

## Product Aspect Ranking Using Sentimental Analysis

*Ashwini Nair, Monika Bhajani, Nameet Nayak*

Department of Computer science,  
St. John College of Engineering and Technology,  
Palghar, India

Department of Computer science,  
St. John College of Engineering and Technology,  
Palghar, India

Department of Computer science,  
St. John College of Engineering and Technology,  
Palghar, India

**Abstract**— There are thousands of products in the market. Every product has an aspect, and some of these aspects of a product are important and some can be ignored, and these important aspect of an product plays a vital role in making decisions so as to buy the product or not, Therefore, identifying these important aspects of a product and ranking them according to there importance and significance considering customer reviews helps to take a firm and better decision. This paper gives an explanation of one of the technique that is Naive Bayes Classifier for identification of the product aspects and its classification. This paper explains how to identify an aspect and to rank them in such a way that the customers find it easy to buy the product. These aspects are identified using (a)Vital aspects generally commented by a lot of customers (b) And there views on how these aspects greatly influence the product.

**Index Terms**—Product Aspects, Aspect Ranking, Aspect Identification, Sentiment Classification, Consumer Review.

### I. INTRODUCTION

INTERNET and e-commerce is being used widely today. Many products are available online. There are many websites like flipkart.com, amazon.com etc. which permit buying products online.

These website encourage feedback of the consumer which gave rise to numerous consumer reviews being collected on day to day basis.

These reviews contain rich and valuable knowledge and have become an important resource for both consumers and firms. Consumers commonly look for quality information from online reviews before buying a product and firms can use these reviews as feedback for better product development, consumer relationship management and for the development of new marketing strategies. A product may have hundreds of aspects. Some of the product aspects are more important than the others and have strong influence on the eventual consumer's decision making as well as firm's product development strategies. Identification of important product aspects become necessary as both consumers and firms are benefited by this. Consumers can easily make purchasing decision by paying attention to the important aspects as well as firms can focus on improving the quality of these aspects and thus enhance product reputation efficiently.

A sample review "The battery life of Samsung Galaxy J7 is amazing." reveals positive opinion on the aspect "battery life" of product Samsung Galaxy J7. There are numerous sites that provides reviews on products and also allows consumer to give reviews on their desired products and those sites are cnet.com, viewpoints.com, revoo.com etc.

A product may have innumerable aspects. For instance a phone can have many aspects such as camera, RAM, storage

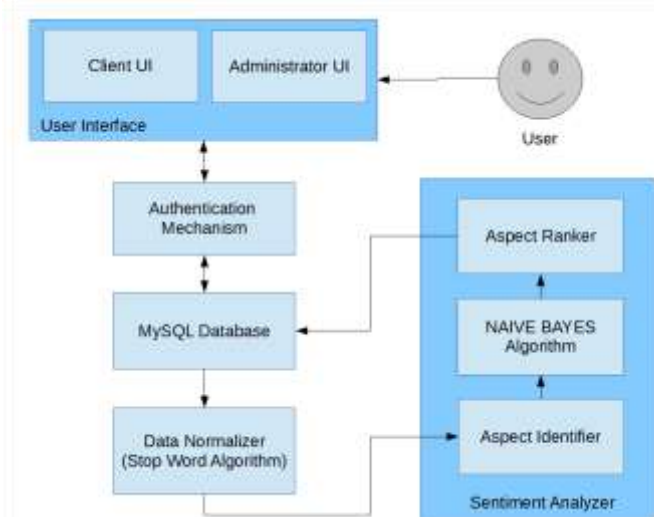
etc. A laptop can have aspects like processor, RAM, screen size etc. Consumers have their own priorities when it comes to aspects of a product. Not every consumers priority is Camera in a phone. Hence a system that suggests the important aspects of a system is required. Thus a system which ranks the aspects is what this paper deals with. In this paper, we present the details of Product Aspect Ranking framework. After aspect identification and sentiment classification, it is tantamount to calculate the Significance of the aspect, and rank them accordingly. After analyzing the aspects we take into account aspect frequency and the Influence of the customer opinions given to each aspect over their overall opinions. Customer expresses his/her opinions regarding the aspects of a product and Rank the product. We then normalized the ranking between [0,1]. The reviews from different websites might be different, for example, some might use the scale of 1 – 10, and others might use the scale of 1-5. This technique is to form a standard rating scale over the disparate websites. After the calculation from sentimental analyzer and the weights, we calculate the sum of all the aspects with their weights. In the same pattern we calculate for each aspect and rank the aspect of the product.

