

Sentiment Analysis using Fuzzy based SVM Classifier

Er. Divya Kadyan

Computer Science Engineering, IET Bhaddal Technical Campus, Rupnagar, India

Dr. Sanjay Singla

Computer Science Engineering, IET Bhaddal Technical Campus, Rupnagar, India

Abstract : In this work, we proposed two techniques for feature extraction as well as classification. For this purpose, we have collected data for over 3 weeks. We considered the sentiments like angry, fear, joy etc. Based on feature extraction, application of fuzzy logic will be done and then classification will be done using neural network. The main advantage of proposed method is that it has high accuracy and low error rate. From the experiments conducted for this research, it is observed that about 97% of the accuracy has been found out with 1 to 3 error rate.

Keywords: Sentiment Analysis, social websites, Feature extraction, fuzzy logic and Support Vector Machines.

I. INTRODUCTION

In today's associated world, clients can send messages at any time. Online networking is not just utilized as an easygoing device for informing and sharing private things and considerations; it is likewise utilized by columnists, legislators and open figures, arrangement of organizations and colleges who need to be more open to people in general, share their reveries and appreciate supposition of people [1, 2, 3]. The following of nationals' responses in online networking amid emergencies has pulled in an expanding level of enthusiasm for the exploration group [4].

Experts take note of that the billions of productions left by individuals month to month that cannot be handled physically by holding popular sentiment surveys. This reality highlights the requirement for mechanized techniques for scholarly investigation of content data, what permits in a brief timeframe to process a lot of information and to comprehend the significance of client messages [5, 6, 7]. This comprehension of the significance of messages is the most essential and complex component of the mechanized preparing. Existing sentiment analysis occurs from the fields of normal dialect handling, computational phonetics, content mining, and a reach from machine learning techniques to run based techniques. Machine learning techniques include preparing of models on particular accumulations of records [8, 9].

Numerous analysts manage the assurance of assumption of individuals in different information gathered from online networking. They have utilized surely understood machine learning strategies for matching and classification information [10].

II. TEXT CLASSIFICATION PROBLEM

Problem of text classification has been shown below [11];

X= document

S= document space

C= set of document classes

T= classifier

$X \rightarrow C$. is the matching of text to classifier

Where $C = \{ \text{positive, negative etc} \}$

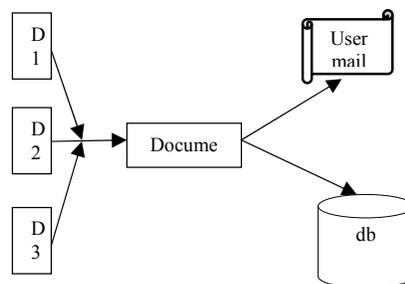


Figure.1 Text classification model

III. LITERATURE REVIEW

Various work has been done in this context. But this section will present recent analyzed work.

As indicated by **Balahur and Alexandra**, recent years are the years of development in the volume of examination in field of estimation investigation, particularly in subjective content sorts (like film or item surveys). The real contrast these subjective writings have with distributed news articles is that their objective is one of a kind and plainly expressed over the content. Taking after various comment endeavors and the examination of the issues experienced, they understood that news assessment mining of enormous information is not quite the same as that of other content sorts [12].

As indicated by **Jebaseeli and A. Nisha**, opinion mining of enormous information or Sentiment Analysis alludes to recognizable proof and arrangement of the perspective or supposition communicated in the content range; utilizing data recovery and computational semantics [13].

Mahalakshmi R and Suseela S, has proposed a strategy for conclusion investigation on twitter by utilizing Hadoop and its biological communities that will procedure the vast volume of information on a Hadoop and the MapReduce capacity will play out the assumption examination [14].

Informal community examination is a procedure for the most part created by sociologists and scientists in social brain science. Informal community examination sees social connections as far as system hypothesis, while singular performing artist being seen as a hub and relationship between every hub are introduced as an edge. Informal community investigation has been characterize in [15].

Horakova and Marketa, presented a model which gathers tweets from long range interpersonal communication locales and along these lines give a perspective of business insight. In the structure, there are two layers in the slant investigation apparatus, the information handling layer and notion examination layer. Information handling layer manages information accumulation and information mining, while feeling investigation layer utilize an application to display the consequence of information mining [16].

IV. SYSTEM ARCHITECTURE

Figure 2 shows the architecture of working model that has been proposed. It consists of training and testing phase.

The proposed methodology in this work will utilize the common dialect handling techniques like fuzzy logic and SVM to extract emotions from text present in various blogs. Fuzzy logic is easy to apply and understand. Mathematical concepts of fuzzy logic are also very simple. It is based on the natural language. Also SVM provides good training.

In training phase the feature extraction will be done using fuzzy logic ad classification will be done using SVM method. The whole workmodel is dependent on three types of text emotions i.e. happy, ad and angry.

