

# An Approach to Product Rating based on Aspect Ranking in Opinion Mining

P. Thamizharasi\* and R. Sathiyavathi

Sathyabama University, Chennai - 600119, Tamil Nadu, India; thamizh18@gmail.com, sathyakrish2723@gmail.com

## Abstract

**Objectives:** The main objective of this paper is product ranking framework that automatically generate from the product review given by own individual user's opinion by sentiment analysis. The additional work conducts product ranking that is spawned in a source of a URL in a large scale. This aspect ranking from the products is used for analyzing the feedback in online marketing websites and rating of a product. **Methods/Analysis:** Bag-of-words (BOW) is the popular way to analyze the text and perform sentimental analysis. Our proposed system focuses on unigram, bigram and neutral polarity. Our proposed system classifies the polarity of each opinion words. In this the dependent and independent word will be given as input and stored in the database. We classify the meaning of the word, in some cases these words will give different meaning or contrast polarity. So it is very important to train the polarity of the words accurately. The classifier is used to predict the sentiment on each aspect by analyzing the positive and negative words separately. Finally aspect ranking is performed to obtain the final graphical result. Most of the time, the sentiments are classified as positive or negative. There are also some neutral words which should also be categorized. So, it's important to train the classifier to predict the positive, negative and neutral words. Finally, aspect ranking of the features is performed to rate the products. **Findings:** This paper has analyzed and categorized the reviews as positive, negative and neutral. It is found that the neutral reviews which are omitted in most of the websites play an important role in deciding the rating of the product. Ten different products belonging to five different categories are considered from a popular website and the output is presented in the form of a 3D graph. **Improvement:** Numerous product reviews are available online in the internet. These reviews enable the users to choose between different products. Since, the reviews are many in number, valuable time of the consumer is wasted in going through each of the reviews and finally arriving at a conclusion. This paper aims at providing a snapshot of the overall product rating based on the key aspects which determine the product features. The ranking algorithm is used to rate the products in a scale of ten as an improvement to the star rating provided by most websites.

**Keywords:** Aspect Ranking, Opinion Mining, Product Aspect

## 1. Introduction

The recent years have seen a tremendous growth of E-commerce and people have opted to buy things online from the websites. From electronics to household articles, the websites provide a wide range of articles for the user's choice. These people depend primarily on the reviews written by users who have purchased the product previously. In order to buy the products, the aspects which define their quality and usability have to be determined. Aspects differ for each product category. For a mobile, the camera quality, screen resolution, storage space are the

important aspects. For a household article like a mixer, the warranty, speed and price are important aspects, for clothes and other utilities, the quality of the fabric and the comfort plays an important role. These aspects have to be identified and extracted from the reviews to determine the exact opinion.

The reviews may be positive, negative or neutral. A positive review about a Dell laptop may be like 'This laptop is good and the speed is excellent'. A negative review may be like 'The laptop is not sturdy at all'. A neutral review may be like 'The speed of the laptop is somewhat good'. Here in order to extract the aspects the unigrams

\*Author for correspondence

such as 'speed', 'sturdy' are extracted for positive and negative reviews. For the neutral reviews, bigram which is 'somewhat good' has to be extracted. The reviews are parsed using a parser and then classified using a classifier.

In this modern era, the emergence of real-time review were rapidly increases the growth of internet. This will be the prime to a special text mining for governing the prejudiced approach (i.e. sentiment). Owing to the recent admiration of internet, individuals have been able to provide countless information to the public easily. The sentences include opinion or sentiments which lead to the sentimental analysis concept. The sentimental analysis (opinion mining) mentions the use of text analysis and computational etymology to classify the mining of biased information. The basic duty of sentiment analysis is categorizing the polarity of an assumed text at the document, sentence or aspect near. Bag-of-words (BOW) is a typical model used for representing text review by vector independent words. However BOW is suitable for text review it is not effective when dealing with sentiment classification, because it dislocates the word order and made some disruptions in the syntactic structure and also abandon some semantic information. Due to scarcities in BOW there is some inadequate in improving the word accuracy, therefore there arises polarity shift problem.

