International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Aspect Based Sentiment Analysis for E-Commerce Websites with Visualization through Machine Learning Algorithm

Nandini S, Yathish D P

Abstract: ***E-commerce is evolving at a rapid pace that new doors have been opened for the people to express their emotions towards the products. The opinions of the customers plays an important role in the e-commerce sites. It is practically a tedious job to analyze the opinions of users and form a pros and cons for respective products. This paper develops a solution through machine learning algorithms by pre-processing the reviews based on features of mobile products. This mainly focus on aspect level of opinions which uses SentiWordNet, Natural Language Processing and aggregate scores for analyzing the text reviews. The experimental results provide the visual representation of products which provide better understanding of product reviews rather than reading through long textual reviews which includes strengths and weakness of the product using Naive Bayes algorithm. This results also helps the e-commerce vendors to overcome the weakness of the products and meet the customer expectations.***

Keywords : Natural language Processing, Opinions, SentiWordNet, Visual representation

I. INTRODUCTION

Opinions play a vital role in recent years for both individuals and for organizations. Traditionally people used to rely on friends or relatives to know the quality of products. Even the organizations used to find the opinions by assigning people to conduct surveys regarding the quality and demand of products which helped to increase the market value. But due to tremendous development in technology, people started using internet which lead e-commerce sites to post individual opinions with respect to products. Recently if an individual wants to purchase a product, it is no longer necessary to enquire friends or relatives. Because there are numerous reviews, blogs, sites available which are given by existing users. By analyzing these reviews an individual can make a firm decision. Hence opinions play a prominent role for both individuals and organizations. Opinion mining is a field of data mining that tract individuals feel, views, emotions towards a particular product. Opinion mining sometimes also known as Sentiment Analysis or Review Mining. There are three components of opinion mining: Opinion Holder, Opinion Object and Opinion Orientation. Opinion Holder can be any person or organization who expresses his/her feelings towards the object. Opinion Orientation represents positive or negative feeling towards the product ie. View of an Opinion .[1] For Example: According to XYZ, *"This iphone is having excellent battery backup"*. Here, XYZ is an Opinion Holder expressing view towards iphone object. Opinion Orientation is Excellent. Here the person is giving positive orientation related to object.

A. Levels of Opinion Mining

Mining of opinions can be classified into three different levels: Sentence Level, Document level and Aspect/Feature Level[2][3]

Sentence Level: In this level, the document is broken into sentences. Each sentence in a document is analyzed to categories a sentence as positive or negative.

Document Level: In this level, the complete document is analyzed to categorize as positive or negative. Here document will be categorized.

Both sentence and document level does not give a clear information towards the opinion object. A positive archive does not imply that opinion holder has positive perspective for the product. Similarly, a negative document does not imply that the opinion holder doesn't like the product. In order to get a clear picture of opinions given by opinion holders for the product, Feature/aspect level classification is used[2][11].

Feature/Aspect Level: In this level, features of the products are identified. Based on these features, the products can be analyzed.

This paper mainly focus on feature level of mining as a finer level of granularity can be obtained from aspect level

B. Challenges of Opinion Mining

1 Usage of abbreviations and short forms: People use more of short forms and abbreviations to express their views towards the products. Because of this, it is very crucial to understand and score a opinion. Example: Usage of *'Gud'* instead of *'Good'*.

2 Spelling Mistakes: People make spelling mistakes while typing in order to express their reviews. It is a challenging task, since sometimes a word can be eliminated during pre-processing(i.e, extraction of adjectives or adverbs)

Revised Manuscript Received on March 30, 2020.
* Correspondence Author
Ms. Nandini S*, Assistant professor in Computer Science and Engineering Department, School of Technology, Bangalore.
Mr. Yathish. D. P, Assistant professor in Computer Science and Engineering Department, School of Technology, Bangalore.

3 Implicit Sentences: For analysing opinions, all sentences are considered to be important because they determine the actual meaning. Due to usage of implicit sentences the task becomes difficult.

