SENTIMENT ANALYSIS AND AUTHENTICATION OF USER   RATINGS IN E-COMMERCE APPLICATION

Aishwarya B K[1] , Nirmala S[2] , Shruthi T[3] , Srinidhi R[4],[5] Shilpa B L *Assistant Professor*
[1, 2, 3, 4,5] *Computer Science and Engineering*
[1,2,3,4,5]*Vidya Vikas Institute of Engineering and Technology, Mysore*
(E-mail: [1]*aishwaryak28@gmail.com* ,[2]*nirmalasuresh38@gmail.com,* [3]*shruthishru0003@gmail.com,*
[4]*nidhinavaya1997@gmail.com* ,[5]*sbl.vviet@gmail.com*)

Abstract—Now a day's shopping of products through E-commerce website has become popular. It's the common nature of human being to get influenced by the opinionated materials they found on websites. This is especially true for product reviews, which have been shown to influence buying behavior. In this paper we discuss about a system to provide the accurate and genuine ratings in E-commerce websites, where sentiments are gathered, analyzed and aggregated from text, this is the main attention of sentiment analysis. In order to get the successful analysis of sentiment(texts) data mining technique is used in this system. In data mining technique different algorithms are used in order to clustering, classification and regression of the sentiments. K-means is used for clustering of the words from the sentences. SVM and K-nn are used for classification and regression. Sentiment classification is a basic task in sentiment analysis, with its aim to classify the sentiment as either positive or negative of a given text.

Keywords—*Sentiment Analysis, K-Means, Data Mining, SVM, K-NN, Regression.*

INTRODUCTION

The digital age, also referred to as the information society, is characterized by ever growing volumes of information. Driven by the current generation of web applications, the nearly limitless connectivity, and insatiable desire for sharing information, in particular among younger generations, the volume of user-generated social media content is growing rapidly and also increases even more in the near future.

In recent years, with the increasing use of internet, shopping products through E-commerce websites is also growing. Online shopping has become very popular now days because of its diversity in products. Also E-commerce websites provide many offers to attract more number of buyers. E-commerce websites are not limited to single type of products, we can purchase anything without going out of home. In E-commerce websites there is a well-known feature of giving reviews to the products. As reviews are means of opinions for the products, checking the reviews before purchasing any product is prevalence. Also to improve the quality of products, manufacturer can consider the comments given by user to improve their production. With the growing volume of reviews available on the websites it is not easy to analyze and conclude. This tends the use of many techniques to classify as weather it is positive or negative. To extract information and analyze and to build the objective results from the massive textual information involves the sentiment analysis and opinion mining technique. Therefore sentence level analysis is the major task. Sentiment analysis and opinion mining, of a special text mining task for determining the subjective attitude (i.e., sentiment) expressedby the text, is becoming a hotspot in the field of data mining. Sentiment classification is a basic task in Sentiment Analysis, which aims to classify the sentiment (e.g.,positive or negative) of a given text.

This paper is organized as follow: Section II describes the related work. Section III briefs about the existing system. Section IV describes the proposed system. Section V has the methodology description. Section VI gives the experimental analysis. Section VII describes the expected results.

RELATED WORK

The paper [1] explains the new type of word-of-mouth information, online consumer product review is an emerging market phenomenon that is playing an increasingly important role in consumers' purchase decisions. This paper argues that online consumer review, a type of product information created by users based on personal usage experience, can serve as a new element in the marketing communications mix and work as free "sales assistants" to help consumers identify the products that best match their idiosyncratic usage conditions.

The paper [2] describes, in the recent years, Sentiment Analysis and Opinion Mining becoming increasingly popular topics in Information Retrieval and Web data analysis. With

the rapid growth of the user-generated content represented in blogs, wikis and Web forums, such an analysis became a useful tool for mining the Web, since it allowed to capture sentiments and opinions at a large scale. Opinion retrieval has established itself as an important part of search engines. Ratings, opinion trends and representative opinions enrich the search experience of users when combined with traditional document retrieval, by revealing more insights about a subject. Opinion aggregation over product reviews can be very useful for product marketing and positioning, exposing the customers' attitude towards a product and its features along different dimensions, such as time, geographical location, and experience. Tracking how opinions or discussions evolve over time can help us identify interesting trends and patterns and better understand the ways that information is propagated in the Internet. In this study, they have reviewed the development of Sentiment Analysis and Opinion Mining and also discussed the evolution of a relatively new research direction, namely, Contradiction Analysis. They have given an overview of recent advances in these areas.

