

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.017

*IJCSMC, Vol. 7, Issue. 4, April 2018, pg.87 – 96*

# AN ENHANCED APPROACH FOR PRODUCT ASPECT RANKING ON THE CONSUMERS REVIEWS

**M. Lovelin Ponn Felciah<sup>1</sup>**

*Asst. Professor in Dept. of Computer Applications  
Bishop Heber College (Autonomous)  
Tiruchirappalli, India  
[lovelinsathya@gmail.com](mailto:lovelinsathya@gmail.com)*

**R. Anbuselvi, Ph.D.,<sup>2</sup>**

*Asst. Professor in Dept. of Computer Science  
Bishop Heber College (Autonomous)  
Tiruchirappalli, India  
[r.anbuselvi@gmail.com](mailto:r.anbuselvi@gmail.com)*

---

*Abstract: Now a days, internet is the main source of information. Several ecommerce websites available where people discuss on different issues of product. Every product has an feature, and some of these features of a product are important and some can be unnoticed, and these important feature of an product plays a dynamic role in making decisions so as to buy the product or not. Therefore, identifying these important features of a product and ranking them according to their importance and significance considering customer reviews helps to take a firm and better decision. The objective of this work is to propose an efficient product aspect ranking framework for mobile phones, which automatically determine the important aspects of mobile phones from online consumer reviews, aiming at improving the usability of the countless reviews and to rank them in such a way that the customers find it easy to buy the product. These aspects are identified using vital aspects generally commented by a lot of customers and their views on how these aspects greatly influence the product. The system classifies the reviews on the basis of aspects and then the aspects are ranked with probability ranking algorithm. Millions of reviews about mobile phones from various websites are grouped and made available within each website by means of graphical representations of each aspect of different products.*

*Keywords: ecommerce, feature, mobile phones, aspect ranking, website.*

---

## I. INTRODUCTION

Recent year's there are so many websites are available for online shopping for e.g. ebay.com, amazon.com, shoppers.com, snapdeal.com, flipcart.com etc. Many marketing websites inspires customers to write reviews to express their opinions on various features of the products [1][11]. Large number of reviews are easily available online and enhancing more and more trust on products available for online purchases. Consumer reviews contain rich and valuable knowledge for both firms and users. However, the reviews are often disorganized, leading to difficulties in information navigation and knowledge acquisition [3]. The product aspect is nothing but the feature of product refers to component or attribute of certain product. A sample review "the battery backup of Samsung is amazing" reveals positive opinion on aspect Battery backup of Samsung. Several forum websites are available online for consumer to post their reviews on millions of products. Many reviews contain worth full knowledge and have become some valuable resource for both customers and business which are helpful to make consumers purchase decision. Consumers commonly search for quality information from that customers can earlier make their purchasing product decision, product development and marketing and consumer relationship management [2].

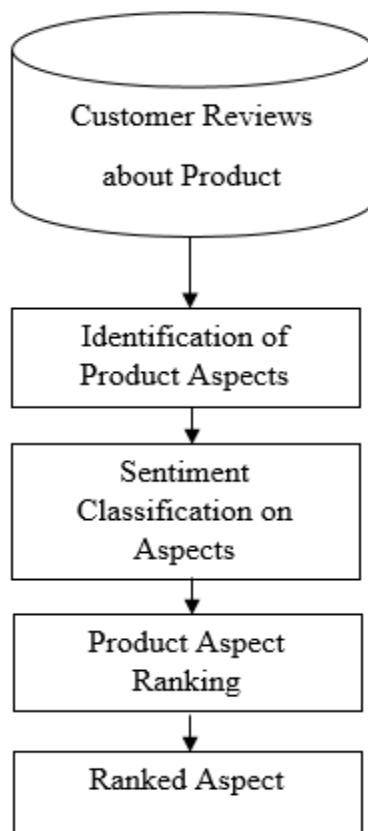
Generally product may have many aspects depending on product to product. For e.g. mobile phones have many aspects such as "battery backup", "camera quality", "design", "speed", "display" here some aspects have strong influence on the customers decision making as well as firms product development [4]. Hence, identifying important product aspects will improve the usability of several reviews and it gives benefit to both consumers and firms. Consumers can conveniently make wise purchasing decision by giving more attentions to the significant aspects, whereas firms can concentrate on improving the quality of these aspects and thus improves product brand effectively [5].