4 Analysing multi-aspect: A single sentence which includes multiple aspects or features is difficult to determine the meaning.

Because it includes one positive aspect and other negative aspect. Example: *"This phone is good, but camera is bad"*.

5 Question based Sentence: In this, the customer enquiries regarding a product in the reviews section. So this sentence ahs to be eliminated[4]. For example: *"Is the battery quality of Samsung phone is good"?*.

6 Language Based Sentence: Sometimes, customer expresses their views in the regional language, which cannot be considered and has to be eliminated[4].

II. LITERATURE SURVEY

In [5], the authors focus on building a tree kernels for sentiment analysis rather than machine learning algorithms for extraction of features. In this paper, authors defined several tree kernels for various subtasks of opinion mining. With tree kernel development, the complexity in extraction of features was reduced but designing the tree kernel evolved to be a major issue. In this paper, the authors designed four tree kernels such as feature sentiment tree(FST), generalised feature sentiment tree(GFST), boundary marked GFST and generalised feature-sentiment tree with polarity labels. The major drawback found is dealing with the negative reviews, as in some applications or products negative reviews plays a major role for making a decision.

In paper[6], the authors focus on phrase-level/feature level of opinion mining. Supervised algorithm known as Naive Bayesian is used to identify the polarity of sentences. In this, the authors have collected only few camera review data sets which does not provide the overall view of the customers.

In [7], the attributes of product are considered. By considering product attributes, the summarization of product is given along with advantages and disadvantages. This summarization again is considered to be a tedious task, since organizations need to read through and it's a time consuming task.

In [8], the authors focused on data mining techniques extracted from twitter. The data set or reviews is derived from twitter regarding e-commerce. A tool known as Rapidminer is used to analyse the comparison of three different classification. In this, emotion conversion is also considered for polarity of sentences during data pre-processing. According to classification, Naive Bayes Classifier is the best classifier with respect to accuracy and precise predictions.

In [8], the visual summary of reviews is missing and hence the customers has to go through large data to come to a conclusion before purchasing the product

In [9], the authors focused on aspects e restaurant application where the features of restaurant like food, service, staff, atmosphere are considered. In this, the authors have used SentiWordNet, dependency passing. But only explicit sentences are considered for the analysing the text sentences.

In [10], the authors have used convolution neural networks approach. The inputs to this convolution neural network are given through Word2Vec pre-trained vectors. The authors have considered the Russian e-commerce websites to extract the reviews along with emoticons model.

III. PROPOSED WORK

Sentiment Analysis is a Natural Language Processing technique because it analysis textual content. It extracts individual emotions or views out of textual content. Since in this paper, we are dealing with textual data, there is a lot of pre-processing which has to be achieved before classification.

The algorithm used for the proposed work can be summarised as below in figure 1

Step 1: Retrieving of all text reviews through URLs from e-commerce websites

Step 2: Apply Pre-Processing Steps

i Lemmatization
ii Tokenization
iii Removal of stop words

Step 3: Extract only the featured sentences

Step 4: Using SentiWordNet score the featured word

Step 5: Apply Naive Bayes algorithm to calculate the orientation and calculate the accuracy

Step 6: Graphical representation

i import scores
ii Plot graph with abscissa containing features and ordinate containing score

Figure 1: Alogithm of proposed work

The below figure 2 depicts the proposed work flow in brief.

A. Review Extraction

The customer reviews are extracted through URLs from Amazon website. Only textual reviews are extracted and eliminated the images which are sometimes used by customers to express their views. The extracted reviews are stored in the data set[5].

B. Pre-Processing

The pre-processing step plays a significant role, as it removes unnecessary words or non-alphabetic characters.

The pre-processing can be done with three steps: Removal of stops words, Lemmatization, tokenization.

