

Result Comparison Of Naive Bayesian And SVM Opinion Mining Algorithm With Mobile Computing Multilingual Opinion Mining Algorithm

N.S. Patel^{1*}, H.B. Bhadka², D.B. Shah³

¹B.N. Patel Institute of Paramedical And Science, C.P. Patel and F.H.Shah Commerce College, Anand, Sardar Patel Education Trust, Sardar Patel University, Anand, Gujarat, India

²Computer Science , C.U.Shah University, Surendanagar, Gujarat, India.

³G.H.Patel Post Graduate Department of Computer Science & Technology, Sardar Patel University, V.V.Nagar, Gujarat, India.

*Corresponding Author: nspatel99@gmail.com, harshad.bhadka@yahoo.com, dbshah66@yahoo.com

Available online at: www.ijcseonline.org

Accepted: 15/Aug/2018, Published: 31/Aug/2018

Abstract—Opinion mining is a personal belief, view, perception, comments or expression on some topic ,person, product etc. In a current digital word every information is on Word Wide Web or digital form. With adventure of 4G mobile technology people use and share data or information on social media like twitter, face book, whatsapp application via mobile or computer. In order to take very quick and reasonable decision, it has become a common practice for decision maker to let their users review or express opinions about particular domain. A common user feels comfortable with the Internet, more and more opinions are writing. The number of opinions users receives is growing fast. Some popular products can get hundreds or thousands of reviews on some good trading venues. This makes it very difficult for a potential customer to read it to help determine if the product should be purchased or not. Therefore in this paper we introduce a simplified mobile computing opinion mining model which handles multilingual voice to text and text processing opinions. Moreover we compare accuracy result of existing Naïve bayes and SVM opinion mining algorithm result with our implemented model.

Keywords—Opinion, Sentiment Analysis, Supervise Learning, Naïve Bayes, SVM, OM

I. INTRODUCTION

Opinion mining is a personal belief or expression on some topic, about particular product, about particular person and so on. Example of opinions are “Education leads to person for good moral”, “Lenovo mobile gives good battery life” , “Virat and Rohit are excellent Indian cricketers”, “India is great country”, I love my India etc. In the current digital based economy, a large amount of information is available in the form of textual data i.e. in digital form which can often be used more easily if it is categorized or classified into some predefined classes. In any business or industrial environment corporate information may be available in multiple different formats, about 80% of which is in text documents. This information exists in the form of descriptive data format with long paragraph which include service reports ,about personal information, manufacturing quality, documentation, customer help desk notes and product reviews and opinions. In order to enhance customer satisfaction and their shopping experiences, it has become a common practice for online traders to enable their customers to review or to express opinions on the products

that they buy. A common users becoming comfortable with the Internet, an increasing number of people are writing reviews. The number of reviews that a product receives grows rapidly. Some popular products can get hundreds or thousands of reviews at some large trading sites. This makes it very hard for a potential customer to read them to help to make a decision on whether to buy the product. In our implemented algorithm we build mobile computing opinion mining algorithm which can find out polarity of text in form of positive, negative and neutral opinion. Presently we accept voice input in Gujarati, Hindi, English and Marathi languages.

II. RELATED WORK

Patel and Shah in 2013[1] Introduce basics of opinion mining, sentiment analysis and the challenges. Singh and Husain 2014[4] explain the Classification Techniques; categorization of work done for feature extraction and classification in opinion mining and sentiment analysis. In addition to this, performance analysis, advantages and disadvantages of different techniques are also appraised like

Naïve Bays Classifier, Support Vector Machine and Multilayer Perceptro by Valsamidisa, Theodosioua, Kazanidisa, Nikolaidisa. [2] and Samhaa R. El-Beltagy et al. [3] have discussed Testing and implementation methodology. They also present the various testing methodology, the challenges and opportunities of the blogs for agriculture in terms of analyzing the information which is stored in them. They used techniques in an experiment blog with the aid of the Rapid Miner software for opinion mining. This framework may thus helps in establishing baselines for opinion mining tasks in agriculture. Trivedi Khushboo N et al. [5] has show term counting based approach, in which total no of negative and positive words are count and then compared. If the dictionary is good then, it really gives good result. They used naïve Bayesian algorithm, which is supervised. And for increasing the accuracy of this algorithm, it is changes in the terms of parameter which are passed to the algorithm. Vidisha M. Pradhan et al. [6] they discuss various algorithms for sentiment analysis are studied and challenges and applications appear in this field are discussed. Dictionary based approach takes less processing time than supervised learning approach but accuracy is not up to the mark. Supervised learning approach provides better accuracy. From their survey, it can be concluded that supervised techniques provide better accuracy compared to dictionary based approach.