It is difficult task for retrieving reviews and identifying the important aspects of product from numerous reviews, if they would done by manually and also it is quite difficult to identify key aspects from textual reviews written by consumers [7]. Hence, an approach to automatically identify the important aspects is very much necessary. In A straightforward frequency- based solution considers the aspects which are frequently commented in consumer reviews as important aspects [8] . However, consumer's opinions on the frequent aspects may not determine their overall opinions on the product, and would not influence their purchasing decisions [6]. For example, most consumers has commented frequently that "there is bad signal connection" of iPhone4, but they may still give high overall ratings to iPhone4. On the contrast, some aspects such that 'design' and 'speed' may not be frequently commented, but generally are more important than " signal connection ". Hence, the frequency-based solution is not able to identify the truly important aspects [9].

## II. PROPOSED SYSTEM

### A. System Overview

Product aspect ranking frame work is beginning with an overview of the three major components that will be Product Aspect Identification, Sentiment Classification on Aspects and Probabilistic Aspect Ranking. First of all recognize the product aspects from the consumer reviews and then examine consumer reviews on the aspects by making use of the sentiment classifier.



*Figure 1:* Workflow of Proposed System

### B. Product Aspect Identification

The customer reviews generally comprises of pros and cons reviews, free text reviews, ratings, over all reviews and so on. The proposed approach working with all kinds of reviews. In the case of free text reviews, first split the reviews into sentences and split each sentence into words, then the frequent noun terms are refined and grouped together. In the case of pros and cons reviews, the aspects are represented in a vocabulary for identification of product aspects from free text reviews, and utilize every aspect to determine the Support Vector Machine (SVM). The SVM is used to identify the noun terms.

### C. Sentiment Classification on Product Aspects

The product aspects are examined by sentiment classification using the existing techniques such as supervised learning and lexicon based approaches. After the identification of the product aspects from the customer reviews the reviews are classified on the basis of the product aspects and also the reviews are classified into their polarities like positive and negative. The supervised learning methods train a sentiment classifier based on training quantity. The classifier is then used to predict the sentiment on each aspect. The reviews are classified using many learning-based classification models such as Support Vector Machine (SVM), Naive Bayes etc.

### D. Product Aspect Ranking Algorithm

Finally, the proposed Product Aspect Ranking Algorithm is used to detect the significant aspects of a product from number of customer reviews. The opinion in a review is a collection of expressions given to specific aspects in the review. In order to take purchase decisions by the consumers based on the aspects that are frequently commented by the consumers. Consumer's opinions on the specific product aspects influences the overall opinions of the product. There are the various aspects that are commented and the importance score is computed with the Probabilistic Aspect Ranking Algorithm. The reviews on the important aspects have strong effect on the overall opinion.

To obtain this overall opinion, the proposed framework compute the Overall rating or in every review  $r$  is generated from the weighted sum of opinions on particular aspect as follows

$$\sum_{k=1}^m W_{rk} O_{rk}$$

Where  $W_{rk}$  is the weight, the no. of opinion on the aspect  $a_k$  and  $O_{rk}$  is the opinion on the aspect  $a_k$ . Larger  $W_{rk}$  means  $a_k$  is more important and vice versa.  $W_r$  is vector of weights and  $O_r$  is a vector of opinion on specific aspect. Overall ratings are generated by the Gaussian distribution and probabilities are generated.  $\{W_r\}_{r=1}^{|R|}$  and  $\{\mu, \Sigma, \sigma^2\}$  are model parameters. While  $\{\mu, \Sigma, \sigma^2\}$  can be calculated from review corpus  $R = \{r_1, \dots, r | R\}$  using maximum likelihood.  $W_r$  in review  $r$  can be optimized through Maximum posteriori (MAP) estimation.  $W_r$  and  $\{\mu, \Sigma, \sigma^2\}$  optimized by proposed algorithm.