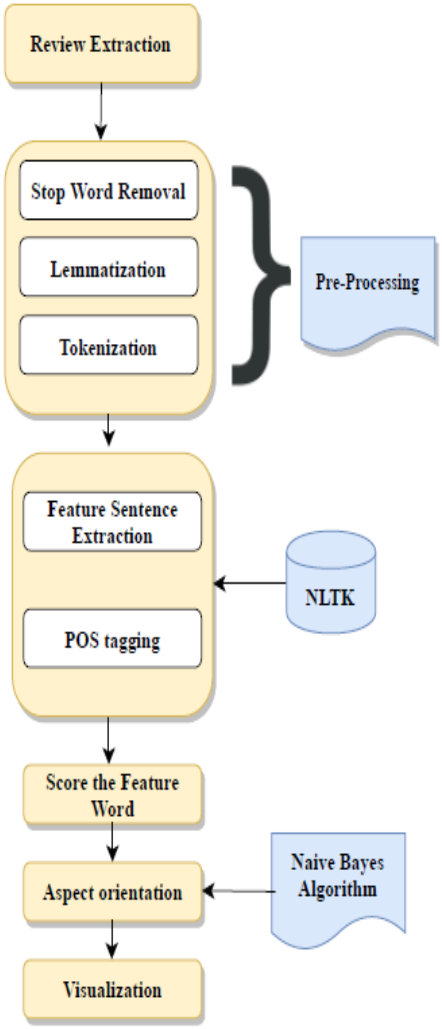

Figure 2: Flow of the proposed Word

i *Stop Word Removal*

In English language, most frequently used words which carry no information are known as stop words. In English Parts-of Speech(PoS) pronouns, prepositions and conjunctions may be considered as stop words and hence these words can be eliminated. To reduce the overhead caused during mining, words like THE, IS, ARE, WAS etc can be removed. A separate list of PoS is maintained and the extracted reviews are compared with this list to remove stop words[2].

ii *Lemmatization*

Stemming or Lemmatization is a process to determine the root word. For example: *troubling, troubles ,troubled* can be stemmed to basic root word "*trouble*".

iii *Tokenization*

Tokenization is a process of breaking sentences words and assigning PoS to each word. PoS tagger is available which tags all words. After tokenization, pre-processing data is stored in another text file.

This _DT Phone _NN have _VBP a _DT good _JJ Camera _NN

C. Featured Sentence Extraction

Feature Extraction is an important step in aspect/feature level of Sentiment analysis[6]. The aspect can be a single word or a phrase. After tokenization, the common aspects are nouns and noun phrases. Hence in order to extract features, we need to search for nouns and noun phrases. In this paper, we are concentrating on mobile aspects. The six aspects considered are Camera, Battery, Processor, RAM, Audio, Display.

D. Extraction of Opinion words

The opinion words used are usually adjectives, verbs, adverb adjective and adverb verb combinations for expressing the views.

E. Score the feature word.

In feature/aspect level, the orientation of the feature can be calculated by combining opinion word scores in the respective sentence. Here, orientation refers to the polarity i.e, the positive, negative or neutral view of the user. The proposed method uses SentiWordNet to assign the polarity. SentiWordNet contains the dictionary of words. The positive score represents the positive opinion, whereas the negative score represents the negative opinion. Figure 3 depicts the extracted feature word

```
bad,neg
good,pos
amazing,pos
wonderful,pos
horrible,neg
best,pos
better,pos
worst,neg
wierd,neg
poor,neg
beautiful,pos
defective,neg
imperfect,neg
damage,neg
difficult,neg
worse,neg
awesome,pos
fantastic,pos
clear,pos
```

Figure 3: Feature Word Extraction

F. Aspect and Sentence Orientation

The next step is to calculate the frequency of positive and negative opinions from extracted opinion words. By using Naive Bayes algorithm, both sentence and aspect orientation can be calculated.

