includes various sub-steps like data extraction from Facebook Posts of youngsters collected over a period of three months. The next step is the creation of Emotion Lexicons (Seed Words). The third step is the selection of relevant features from the group of tokenized data and the last step is assigning final weights to selected features (emotion lexicons) by taking into consideration their discourse relation like negation, double negation, conjunction, use of intensifiers and diminishers present in the sentence. The second phase starts by feeding this training dataset to SVM classifier for identifying underlying sentiments as well as the degree of sentiments. Emotion scoring, and the results obtained after this phase are used to find the level of emotional distress. The degree of emotional distress is further divided as positive, medium, negative and neutral. To perform our experiment, we have used WEKA machine learning toolkit, version 3.8. The data is

divided into training dataset with 60% data and test dataset with 40% data. Information Gain and Ranker algorithms are used for feature selection. The dataset is trained using Support Vector Machine (SMO) classification algorithm implemented with 10-fold validation. We evaluated the performance of the hybrid analysis using various parameters like the presence of emotion lexicons, frequency of emotion lexicons, and the discourse relation. These features are the deciding factors for predicting the level of sentiments expressed through the Facebook post.

V. PERFORMANCE EVALUATION

With lexical analysis, we have received 78.05% accuracy in sentiment polarity and 70.41% accuracy in sentiment degree. When we have analyzed the dataset using the SVM algorithm without considering the discourse features found in the post, we have got sentiment polarity with an accuracy of 96% and sentiment degree with 95.41% accuracy. After applying the hybrid analysis for various parameters, we have achieved the results as shown below.

Table I. Comparison of Hybrid Analysis with presence and frequency of emotion lexicons for predicting Sentiment Polarity

Metric	Sentiment Polarity	Degree of Sentiment
Hybrid analysis by considering the presence of emotion lexicon	94.54%	-
Hybrid analysis by considering the frequency of emotion lexicon	94.50%	90.17%

The results shown are in Table I indicate that the mere presence of emotion lexicon or its frequency in the post does not affect the polarity of sentiment expressed through the post. Based on the presence of emotion lexicon alone, it is difficult to predict the degree of sentiment.

Table II. Comparison of Hybrid Analysis with emotion lexicon score and different features for predicting Sentiment Polarity

Metric	Sentiment Polarity	Degree of Sentiment

Hybrid analysis by considering negation, double negation, and conjunction	87.95%	85.58%
negation, double negation, conjunction, intensifier, and diminishers	88.15%	85.21%

The hybrid analysis approach has detected sentiment polarity with 87.95% accuracy when improved lexicon scores coupled with language discourse relation are taken into consideration. These results are further improved with the incorporation of intensifiers and diminishers.

Improvement in the prediction of True Positives

Table III. depicts that there is a decrease in false negatives with the incorporation of the frequency of emotion lexicons than just their presence in the posts.

Predicting Degree of Sentiment using different sentiment levels

To predict the degree of sentiment exhibited by the post, we have categorized sentiments into five different levels namely very positive, positive, neutral, negative and very negative. The results are shown in Table V also depict that the degree of sentiment is better predicted by sentiment frequency than its score. On the contrary, the results obtained using frequencies are misleading because they simply represent the occurrence of emotion lexicons in the given post. They do not consider the SentiWordNet score assigned to these lexicons. It is also noted that classifier is not able to categorize posts into very positive or very negative categories successfully because of their very less percentage in the dataset. Hence, we have decided to consider only four levels namely positive, moderate, negative and neutral for further analysis.

Our study also aims to analyze the predominant emotions expressed through these Facebook posts of young people. Generally, Facebook posts of youngsters of this age are found to express their feelings of love and affection apart from self-expression. Hence, we have also incorporated ‘love’, a complex emotion obtained by combining emotions of joy and trust in our study. Fig. 2 represents the distribution of emotions found in these posts. The emotion of love leads the other emotions which are mostly positive emotions. The other notable emotions are joy, sadness, and anger. The emotions of anticipation, trust, surprise, fear, and disgust are found in a handful of posts.

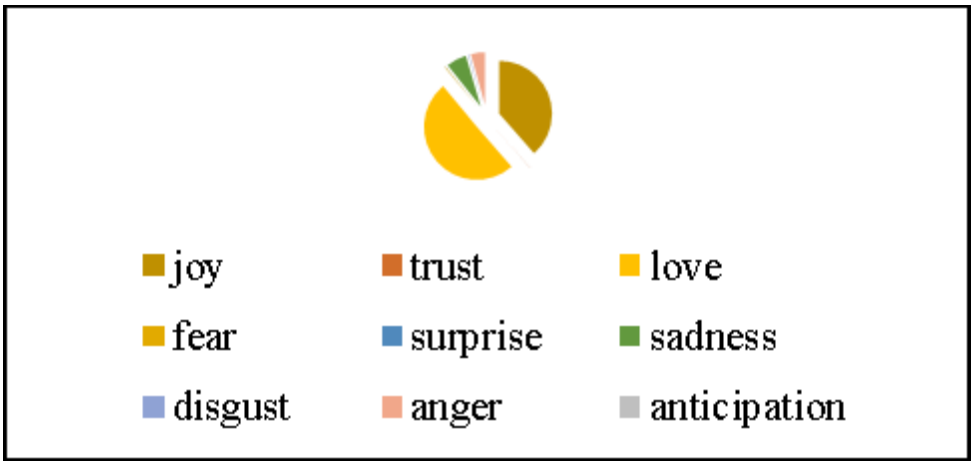

Fig. 2. Distribution of Emotions found in Facebook Posts

The underlying basic emotions found in Facebook Posts when detected using hybrid analysis are as shown in Table VII. The analysis is performed using various parameters like presence, frequency, language discourse features and the intensity of emotion lexicons.