Bo Pang et.al. [3] covers techniques and approaches that promise to directly enable opinion-oriented information seeking systems.An important part of information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people now can, and do actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object.

The paper [5] explains machine learning based method of sentiment classification of sentences using word-level polarity. The polarities of words in a sentence are not always the same as that of the sentence, because there can be polarity-shifters such as negation expressions. The proposed method models the polarity-shifters. Here model can be trained in two different ways: word-wise and sentence-wise learning. In sentence-wise learning, the model can be trained so that the prediction of sentence polarities should be accurate. The model can also be combined with features used in previous work such as bag-of-words and n-grams. We empirically show that our method almost always improves the performance of sentiment classification of sentences especially when we have only small amount of training data.

EXISTING SYSTEM

The existing websites use random rating process where the product rating is just given by selection of stars. The genuine or non-genuine customer registered in these websites go for

product rating without having any knowledge of the product and also rating the product without purchasing it by simply selecting traditional star approach. Though the websites have the description section to describe the product good and bad most of the fake reviewers never use this approach, users just simply select 5 star for good and single or less stars for bad representation.

DEMERITS OF EXISTING SYSTEM

Using random rating (star selection approach) may miss lead the product ratings, where in case a good product might receive single or less stars and a bad product might receive high stars or 5 stars.

This approach might not clearly clarify the good and bad of the product. Where the preceding customer who wants to purchase might get in confusion in what advantages made this product good or what disadvantage made this product as bad or less rating.

The traditional approach might misguide the buyer like a good product might receive bad rating thus customers fail to purchase it and bad product might receive good star ratings and customers get cheated.

PROPSED SYSTEM

This system is proposing sentiment analysis approach for user ratings in E-Commerce websites using data mining technique. Sentiment analysis of natural language texts is a large and growing field [5]. Sentiment analysis or Opinion Mining is the computational treatment of opinions and subjectivity of text. Sentiment analysis is an Information Extraction task that intends to acquire writers feelings expressed in positive or negative comments, after analyzing the documents recorded which has been given by the user. Sentiment analysis is used in extracting the recommendations related to the user's search. The traditional random rating (Stars selection) approach is removed in this proposed system.This is a simple E-Commerce distributed web application where the customers can buy the products view the products and its description and ratings. Here only genuine buyers can give the ratings for the product.Whenever User / customer buys the product, after successful purchase the user is provided with the unique key, using random key generation method. User / Customer must have the purchase key in order to rate the product. It is show in the below diagram.

Fig 1: key generation for the genuin
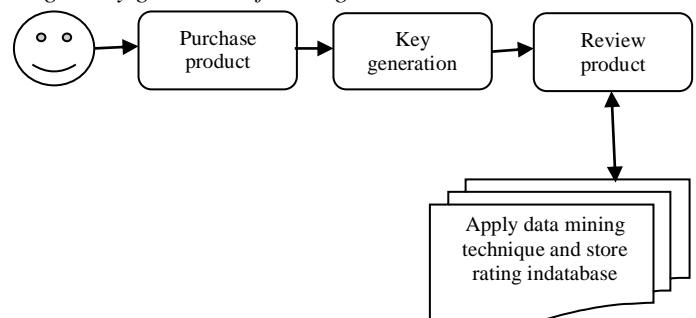
Here data mining concept is used for rating purpose where the customer who has purchased the product should describe his words in order to give ratings. The text entered by the user is considered as the sentiments and analysis of these text inputs results as either the text (sentence)is positive or negative. User authentication is the new approach where it is not implemented in any of existing system.

NEED FOR THE PROPOSED SYSTEM

There is no random rating process, Users can view the product but cannot rate or comment about the product unless he/she has purchased and used it. If the user purchases the product the system generates purchase key and only a buyer can review or rate the product.

The product rating is increased or decreased based on the reviews given by the user.

The ratings are shown in the data analytics which shows positive rating, negative rating and average rating.