The polarity shift is a phenomenon in which the sentiment text gets reversed (i.e. Positive to Negative and vice versa). This leads to polarity shift elimination-negations. Sentiment classification is a distinctive task of text classification which leads to permitting the sentimental polarities of opinions, example comfortable, uncomfortable, positive to negative and vice versa. A document is modeled as a Bag of Word through machine learning approach, i.e. a set of gratified words without any word order or syntactic relation information. Example: 'The sofa is not comfortable', the word 'comfortable' is positive, but the polarity of the whole sentence is reversed because of the negation word not. This is what sometimes the machine learning approaches get fail under certain circumstances. In general, there are two steps to include negation statistics into a system negation detection and negation classification. First for negation detection some negations prompt words 'no', 'not', 'never' are generally applied to diagnose negation phrases or sentences. And second, the negation classification is the one way to ingress negation information directly to the reverse polarity of the words which contain negation trigger words.

In<sup>1</sup> examines combination of various semantic models like heuristic and machine learned to improvise the sub sentential sentimental analysis.

In<sup>2</sup> identified a model for sub sentential sentimental analysis which predicts polarity in dependency graphs. Thus the polarity shift challenges were found in corpus and lexicon methodologies.

In<sup>3</sup> proposed a new method by enhancing a word "NOT" to the words. So the scope of negation will be changed.

In<sup>4</sup> proposed more linguistic features or lexical resources to improvise the sentimental classification accuracy.

In<sup>5</sup> examined a new machine learning method to extract polarity shifters for word and sentence wise sentiment classification.

In<sup>7</sup> uses the HMM model and conditional random field in supervised learning technique to learn the extractor. These are the techniques which collect the reviews to train a model for extraction. This model is termed as an 'Extractor' and is used for aspect identification.

In<sup>8</sup> use the unsupervised method and the demerit is that it may contain noise. In<sup>9</sup> uses a phrase dependence parsing. It takes a sentence as input and segments it into phrases. It gives importance to phrases and not on single word inside phrase. In this method, the aspects are nouns or noun phrases. The frequency of the occurrence of the noun and noun phrases is calculated and considered.

In<sup>8</sup> use this technique for sentiment classification. They use synonym/antonym relations defined in WordNet to relate the word set to obtain a sentiment lexicon. Three approaches are there to collect the list of opinion words. They are manual, dictionary and corpus based approaches.

Manual approach is not alone and is combined with the other two techniques.

Dictionary based approach uses a small set of words collected manually and these synonyms and antonyms of these words are added to this set by searching the words in the well-known Word Net. The main strategy for dictionary based approach is present in<sup>10,11</sup>.

## 2. Methods

For sentiment classification our paper uses the method called Extractive review summarization. In this an abstractive summary and extractive summary is obtained. An abstractive summary helps in identifying the major

topics in the review document and then expressing the topics in a clear manner. An extractive summary consists of identifying the important sentence and the paragraphs from the online review.

Figure 1 shows how the documents are passed through a polarity detector and finally to a classifier, wherein the documents are classified as positive, negative and neutral.

The impulse of our approach is to improve the performance of sentiment classification by polarity shifting between sentences. For interpretation, the training data used for polarity detection and polarity classification are referred to as the polarity detection training data and the polarity classification training data. Figure 1.1 shows that the documents are divided into document level classification and extractive review summarization. This incorporates into a sentiment shifting information as positive, negative and neutral. The above framework is a general one, that is, different polarity shifting takes place in accordance with polarity words or phrases.

### 3. Proposed System

We summarize our project into six steps and then review the techniques' involved.

There are six steps in our proposed work:

- Consumer reviews are collected for aspect ranking and product rating.

