# Effective Product Ranking Method based on Opinion Mining

Madhavi Kulkarni
Student
Department of Computer Engineering
G. H. Raisoni College of Engineering &
Management
Pune, India

Mayuri Lingayat
Asst. Professor
Department of Computer Engineering
G. H. Raisoni College of Engineering &
Management Pune, India

## ABSTRACT

As internet is spreading out its bound, the demand of online transaction is also getting considerably increased. Now everyone wants fast and direct to home service without tacking any efforts. Online shopping is a way of effective transaction between money and goods which is done by end user without spending a large time span. Every product on online shopping website is associated with reviews which represents quality of that particular product. Every time the consumers are purchasing the product online by reading the product review. But reading all these reviews before buying product, consumes more time. Hence there is need of some systematic analysis of product reviews which helps to the consumer to find effective product among millions of the products. Here we have proposed a novel approach to rank the product efficiently by mining the genuine reviews of the product. But major problem arises when there is assignment of fake review given by anonymous user. So this system will provide methodology which will allow only those users to give review who have purchased product from that website. Other users are not allowed to give review. This will reduce the wrong reviewing of product and customer will get reliable product.

## General Terms

Opinion mining

## Keywords

Reviews, product ranking, opinion mining, POS, xml documents

## 1. INTRODUCTION

In old days shopping was a concept in which a customer used to buy a product from a mall or from shop. And he was paying money to supplier at the time of shopping. For traditional shopping, customer needs to be physically present at shop or at mall. Also there was no review system available to describe the quality criteria of product. Customer used to buy product on the basis of retailers opinion. Sometimes retailers give fake feedback to sell their product. But in early days internet has done massive amount of evolution. Every activity is getting associated with internet. So for the sake of customer and his need concept of online shopping is introduced.

In online shopping customer can buy or sell his products by sitting at home and using digital devices like smartphones, laptops, computers etc. Here user is doing payment by means of credit card or net banking systems. There is no need of customer to be physically present over shop or at mall for purchasing product and paying money. For selecting good quality products among all products, online shopping provides

review of each product given by various customers. Normally customer refers these reviews before buying any product.

But it consumes more time of customer to read each and every review of the product given by the other customers and then take decision for purchasing product. As some reviews are good and some review are bad so customer has to examine each n every review before selecting that product. So we have proposed a methodology to guide customer for choosing a best one products. Here we are going to sort out the useful reviews of particular product by using opinion mining concept.
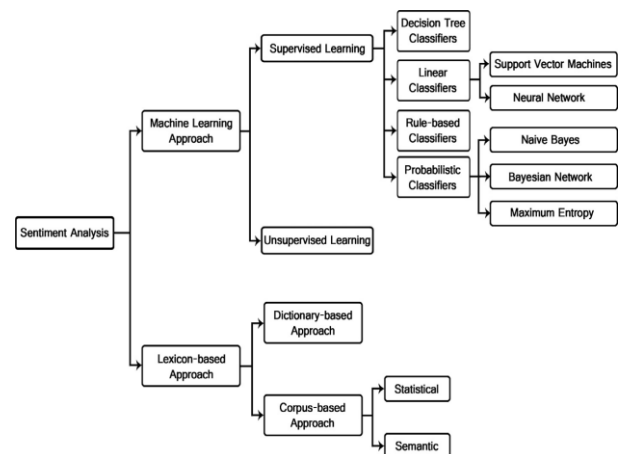


**Fig 1. Opinion mining techniques**

Another problem arises when there is allocation of fake review to any product. For example, if one mobile phone is available for selling on two different e shopping website like A and B. The B website can give fake negative feedback to the phone selling at website A. due to which customer will reject that phone although it is having good quality parameters. To avoid this we are going to design the mechanism which will accept review only from those customers who really have brought that product. This will minimize fake reviewing of products done by challengers.