The probabilities or count can be calculated by using Naive Bayes Classifier, steps are as follows in figure 4

```
Initially, let the probabilities of positive and negative count be zero[pos_count=0, neg_count=0]

if word is a positive word in SentiWordNet then
        mark(word)
        orientation ← count positive
        increment pos_count
end if
if word is a negation word in SentiWordNet then
        orientation ← count negative
        increment neg_count
end if
return orientation
```

Figure 4: Naive Bayes for calculating propabilities

G. Visualization

By calculating the count, the visual summary is provided by bar graph with X-axis containing features and Y-axis containing scores.[11]

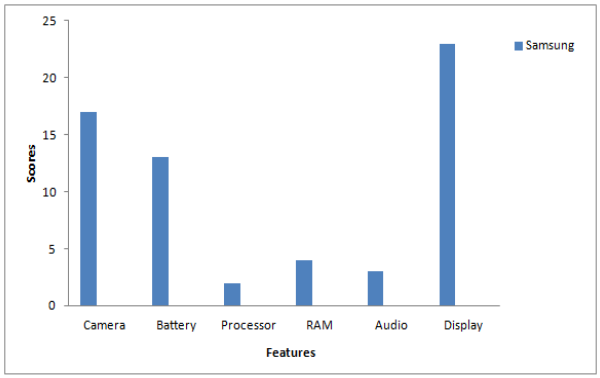

Figure 5: Positive opinions of customers

The above figure 5 depicts the positive opinions of customers based on features. By calculating scores for the respective features the visual summary can be shown as in the figure 5

The proposed work can also be extended for comparing two phones based on the features[11]. The results are shown in the below figure 6

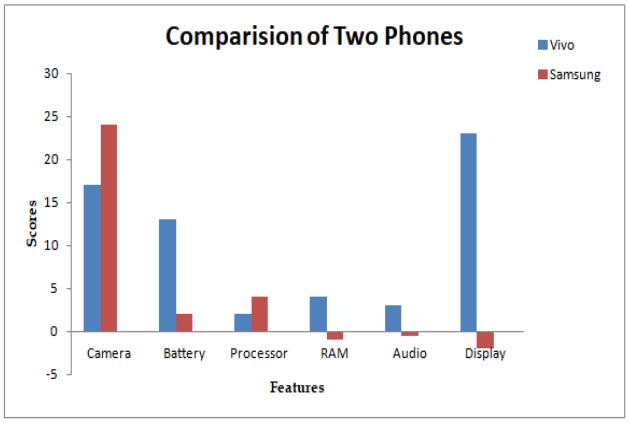

Figure 6: Comparison of two phones

IV. RESULT AND DISCUSSION

Finally the no. of positive and negative opinions are extracted using Naive Bayes algorithm.

The performance measure using Naive Bayes algorithm are given by the following parameters, with following formulae:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

In the above formula, TP represents true Positive. TN is true negative, FP is false positive and FN represents false Negative instances.

True Positive(TP): It is the no. of positive instances that are correctly categorised as positive

True Negative(TN): It is the no. of negative instances that are correctly categorised as negative.

False Positive(FP): It is the no. of positive instances that are wrongly categorised as positive.

False Negative(FN): It is the no. of negative instances that are wrongly categorised as negative.

The table I represents the results of naive Bayes classifier

TABLE I: Results of Naive Bayes Classifier

Classifier	Accuracy	Precision	Recall
Naive Bayes	0.754568	0.585714	0.72329

V. CONCLUSION AND FUTURE WORK

Due to explosive usage of internet and social media, one can feel free to express their opinions or views towards the products. This helps in making a decision before purchasing any products since the merits and demerits are provided by the existing users. In " *Aspect Based Sentiment Analysis for E-Commerce Websites With Visualization through Machine Learning Algorithm* ", we have proposed a fine grained feature level of opinion mining. It is helpful as a customer or organization will have a clear view of products based on features. This method provides a visual representation which is very easy to conclude rather than going through lengthy textual data. Comparison of two phones based on features is also possible so that the vendors can improve the quality of products and also individual can have a clear picture of products before buying. Implicit sentences are also considered during pre-processing. This work can be extended by using other machine learning algorithms for better accuracy. This work can also be extended for an automated feature extraction for any applications using machine learning algorithms.

