

A Survey on Sentimental Analysis of Online Product Reviews for Various E-Shopping Websites

Punitha Devi C^{*1}, Vijayalalitha V², Raja Rajeswari S², Lendaa V²

^{*1}Professor, Department Of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry, India

²B.Tech, Department Of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry, India

ABSTRACT

The rapid growth of internet is a wide source of learning and decision making. Thousands of reviews are generated each day for the online products. Analyses of these opinions are very important because it is helpful in determining the users' perspective about the product and also for making business decisions in the organizations. The sentiment analysis system is build on the basis of identifying the words and classifying its sentiments from the e-shopping websites using data mining approach. The proposed system selects the words from the learning dataset and then classifies them into objects and sentiments based on a feature extraction methodology. A Quality Data miner technique has been adapted to study about the polarity of the each review in both sentence and aspect level. A featured specific comparison has been made with the results that are retrieved using the QD technique and the product with the desired feature has been recommended to the user. The QD miner technique is able to classify user opinions with good precision and accuracy.

Keywords: QD Miner, Sentiment Analysis, Product Reviews, E-Shopping Websites, Review Analysis

I. INTRODUCTION

Sentimental analysis is a technique of data mining which is a way of identifying the opinion or the key idea of the user in a piece of text, comments of the user, reviews given in the websites. It also expresses the polarity of the sentence in a positive, negative and in a neutral manner. The analysis can be done in three different levels such as word level, sentence level and in document level [7].

The sentimental analysis undergoes a sequence of steps such as it tracks the direct opinions followed by the comparison of the opinions and then detect the spam reviews. The classification of texts are done by three different approaches like machine learning techniques using a classifier like SVM, Naïve Bayes, Maximum entropy and Baseline technique or else by

using the polarity of the document or by using the SentiWordNet which is a dictionary that are available in the library. The reviews are helpful for the customer to buy a product.

The reviews are not that easy to analyse, some are given in a tedious manner and some are given directly. The phases used for analysis of the reviews [16]:

- **Pre-processing phase:** The dataset is first pre processed by cleaning and reducing the noise in the given input and the review is broken into set of words.
- **Feature extraction phase:** This phase is used to analyze the given data by extracting the keywords and checks for the matching words in the dictionary.

- **Classification phase:** Based on the classifier algorithm the extracted data has been classified to determine the polarity (positive, negative, neutral) of the sentence.

Hence it is necessary for the summarization of the reviews. Then they are analysed by sentimental analysis. Sentimental analysis is used to extract the subjective information or the opinion by using the quality term miner technique. Therefore, it thus helps the customer to visualize the best quality products at a better accuracy rate.

II. LITERATURE SURVEY

A survey has been undergone for the sentimental analysis of online sites and the challenges faced by them has been mentioned below

- **Esha Tyagi et al, (2014) [11]** proposed a model by using SVM algorithm. The primary concern of the model is to determine the polarity of the ambiguous sentiment or reviews. The resulting performance is tested to measure the accuracy of the sentiment as positive, negative or neutral.
Challenges: The accuracy that has been obtained by the model is enhanced for the inclusion of more sentence forms.
- **Haseena Rahmath et al, (2014) [15]** used a methodology of natural language processing in the segmentation of the online product reviews to state its polarity. The system used a set of fuzzy functions to classify the words on the basis of their functionality to make the system more accurate in sentiment classification and categorization of reviews.
Challenges: A rule set can be refined to extract more dependency relations from dataset in order to improve the precision values of the dataset.
- **Manvee Chauhan et al, (2015) [12]** proposed a model by using Naïve Bayes and SVM techniques. The system can be used to the text files that are large in size and for the ones that take more computational time. It is determined

that Naïve Bayes gave a better accuracy rate when compared to SVM.

Challenges: The training labels that are given as inputs which maps the output needs human experts to supervise the model.

- **Xing Fang et al, (2015) [13]** refined a work based on the researches perspective of reviews on various APIs by using random forest technique. The data that has been collected are categorized as sentence-level and review-level for the determination of polarity of the reviews using a transformation vector called feature vector.