III. METHODOLOGY

For opinion mining or sentiment analysis presently two well known algorithm is used namely they are describe as a Naïve Bayes and Support Vector Machine let us understand it briefly.

A naïve bayes classifier is an algorithm based on simple probability. It uses Bayes theorem, but assumes that the cases are independent of each other, which is an unrealistic assumption in practical naïve Bayes classifiers. The world works well in complex real-world situations. The naïve Bayes classification algorithm can be trained highly in supervised learning, such as an insurance company that wants to promote a new policy to reduce advertising costs. The company wants to reach the most likely prospects that the company can gather. including income levels, number of current insurance policies, the number of cars owned, money invested and information about whether a customer recently exchanged insurance companies. With the help of naïve classifiers, the company can predict the likelihood of a customer responding positively to an offer policy. With this information, the company can lower your advertising costs limiting promotion to the most likely customers. The naïve Bayes algorithm provides the model for rapid construction and punctuation binary situations and multiclass to relatively low data volumes. [7]

Support vector machine support vector Machine (SVM) is a set of related supervised learning methods used for classification and regression. They belong to a family of linear classifications widespread. In other words, Support Vector Machine (SVM) is a classification and prediction tool that regression uses machine learning theory to maximize the accuracy of predicting automatically, preventing over Adapting to data from Support vector machine can be defined as system as space Use the hypothesis of a linear function in a high-dimensional function space, trained with an algorithm of optimization theory that implements a learning perspective derived from the theory of statistical learning. The support vector machine was originally popular with the NIPS community and is now an active part of machine learning research around the world. SVM becomes known when using pixel maps as inputs; provides comparable accuracy to sophisticated neural networks with features developed in a task he handwriting recognition is a translation that will be used for many applications such as handwriting analysis, face analysis and so on, especially for the design and regression based application foundation. The Support Vector Machines (SVM) has gained popularity due to many promising features such as the best empirical formulation. The performance uses the structural principle of Risk Minimization (SRM), which has proven to be superior to traditional empirical principle risk minimization (ERM), used by conventional neural networks. SRM minimizes an upper limit for the expected risk, where as ERM minimizes the failure in training data. This is this difference as SVM, more likely to generalize, which the target is in. [7]

Comparision of an Existing Algorithm and implemented Algorithm with the help of all literature survey we have to combine the features of naïve Bayesian algorithm and support vector machine algorithm. It is also noted that both existing algorithm use approaches like tokenization , stemming ,stop words etc. let us we compare the existing opinion mining algorithms and our modified OM(Opinion Mining algorithm) for polarity identification. In figure 1 it is to be clear that in both the existing algorithm there is use of probability label. For each sentence mining, developer has apply tokenization, stemming and removal of stop word where as new developed algorithm means in modified algorithm use supervise term rank base approach. In which large corpus of more than 5000 words is to be taken and each word have a rank like +1, -1, 0, +2, -2 etc. According to weight of word and the result of a polarity is to be generated from overall score of all words from particular sentence. It search all opinions of particular domain and then indicates your final score or result, it means if your result is to be in plus then your review or opinion is taken as a positive. If your result is to be in minus then your review or opinion is taken as a negative. If your result indicates exactly zero then opinion or review is taken as a neutral.