Opinion mining is the methods of determining the approach of the speaker with respect to the product. In general opinion of the user is most important for organization or to individuals to increase the performance of the service. So opinion mining is the technique to extract the information about particular things based upon the reviews. The opinion mining is becoming very interesting area of research due to the emerging web technology. The machine learning is used to classify the opinion text. In the section two we will see the different types

of machine learning techniques. In the literature survey section we have explained inn details various types of machine learning and sentiment analysis.
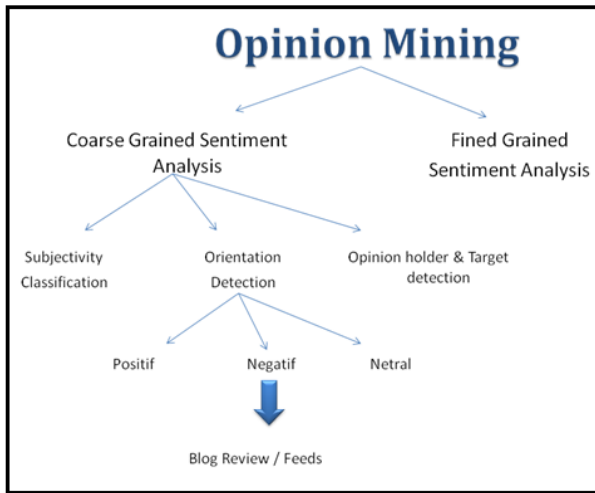


**Fig.2 Opinion Mining**

## 2. LITERATURE SURVEY

Much work has recently been undertaken in opinion mining over the last few years. The most related work in the field of opinion mining using various methods of machine learning. These methods are having varying degree of effectiveness. Authors in [2] presented the FP-growth method for frequent pattern mining from review documents which act as a backbone for mining the opinion words along with their relevant features by experimental data over two different domains which are very different in their nature. But they are trying to add more rules to improve the value of recall. Abd. Samad Hasan Basari, Burairah Hussin, I. Gede Pramudya Ananta and Junta Zeniarja [3] proposed their research on improving the efficiency of SVM classifier by using particle swarm optimization technique. SVM classifies the reviews into two classes as positive and negative. The positive class shows good message opinion otherwise the negative class shows the bad message opinion of certain movies. It is based on the accuracy level of SVM with the validation process uses 10-Fold cross validation and confusion matrix. Proposed hybrid Partical Swarm Optimization (PSO) is used to improve the election of best parameter in order to solve the dual optimization problem. Po-Wei Liang and Bi-Ru Dai [4] proposed a method for opinion mining on social media data. In their paper, they proposed a new system architecture that can automatically examine the sentiments of these messages. They combine their system with manually annotated data from Twitter, for the task of sentiment analysis. In their system, machines can learn how to automatically extract the set of messages which contain opinions, filter out non opinion messages and determine their sentiment directions (i.e. positive, negative). There is still need to improve and refine their system for the accuracy. So their future work is to work on such methods. In [5][6] also proposed methods for sentiment classification by using Naïve Bayes and SVM methods for social media facebook and twitter. Authors in [7] proposed a rankbox method which is ranking system for mining complex relationships on semantic web. They used SVM algorithm for classification. Darena, F., Zizka, J., Burda, K. [8] and Jianwei Wu, Bing Xu and Sheng Li [13] used unsupervised learning approach in their proposed method. In

[9] and [10], authors presented the pattern discovery and statistical data based polarity classification respectively. Authors in [16] proposed a product ranking method based on mining Chinese reviews. They used syntactic parser for finding polarity. In [14][15]and [20], they presented their work in online product ranking system. This is some related work where machine learning approaches are used in the field of opinion mining and sentiment classification. Peng Jiang, Chunxia Zhang, Hongping Fu, ZhendongNiu, Qing Yang [14] defined several tree kernels for sentiment expression extraction and sentiment classification, which are subtasks of opinion mining. Their proposed tree kernels encode not only syntactic structure information, but also sentiment related information, such as sentiment boundary and sentiment polarity, which are important features to opinion mining. Authors in [17] proposed a mining on online product reviews. They proposed extraction approach is different from the previous methods because they only mine the features of the product in opinion sentences which the customers have expressed their positive or negative experiences on. In order to find opinion sentence, they proposed a SentiWordNet-based algorithm. Yin-Fu Huang and Heng Lin [1] proposed a method of product ranking system using opinion mining techniques, where users can specify product features to get back the ranking results of all matched products. They overcome the disadvantage of traditional ranking system which is publicly used and not for a particular individual and it is based only on product sales. They used three more factors for calculating the product score.

