# An Efficient Product Aspect Ranking and its Application: A Review

**Shahuraj Patil[1], Jyoti Raghatwan[2]**

[1]Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

[2]Professor, Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

**Abstract:** *The purpose of the work we are going to find import aspect of the product and its rank this aspect by using numerous consumer reviews. The consumer reviews contain rich and valuable knowledge about the product. And this knowledge is very useful for both consumer and firm. Consumer can make wise purchasing decision by paying more attention towards important aspect or feature. And firm will concentrate on important features or aspect while improving the quality of the aspect. So in this proposed framework, this will identify the important aspect of product from online consumer reviews. The important aspects are commented again and again in consumer review and the consumers opinions on the important aspects are greatly influence their overall opinions on the product. From the consumer reviews the important aspect are identified by using NPL tool, and will classify the sentiment on that aspect, and finally we are going apply the ranking algorithm to determine the particular product rating.*

**Keywords:** e-commerce, online shopping, Consumer review, Sentiment classification, Product aspect, ranking algorithm.

## 1. Introduction

In the recent years the peoples trend towards online shopping increases day by day. There is a rapid expansion in e-commerce. Many retail websites are available on online shopping. The famous website in online shopping Amazon.com offer 36 million product.Shoppers.com offers 5 million products and many websites are available on website such as ebay.com, flipkart, mantra,snapdeal and so on. most retail website provide platform to post reviews on millions of product and encourage people to post their opinion or sentiment on various aspect of the product. The reviews play very important role while purchasing product, because consumer can make wise purchasing decision by paying more attention towards important aspect and firm will concentrate on important aspect while improving the quality. The reviews on product it will be positive review or it will be negative review or it will be neutral. Lets see a sample review "The battery life of blackberry curve is amazing.""The picture quality of the camera is too good.""the battery life of micromax mad a94 is not good."The above reviews contain opinion about battery and camera. The first review about battery is a positive opinion and the last review about battery is a negative opinion. About the camera there is a positive opinion. There are millions of review on millions of product that are available on various website.CNet.com, Pricegrabes.com are the famous web sites that contain millions of review on various product. The consumer can post there review on these websites. And these reviews are useful for consumers who are going to buy new product. Generally product may have numbers of aspect. for example mobile has aspect like usability,design,applications,network,battery etc.for laptop, the aspect such as hard disk,RAM,Graphics card,screen,Battery etc. The product aspects are greatly influenced on product quality.
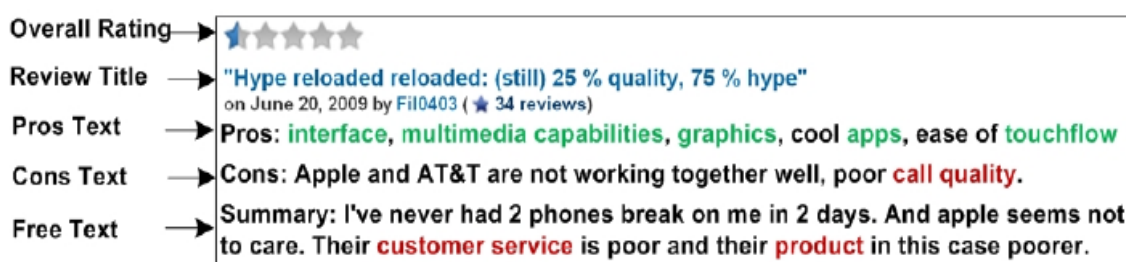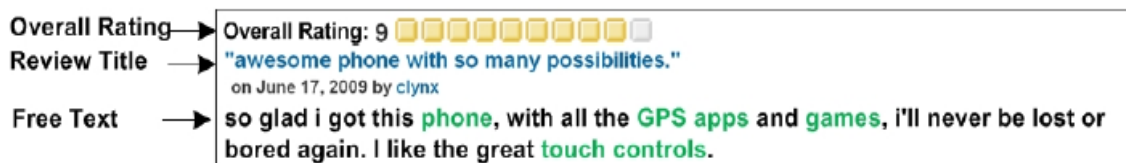


**Figure 1:** Reviews on *iPhone 3GS* product

So aspect identification is very important task while buying product. Paying more attention to the important aspect is very useful while taking decisions about product. Firms can focus on improving and enhancing the quality of product aspect and enhance the reputation of the product more effectively. It is very difficult for customer to identify the important aspect of product from various websites. And also the reviews are often disorganized it causes problem while knowledge acquisition about product. Therefore we develop an approach to automatically identify the important aspect is highly demanding.

In our proposed work, based on consumer reviews of product first we identify the important features of product. Then we classify the sentiment on that aspect. And then we develop the aspect ranking algorithm for providing the rating to particular product.

Methodology:
Following four methodologies are used in proposed work.
1) Reviews extraction and Preprocessing.
2) Aspect Identification of the product
3) Classify the positive and negative reviews of product by sentiment classifier.
4) The probabilistic ranking algorithm used for product ranking.