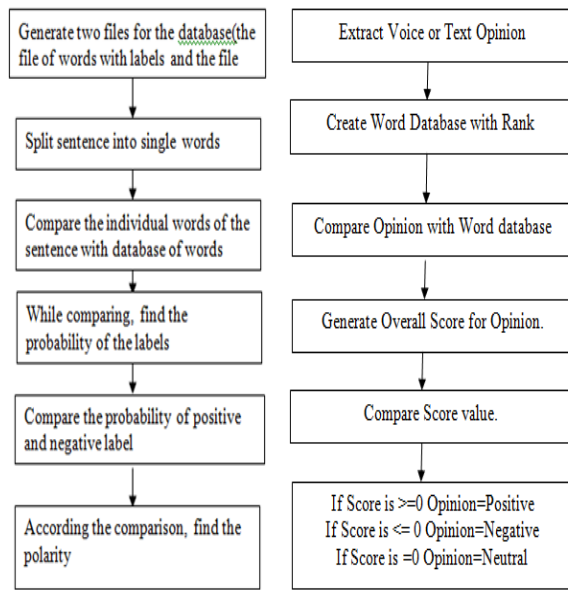


Figure: 1 Comparisons of Existing Algorithm and Proposed Algorithm

IV. RESULTS AND DISCUSSION

The execution of the experimental models is measured by different evaluations parameters which are accuracy, precision, recall and F-measure. All are criteria for measure machine learning algorithm. Let us understand each evaluation parameters of a machine learning algorithm and it is the evaluation standard for opinion mining output [8]

A. Accuracy :

Accuracy is the part of all predicted events that are true to all predicted events. An Accuracy of 100% means that the predicted instances are exactly the same as the actual cases.

$$\text{Accuracy} = \frac{\text{TPO} + \text{TNO}}{\text{TCO}} \times 100$$

Where

TPO = Total Numbers of Positives Opinions.

TNO = Total Numbers of Negatives Opinions.

TCO = Total Collected Opinion

B. Precision :

Precision is the part of predicted true positive cases in relation to all expected events. Precision indicates that a system can retrieve only relevant aspects and opinions. It is the fraction of aspects and opinions collected that are relevant to each product.

$$\text{Precision} = \frac{\text{TPO}}{\text{TCO}} \times 100$$

Where

TPO = Total Numbers of Positives Opinions.

TNO = Total Numbers of Negatives Opinions.

TCO = Total Collected Opinion

C Recall:

Recall is the part of instances that are expected to be true negative in relation to all total collected opinion. Moreover recall indicates that a system can retrieve only relevant features and opinions. That is the fraction or part of the total number of relevant features and opinions that have been part of this system.

$$\text{Recall} = \frac{\text{TN}}{\text{TCO}} \times 100$$

Where

TN = Total Numbers of Neutrals Opinions.

TCO = Total Collected Opinion

D F-measure:

F -Measurement is another way of measuring accuracy. It is calculated based on summation of accuracy and recall. There is a direct link between the value of the F measure. If the accuracy is high, The F measurement will also be high. The measure is calculated as follows:

$$\text{F Measure} = \text{Accuracy} + \text{Recall}$$

Where

TPO = Total Numbers of Positives Opinions.

TNO = Total Numbers of Negatives Opinions.

TN = Total Numbers of Neutrals Opinions.

TCO = Total Collected Opinion

Results Comparison Of Different Opinion Mining Algorithms

The case study is carried out using farmer ratings of various government agriculture schemes. during this research work All opinions were collected by developed mobile application. This application store a numbers of farmers opinions of different government schemes. The result shows the polarity of each sentence, that is, whether an opinion is positive, negative or neutral. The final results are shown in the diagram. As we all know, the man always gives the correct opinion of carefully evaluating the proposed system in comparison with the human decision, we manually read all the sentences and determine if the sentences are positive, negative or neutral. The same work was done on another "AIRC Sentiment Analyzer" system, available online. Finally, the results are compared and it seems that the proposed system performance is better than the "AIRC Sentiment Analyzer". [8] in figure 2 and figure 3 we compare implemented results of naive Bayesian algorithm and OM ,OM and SVM opinion mining machine learning algorithm.

A. Comparison of Naive Bayesian and OM Algorithm:

In a given below figure we explain result analysis of Naive Bayesian algorithm and our implemented OM (Opinion Mining) algorithm which clearly shows that accuracy of our mobile computing algorithm gives more accurate result than existing machine learning algorithm. in figure it is show that

naive Bayesian gives us 76.00% accuracy while mobile computing OM algorithms gives us 88% accuracy.