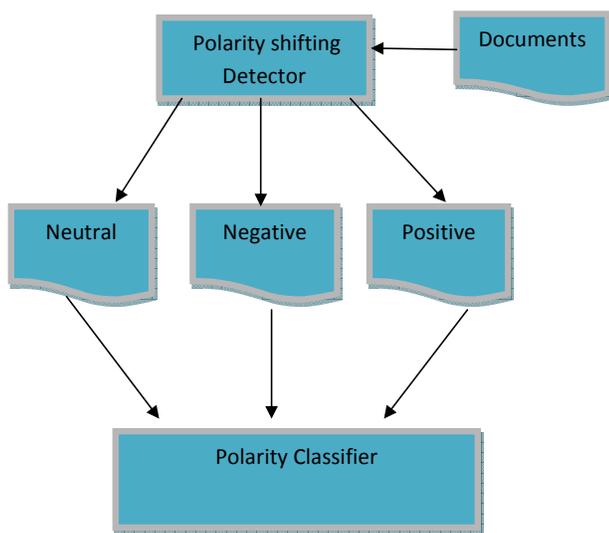


Figure 1. Sentiment classification using document.

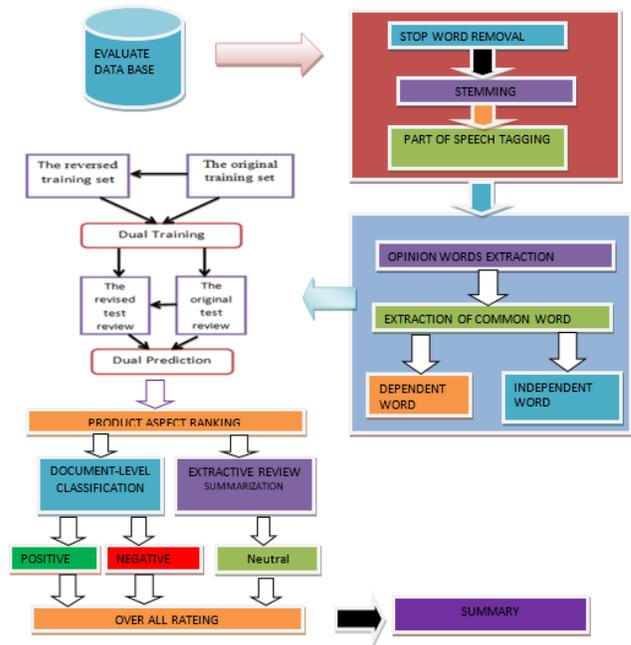


Figure 1.1. Architecture diagram.

- These reviews are preprocessed and the important opinion words are identified.
- The aspects present in the opinion words are identified.
- The opinion words are classified based on polarity as positive, negative and neutral.
- Aspects are ranked according to their importance.
- The product is rated in an overall manner.

#### 3.1 Pre-Processing of Review

The user reviews are crawled from the users by adding product review and product domain as input. This input is stored as a text format in the database. From the database the review is given as input in stop word removal for removing the stop word from user given review. In stop word we perform some following functions they are, the words which are like is, was, this, that, at, it must be removed for easy analysis and meaningless words are removed in the review. The gradient of stop word methodical is based on alphabetical order so it is considered to be a single array for quick accessing. Finally the given user input will undergo stop word process and it is removed. Next step in the pre-processing of review is Stemming.

Stemming algorithm deals with the removal of prefixes and suffixes for the given review, in which variant sentences are reduced to form a common sentence (e.g.

Connections, Connectives, Connected as Connect). In grammatical stemming algorithm performs many mistakes, but in concept we need the exact word for exact review ranking.

The last step in the pre-processing of review is the parts of speech tagging. In POS tagging each and every user reviews are taken as opinion word and formed into noun, verb, adjective and adverb. Tagging is done so that we can easily categorize the features of the given product.

### 3.2 Opinion Review Word Mining

This module classifies the word as dependent and independent. From the user's opinion word the tagged sentence will form the unigram and bigram. Unigram is a single word and bigram is a combination of unigram. In our project we consider only the adjectives since it express the attitude and feeling of the opinion holder.

In the formation each word are described as unigram and if it combines into two words it will be as bigram. For example, 'The camera gives good quality' the word "good" will give a positive polarity but when it is combined with some other words like "not good" will give a negative polarity and combine with words like "somewhat good" gives a neutral polarity. So it is necessary to form a unigram and bigram.