Optimizing  $W_r$  given  $\{\mu, \Sigma, \sigma^2\}$

$$\bar{W}_r = \left( \frac{O_r O_r^T}{\sigma^2} + \Sigma^{-1} \right)^{-1} \left( \frac{O_r O_r}{\sigma^2} + \Sigma^{-1} \mu \right) \quad \dots (1)$$

Optimizing  $\{\mu, \Sigma, \sigma^2\}$  given  $W_r$

$$\bar{\sigma}^2 = \frac{1}{|R|} \sum_{r \in R} (O_r - W_r^T)^2 \quad \dots (2)$$

**Algorithm: EPARBF Algorithm**

*Input* : Consumer review corpus R, each review  $r \in R$  is associated with an overall rating  $O_r$  and a vector of opinions  $O_r$  on specific aspects.

*Output* : Importance scores  $\bar{W}_k$  for all  $m$  aspects.

1. *Get the number of reviews*
2. *Create  $F = \{f_1, \dots, f_n\}$  in reviews.*
3. *for each  $f$  in  $F = \{f_1, \dots, f_n\}$  do*
4. *while not converged do*  
     *Update  $\{W_r\}_{r=1}^{|R|}$  according to Equation (1);*  
     *Update  $\{\mu, \Sigma, \sigma^2\}$  according to Equation (2);*  
     *End while*
5. *Compute aspect importance scores  $\bar{W}_k \mid_{k=1}^m = 1$ .*
6. *Calculate the weight of each aspect*
7. *Rank the aspects based on Frequency*
8.  *$P \leftarrow$  find Polarity ( $P[f]$ )*
9.  *$\{(f, o_i, \dots, o_j) \dots\} \leftarrow$  Output Tuples( $P(f) \in R$ )*
10.  *$tfidf_{i,j} = tf_{i,j} * idf_i$*
11.  *$apply \leftarrow$  classifier*
12. *Calculate  $T(SA)$  is total score of each feature*

### ***E. Mathematical Model***

Let  $S=I, P, O$

Where ,

I represents the set of reviews which are input to the Product aspect ranking system.

P represents the set of processes that are used for the sentiment classification and aspect ranking.

O represents the set of output for review processing.

I = I1, I2

P = P1, P2, P3

IO = IO1, IO2

O = O

Where,

I1 = Pros & Cons Review,

I2 = Free Text Review,

P1 = Process for identifying the Product Aspects,

P2 = Sentiment Classification,

P3 = Ranking the Aspects,

IO1 = Identified Product Aspects,

IO2 = Classification Sentiments,

O = Ranked Aspects.

## **III. IMPLEMENTATION DETAILS**

The proposed framework is divided into three modules. Three modules are successfully completed and gives desired output. First module of the proposed system is Product aspect identification in which consumer's reviews are input to this module and output of the system is identified aspects. Second module is Sentiment classification using SVM classifier on the basis of aspects. Third module is product aspect ranking in which the aspects that are identified from the reviews those aspects are ranked.

### ***A. Product Aspect Identification***

Consumer reviews are of in different formats. The consumers to give an overall rating on the product, describe concise positive and negative opinions (i.e. Pros and Cons) on some product aspects, and write a paragraph of detailed review in free text. Some websites. In summary, besides an overall rating, a consumer review consists of Pros and Cons reviews, free text review, or both.