Table VII. Comparison of Hybrid Analysis for predicting Emotions

Metric	Emotion Prediction
Hybrid analysis with the presence of emotion lexicon	94.69%
Hybrid analysis with the frequency of emotion lexicon	94.38%
Hybrid analysis with language discourse relation of emotion lexicon	71.25%
Hybrid analysis with language discourse relation of emotion lexicon and its intensity	86.52%

The emotions are predicted more accurately when the hybrid analysis is performed using language discourse relation along with the intensity of emotion lexicon present in the post. Table VIII represents the results using the confusion matrix. The results obtained based on the presence or frequencies of the emotion lexicon in the post are ambiguous. The classifier has not predicted emotions of anger and sadness accurately. The probable reason could be their lesser number in the dataset.

VI. CONCLUSION

We have proposed a hybrid method of SVM and lexical method coupled with language discourse relation to detect eight primary emotions from Facebook Posts based on Plutchik's wheel of emotions. We have analysed the performance of the proposed method by using different parameters and have found that it has given better results as compared to lexical or machine-based analysis. Detection of emotions from Facebook posts has given promising results when emotion scores of lexicons were taken into consideration.

REFERENCES

1. F. Cheong, and C. Cheong, "Social Media Data Mining: A Social Network analysis of Tweets during the 2010-2011 Australian floods," *Information Systems* (PACIS), 2011[*15th Pacific Asia Conf.*, July 7-11, 2011, pp. 1-16].

2. D.I. Kramer, J. E. Guillory, and J. T. Hancock, (2014, June). Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*, 111 (24), pp. 8788-8790. Available: https://www.pnas.org/content/111/24/8788

3. R. Lin, S. Utz (2015, November). The emotional responses of browsing Facebook: Happiness, envy, and the role of tie strength. Computers in Human Behavior, 52, pp.29-38. Available: https://doi.org/10.1016/j.chb.2015.04.064

4. M. Settanni, D. Marengo (2015, July). Sharing emotion online: studying emotional well-being via automated text analysis of Facebook post. Frontiers in Psychology, 6, pp.1045. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4512028/

5. B. Liu*, "Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing,", 2nd ed. Chapman and Hall: Florida, 2010.*

6. B. Liu, *Sentiment Analysis and Opinion Mining. Morgan & Claypool, 2012.*

7. E. Cambria, B. Schuller, Y. Xia, and C. Havasi (2013, March). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), pp. 15-21.

8. J. K. Akhter, S. Soria, *Sentiment Analysis: Facebook Post Messages*. Final Project CS224N, Spring, 2010.

9. R. E. Wilson, S. D. Gosling, and L.T. Graham (2012, May). A Review of Facebook Research in the Social Sciences. Perspectives on Psychological Science, 7(3), pp. 203 –220.

10. S. Mukherjee, and P. Bhattacharyya, "Sentiment analysis in twitter with lightweight discourse analysis," Proceedings of COLING 2012: Technical Papers, Dec. 2012, pp.1847–1864 [24th Int. Conf. Computational *Linguistics*, Mumbai, 2012, pp. 1847-1864].

11. G. Farnadi, S. Zoghbi, M. Moens, and M. De Cock, "Recognizing personality traits using Facebook Posts," AAAI Workshop Technical Report, Computational Personality Recognition, 2013, pp.14-18 [7th Int. AAAI Conference on Weblogs and Social Media, 2013].

12. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro*, "Sentiment analysis of Facebook Postes using Naïve Bayes classifier for language learning," IEEE, Oct. 2013, pp. 1-6 [IISA 2013, Greece, 2013, pp. 1-6]*.

13. S. B. Hamouda, and J. Akaichi (2013, May). *Social Networks' Text Mining for Sentiment Classification: The case of Facebook' Postes updates in the "Arabic Spring" Era". International Journal of Application or Innovation in Engineering & Management, 2(5), pp. 470-478.Available: https://pdfs.semanticscholar.org/aff6/bee30fcbafdc27132d5ad4e00a666f2fee3b.pdf*

14. Ortigosa, J. M. Martín, and R. M. Carro (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior, Elsevier, 31, pp.527–541.

15. J. A. McCart , D. K. Finch, J. Jarman, E. Hickling, J. D. Lind, M. R. Richardson, D. J. Berndt, and S. L. Luther (2012, January). Using ensemble models to classify the sentiment expressed in suicide notes. Biomed Inform Insights, 5(Suppl. 1), pp. 77-85.

16. R. Prabowo, M. Thelwall (2009, April). Sentiment Analysis: A Combined Approach. Journal of Informetrics, Elsevier, 3(2), pp. 143-157.

17. M. El-Din *(2016). Enhancement Bag-of Lexicons Model for Solving the Challenges of Sentiment Analysis. International Journal of Advanced Computer Science and Applications, 7(1), pp.* 244-252.

18. P. D. Blinov, M. V. Klekovkina, E. V. Kotelnikov, and O. A. Pestov. Research of lexical approach and machine learning methods for sentiment analysis. Vyatka State Humanities University, Kirov, Russia.

19. R. Plutchik. *A general psychoevolutionary theory of emotion*, New York: Academic Press, 1980.

20. Asmi, and T. Ishaya, "Negation Identification and Calculation in Sentiment Analysis",2012, pp. 1-7 [*2nd Int. Conf. Advances in Information Mining and Management*, 2012, pp. 1-7].