Challenges: The scheme does not work well for the reviews that contain neutral or implicit sentiments and the accuracy keeps on varying in the range of 0.5.

- **Raheesa safrin et al, (2017) [14]** proposed a working model in the field of natural language processing where the review has been classified as sentence level categorization. The review is drawn into adjectives, adverbs and various parts of speech and compared with the dataset for the determination of polarity as positive, negative or neutral reviews.

Challenges: The system does not work for the reviews that consist of both positive and negative meaning words.

III. EXISTING SYSTEM

A user senti-word measurement approach, which is based on the mined sentiment words and sentiment degree words from user review. Keywords mentioned in the customer review will be mined and will be matched with the words given in the dictionary, and based upon the comparison the system will rate the product. We make use of Sentiment for rating prediction. User sentiment similarities will be focusing on the user interest and their preferences which is used for analysis. User sentiment influence reflects how the sentiment spreads among the trusted users. The existing system deals with the individual study of reviews on buying

a product in the e-shopping websites. This system supports only English language not any other language. The review of a product 'it is not much good' is considered to be a positive review which is actually a negative one. The words that are not matched in the dictionary is not intimated.

IV. APPLICATION OF SENTIMENTAL ANALYSIS

The below table describes about the major applications which is used for sentimental analysis to predict the best product. In movie review will improve the standard of films. In hotel reviews it will help to improve the services and to satisfy the customer needs. In such a way it helps to improve services in various fields.

Table 1. Application Of Sentimental Analysis For Product Reviews

| S.N O | AUTHOR NAME AND YEAR | TECHNOLOGY/ ALGORITHM USED | PARAMETERS USED FOR ANALYSIS | ACCURACY |
|-------|--|--|---|----------|
| 1. | Nagamma P et al,2015[1] | clustering algorithm (Movie Reviews) | Actors Rank of sale Release date Story Audience reviews | 89.64% |
| 2. | Walter Kasper et al,2011[4] | BEASHOT System (Hotel Reviews) | Services Cost Parking facilities Rate View | 74.06% |
| 3. | Andy Moniz et al,2016[2] | Naive bayes classifier (Firm Reviews) | Firm outlook Opportunities Career Salary | 78.04% |
| 4. | Rekha et al,2015[5] | Opinion mining and MATLAB Software (Mobile Phone Reviews) | Camera Battery Screen Sounds Performance | 68% |
| 5. | Michael Meidl and Steven Lytinen et al,2014[3] | Vector space approach (Games Reviews) | Graphics Game play Sound track Character used Story | 56.02% |
| 6. | Ana Valdivia et al,2017[8] | Automatic sentiment detection algorithm (Tourism Reviews) | Location Routes Hospitality Trip advisor | 48.08% |

V. TECHNOLOGY USED

Sentimental analysis can be broadly classified into three main categories such as Machine learning approach, lexicon based approach .[9][10]

MACHINE LEARNING APPROACH

This technique is used for classification of words into positive, negative and neutral words. It requires many training datasets used for prediction. They are classified into two types supervised and unsupervised learning [9][10].