#### ***The Pros and Cons reviews***

From the Pros and Cons reviews, the proposed framework identify the aspects by extracting the frequent noun terms in the reviews. The proposed system can obtain highly accurate aspects by extracting frequent noun terms from the Pros and Cons reviews and which obtains the vocabulary for the identification of product aspects.

### ***The Free Text Review***

For identifying aspects in the free text reviews, free text review is used as an input to the system. Then Pros and cons reviews are used to generate the vocabulary of the product aspects and that vocabulary is used for the identification of the product aspects. First free text reviews are split into the sentences and then each word of the sentence is parsed for the identification of the aspects of product.

### ***Parser***

In order to obtain more precise identification of aspects, system here proposes to exploit the Pros and Cons reviews as auxiliary knowledge to assist identify aspects in the free text reviews. In particular, a system first splits the free text reviews into sentences, and parses each sentence using parser. The frequent noun phrases are then extracted from the sentence parsing trees as candidate aspects. Since these candidates may contain noises, System further leverage the Pros and Cons reviews helps to identify aspects from the words. System collects all the frequent noun terms extracted from the Pros and Cons reviews to form a vocabulary then represent each aspect in the Pros and Cons reviews. This vocabulary is used in the SVM classifier for identification of product aspects.

### ***SVM Classifier***

The resultant classifier is in turn used to identify aspects in the words extracted from the free text reviews. It removes the noises from the reviews. The stop words are neglected from the free text reviews. It uses vocabulary of product aspects. Product aspect vocabulary is generated from the pros and cons reviews.

### ***B .Sentiment Classification on the basis of Aspects***

Reviews are classified on the basis of the aspects of the product aspects. Lexicon based approach is used to classify those aspects.

### ***C. Product Aspect Ranking***

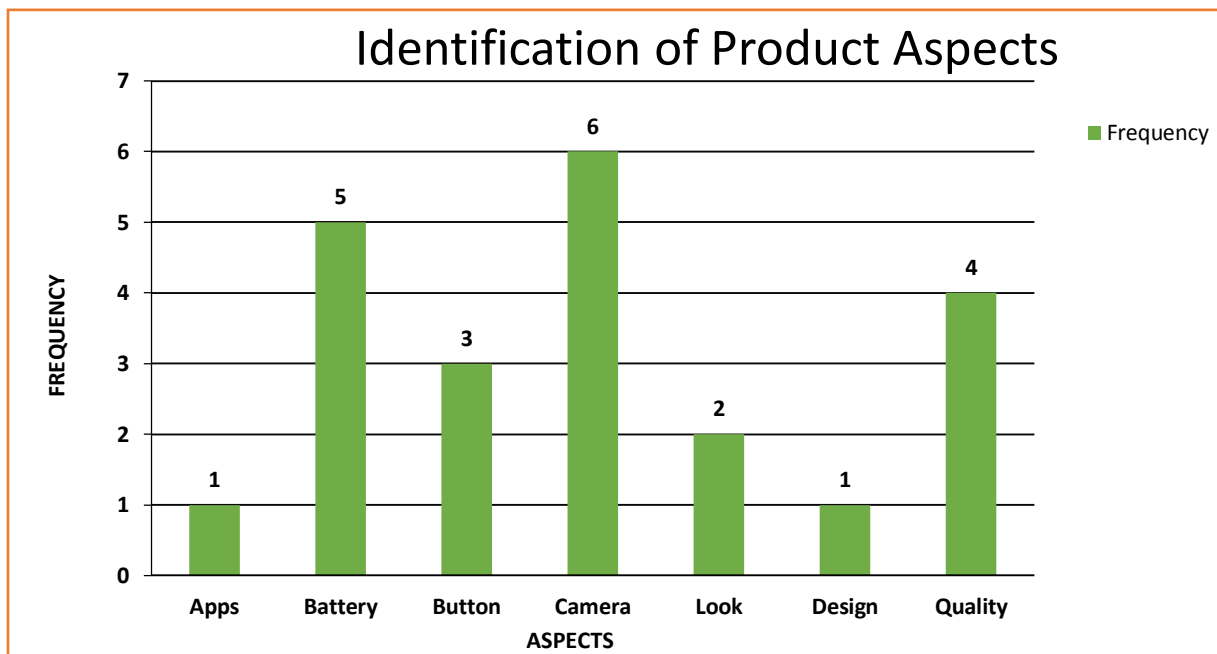
Frequency of each aspect is computed by the system and on the basis of the frequency of aspects the aspects are ranked. Frequency means how many times the particular aspect is reviewed. So the popularity of a specific aspect among the number of aspects can be determined and the graphical representation of its result is shown by the system.

