International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 5, Issue 11, ISSN No. 2455-2143, Pages 274-278
Published Online March 2021 in IJEAST (http://www.ijeast.com)

AN EXPERIMENTAL ANALYSIS ON E-COMMERCE REVIEWS,WITH SENTIMENT CLASSIFICATION USING OPINION MINING ON WEB

Latha S S
Assistant Professor
Department of Information Science & Engineering
MVJ College of Engineering, Bengaluru, INDIA

*Abstract***— Sentiment analysis is a big branch in the field of natural language processing. Sentiment analysis mainly text based analysis, but there are some challenges that make it difficult as compared to traditional text based analysis. This paper empathizes on the need of an attempt to improve research process and progress of sentiment analysis on the basis of investigation. Outcome of the analysis are summarized in this paper.**

This paper analyze the reviews of products manually by collecting data in the form of a excel file. Then it will produce and classify the reviews as positive or negative comments to get the best product. Now it's more relevant to automate reviews data it is growing exponentially. This method works by web scrapping reviews from e-commerce website. Data cleaning is applied to remove the unwanted data known as stop words. The features are identified. The feature can be camera, battery life etc. Obtain frequency across all the products and for all the reviews per feature. The intended work is to extract the features from the reviews and detecting the polarity for each aspect, thus resulting in feature extraction matrix (FEM). FEM matrix has each row as an observation for a product and each of the columns represent the feature. List of Products based on highest value of FEM for searched features and product recommendations are generated based on the user searched feature.

Keywords — Sentiment analysis, Data cleaning, Feature extraction matrix, Polarity, Features, Opinion mining.

I. INTRODUCTION

Sentiment analysis is the process of detecting positive or negative sentiment in text. We have huge amount of data available on web. Majority of the customers writes reviews about products online. People give opinions on multiple products and also compare them. It's very important to identify the sentiment of the each costumer. If the sentence incorporates the product names, they need to be selected. These types of problems are significant because we have to know what products in each sentence speak about the opinion mined from the sentence. Our algorithm should think like human beings.

Businesses and consumers buying and selling products in online refers to the e-commerce. The most of the e-commerce websites sell products to the public directly. A review refers to the evaluation of a service, publication, review of movies, video game review, review of a music composition or music recording, book review, hardware piece like a car or computer, performance of an event, such as a live music concert, play, musical theatre show, dance show or exhibition of art. The previous system just finds reviews of previous product by collecting data from the web and classifies the reviews as positive and negative by considering selected attributes to get the opinion of the product. The previous system does not consider the features of the product.

We have large volume of data available on web. The inception of the sentiment analysis coincides with those of online data in form of reviews. Without the availability of that data, the research on sentiment analysis could not have been possible. Here, the focus is, not only to detect the polarity of the product reviews but to resolve the sentiment at more detailed level. The intended work is to extract the features from the review and detecting the polarity for each aspect, thus resulting in feature extraction matrix (FEM). The work is also concerned with calculating the sentiment score associated with features of entity. Sentences contain the sentiment bearing words will be considered for analysis. Stop words are removed at data pre-processing and Data cleaning step. Final results will be converted in to feature extraction matrix. By representing the feature extraction matrix, it would be easier for other readers to understand on what features the opinion holder has commented upon.

II. RELATED WORK

Sentiment classification comes under the problem of text classification. Previously, text classification incorporated a work of classifying a document topic wise, e.g., sports, and sciences. Key features help in detecting the theme of the document. As far as sentiment analysis is concerned, sentiment words or opinion words, for example good, excellent, amazing, bad etc, play a significant role in classifying a document. These are the words that help in deciding the polarity of the reviews. It is necessary to identify the sentiment related to each aspect of entity, when review discusses about several aspects of the entity. There is a need to calculate the overall sentiment of the product, more accurately. A feature extraction matrix (FEM) would be generated as the result of the proposed work to determine sentiment related to features of the product was studied in paper [1].

In [2], for each review, number of features is extracted and desired score get associated with the feature. Binary distribution is used in generating the feature extraction matrix (FEM). The state of the feature is set to 1 by the algorithm, if it is present in the review otherwise, it is 0.