## 2. Literature Survey

In this section, we review existing works related to the Proposed framework, and the two evaluated real-world applications. We start with the works on the aspect identification from the consumer review. Existing methods for aspect identification based on supervised and unsupervised methods. Supervised method learns an removal model from a collection of labeled reviews. The extraction model, or called extractor, is used to spot aspects in new reviews. Most existing supervised methods are based on the sequential learning technique [1] These supervised methods require enough labeled samples for training. It is time-consuming and labor-intensive to label samples. On the other hand, unsupervised methods have emerged newly. The most notable unsupervised approach was proposed by Hu and Liu [4]. They tacit that product aspects are nouns and noun phrases. The approach first extracts the nouns and noun phrases as candidate aspects. The rate of frequencies of the noun and noun phrases are counted, and only the frequent ones are reserved as aspects. Popescu and Etzioni [12] developed the OPINE system, which extracts aspects based on the KnowItAll Web information extraction system . Sentiment analysis or Opinion mining is a type of natural language processing used for tracking the mood or polarity of public about product. Sentiment classification aims to classify the given text to one or more predefined sentiment categories. Such as Positive,Negative,Neutral.There are various classification techniques are available.Genre classification classifies text into different style such as"editorial","novel","poem"etc.They do not tell the sentiments are positive or negative[5]. The lexicon-based approaches are typically unsupervised.The lexicon based methods utilize a sentiment lexicon consist of list of

sentiment words, phrases and idoms,to determine sentiment orientation on each aspect[9].

## 3. Methodology

Following four steps are used in proposed work.

1) Reviews extraction and Preprocessing.
2) Aspect Identification of the product
3) Classify the positive and negative reviews of product by sentiment classifier.
4) The probabilistic ranking algorithm used for product ranking.

### 3.1 Reviews Extraction and Preprocessing

Before the Product Aspect Identification task there is a very important task called data preprocessing. Compared to regular text document the reviews are generally less formal and written in an ad hoc manner. If the sentiment analysis applied on raw review often achieve very poor performance in most case. Therefore the preprocessing techniques on reviews are necessary for obtaining satisfactory result on sentiment analysis. There are various data preprocessing methods are available [1].
1) Stemming: In stemming we will remove the postfix from each word such as ing,tion etc. Eg. Running will become run after stemming.
2) In tokenization we will tokenize each sentence by space. Means we will remove the spaces. Also we can remove emotion icons such as smilies. Stop word removation like a, an, the etc.
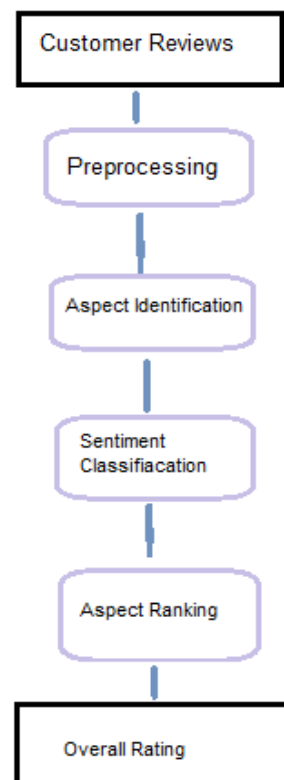


**Figure 3.1:** Architecture Diagram

### 3.2 Aspect Identification of the product

In Aspect identification we identify aspect from numerous consumer reviews. The reviews are available different forum websites. But consumer reviews are composed in different formats on various forum websites. Consumer review consists of positive and negative reviews. On some website there is a overall rating on the product, some website the reviews are in paragraph in free text.CNet.com, Viewpoints.com,Reevoo.com are the various websites for reviews and has different format. From these reviews the aspect are identified as a frequent Noun term. Previous study on aspect identification shows that aspects are usually noun or noun phrases[2].We can get accurate aspect by extracting frequent noun from the positive and negative reviews[3].Hu and Liu proposed most notable approach for aspect identification. In this approach it first identifies the noun and noun phrases in the document. The occurrence frequency of noun and noun phrase are counted then only the frequent noun terms are kept as the aspect [4].Phrase dependency parser used to extract the noun phrase by Wu et al.To filter the noise they uses language model by an intuition tjat more likely a candidate to be an aspect, and more closely it related to the consumers reviews. Language [5] was built on the product reviews. And this language model used to predict the related scores of the candidate aspect. There is a filtration of candidate with low score. Some aspect may contain synonym term such as "headphone" and "earphone."In such situation we perform synonym clustering to obtain unique aspect. These synonym terms are collected from synonym dictionary Websites.