Considering the example of Samsung Galaxy J7 for product aspect ranking. It has various aspects such as Network, Body, Display, Platform, Memory, Camera, Sound, Communication, Battery, Browsers, Sensors, WLAN etc. Cost also turns out to be an important aspect in such kind of purchase. Hence Galaxy J7 which is available for a reasonable cost in the market has quite an average rating for cost. With an astounding 13MP back camera and 5MP front camera with flash on both sides gives the camera feature of J7 a very high rating and hence bringing it to the topmost position. In this way considering all the customer reviews and analyzing the rank using the proposed technique in the paper we get the ranked aspects and hence making it easy for the consumer to buy a product according to their aspect preference.

This approach can be used in any department irrespective of its diversities such as hotels , hawkers etc. The rest of the paper is organized as follows Section A deals with Aspect identification framework, Section B deals with Feature extraction framework , Section C deals with Sentimental Analysis Framework and Finally Section D deals with the Aspect ranking framework.

## II. IMPLEMENTATION

In this section, we present the details of the proposed Product Aspect Ranking framework. We start with an overview of its pipeline (see Fig. 2) consisting of three main components: (a) aspect identification; (b) sentiment classification on aspects; and (c) probabilistic aspect ranking. Given the consumer reviews of a product, we first identify the aspects in the reviews and then analyze consumer opinions on the aspects via a sentiment classifier. Finally, we propose a probabilistic aspect ranking algorithm to infer the importance of the aspects by simultaneously taking into account aspect frequency and the influence of consumers' opinions given to each aspect over their overall opinions.



### A. Aspect Identification

Consumer reviews are composed in different formats on various forum Websites. The Websites such as CNet.com require consumers to give an overall rating on the product, describe concise positive and negative opinions (i.e. Pros and Cons) on some product aspects, as well as write a paragraph of detailed review in free text. In summary, besides an overall rating, a consumer review consists of Pros and Cons reviews, free text review, or both. For the Pros and Cons reviews, we identify the aspects by extracting the frequent noun terms in the reviews. Previous studies have shown that aspects are usually nouns or noun phrases , and we can obtain highly accurate aspects by extracting frequent noun terms from the Pros and Cons reviews. For identifying aspects in the free text reviews, a straightforward solution is to employ an existing aspect identification approach. One of the most notable

existing approach is that proposed by Hu and Liu . It first identifies the nouns and noun phrases in the documents. The occurrence frequencies of the nouns and noun phrases are counted, and only the frequent ones are kept as aspects. Although this simple method is effective in some cases, its well-known limitation is that the identified aspects usually contain noises. Recently, Wu et al. used a phrase dependency parser to extract noun phrases, which form candidate aspects. To filter out the noises, they used a language model by an intuition that the more likely a candidate to be an aspect, the more closely it related to the reviews. The language model was built on product reviews, and used to predict the related scores of the candidate aspects. The candidates with low scores were then filtered out. However, such language model might be biased to the frequent terms in the reviews and cannot precisely sense the related scores of the aspect terms, as a result cannot filter out the noises effectively. In order to obtain more precise identification of aspects, we here propose to exploit the Pros and Cons reviews as auxiliary knowledge to assist identify aspects in the free text reviews. In particular, we first split the free text reviews into sentences, and parse each sentence using Stanford parser2. The frequent noun phrases are then extracted from the sentence parsing trees as candidate aspects. Since these candidates may contain noises, we further leverage the Pros and Cons reviews to assist identify aspects from the candidates. We collect all the frequent noun terms extracted from the Pros and Cons reviews to form a vocabulary. We then represent each aspect in the Pros and Cons reviews into a unigram feature, and utilize all the aspects to learn a one-class Support Vector Machine (SVM) classifier. The resultant classifier is in turn used to identify aspects in the candidates extracted from the free text reviews. As the identified aspects may contain some synonym terms, such as "earphone" and "headphone," we perform synonym clustering to obtain unique aspects. In particular, we collect the synonym terms of the aspects as features. The synonym terms are collected from the synonym dictionary Website.

