Efficient Implementation of Energy Saving Technique for Data Mining Applications

Kainat Ayub[1], Dr. Sunil Patil[2], Abhinav Shukla[3]
M.Tech (CSE) Scholar, Department of CSE, RKDF University Bhopal, India[1]
Professor, Department of CSE, RKDF University Bhopal, India[2]
Head of Department, Department of CSE, RKDF University Bhopal, India[3]

Abstract - *Websites for online shopping is becoming more and more popular nowadays. Companies are eager to know about their customer buying behavior to increase their product sale. Extracting knowledge from large database, Data Mining is the key approach to use for accurate result. But in our context, we have to process customer reviews from large E-commerce, database for which Opinion Mining is the best approach for mining customer reviews about the product. The widely available internet resources are letting the users to shop any products anywhere, anytime at any cost. With the brisk development in the 3G and 4G we can expect a tremendous development in the area of M-commerce and E-commerce. In existing papers, opinion mining is used to process the online product reviews, feature and recommend the best product among others. Natural Language Processing (NLP) and Naive Bayes classification both are used to determine the polarity of reviews (obtain a polarity score from negative review and positive review). In this paper a novel technique is proposed for opinion mining and feature extraction of product reviews. The objective is to encourage the customers and assist them in choosing the right product. It is based on natural language processing, opinion mining and AdaBoost classifier. Results indicate that the proposed methods are highly effective and efficient in performing their tasks. We will also aim at improving the accuracy of our opinion polarity detection and feature extraction among other techniques.*

Key Words: **Opinion Mining, Part-of-speech (POS) Tagging, Natural Language Processing (NLP), Sentiment** Analysis, Naïve Bayes Classification and AdaBoost classifier.

1. INTRODUCTION

As we all know very well that E-Commerce sites are gaining popularity across all over the world. Customers are migrating towards online purchases more instead of going to the markets because of its easiness, convenience, reliability, and rapidness. There are a number of Online shopping websites that are available on the internet, such as Amazon, Flipkart, Snapdeal, Jabong, Myntra, Paytm, Zovi, etc. These websites allow the users to buy products with ease and lesser prize. A lot of attractive and day-to-day useful products like books, electronic goods, home appliances, clothing, and footwear are sold from these sites. These websites provide an option to the customers to write their review about their product that they buy from these sites. These reviews or opinions are very helpful to the users, manufacturers of the product as well as the developers of the website. The users who are in quandary to buy a product can read the reviews about the particular product from these websites so that they can have a view about their product before buying it and also know which is on the 1st position. Potential buyers can make decisions based on the reviews of customers who have purchased and experienced the product. The manufacturers of the product will be able to know the minor or major drawbacks of the product from the reviews which helps the manufacturers to get a chance to release the updated version of the product which satisfies the reviews that are mentioned in the websites. Hence online reviews play a significant role in understanding the

customer's voice. Sentiment analysis and opinion mining through machine learning algorithms offer a great possibility in automating the process of gathering, processing, and making sense of the data. By studying the reviews of customer about products helps both shoppers as well as E-commerce companies too.

As we know that how these reviews are important, but there is a problem which may be faced by an individual. It may be possible that a product is having 10 reviews, but it is also possible that a product having 3900 reviews with nearly 4800 ratings [1]. To overcome this problem E-commerce sites provide as many details about the products as possible on their web page. To make it easier for customers to make decisions E-commerce site provides big information in the form of these following aspects:-

- Star rating of products (like we observe movies reviews)
- Product specifications and price
- Age of reviews
- Overall score of a product with its image.
- Product Features in the review mention by customer.

2. LITERATURE SURVEY

E-commerce sites are growing rapidly; similarly opinion mining is also grabbing the rapidness as well because increase in sales and profit of a company is main motive and for that they must have the knowledge about their customer buying behavior.

In this unit, to grab the accurate knowledge about customer needs, there are so many different tools and techniques existing for processing customer reviews about products and score of products with product comparison. This section shows related work in context to mine opinions of customer's reviews and problems in the existing approaches. As to analyze the reviews of products, opinion mining is the key approach.

2.1 Related Work

There are so many tools and techniques; these are used for mining large database but for mining opinions of customer about products a unique technique is necessary. So, many authors proposed technique accordingly let's discuss as follow:-

- Author Pang and Liu have extracted reviews and done an opinion mining analysis and sentiment analysis [4], [5].
- Blessy founds that Naïve Bayes Classifier can also use in processing customer reviews [2].
- Liu, Bing, and Zhang used POS Tagging to identify the phrase of words. By tagging the words with its phrase as a word is a noun, adjective or verb [5], [6].
- Jayashri and Mayura have proposed a system in which Support Vector Machine (SVM) using for mining [7].
- Rajeev and Rekha have developed a system in which by using tools and python code customer reviews extracted and processed based on that product recommended [1].