Average semantic orientation of the phrases was used for classifying the reviews. The PMI-IR algorithm was used to determine the semantic orientation. The algorithm worked in three steps. First, it would extract all the phrases that contain adverbs or adjectives. In second step, it will predict the semantic orientation for each phrase. Third step is to classify the review on the basis of the orientation result. The second step of the algorithm is the most important step. The semantic orientation is predicted here as: SO (phrase that contain verbs or adjectives) = PMI (phrase that contain verbs or adjectives, “excellent”) - PMI (phrase that contain verbs or adjectives, “poor”). Mining the various product features and customer reviews summarization was studied in [3]. Thought about assemble features and advancement, like what we have used in positioning computation. A company must know which highlights of a particular item are essential and which highlights should be magnify to expand consumer loyalty. The main features on which opinion conveyed are selected and the reviews are extracted based on the features recognition.

In [5], a method based on bootstrapping was employed for studying targets and opinions on them. The method was called double propagation method. The method focus at determining the linking between the opinion words and targets.

Soft computing methods are strenuously conveyed in E-commerce businesses as information warehousing, and “soft computing” is the core of information warehousing or of some other drive innovations today was studied in [6]. According to the ACM communications magazine [7], many open challenges still existing within the field of sentiment analysis. And hence to sharpen the prevailing techniques and to provide satisfactory solutions to those challenges, lot of work is want to be done in this field.

We must identify the sentiment attached to each aspect of an entity, document discusses about many feature of the entity. There is a need to calculate the combined sentiment of the product review acuuratly.

In Parashar & Gupta’s study [8], the consumer of an E-commerce site has no actual way to evaluate the quality of an acknowledged item to peruse huge number of reviews. This exploration work centres on building up a basic leadership computation which can evaluate the nature of an item by classify past surveys on a size of numbers and showing it on an E-commerce site. Purchaser can make utilization of these positions furnished with items over any E-commerce site to fix on their own options.

III. FRAME WORK FOR E-COMMERCE REVIEW MANAGEMANT SYSTEM

Model to analyse E-commerce Website is shown in Fig.1. First collects the reviews of products from the web and then parse the reviews to clean collected information. Cleaned data are divided to determine tokens. Once the token is identified it computes the frequency of identified keywords. The frequencies of keywords are used to represent features in our proposed model. The FEM matrix is constructed by using the list of Features to find the rank of product. In this model we have considered various features such as Screen, Price, Speaker, Battery, Camera and Quality and then provide the overall sentiment to analyse the reviews and comments of customers in an E-commerce website.

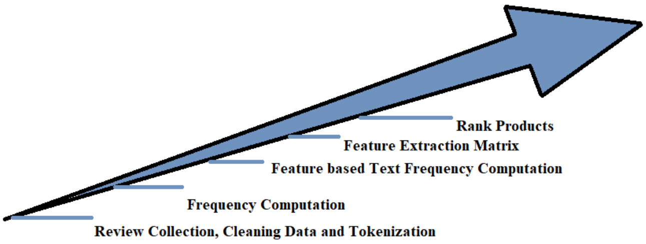
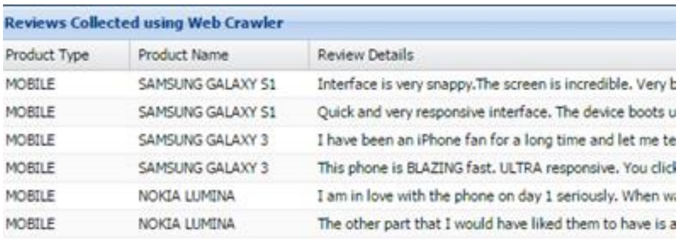

Rank Products
Feature Extraction Matrix
Feature based Text Frequency Computation
Frequency Computation
Review Collection, Cleaning Data and Tokenization

Fig.1.1. Model to analyse E-commerce Website

The following goals are defined

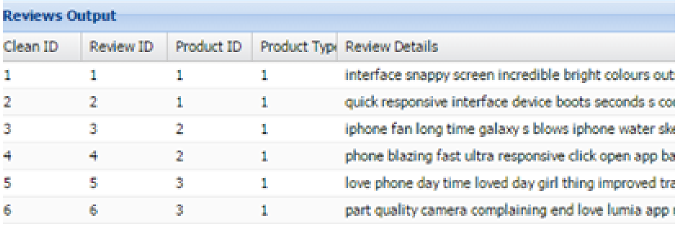
Review Collection, Cleaning Data and Tokenization: The Review Collection, Cleaning Data and Tokenization involves the data pre-processing and cleaning of the data set. First collect reviews from web to obtain real time and non-real time reviews for the products and collect it from e-commerce website such as flipkart. Pre-processing steps include removing of information about the reviews that are not required for performing sentiment analysis such as date, stop word and time of a review. After that obtains all the keywords in the cleaned reviews. The review analysis is used to identify the necessary information including opinions and product features. Opinions and features are extracted from this step.