Table 1 shows the aspects and frequency of product. Table 2 shows ranking of aspects based on frequency.

**Table 1: Identification of Aspects**

S.No.	Aspects	Frequency
1.	Apps	1
2.	Battery	5
3.	Button	3
4.	Camera	6
5.	Look	2
6.	Design	1
7.	Quality	4

Figure 2 shows the identification of product aspects based on Frequency. Figure 3 shows the ranked aspects of product based on the frequency.



**Figure 2 :** Identification of Aspects based on Frequency

The performance of the Product aspect ranking system are computed by the following formulas.

$$Precision = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of documents retrieved}}$$

$$Recall = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of relevant documents}}$$



$$Frequency = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

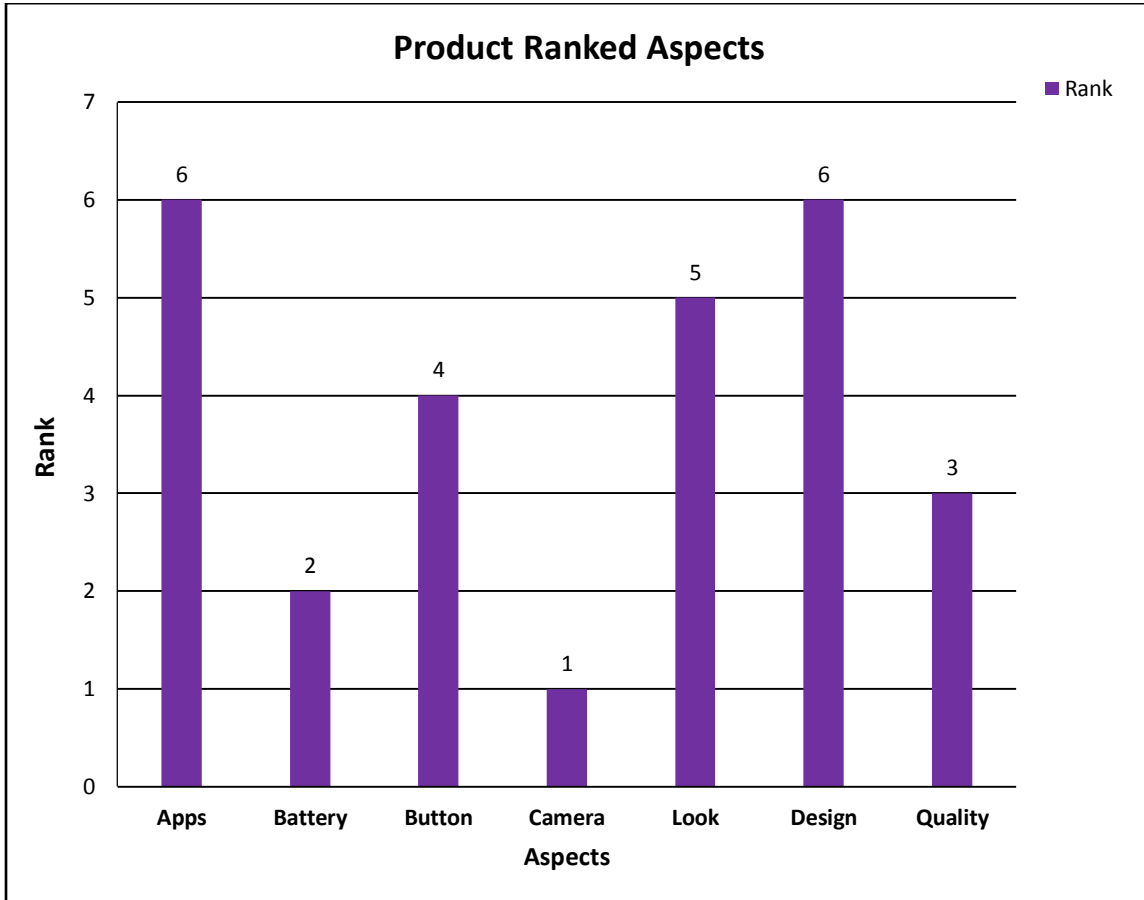


Figure 3: Ranked Aspects based on Frequency

#### IV. CONCLUSION

In this paper, the proposed framework is efficiently identify the important aspects of products from different consumer reviews. The module contains three main modules, i.e., product aspect identification, aspect sentiment classification, and aspect ranking. First, the reviews are separated as Pros and Cons reviews to improve aspect identification and sentiment classification on free-text reviews. Then established aspect ranking algorithm to realize the significance of different aspects of a product from various reviews. The algorithm at the same time discovers the aspect frequency and the influence of consumer views given to each aspect over the overall reviews. The product aspects are finally ranked according to their importance scores. Experimental results have proved the effectiveness of the proposed approaches. Significant performance improvements have been obtained with the help of product aspect ranking.