### 3.3. Sentiment Classifier

Sentiment analysis or Opinion mining is a type of natural language processing used for tracking the mood or polarity of public about product. Sentiment classification aims to classify the given text to one or more predefined sentiment categories. Such as Positive, Negative, Neutral.There is various classification techniques are available. Genre classification classifies text into different style such as "editorial","novel","poem"etc.They do not tell the sentiments are positive or negative[5][6][7].There is another approach for detecting sentiment in text present in literature concern the use of lexical resources such as a dictionary of opinionated terms.SentiWordNet is one such resource that contain opinion information on terms extracted from WorldNet database and it is available to all for research purpose. The SentiWord is built via supervised method[8].There are two types of learning supervised learning and another is unsupervised learning. The lexicon-based approaches are typically unsupervised. The lexicon based methods utilize a sentiment lexicon consist of list of sentiment words, phrases and idoms,to determine sentiment orientation on each aspect[9].the performance of supervised learning dependent on training data. It can not perform well without sufficient data. Supervised learning method train a sentiment classifier based on training corpus. The classifier is used to predict the sentiment on each aspect. There are many learning based classification models are available. Support Vector Machine (SVM), Naive Bayes, and Maximum Entropy (ME)model these are the learning based

classification model[10].The NPL techniques are used to find out the the consumer reviews from their own languages and make it into understable format.

### 3.4 Aspect Ranking Algorithm

This proposed product aspect ranking framework, which will identify the important aspect of product from online consumer reviews. The important aspects are commented again and again in consumer review and the consumers opinions on the important aspects are greatly influence their overall opinions on the product. The overall opinion in a review is an aggregation of the opinions given to specific aspects in the review, and various aspects have different contributions in the aggregation. That is, the opinions on (un)important aspects have strong (weak) impacts on the generation of overall opinion.

## 4. Applications

Aspect ranking is useful to a wide range of real world applications. We here investigate its ability in two applications, i.e. document-level sentiment classification on review documents, and extractive review summarization. Paying more attention to the important aspect is very useful while taking decisions about product.Firms can focus on improving and enhancing the quality of product aspect and enhance the reputation of the product more effectively.

## 5. Conclusions and Future Works

In this paper,we have surveyed the reference paper releated to Aspect identification,Sentiment classification. We have planned to identify the important aspects of a product from online consumer reviews. Our supposition is that the important aspects of a product should be the aspects that are frequently commented by consumers and consumers' opinions on the important aspects greatly pressure their overall opinions on the product. Based on this assumption, we will try to develop an aspect ranking algorithm which will identify the important aspects by concurrently considering the aspect frequency and the pressure of consumers' opinions given to each aspect on their overall opinions.

## References

[1] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, Tat-Seng Chua," Product Aspect Ranking and Its Applications".IEEE2014.
[2] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing*, New York, NY, USA: Marcel Dekker, Inc., 2009.
[3] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proc. 14th Int. Conf. WWW*, Chiba, Japan, 2005, pp. 342–351.
[4] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. SIGKDD*, Seattle, WA, USA, 2004, pp. 168–177.
[5] Finn, A., Kushmerick, N., and Smyth, B. 2002. Genre Classification and Domain Transfer for Information

Filtering. In *Proc. of European Colloquium on Information Retrieval Research*, pages 353-362.

[6] Karlgren, J. and Cutting, D. 1994. Recognizing Text Genres with Simple Metrics using Discriminant Analysis. *COLING'94*.

[7] Kessler, B., Nunberg, G., and Schutze, H. 1997. Automatic Detection of Text Genre. In *Proc. of 35th ACL/8th EACL.*

[8] Esuli A, Sebastiani F. (2006). SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings from International Conference on Language Resources and Evaluation (LREC)*, Genoa, 2006.

[9] B. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," in *Proc. IT&T Conf.*, Dublin, Ireland, 2009.

[10] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. EMNLP*, Philadelphia, PA, USA, 2002, pp. 79–86.

[11] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in *Proc. ACL*, Singapore, 2009 pp. 1533–1541.

[12] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proc. 14th Int. Conf. WWW*, Chiba, Japan, 2005, pp. 342–351.

[13] A. M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Proc. HLT/EMNLP*, Vancouver, BC, Canada, 2005, pp. 339–346.

[14] Bing Liu" Sentiment Analysis and Subjectivity" *Handbook of Natural Language Processing*, Second Edition,2010

[15] Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews"by Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, Tat-Seng Chua Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 1496–1505, Portland, Oregon, June 19-24, 2011.

## Author Profile

**Prof. Jyoti Raghatwan** received the B.E. Degrees in Computer Engineering from Pune University, Pune. She is working as Assistant Professor in department of Computer Engineering, RMD Sinhgad School of Engineering Pune, India. She is having more than six year experience. Her research interest is in Information security.

**Shahuraj Patil** Reseach Scholar RMD Sinhgad of Engineering, University of Pune. He has received B.E. in Copmputer Engineering from Mumbai University, Mumbai. Currently he is pursuing M.E. in Computer Engineering from RMD Sinhgad School of Engineering, Pune, Savitribai Phule, Pune University, Pune.