Reviews Collected using Web Crawler

Product Type	Product Name	Review Details
MOBILE	SAMSUNG GALAXY S1	Interface is very snappy.The screen is incredible. Very b
MOBILE	SAMSUNG GALAXY S1	Quick and very responsive interface. The device boots u
MOBILE	SAMSUNG GALAXY 3	I have been an iPhone fan for a long time and let me te
MOBILE	SAMSUNG GALAXY 3	This phone is BLAZING fast. ULTRA responsive. You clic
MOBILE	NOKIA LUMINA	I am in love with the phone on day 1 seriously. When w
MOBILE	NOKIA LUMINA	The other part that I would have liked them to have is a

Fig.1.2. Data Collection



Reviews Output

Clean ID	Review ID	Product ID	Product Typ	Review Details
1	1	1	1	interface snappy screen incredible bright colours out
2	2	1	1	quick responsive interface device boots seconds s co
3	3	2	1	iphone fan long time galaxy s blows iphone water sk
4	4	2	1	phone blazing fast ultra responsive click open app ba
5	5	3	1	love phone day time loved day girl thing improved tra
6	6	3	1	part quality camera complaining end love lumia app

Fig.1.3. Cleaned Reviews

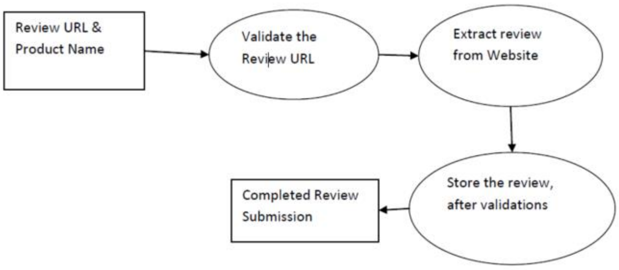

Review URL & Product Name
Validate the Review URL
Extract review from Website
Store the review, after validations
Completed Review Submission

Fig. 1.4 illustrate the sequence of Review Collections

Frequency Computation: This module is computes the text frequency. It indicates that how many times each keyword is appeared in each review and assign them a unique frequency ID. Frequency ID is unique for all the tokens in the frequency matrix.

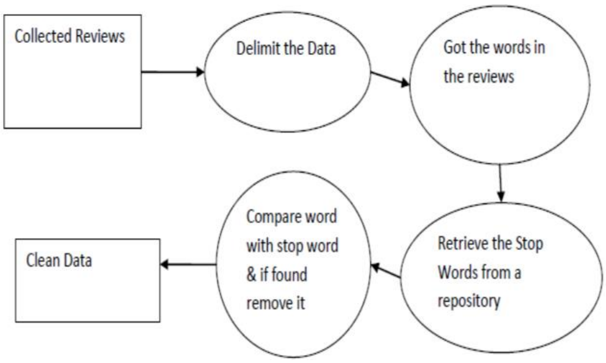

Collected Reviews
Delimit the Data
Got the words in the reviews
Retrieve the Stop Words from a repository
Compare word with stop word & if found remove it
Clean Data

Fig. 1.5 Describe the sequence of Cleaned Data

Feature based Text Frequency Computation: Single product contains many reviews. So this module is used to obtain Computation of Frequency across all products and all reviews per feature.

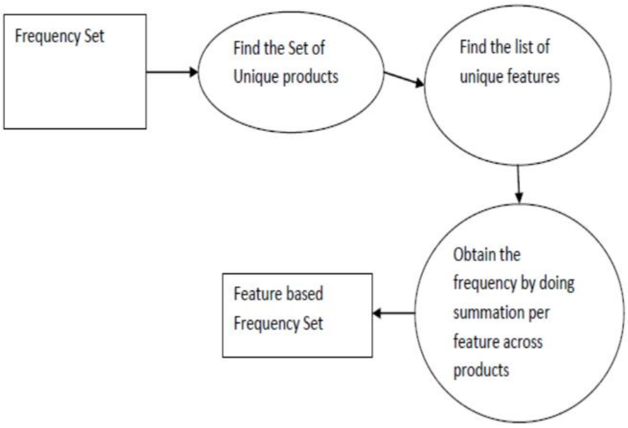

Frequency Set
Find the Set of Unique products
Find the list of unique features
Obtain the frequency by doing summation per feature across products
Feature based Frequency Set

Fig.1.6 illustrate the sequence of Feature based Frequency

Feature based positive frequency: This module is used to computes positive tagging, negative tagging and neutral tagging.