METHODOLOGY USED

In the proposed system data mining methodologies are used. As machine learning approaches allow many features, they combine multiple lexicons, adding sentiment information from both the general inquirer lexicon as well as from the polarity lexicon.

A. K-Means Algorithm

K-Means algorithm [6][7] is used in this system to create k groups from set of dictionary words. K-means clustering that is also known as nearest centroid classifier algorithm is a method of vector quantization that is considerably popular for cluster analysis in data mining. K-means is used to create k groups from a set of objects so that the members of a group are more similar. It's a well known popular cluster analysis technique used for exploring a dataset.

Pseudo code:
 Input: k (The number of clusters),
        D (A set of lift ratios)
Output: A set of K clusters
Method: Arbitrarily choose k objects from D as the initial cluster centres
    Repeat:
    1. (re)assign each object to the cluster to which object is the most similar ,based on the mean value of the object in the cluster;
    2. Update the clusters names that is calculate the mean value of the object for each cluster

Until no change;

B. SVM Algorithm

Support vector machines [8][9] that are also known as support vector networks are basically supervised learning models that are associated with learning algorithms which then analyze data that are used for the analysis of regression and classification. An SVM model is created that is a representation of the examples as points in space, that are further mapped so that the examples of the separate categories are then divided and classified.

Formula:
$h(x_i) = sign(\sum_{i=1}^{s} A_i Y_i K(x_j, x_i) + b)$
$K(V, V^{`1}) = exp(\|v-v1\|2)$

$$K(v, v') = exp(\frac{\|v - v'\|^2}{2\gamma^2})$$

Pseudo code:
Initialize data
While Termination   Not Met do
Select First Node Randomly
{Node-Method, End-Node-Method}
While Number of visited Node<Number of Node do
        If there are Univisitedneighbours
                Select One Node Randomly
        Elseif there are visited neighbours
                Select One Node Randomly
        Elseif there are Taboo neighbours
                Select One Node Randomly
        Endif
endwhile
endwhile
Display Solutions

C. K-NN Algorithm

The K-Nearest Neighbors algorithm (K-NN) [10] is a type of lazy learning or instance-based learning and is considered as a non-parametric method that is used for classification and regression. In both the mentioned cases, the input consists of the k closest training examples in the space and the output depends on whether the algorithm is being used for classification or regression. This k-NN Algorithm is considered and is also among the simplest of all machine learning algorithms.

Pseudo code:
k-Nearest Neighbor
Classify(X,Y,x)//X: training data, Y: class labels of X, x:unknown sample
For i=1 to m do
Compute distance d(Xi, x)
End for
Compute set I containing indices for the k smallest distances d(Xi, X).
Return majority label for { Y_i where i ∈ I}

EXPERIMENTAL ANALYSIS

The text entered by the user in the comment section will be considered as sentiments and clustered into different similar group of words.Then using SVM algorithm all word will classify into positive set, negative set and general set. K-NN is used to generate the keywords. This classification is then assigned with values with the help of dictionary dataset in order to represent the positive ratings and negative ratings in graph (bar charts).

The customer who purchases the product will be generated with unique key to give comments to that particular product. In comment section user has to write his opinion in text form. For example, user can type

1. The product is very good.
2. The battery is good.
3. Worst product.
4. Am not at all happy with it.

These sentences will be the dataset from which the emotion behind it can be extracted.

EXPECTED RESULTS

The comments from the user ratings are analyzed and the values are assigned for each sentence. Using these values a graph representation of the positive and negative bar charts should be displayed. Average of positive and negative bars should also be displayed in bar chart. Fig 2 describes how the user can view the ratings of the products.
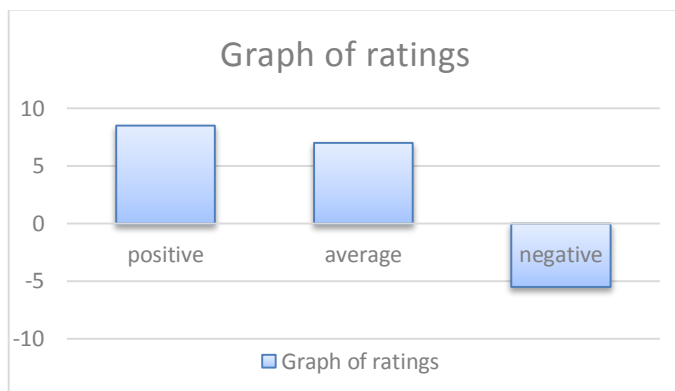