### B. Sentimental analysis

#### i) Pre-Processing:

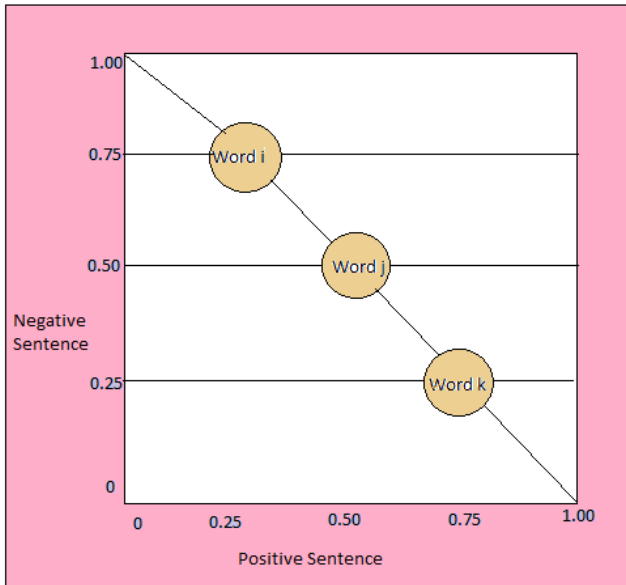
The reviews are passed through a stopwords list where the stopwords in the reviews are been removed i.e is, and, -, (, /, etc as these stopwords do not contribute to any sentiment and so there is no point in keeping them. After the stopwords are removed what is left are the feature words which are included in the feature vector and these words in the feature vector contribute to the sentiment of a review and this feature vector is passed to the classifier for training purpose.

#### ii) Classification:

Sentiments are expressed on the aspects and analyzing them is called aspect-level sentimental classification. So, there are two basic techniques for classification Supervised technique and the Unsupervised technique also known as lexicon-based approach. In the lexicon based approach list of sentiment words, phrases, idioms are used to determine the sentiment related to a particular aspect and this method is easy to implement as their performance is based on the quality of lexicon and on the other hand in supervised learning technique a sentiment classifier is trained based on the training data. The sentiment classifier that is trained is then used to predict the sentiment related to an

aspect. There are many learning-based classification models available eg: Maximum Entropy (ME), Naive bayes and Support Vector Machine(SVM).

Supervised Learning depends on the trained data and so it will give satisfactory results only if it has sufficient amount of training samples. Labelling the training data is labour-intensive and time consuming .



The reviews are categorized into Positive and Negative opinions on the aspects and it is done explicitly by Pros and Cons reviews. And these Pros and Cons reviews are the training samples for training the classifier. After training the classifier, the classifier in turn takes the free text reviews of the customers and determines whether an aspect has a Positive or a Negative reviews. In this paper we are using Naive Bayes classifier , For the Naive Bayes classifier there is a feature dictionary that has all Negative and Positive to distinguish a Positive and Negative sentiment. The feature vector is passed through this classifier and this classifier distinguishes the Positive, Negative and Neutral reviews and this is how the opinion of an aspect is determined.