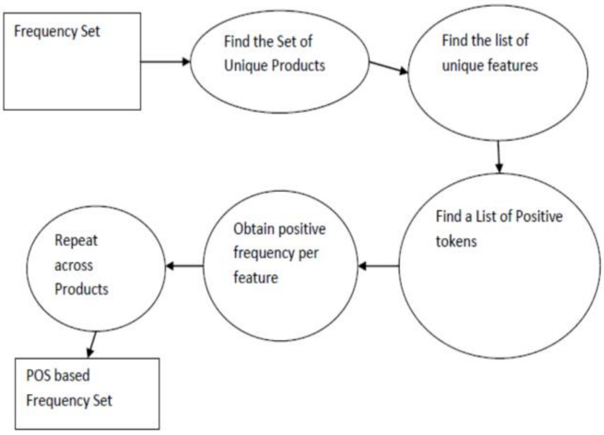

Frequency Set
Find the Set of Unique Products
Find the list of unique features
Find a List of Positive tokens
Obtain positive frequency per feature
Repeat across Products
POS based Frequency Set

Fig. 1.7 Describe the sequence of POS based Frequency

Feature Extraction Matrix: Generation of FEM matrix which has one row per product and one column per feature. Products are mobiles and features are Screen, Price, Speaker, Battery, Camera and Quality.

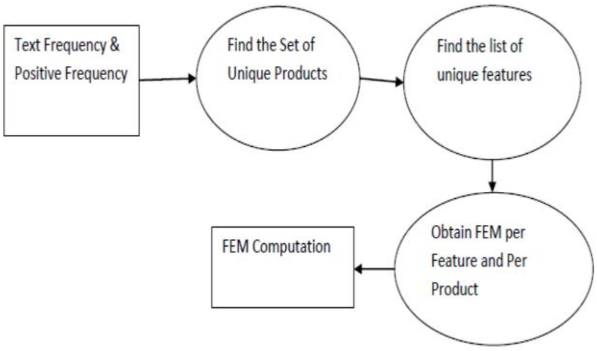

Text Frequency & Positive Frequency
Find the Set of Unique Products
Find the list of unique features
Obtain FEM per Feature and Per Product
FEM Computation

Fig.1.8 illustrate the sequence of FEM Matrix Computation

Rank Products: List of products based on highest value of FEM for searched features and product recommendations are generated based on the user searched feature. We are looking for a product having positive polarity maximum, negative polarity minimum and Neutralpolarity maximum.

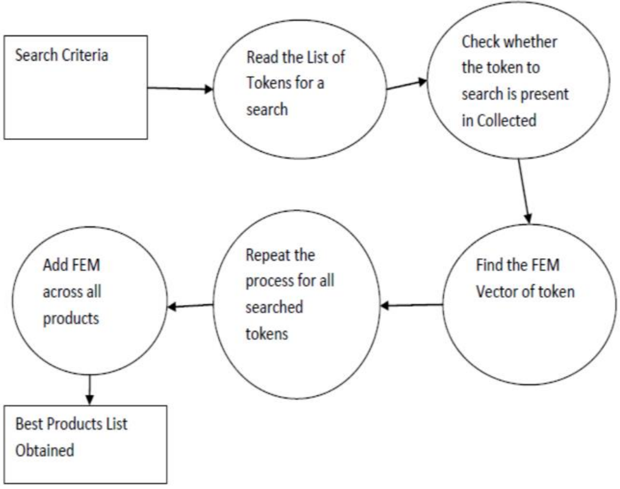

Search Criteria
Read the List of Tokens for a search
Check whether the token to search is present in Collected
Find the FEM Vector of token
Repeat the process for all searched tokens
Add FEM across all products
Best Products List Obtained

Fig.1.9 Describe the sequence of Ranking Products

IV. EXPERIMENTAL RESULTS

4.1 FEM Computation Algorithm

Step 1: Start
Step 2: Count the number of reviews (Nreview).
Step 3: *Retrieve the first review.*
Step 4: Convert the review into array using splitter.
Step 5: Count the Review array (Ncount).
Step 6: Retrieve the first token in review array.
Step 7: Compare review token with Stop word.
Step 8: If first token found in the stop word, remove that token otherwise store in the wordbuffer.
Step 9: Repeat the above step for all the tokens and reviews.
Step 10: Measure frequency for that tokens.
Step 11: Repeat the steps 10 and 11 for all tokens in the reviews.
Step 12: End

4.2 Algorithm for extraction of features

The FEM computation starts after Tokenization. Tokenization is the process of dividing tokens in the Cleaned Data. For every tokens Feature and Frequency is measured.