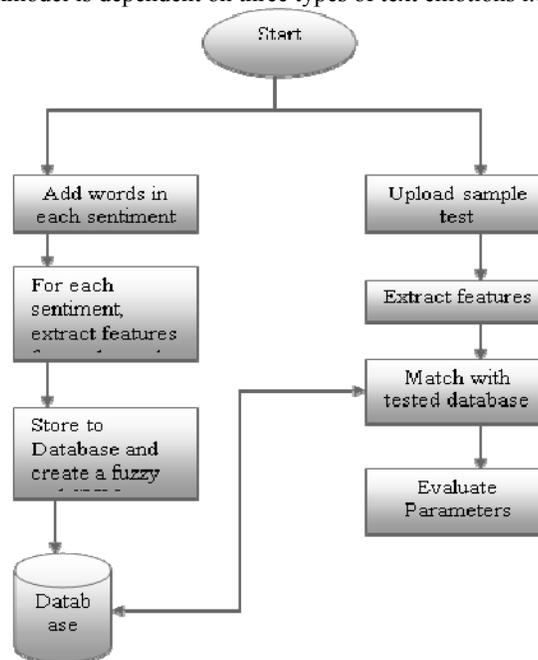


Figure. 2 Proposed WorkFlowchart

V. MATERIALS AND METHODS

In proposed work a new sentiment classification system has been built using fuzzy and SVM method. The steps of proposed algorithm are;

- ⇒ Add words to dictionary, Input tweets from social website
- ⇒ Read one document
- ⇒ Perform feature extraction
- ⇒ Apply fuzzy rules for expressions
- ⇒ Testing will be done using SVM
- ⇒ Read query from users.
- ⇒ Identify the sentiments
- ⇒ Put users in proper group
- ⇒ Validate results

VI. RESULT ANALYSIS

Proposed work is implemented in MATLAB 2010 environment. The experiments have been carried out on real collected dataset.

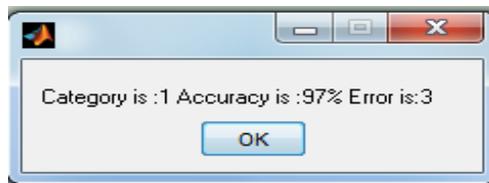


Figure.3 Parameter Evaluation

Above window shows the evaluation of the parameters and the obtained values are accuracy = 97% and error rate = 3 for category 1 for 1 iteration.

Table.1 Parameter Evaluation

| Iteration no. | Category | Accuracy (%) | Error rate |
|---------------|----------|--------------|------------|
| 1 | 1 | 97.11 | 2 |
| 2 | 2 | 97.22 | 2 |
| 3 | 1 | 97.11 | 1 |
| 4 | 1 | 97.22 | 2 |
| 5 | 2 | 97.24 | 1 |

Two metrics has been chosen for validation of the proposed work model i.e. accuracy and error rate. For good working model there should be high accuracy and less error rate.

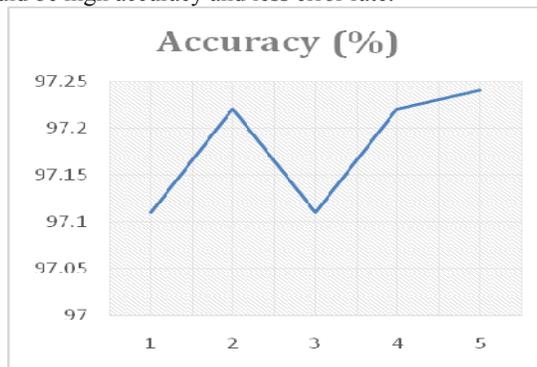


Figure.3 Accuracy Evaluation for proposed Method

Accuracy is the measure of the efficiency of the work model. It must be high for good model. In proposed work average accuracy of 97% has been achieved for 5 iterations.

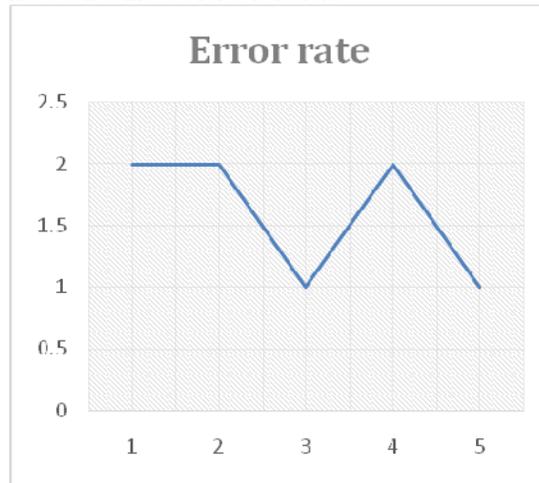


Figure.4 Error rate for proposed Method

Error rate is also the measure of the working efficiency of the work model. It must be low for good presentative model. In proposed work average error rate has been found to be lies between 1- 2for 5 iterations.