However the reviews used as a dataset in this paper [1] are taken from the Amazon site directly. But the product reviews on a website are submitted by customers with no record of ever buying the product they are reviewing. So this could affect the ranking result. So in our proposed system, we are working on only the genuine reviews as a dataset. For that the customer has to purchase the product and then he will get the purchase id of this transaction. And then only he is able to write the review about a particular product.

## 3. EXISTING SYSTEM

Initially there was an issue of referring useful review among many. This problem was solved in existing system by means of opinion mining. User has to read all possible reviews for selecting that product. Opinion mining techniques recognizes the polarity of each sentence in all reviews given to product, and then computes the total of all similar products using the standard Functions.

Existing system has provided one feature that anyone can give feedback about any product. This causes limitation for this framework. The person from challenging e-shopping website can give fake feedback to the original website. This can be done for achieving popularity in internet marketing.

There is no special functionality describe in existing system for avoiding this fake reviewing. So existing system is able to mine negative and positive feedback but it fails to identify real one and fake one review in list of product review. If any product is having good quality but challenger user has enter inaccurate or negative review about product then other customers avoid to buy that product though the product is best. It will create a big loss for hosting website in terms of money, market position and customer feedback.

# 4. SYSTEM FRAMEWORK

As discussed earlier existing system is having some major drawbacks which need be overcome for effective re-ranking of reviews. Here we have proposed a framework, as shown in Fig. 3 for avoiding acceptance of fake feedback. In this we are using predefined dataset for review. Then we are going to do extraction of opinion from that reviews. After that we will remove the common words which create effect of common objective. Then depending upon strength of remaining objective we are computing degree of adverbs. Next step is to check sentence polarity. Then sentence polarity is integrated into the XML files. Then we will describe only those feedbacks which are given by customers who have already purchased the product. Our proposed system consists of following modules.

## 4.1 Selection of Dataset:

Here we are using pre-defined dataset of particular e shopping website. On the website user can specify his product brand and information of the product and download his brand and give the product review by using the unique id get after purchasing that product. These all data is being saved into the database. Dataset is a part of database and we are going to use this dataset as an input to system.
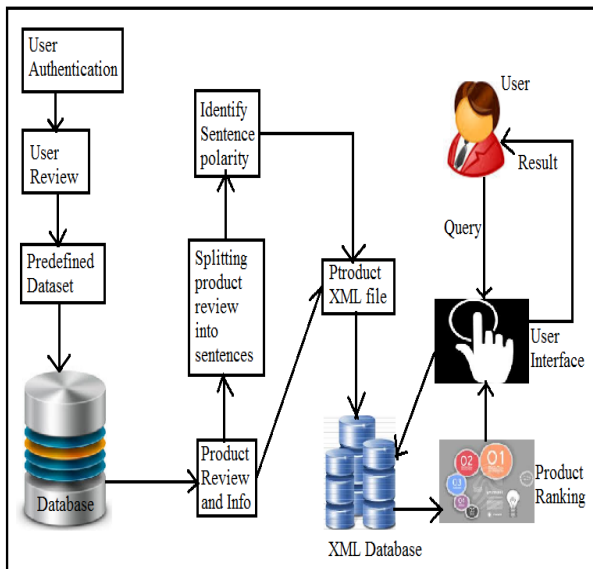


**Fig. 3 Proposed System Architecture.**

## 4.2 Mining of Opinion words:

We extract the opinion word by sentiment dictionary based approach. After mining the opinion words then calculate the opinion strength by using following formula,

$$P\{\widehat{W|C}\} = -\frac{Count(w,c)+1}{Count(c)+|V|} \quad (1)$$

Where, P is the set of polarity it may be positive or negative depending upon the value of word w and count c. For example polarity of adjective good is positive and the polarity of word bad is negative. The above figure shows the working of the proposed system.

## 4.3 Calculation Document Strength:

The opinion strength may be very high if P is a common adjective. So there is need of removing the effect of common adjective. This can be calculated as follow,

$$Cnb = argmax_{Cj \in C} P(Cj) \prod_{i \in positions} P\ \{Wi|Cj\} \quad (2)$$

Where, $C_{nb\ is}$ count of Native Bayes and set of P is the set of polarity. The normalization formula makes the value of P in between [0, 1]. Here w denotes word and c denotes number of count of occurrences word.