REFERENCES

1. Review on Techniques and Tools used for Opinion Mining", Asmita Dhokrat, Sunil Khillare, C. Namrata Mahende, International Journal of Computer Applications Technology and Research Volume 4– Issue 6, 419 - 424, 2015, ISSN:- 2319–8656
2. "Aspect-Level Sentiment Analysis on E-Commerce Data", Satuluri Vanaja, Meena Belwal, Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018) IEEE Xplore Compliant; ISBN:978-1-5386-2456-
3. "A Study of Sentiment Analysis Task and It's Challenges", Shubham V. Pandey, A. V. Deorankar, 978-1-5386-8158-9/19/2019IEEE
4. "An Effective Model for Aspect Based Opinion Mining for Social Reviews", Jibran Mir, Muhammad Usman, The Tenth International Conference on Digital Information Management (ICDIM 2015)
5. "An approach Based on Tree Kernels for Opinion Mining of Online Product Reviews", Peng Jiang, Chunxia Zhang, Hongping Fu, Zhendong Niu, Qing Yang, 2010 IEEE International Conference on Data Mining
6. "Extracting Aspects and Mining Opinions in product reviews using Supervised learning algorithms," A.Jeyapriya, C.S.Kanimozhi Selvi, IEEE SPONSORED 2ND INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION SYSTEMS(ICECS '2015)
7. "Fine-grained Sentiment Analysis of Online reviews", Yan Wan, Hongzhurui Nie, Tianguang Lan, Zhaohui Wang, 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)

8. "Sentiment analysis about E-commerce from Tweets using decision Tree, K-Nearest Neighbor and Naive Bayes", Achmad Bayhaqy, Sfenrianto Sfenrianto, Kaman Nainggolan, Emil R. Kaburuan
9. "Syntactic Approach for Aspect Based Opinion Mining", Chinsha T C. Shibily Joseph, Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)
10. "Sentiment Analysis of Product Reviews in Russian using Convolutional Neural Networks", Sergey Smetanin, Mikhail Komarov, 2019 IEEE 21st Conference on Business Informatics (CBI)
11. "OPINION MINING", Bing Liu
12. "Sentiment Analysis on E-commerce Application by using Opinion Mining", Nitu Kumari, Dr. Shailendra Narayan Singh, 978-1-4673-8203-8/16/c 2016 IEEE

AUTHORS PROFILE

Ms. Nandini S currently associated with GITAM (Deemed to be University) as Assistant professor in Computer Science and Engineering Department, School of Technology, Bangalore. She has B.Tech degree in Information Science and Engineering from Sapthagiri College of Engineering, Bengaluru and M.Tech in Computer Network and Engineering from Nitte Meenakshi Institute of Technology, Bengaluru. She has professional experience of Three years of teaching. Her areas of interest are Computer Networks, Software Engineering, Artificial Intelligence and applications, Data mining and Data Warehouse, Computer Organization and Architecture, Machine Learning, Real Time Systems, Programing Language and Pragmatics, Operating Systems Robotics, Internet of Things and Sensor Guided Robotics.

Mr. Yathish.D.P is currently associated with GITAM (Deemed to be University) as Assistant professor in Computer Science and Engineering Department, School of Technology, Bangalore. He has B.Tech degree in Computer Science and Engineering from Channabasaweshwara Institute of Technology, Gubbi,Tumakuru and M.Tech in Computer Science and Engineering from Siddaganga Institute of Technology, Tumakuru. He has professional experience of Three years of teaching. His areas of interest are Compiler Design, Formal Language and Automata Theory, Software Engineering, Artificial Intelligence and applications, Machine Learning, Web Technology and its real time applications, Programing Language and Pragmatics, Operating Systems Robotics, Internet of Things and Sensor Guided Robotics.