Figure 2: Result Comparison of Naïve Bayesian OM (Implemented) algorithm.

B. Comparison of SVM and OM Algorithm:

In a given below figure we explain result analysis of SVM(Support Vector Machine) algorithm and our implemented OM algorithm which is clearly shows that accuracy of our mobile computing algorithms is better than existing algorithm. in a figure it is show that Support Vector Machine gives us 62.67% accuracy while mobile computing OM algorithms gives us 88% accuracy.



Figure 3 : Result Comparison of SVM and OM (Implemented) Algorithm.

V. CONCLUSION AND FUTURE SCOPE

For polarity identification i.e. count occurrences of positive, negative and neutral opinions there are many opinion and text mining system and many algorithms are exist. In this paper we discuss naive Bayesian, support vector machines (SVM) algorithm. We compare this two algorithms accuracy with our implemented mobile computing multilingual opinion mining algorithm which gave high accuracy and powerful performances. Moreover it supports multilingual voice processing presently our mobile computing algorithm is work with voice input in Gujarati, Hindi, English and Marathi languages.

REFERENCES

- [1]. Rushabh Shah and Bhoomit Patel ” Procedure of Opinion Mining and Sentiment Analysis: A Study.” International Journal of Current Engineering and Technology Vol.4, No.6.
- [2]. Stavros Valsamidisa, Theodosios Theodosioua, Ioannis Kazanidisa, Michael Nikolaidisa”A Framework for Opinion Mining in Blogs for Agriculture. 6th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2013).
- [3]. “Samhaa R. El-Beltagy*, Ahmed Rafea_, Said Mabrouk_ and Mahmoud Rafea” An Approach for Mining Accumulated Crop Cultivation Problems and their Solutions. Association for the Advancement of Artificial Intelligence .
- [4]. “Pravesh Kumar Singh, Mohd Shahid Husain “Methodological Study Of Opinion Mining And Sentiment Analysis Techniques.” International Journal on Soft Computing (IJSC) Vol. 5, No. 1, February 2014.
- [5]. Trivedi Khushboo N, Swati K. Vekariya, Shailendra Mishra, “Mining of Sentence Level Opinion Using Supervised Term Weighted Approach of Naïve Bayesian Algorithm”, International Journal of computer Technology & Applications, Vol 3 (3), 987-991 ISSN:2229-6093,May-June 2012.
- [6]. Vidisha M. Pradhan ,jay vala and prem balani. “A Survey on Sentiment Analysis Algorithms for Opinion Mining” International Journal of Computer Applications (0975 – 8887) Volume 133 – No.9, January 2016.
- [7]. Y nikhil ,P sneha ,S prithvi raj , ajay ram ,e rajiv “sentiment analysis of mobile reviews using supervised learning methods” a dissertation submitted in partial fulfillment of the requirements for the award of the degree of bachelor of technology in computer science and engineering department of computer science and engineering university college of engineering

jawaharlal nehru technological university kakinada,
kakinada – 533003, a.p 2011 – 2013.

- [8]. Richa Sharma, Sweta Nigam , Rekha Jain. “ Polarity Detection at Sentence Level” International Journal of Computer Applications (0975 – 8887) Volume 86 – No 11, January 2014.
- [9]. Mr.Nayan S. Patel, Dr.D.B.Shah- “An Algorithm for Mobile Computing Opinion Mining In Multilingual Forms By Voice and Text Processing” International Journal of Scientific Research in Science and Technology ISSN: 2395-602X Volume 4 | Issue 5, 1262-1266 issue March 2018.

Author’s profile

Mr. Nayan S Patel is a research scholar in computer science at faculty of computer science , C.U.Shah University, Surendrangar, Gujarat. he did his Master of Computer Application From Department of Computer Science, Sardar Patel University(SPU),V.V.Nagar, Gujarat. He has published seven research paper in reputed international journal. His main research work focuses on Opinion Mining, Mobile Computing and Natural Language Processing. He has four years of research experience.