### C. Aspect Ranking

In this section, we propose a way to rank the aspects of a product. We can use various methods to rank the aspect. The characteristics possessed by the important aspects are (a) Reviews about these aspects are frequently commented and (b)The opinion of the consumer that greatly influence the overall opinion. In this paper we propose simple yet efficient way to rank the aspect of a product. The importance of an aspect can be found by calculating the number of positive reviews the aspect has. Greater the positive review, higher is the rank of the aspect in that particular product. For example Samsung GALAXY J7 Frontcamera , its spell is improved by front flash and so it is considered as one the main aspect which has the highest number of positive reviews and hence Camera is ranked the topmost. This is one of the simple yet efficient method to rank the aspects of a product. But sometimes two or more aspects can have same number of positive reviews and this may result to scepticism. Hence to avoid such confusions

we use SentiWordNet scores. SentiWordNet is a dictionary available that gives positive and negative score for a word in a sentence. Using SentiWord we can calculate the total positive score of ll the aspects that we found for a particular product in the Aspect identification module. After calculating the score of each aspect we can compare the scores of all aspects and then rank accordingly. A simple algorithm for the above process is as follows.

#### ALGORITHM 1:

Input: Document with feature words from the Feature word extraction module with their specific aspects.

Output: Output file will consist of all the aspects ranked in proper order with their specific reviews.

While zero reviews left do

    Calculate the total positive and negative score using SentiWordNet.

End while

    Compare the scores calculated and rank them in order.

Apart from the above mentioned methods we also have a probabilistic aspect ranking algorithm [4]. In this method we also deal with weightage of a particular review. Weightage deals with the impact of that particular review in the overall review set. In this approach it is also considered that various aspects have different impact in the aggregation. Hence inorder to deal with these kind of aggregation we have overall rating of each review which can be calculated using Gaussian distribution. In order to deal with uncertainty in such a approach we use Multivariate Gaussian Distribution. A simplified algorithm for probabilistic aspect ranking is as shown below.

#### ALGORITHM 2:

Input: Consumer review corpus where each review is associated with an overall rating, and a vector of opinions or on specific aspects.

Output: Importance scores for all the aspects.

while not converged do

    Update weights according to Equation mentioned in [4];

    Update parameters using Maximum Likelihood for calculating further weights according to Equation mentioned in[4];

end while

    Compute aspect importance scores for all aspects

### III. CONCLUSION

In this paper, we have proposed a product aspect ranking module to identify the important aspects of products from different consumer reviews. The module contains three main sud modules, i.e., product aspect identification, aspect sentiment classification, and aspect ranking. First, we

separated the Pros and Cons reviews to improve aspect identification and sentiment classification on free-text reviews. We then developed a probabilistic aspect ranking algorithm to deduce the significance of various aspects of a product from various reviews. The algorithm at the same time explores 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. We have conducted extensive experiments to systematically evaluate the proposed framework. The experimental corpus contains 94,560 consumer reviews of 21 popular products in eight domains. This corpus is publicly available by request. Experimental results have demonstrated the effectiveness of the proposed approaches. Moreover, we applied product aspect ranking to facilitate two real-world applications, i.e., document level sentiment classification and extractive review summarization. Significant performance improvements have been obtained with the help of product aspect ranking.

#### IV. REFERENCES

- [1] B. Snyder and R. Barzilay.: Multiple Aspect Ranking using the Good Grief Algorithm. in Proc. of HLT-NAACL, pp. 300-307. New York, USA. 2007
- [2] G. Paltoglou and M. Thelwall.: A study of Information Retrieval Weighting Schemes for Sentiment Analysis. in Proc. of ACL, pp. 1386-1395. Uppsala, Sweden. 2010.
- [3] J. Yu, Z.-J.Zha, M. Wang, and T. S. Chua.: Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews. in Proc. of ACL, pp. 1496-1505, Portland, USA. 2011
- [4] A. M. Popescu and O. Etzioni.: Extracting Product Features and Opinions from Reviews. in Proc. of HLT/EMNLP, pp. 339-346, Vancouver, Canada. 2005
- [5] ComScore Reports  
[http://www.comscore.com/Press Events/Press Releases](http://www.comscore.com/Press%20Events/Press%20Releases), 2011.
- [6] Zheng-Jun Zha Member, IEEE Jianxing Yu, Meng Wang Member, IEEE and Tat-Seng Chua: Product aspect ranking using sentimental analysis, Singapore. 2013.