# REFERENCES

- [1] Manvee Chauhan and Divakar Yadav, “*Sentimental Analysis of Product Based Reviews Using Machine Learning Approaches*”, Journal of Network Communications and Emerging Technologies (JNCET), ISSN: 2395-5317, Volume 5, Special Issue 2, December (2015).
- [2] Feldman and Ronen, “*Techniques and applications for sentiment analysis*”, Communications of the ACM 5, Volume 56 Issue4, pp.82- 89, 2013.
- [3] Hema Krishnan, M. Sudheep Elayidom and T. Santhanakrishnan, “*Sentiment Analysis of Tweets for Inferring Popularity of Mobile Phones*”, International Journal of Computer Applications (IJCA ISSN: 0975 – 8887, Volume 157 – No 2, January 2017.
- [4] Hema Krishnan, M. Sudheep Elayidom and T. Santhanakrishnan, “*Impact and Application of Sentiment Analysis using Twitter: A Survey*”, International J. of Advanced Research in Computer and Communication Engineering, (IJARCCE), Vol. 4, Special Issue 1, 2015, pp. 18–21.
- [5] B. Ohana and B. Tierneyi, “*Sentiment classification of reviews using SentiWordNet*”, in Proc. IT and T Conf., Dublin, Ireland, 2009.
- [6] O. Etzioni et al, “*Unsupervised named-entity extraction from the web: An experimental study*”, J. Artif. Intell., vol. 165, no. 1, pp. 91- 134. Jun. 2005.
- [7] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang and Tat-Seng Chua, “*Product Aspect Ranking and Its Applications*”, IEEE transactions on knowledge and data engineering, Vol.26, No.5, May 2014.
- [8] Chetan Mate, “*Product Aspect Ranking using Sentiment Analysis: A Survey*”, International Research Journal of Engineering and Technology (IRJET), ISSN: 2395-0072, Volume: 03 Issue: 01, Jan-2015.
- [9] F. Li et al., “*Structure-aware review mining and summarization,*” in Proc. 23rd Int. Conf. COLING, Beijing, China, (2010), pp. 653- 661.
- [10] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang and Tat-Seng Chua, “*Product Aspect Ranking and Its Applications*”, IEEE2014.
- [11] Bing Liu, “*Sentiment Analysis and Opinion Mining*” pp.7-140, 2012.