## 4.4 Degree of Adverbs:

**Table 1 Levels of adverbs**

| High Level (0.6) | Medium Level (0.5) | Low Level (0.4) | Negative |
|---|---|---|---|
| Very, much, to so, large, too, completely … | Partly, almost, rather, half… | Somewhat, slightly, little bit, half… | Not… |

The adverb can be modified to adjective that is (adverb + adjective) and increase weakness or strength of the adverb. It not only increases or decreases the strength but also change the polarity of the adverb by modifying the polarity of the adverb.

## 4.5 Calculating Sentence Polarity:

We are using three values to calculate the sentence polarity which we have calculated in the above sections.

$$P = \frac{P(w|c)}{P(w)} \quad (3)$$

In the above equation sentence P is the sentence containing the adverb p, numerator is the opinion strength of the P, denominator is the polarity frequency, and Degree of P is the degree of adverb modifying P.

## 4.6 Building Product XML Files:

In this section product information and sentence polarity is integrated into the XML files. This XML file consist of three sections one is product information , second is review section which describe the review of the product and product polarity, third is the specification section which describe the product specification.

## 4.7 Product Ranking:

The proposed system is rank only product selected by user not for all products. User can specify his brand, color, cost, etc. about the product. Then proposed system search the product with specification specified by the customer. Then system will find the product with specified product and with rank.

$$PMI(X,Y) = log_2 \frac{P(x,y)}{P(x)P(y\ )} \quad (4)$$

Where PMI is point wise mutual information. Equation describes that how much more do events x and y co-occur than if they were independent.

# 5. RESULTS AND ANALYSIS

The proposed system takes the dataset from the random online shopping website where the reviews about the two products i.e. camera and mobile phones are taken. Both the product reviews are worked on and according to polarity of the terms used in the reviews and comments, the products are ranked. Generally, the product having highest rank may not necessarily have good qualities, or might not be the best one, but after the use of product, the reviews that are given to the

product are also the very important factor that matters a lot while giving reviews about the product. People now a days, don't waste much time in providing the reviews about the products. So they are using shortcut words such as "GUD for GOOD and XLNT against EXCELLENT or LYK INSTEAD of LIKE". As the proposed system works totally on the polarity of the words, if such words are used, we may not be able to get the optimal results. So the proposed system is enhanced with another feature where in the database is trained to accept these kind of words if found in reviews, the words are replaced with the actual words corresponding to it manually during the data training part. The graphical result is shown in the following graph.
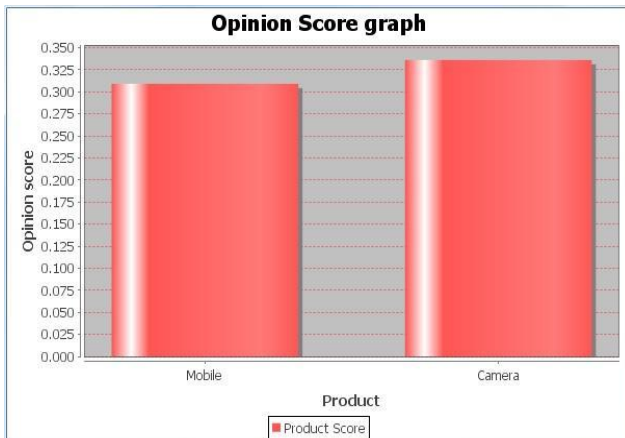


**Fig. 4 Graphical Results**

# 6. CONCLUSION

Now a days online shopping is becoming more popular because large number of sellers are selling product online. The people are purchasing product and give their review according the product quality. Our goal is to find the product with best quality and the quality is predicting from the users reviews about that product. We have proposed product ranking system which can take product information in the form of query and system will provide the product matching with customer's requirements along with product ranking. The experiment will helps to all to find best product online with less time. In future, focus will be on improving the parsing method

# 7. ACKNOWLEDGMENTS

# 8. REFERENCES

[1] Abd. Samad Hasan Basari, Burairah Hussin, I. Gede Pramudya Ananta and Junta Zeniarja, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization", Published by Elsevier Ltd 2013.