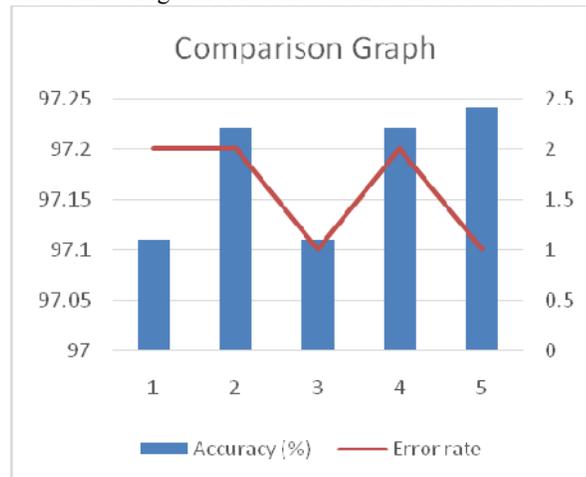


Figure.5 Parameter Evaluation Graphical Representation

Above figure shows the graphical representation of both metrics where blue area is for accuracy and orange area is for error rate. So, it can be seen that obtained error rate is low and accuracy is high for 5 iterations.

VII.CONCLUSION

Sentiment Analysis classification system has been built in proposed work. Training phase is the first phase where we train the system with various categories that will use to classify the user's queries. There are three type of emotion SAD, HAPPY, ANGRY. Different panel that shows the training of their corresponding textual set and knowledge base of system has been built in application model. Then text feature extraction and classification will be done using fuzzy logic and SVM. From observations the 97% of the accuracy has been achieved.

RECOMMENDATIONS

Future research goes in the direction of the utilization of the clustering method in order to further enhance the accuracy.

REFERENCES

- [1] Swati A. Kawathekar and Dr. Manali M. Kshirsagar, "Movie Review analysis using Rule-Based & Support Vector Machines methods", IOSR Journal of Engineering, Vol. 2(3) pp: 389-391, March 2012.
- [2] AlaaHamouda, Mahmoud Marei and Mohamed Rohaim, "Building Machine Learning Based Senti-word Lexicon for Sentiment Analysis", Journal of Advances in Information Technology, Vol. 2, No. 4, pp 199-203, November 2011.
- [3] Archana Shukla, "Sentiment Analysis of Document Based on Annotation", International Journal of Web & Semantic Technology (IJWesT) Vol.2, No.4, pp 91-103, October 2011.

- [4] Aurangzeb Khan, BaharumBaharudin and Khairullah Khan, "Sentiment Classification Using Sentence-level Lexical Based Semantic Orientation of Online Reviews", Trends in Applied Sciences Research, Vol. 6, pp. 1141- 1157, July, 2011.
- [5] DadvarMaral, C. Hauff, and Jong de, Franciska, "Scope of negation detection in sentiment analysis", Dutch-Belgian Information Retrieval Workshop, Netherlands, February 2011.
- [6] AnimeshKar, Deba Prasad Mandal, "Finding Opinion Strength Using Fuzzy Logic on Web Reviews", International Journal of Engineering and Industries, volume 2, Number 1, pp 37-43, March, 2011.
- [7] Adnan Duric and Fei Song, "Feature Selection for Sentiment Analysis Based on Content and Syntax Models", Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Portland, Oregon, USA, pp. 96-103, June, 2011.
- [8] Aditya Joshi, Balamurali A. R., Pushpak Bhattacharyya and RajatMohanty, "C-Feel-It: A Sentiment Analyzer for Micro-blogs", The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA, pp. 127-132. June, 2011.
- [9] V. Rentoumi, S. Petrakis, M. Klenner, G. A. Vouros and V. Karkaletsis, "United we stand: improving sentiment analysis by joining machine learning and rule based methods", 7th International Conference on Language Resources and Evaluation, Malta, pp. 1089- 1094, May 2010.
- [10] Julia Maria Schulz, Christa Womser Hacker and Thomas Mandl, "Multilingual Corpus Development for Opinion Mining", Proceedings of the 7th International Conference on Language Resources and Evaluation, Valletta, Malta, pp. 3409-3412, May 2010
- [11] Alena Neviarouskaya, Helmut Prendinger, Mitsuru Ishizuka, "Affect Analysis Model: novel rule-based approach to affect sensing from text", Natural Language Engineering, Cambridge University, Vol. 17, pp. 95- 135, September 2010.
- [12] Balahur, Alexandra, et al. "Sentiment analysis in the news." arXiv preprint arXiv:1309.6202 (2013).
- [13] Jebaseeli, A. Nisha, and E. Kirubakaran. "A Survey on Sentiment Analysis of (Product) Reviews." International Journal of Computer Applications 47.11 (2012).
- [14] Scholar, P. G. "Big-SoSA: Social Sentiment Analysis and Data Visualization on Big Data."
- [15] Bhumika1, Prof Sukhjrit Singh Sehra2, Prof Anand Nayyar3, "A Review Paper On Algorithms Used For Text Classification", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 3, March 2013.
- [16] Horakova, Marketa. "Sentiment Analysis Tool using Machine Learning." Global Journal on Technology (2015)..