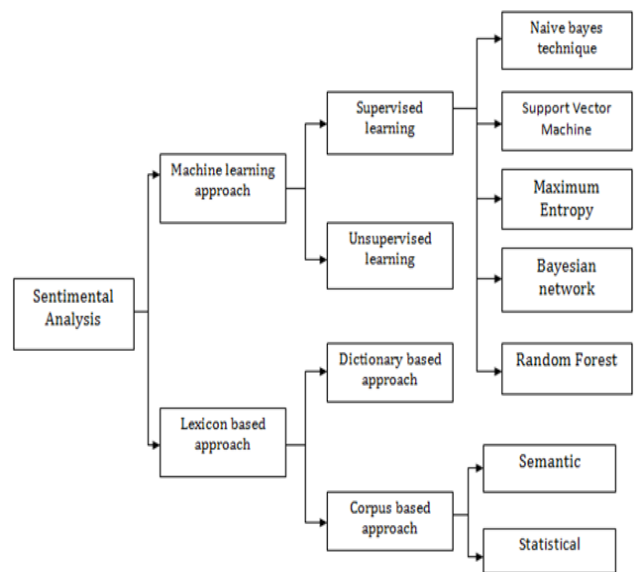


Figure 1.Classification Of Sentiment Analysis

Supervised learning: As the name suggests supervised learning is the process of learning mechanism from the trained data set which are being continuously supervised. The algorithm frequently makes the prediction over the results and the algorithm is terminated when the result reaches the acceptable level.

Unsupervised learning: It is a machine learning approach that is used to inference from the learning dataset that consists of data that are without label or any reference. The most commonly used unsupervised learning is the cluster analysis for finding the hidden patterns.

LEXICON BASED APPROACH

This technique is mainly used for extracting the sentiment from the words or phrases. The classification of this approach is drawn into dictionaries and corpus based approach.

Dictionary based approach: The dictionary in this approach works on by collecting the synonyms and antonyms in the given phrases. The words from the dictionary are used to build the thesaurus.

Corpus based approach: They provide the dictionaries that are related to a specific domain. This methodology grows on by the search of related words based upon the usage of either statistically or semantically.

SVM:

SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. It is used for the recognition of handwritten texts and patterns. It is also used for the classification of reviews based on their quality.

Disadvantage: It retains lack of transparency and computational efficiency. It occupies more memory space and computational time.

NAIVE BAYES CLASSIFIER:

It is used in the real time process for getting accurate results in predictive model. It occupies less memory space and processing time. The classifier requires small amount of dataset to estimate the parameters for evaluation.

Disadvantage: Strong feature independence assumption is made on the data which cause a loss of accuracy.

BAYESIAN NETWORK:

Bayesian network structures are classified by using hierarchical clustering and the results will be formed in the form small state spaces. It is mainly used

mainly for risk analysis, prediction and data collection. Bayesian network can perform computation in places where incomplete data sets are used for analysis.

Disadvantage: It is complex to generate a graph structure from the analysed data. In order to calculate the probability of a single process the entire data set should be calculated.

MAXIMUM ENTROPY:

Maximum entropy is effective by using it in number of natural language processing applications. It provides a proper distribution and it will not assume statistical independences of random variables.

Disadvantage: It does not give expected iteration at each level and it is not constant. It requires more additional resources or annotations.

RANDOM FOREST:

It works upon a tree based approach. As the number of performance increases the accuracy level also increases. The variable that is taken for the classification is automatically generated based upon their relevance.

Disadvantage: Training data includes both the input and the desired results.

SEMANTIC ANALYSIS:

It is a process of building structural concepts from a large set of documents. In semantic analysis indexing is a class of technique which represents documents as vectors in term spaces. Semantic researches are used to increase the system performance.

Disadvantage: Representation of distributional model is very tedious so it is hard to index based upon individual dimension .Technical way of analysing of the problem is difficult.

STATISTICAL ANALYSIS:

It is the ability of extracting statistical regularities around them. It is a generalised mechanism, which was identified in human acquisition language system. It is a way, which formalizes the relationship between the datasets in the form of mathematical equation

Disadvantages: Though the results of the statistical analysis are perfect it produces series of distortions and misleading conclusions.

VI. CONCLUSION

In this proposal, we analyse the reviews and services provided by e-shopping sites for various products. We collect reviews from different sites and analyse the sentiment of words using dictionary data sets and then the words are classified as positive, negative and neutral reviews. Based on these reviews products rating are given and it also describes the best features of the product when compared with other product .It also analyse the native language words used by the user and predict the word sentiment.