Fig 2: graph showing the positive, negative and average ratings

CONCLUSION

The proposed system makes huge difference in E-Commerce like websites. Sentiment analysis using data mining technique extracts the emotion of the comment given by users.The traditional approach is changed to data based rating technique.Rating the product based on description will make sense to the product rating. This system authenticates the user purchase through the unique key generation for every user and each product. This avoids the fake reviews about the product. Only genuine customer can rate the product. Users can't give the random fake ratings since they will not be having purchase key.The purchaser can go through the genuine ratings based on the reviews given by the genuine buyers.Buyer can get better knowledge of the product by considering the reviews of genuine buyers. More improved product rating technology where fake ratings can be eliminated.

In future, the system can be made more robust and secure in order to meet the performance and accuracy. Also the system can make mandatory description in detail to the product in order to make easier for any customer to buy and get practical information of the product.

REFERENCES

[1] Yubo Chen and JinhongXie, "Online consumer review: Word-of-Mouth as a new element of marketing communication mix", 2009.

[2] Mikalai Tsytsarau and Themis Palpanas, "Survey on mining subjective data on the web", 2013.

[3] Bo Pang and Lillian Lee "Opinion mining and sentiment analysis", 2010.

[4] D. Ikeda, H. Takamura, L. Ratinov, and M. Okumura "Learning to Shift the Polarity of Words for Sentiment Classification", 2015.

[5] T Karnan, G. Seenuvasan "Sentiment Analysis in E-commerce using recommendation system", IJCSMC, Vol. 6, Issue. 8, August 2017, pg.55 – 62

[6] Tapas Kanungo, Nathan S. Netanyahu, Angela Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", IEEE Trasactions on pattern analysis and machine intelligence, vol.24, No.7, July 2002.

[7] S. Na, L. Xumin and G. Yong, "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm," 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, Jinggangshan, 2010, pp. 63-67.

[8] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998.

[9] Y. Tang, Y. Zhang, N. V. Chawla and S. Krasser, "SVMs Modeling for Highly Imbalanced Classification," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, no. 1, pp. 281-288, Feb. 2009.

[10] Ahmad Basheer Hassanat, Mohammad Ali Abbadi, Ghada Awad Altarawneh, "Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 12,No. 8, August 2014.

Author Profile:
Ms Aishwarya B K pursuing 8th semester of Bachelor of Engineering in the Department of Computer Science and Engineering from Visvesvaraya Technological University, VVIET, Mysuru. Her main areas of interest includes Sentiment Analysis, Database and Natural Language Processing.

Author Profile:
Mrs. Shilpa B L pursed Bachelor of Engineering from Visvesvaraya Technological University, India in 2009 and Master of Technology from Visvesvaraya Technological University, India in year 2011. She is currently working as Assistant professor in Department of Computer Science and Enginering, VVIET, Mysuru, Affiliated to Visvesvaraya Technological University, India. She is a life member of the IAENG since 2013, ICSES since 2017. Her main research work focuses on Predictive analytics, Big Data Analytics, Natural Language Processing, Data Mining and Machine Learning. She has 7 years of teaching experience and 2 years of Industry Experience.

Author Profile:
Ms Nirmala S pursuing 8th semester of Bachelor of Engineering in the Department of Computer Science and Engineering from Visvesvaraya Technological University, VVIET, Mysuru. Her main areas of interest includes Sentiment Analysis, Database and Natural Language Processing.

Author Profile:
Ms Shruthi T pursuing 8th semester of Bachelor of Engineering in the Department of Computer Science and Engineering from Visvesvaraya Technological University, VVIET, Mysuru. Her main areas of interest includes Sentiment Analysis, Database and Natural Language Processing.

Author Profile:
Ms Srinidhi R pursuing 8th semester of Bachelor of Engineering in the Department of Computer Science and Engineering from Visvesvaraya Technological University, VVIET, Mysuru. Her main areas of interest includes Sentiment Analysis, Database and Natural Language Processing.