2.2 Comparison of Different Systems

Figure 1 shows the comparison of various existing tools and also differentiate our work in this context. These parameters and tools are used here for comparison that gives us a clear idea of where a concrete work can be done and also displays feasible technique.

Fetures/Techniques Other survey papers	Sentiment Classification	NLP	Review Summarization	Naïve bayes classification	Opinion Summarization	Opinion mining/Sentiment Analysis	Parts of Speech tagging	Feature-based Summarization	AdaBoost
Recommender System on Customer Reviews using Sentiment Classification	✔	✔	✖	✖	✖	✔	✖	✖	✖
Feature Based Summarization of Customers' Reviews	✔	✔	✔	✖	✔	✔	✔	✔	✖
Mining and Summarizing Customer Reviews	✔	✔	✔	✖	✔	✔	✔	✔	✖
Challenges in developing opinion mining tools for social media	✔	✔	✖	✖	✖	✔	✖	✖	✖
Implicit Aspect Identification Techniques for Mining Opinions	✔	✖	✖	✖	✖	✔	✖	✔	✖
Existing paper	✖	✔	✔	✔	✖	✔	✔	✔	✖
Proposed paper	✔	✔	✔	✔	✔	✔	✔	✔	✔

Taxonomy Chart

Fig -1: Comparison of Different Systems

3. SYSTEM ARCHITECTURE

In our paper a novel technique is proposed to recommend online products to the customer after comparing products with each other from large E-commerce database. Natural Language Processing (NLP) technique is used to obtain the polarity of the reviews and AdaBoost classifier is used for review processing from different E-commerce sites.

AdaBoost general algorithm:

1. Start

2. Dataset load to system.

   a. $S=\{a1,a2,..,a_n\}$, a dataset

   b. $a_1,a_2...,a_n$ – attributes of dataset (column names)

3. Weight assigns to each attribute according their priorities.

   (Which attribute should take for consideration to find the attackers. Attribute with higher priorities or weight will take first and so on.)

4. Labeling to each review by considering weight of attribute (positive or negative review)

5. The Dataset will be prepared for classification with help of step 1 2 3.

6. Classification is done on basis of label of review.

7. After a degree classification of each label (positive or negative) gets calculated.

8. Compare the degree with a threshold value.

9. Result from step 7 show classification of the dataset.
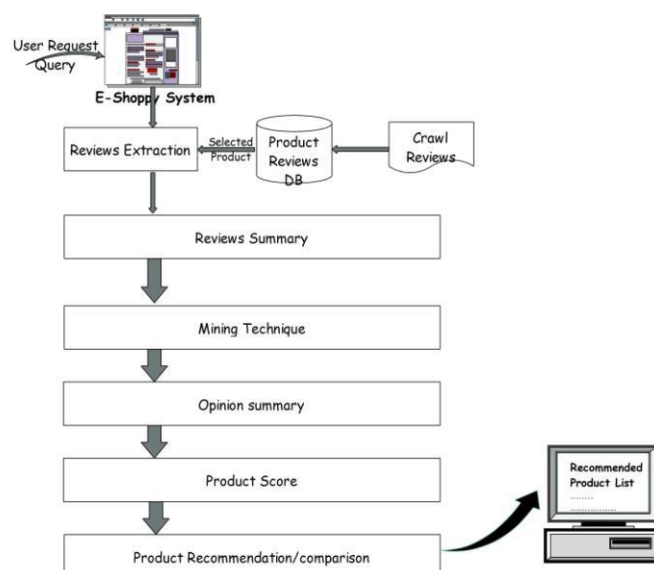
10. Stop.

Fig -2: System Architecture

In our system we select AdaBoost classifier as a mining technique because it gives more than 90% accuracy with required result. Let's understand AdaBoost classifier by splitting its name as Ada or Boost, individually their meaning is Ada- Adapt and Boost- Boosting. Adapt or consider weak learner and boost it to strong learner for better accuracy. Some modules are in our system from which reviews are processed step by step are as follows:-

1. Product selection

In this module customer select product from given category. If selected product is not found in the local database, then failure message is shown to the customer. Initially we add mobile products in the category because mobile phones are one of the most reviewed and sold products on E-commerce sites.

2. Reviews extraction

This module explains itself by its name "review extraction" as customer reviews extracted from large E-commerce site and store into localDB. The Local database is managed by admin in which lots of data related to product reviews are stored; those extracts from different sources like Flipkart; Amazon; and other E-commerce site. But we prefer Amazon because several

products had more number of valid reviews on Amazon more than any other e-commerce site.

3. Reviews Preprocessing

After review extraction step customer reviews are forwarded for processing. Mining has done on reviews to extract more precise reviews, and then sentences are split into words to identify their phrase (POS Tagging) and polarity (NLP). By splitting reviews, feature based opinion and reviews identify easily. These processed reviews are forward for further step.

4. Mining Technique

In this module mining technique using to mine opinions of customer reviews and summary of reviews generate and store for further step processing. Actual mining comes under this step afterwards score will generate based on this review summary.