[2] Yin-Fu Huang and Heng Lin, "Web Product Ranking Using Opinion Mining" Conference on Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on, pages.184 – 190.

[3] Tanvir Ahmad and Mohammad Najmud Doja, "Opinion Mining using Frequent Pattern Growth Method from Unstructured Text", Computational and Business Intelligence (ISCBI), 2013 International Symposium on IEEE Conference Publications, pages. 92 – 95.

[4] Po-Wei Liang and Bi-Ru Dai, "Opinion Mining on Social Media Data", 14th International Conference on Mobile Data Management, 2013 IEEE , pages 91-96.

[5] Jalel Akaichi, Zeineb Dhouioui and Maria José López-Huertas Pérez,"Text Mining Facebook Status Updates for Sentiment Classification" System Theory, Control and Computing (ICSTCC), 2013 17th International Conference, pages. 640 - 645

[6] Akaichi, J., "Social Networks' Facebook' Statutes Updates Mining for Sentiment Classification", Social Computing (SocialCom), 2013 International Conference on, pages. 886 – 891.

[7] Na Chen and Viktor K. Prasanna, "Rankbox: An Adaptive Ranking System for Mining Complex Semantic Relationships Using User Feedback", Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on, pages. 77 - 84.

[8] Darena, F., Zizka, J., Burda, K., "Grouping of Customer Opinions Written in Natural Language Using Unsupervised Machine Learning", Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2012 14th International Symposium on, pages. 265 – 270.

[9] Ningzhong, Yuefeng Li, and Sheng-Tang Wu,"Effective Pattern Discovery for Text Mining", IEEE transactions on knowledge and data engineering, vol. 24, no. 1, January 2012, pages. 30 – 44.

[10] Dominique Ziegelmayer and Rainer Schrader, "Sentiment polarity classification using statistical data compression models", 2012 IEEE 12th International Conference on Data Mining Workshops, pages. 731 – 738.

[11] Balakrishnangokulakrishnan, Pavalanathan Priyanthan , thiruchittampalamragavan , Nadarajah Prasath and ashehan Perera, "Opinion Mining and Sentiment Analysis on a Twitter Data Stream", The International Conference on Advances in ICT for Emerging Regions – icter 2012, pages. 182 – 188.

[12] Krzysztof Jędrzejewski, MikołajMorzy proposed "Opinion Mining and Social Networks: A Promising Match", 2011 International Conference on Advances in Social Networks Analysis and Mining, pages. 599 – 604.

[13] Jianwei Wu, Bing Xu and Sheng Li "An Unsupervised Approach to Rank Product Reviews", 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pages. 1769 – 1772.

[14] Peng Jiang, Chunxia Zhang, Hongping Fu, ZhendongNiu, Qing Yang, "An Approach Based on Tree Kernels for Opinion Mining of Online Product Reviews", 2010 IEEE International Conference on Data Mining, pages. 256 - 265

[15] Weishu Hu, Zhiguo Gong and Jingzhi Guo, "Mining Product Features from Online Reviews", 2010 IEEE International Conference on E-Business Engineering, pages. 24 – 29.

[16] P. Tian, Y. Liu, M. Liu, and S. Zhu, "Research of product ranking technology based on opinion mining" Proc. 2009, The 2nd International Conference on Intelligent Computation Technology and Automation, pages. 239 - 243.

[17] Juling Ding, Zhongjian Le, Ping Zhou, Gensheng Wang, Wei Shu, "An Opinion-Tree based Flexible Opinion Mining Model", 2009 International Conference on Web Information Systems and Mining, IEEE, pages. 149 – 152.

[18] Jung-Yeon Yang, Jaeseok Myung and Sang-goo Lee, "The Method for a Summarization of Product Reviews Using the User's Opinion", 2009 International Conference on Information, Process, and Knowledge Management, pages. 84 – 89.

[19] Alexandra, Balahur and Andres Montoyo, "A Feature Dependent Method for Opinion Mining and Classification", Natural Language Processing and Knowledge Engineering, 2008, NLP-KE '08, IEEE, pages. 1 - 7

[20] Jian Liu , Gengfeng Wu and Jianxin Yao, "Opinion Searching in Multi-product Reviews", Proceedings of The Sixth IEEE International Conference on Computer and Information Technology (CIT'06) 2006, pages. 25