1. In the matrix M[r][f], consider review set R as tuples and feature set as columns.
2. *For* all review ri EXISTS IN R, do
If (SO (*fx*) = exists)
Set value (M[*ri*,*fx*]) = 1
else
Set value (M[*ri*,*fx*]) = 0

4.3 Polarity Computation Algorithm

Step 1: Start
*Step 2:*Retrieve list of Positive keywords, Negative keywords and Neutral keywords.
Step 3: Calculate number of Review.
Step 4: Compute Positive Polarity(If any review is having keywords like e.g. Good, Nice, Awesome, Amazing, Excellent etc.) Negative Polarity(If any review is having keywords like e.g. Worst, Bad, Slow, Very Slow etc.) and Neutral Polarity (If any review is having keywords which are not coming under Positive or Negative then these words are considered as Neutral Keywords).
Step 5: End
This algorithm retrieves list of positive keywords, Negative keywords and Neutral keywords. Compute Positive Polarity, Negative Polarity and Neutral Polarity per sentence.

V. CONCLUSION AND FUTURE ENHANCEMENT

This paper study a prototype system can be used to track and manage customer reviews. With the rapid development of e-commerce, customer reviews will become more and more important for e-commerce enterprises and manufacturers. The prototype system model can be a reference for e-commerce enterprises, which is a cost-effective solution available to manage and analyse online reviews. The current paper first collects the reviews of products from the web and then parses the reviews to clean collected information. Cleaned data are divided to determine tokens. Once the token is identified it computes the frequency of identified keywords. The frequency of keywords is used to represent features in our proposed model. The FEM matrix is constructed by using the list of Features to find the rank of product. In this model we have considered various features such as Screen, Price, Speaker, Battery and Camera and Quality to provide the overall sentiment to analyze the reviews and comments of customers in an E-commerce website.

Our application currently selecting only top reviews. This application can be modified to consider all the reviews and select the best reviews. Current project works only for the mobile products. This paper can also be extended to analyze reviews and comments of other electronic devices.

VI. References

[1].D. Thakur and J. Singh, "The SAFE miner: A fine grained aspect level approach for resolving the sentiment," Proceedings of the 2015 Third International Conference on Computer, Communication, Control and Information Technology (C3IT), Hooghly, 2015, pp. 1-6, doi: 10.1109/C3IT.2015.7060151.

[2] Latha.S.S," Analysing the reviews and comments of customers in an e-commerce websites", JEST-M May- 2017 Issue-5 Page no[1-4],ISSN:: 2394 -6156.

[3].M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04), Aug. 2004.

[4] S. A. Sadhana, L. SaiRamesh, S. Sabena, S. Ganapathy and A. Kannan, "Mining Target Opinions from Online Reviews Using Semi-supervised Word Alignment Model," 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM), Tindivanam, 2017, pp. 196-200, doi: 10.1109/ICRTCCM.2017.66.

[5]. G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion Word Expansion and Target Extraction through Double Propagation," *Computational Linguistics,* vol. 37, No. 1: 9.27, 2011

[6] G. Dubey, A. Rana, and N. K. Shukla, "User reviews data anal-ysis using opinion mining on web,"in 2015 InternationalConference on Futuristic Trends on Computational Analysisand Knowledge Management (ABLAZE), pp. 603–612, Noida,India, February 2015

[7] R. Feldman. (2013) communications of the ACM on techniques and applications for sentiment analysis. [Online]. Available: http://cacm.acm.org/magazines/2013/4/162501-techniques-andapplications- for-sentiment-analysis/fulltext (2013), last accessed 6
march, 2014.

[8] A. Parashar and E. Gupta, "ANN based ranking algorithm forproducts on E-commerce website,"in 2017 Third International Conference on Advances in Electrical, Electronics, Infor-mation, Communication and Bio-Informatics (AEEICB),pp. 362–366, Chennai, India, February 2017.