5. Opinion summary

Overall summary generates in this module. After processing reviews precise and valuable information is summarized. In this overall summary, these attributes are stays: - customer id; product id; review/comments; helpfulness score; a ranking of the product. Based on these attributes product score will generate.

6. Product score

This module includes the star ratings, the polarity of the reviews, and age of the review and the helpfulness product score of the review for calculating the score for a product. The overall product calculation is done and best rated product is shown to the customer with product image.

7. Recommending product list

In this module a product list is recommended to the customer on the customer panel when he/she will request query as product mobile then best suitable product with its score will display.

After processing customer reviews from the above steps, product recommended based on overall score. Product list recommend to the customer on request and make them buy best product and companies having customer knowledge about buying.

4. RESULT

After implementing opinion mining of user reviews by processing it through module wise, best scored product recommend to customer on selected categories. The comparison between existing technique (Naïve Bayes Classifier) and proposed technique (AdaBoost Classifier) is shown by plotting graphs of accuracy v/s category and time v/s category. Comparative graphs are as follows:-
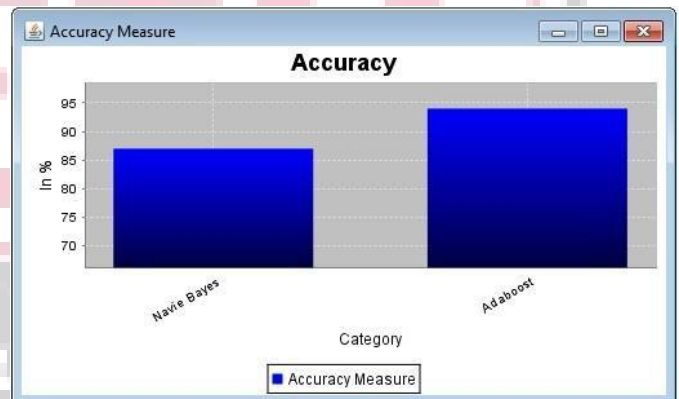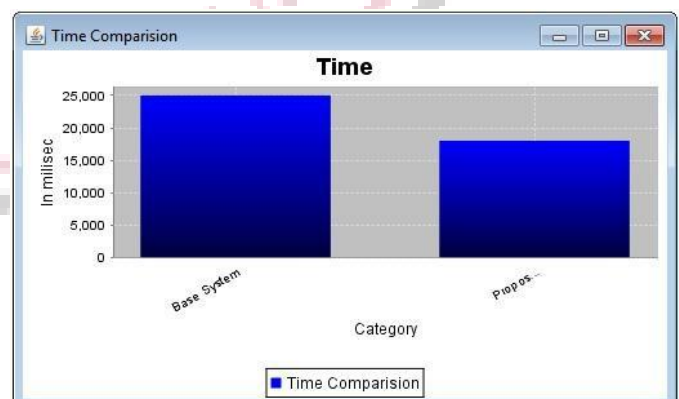

Fig -3: Accuracy Graph

Fig -4: Time Graph

5. CONCLUSIONS AND FUTURE ENHACEMENT

 In this paper, we have shown our work on opinion mining of online customer reviews of mobiles and tablets. We have shown the ranking of products, product score, comparison between more than two products, recommend product list along with its overall score. We also compare existing technique (Naive Bayes Classifier) from the proposed technique (AdaBoost Classifier). The main aim of all this is to collect beneficial information from the thousands of reviews about products and conclude them by recommending best suitable products to the customers.

As future enhancement the system may be extended to compare more than two products and product categories increase in more number of products rather than just single product like mobile phones. The aim of our proposed system is to help the user to select the best product they need.

REFERENCES

[1] Rajeev PV, Rekha VS (2015) Recommending products to customers using opinion mining of online product and features. ICCPCT 2015 IEEE.

[2] Blessy Selvam, S.Abirami A survey on opinion mining framework. IJARCCE, 2013.

[3] Zhan, Jiaming, Han Tong Loh, and Ying Liu. "Gather Customer Concerns From Online Product Reviews A Text Summarization Approach." Expert Systems with Applications, 2009.

[4] Pang, Bo, and Lillian Lee. "Opinion Mining And Sentiment Analysis." Foundations and trends in information retrieval, 2008.

[5] Liu, Bing, and Lei Zhang. "A Survey of Opinion Mining And Sentiment Analysis." Mining Text Data. Springer US, 2012.

[6] Minqing, and Bing Liu. "Mining And Summarizing Customer Reviews." Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004.

[7] Jayashri Khairnar*, Mayura Kinikar, Machine Learning Algorithms for Opinion Mining and Sentiment Classification International Journal of Scientific and Research Publications, 2013.

[8] Maynard, Diana, Kalina Bontcheva, and Dominic Rout. "Challenges In Developing Opinion Mining Tools For Social Media."Proceedingsof@NLP(2012).

n-gl.com