VII. REFERENCES

- [1]. Prof Nagamma P, Pruthvi H.R, Nisha K.K , Shwetha N H "An Improved Sentiment Analysis of Online Movie Reviews Based On Clustering For Box-Office Prediction" International Conference on Computing, Communication and Automation (ICCCA2015) ISBN:978-1-4799-8890-7/15/\$31.00 ©2015 IEEE.
- [2]. Andy Moniz ,Franciska de Jong "Sentiment analysis and the impact of employee satisfaction on firm earnings"36th European Conference on IR Research, Vol. 8416 ,pp. 519-527, 2014.
- [3]. Michael Meidl and Steven Lytinen et al "Using Game Reviews to Recommend Games" Games and Natural: Papers from the AIIDE Workshop <https://www.aaai.org/ocs/index.php/AIIDE/AIIDE14/paper/.../9033,2014>.
- [4]. Walter Kasper , Mihaela Vela "Sentiment Analysis for Hotel Reviews" Proceedings of the Computational Linguistics-Applications Conference ,ISBN 978-83-60810-47-745 ,pp 45-52,2011.
- [5]. Rekha , Dr. Williamjeet Singh "Sentiment Analysis of Online Mobile Reviews" Inventive Communication and Computational Technologies (ICICCT), 2017 International Conference, INSPEC Accession Number: 17042274 DOI: 10.1109/ICICCT.2017.7975199, 2017.
- [6]. Zhang Hailong, Gan Wenyan, Jiang Bo "Machine Learning and Lexicon Based Methods for Sentiment Classification: A Survey" WISA '14 Proceedings of the 2014 11th Web Information System and Application Conference, ISBN: 978-1-4799-5727-9, DOI: 10.1109/WISA.2014.55, pp 262-265, 2014.
- [7]. Walaa Medhata, Ahmed Hassanb, Hoda Korashyb "Sentiment analysis algorithms and applications: A survey "Ain Shams Engineering Journal , Vol. 5 issue-4, pp 1093–1113, 2014.
- [8]. Ana Valdivia, M. Victoria Luzón, and Francisco Herrera "Sentiment Analysis in Trip Advisor " IEEE Intelligent Systems , Volume: 32, Issue: 4, pp 72-75, DOI: 10.1109/MIS.2017.3121555, 2017.
- [9]. Amit K. Burde, Tushar G. Barage, Vikram B. Shevate, Preeti Suryawanshi, Viresh Chapate "Survey on Sentiment Analysis and its Classification Technique " International Journal of Computer Applications , NCAC 2015 - Number 5, pp 21-26, 2015.
- [10]. Vandana V. Chaudhari, Chitra A. Dhawale and Sanjay Misra "Sentiment Analysis Classification: A Brief Review" IJCTA, Vol. 9 Issue -23 , pp 447-454 , 2016.
- [11]. Esha Tyagi, Arvind Kumar Sharma "Sentiment Analysis of Product Reviews using Support Vector Machine Learning Algorithm " Indian

Journal of Science and Technology, Vol 10
Issue -35, DOI: 10.17485 ,2017.

- [12]. Manvee Chauhan , Divakar Yadav” Sentimental Analysis of Product Based Reviews Using Machine Learning Approaches” Journal of Network Communications and Emerging Technologies (JNCET) Volume 5, Special Issue 2, December (2015) .
- [13]. Xiang Fang ,Justin Zhan "Sentimental analysis of product review data "Journal of Big Data (2015) 2:5.<https://doi.org/10.1186/s40537-015-0015-2> ,2015.
- [14]. Raheesa Safrin "Sentimental analysis on online product review "International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 04 ,pp 2395-0072, Apr -2017 .
- [15]. Haseena Rahmath P , Tanvir Ahmad "Fuzzy based Sentiment Analysis of Online Product Reviews using Machine Learning Techniques "International Journal of Computer Applications, Volume 99-Issue No.17, pp 0975-8887, August 2014.
- [16]. Vimalkumar B. Vaghela , Bhumika M. Jadav " Analysis of Various Sentiment Classification Techniques ", International Journal of Computer Applications ,Volume 140-No.3, pp.0975-8887, 2016.