### 3.3 Opinion Word Polarity Classification

Semantic orientation of a word represents whether the word is positive or negative evaluation. This semantic orientation of a word also mentioned as the polarity of a word and also it has some applications in the area of opinion mining, answering and filtering offensive messages. This opinion mining and sentiment analysis of a product review or any biased declaration in a part has received a substantial attention.

So this module will classify the polarity of each opinion word. In this the dependent and independent word will be given as input and stored in database. Based on the database we classify the meaning of the word, in some cases these words will give different meaning or contrast polarity. So it is very important to train a classifier for classifying the polarity of the word accurately.

### 3.4 Clustering of Polarity in Opinion Word

The classifier is used predict the sentiment on each aspect by positive and negative separately. It will group the data based on K-nearest neighbors, he we insert edges between

a node and its K-nearest neighbors. Each node connects at least K nodes. After clustering we generate a summary for easy and understandable representation of reviews.

### 3.5 Aspect ranking

Finally, we put forward a probabilistic aspect ranking algorithm to conclude the prominence of the aspects by simultaneously taking into account aspect frequency and the influence of consumers' opinions to each given aspect and find the overall opinion.

## 4. Experimental Results

Figure 2 depicts the home page of our work where the existing users and administrators can login. New users can also register here.

Figure 3 depicts the login page for the existing users. The users have to login with their user name, password and a secret key which is sent to their mail id. This secret key is sent to their mail id on registration.

Figure 4 shows the new user registration page. Here a new user registers with his details like first name, last name, email id etc.

Figure 5 and Figure 6 shows how the user can add the reviews for each product. In Figure 5 the user can add the product domain and the products manually. Figure 6 shows how the user can add the reviews for the particular product.

Figure 7 shows the amazon.com page being downloaded into our work. The reviews of the particular product chosen by the user can be viewed by the user.

To view the aspect ranking of the product the user can select the particular product and product domain from the dropdown list. This is shown in Figure 8. The overall rating of the product is presented in the form of a 3D graph in Figure 9.



Figure 2. Home page.



Figure 3. Login page for existing user.

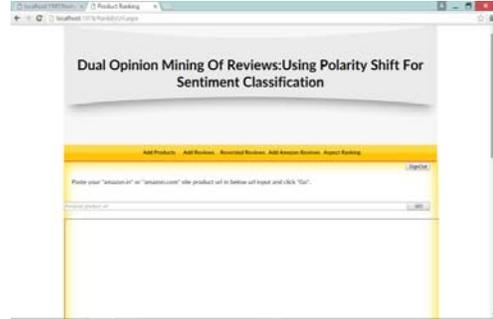


Figure 7. Reviews of different products from Amazon.com.



Figure 4. New user registration.



Figure 8. Aspect ranking of reviews.

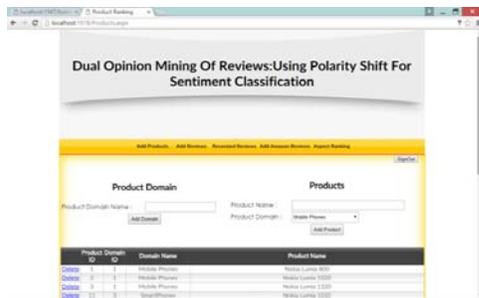


Figure 5. New products and product domain addition by user.

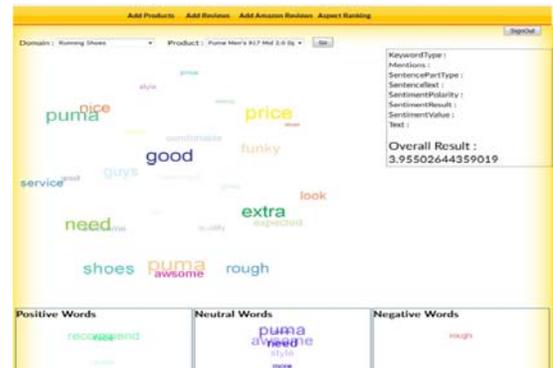


Figure 9. Final output screen.



Figure 6. User reviews of the product based on domain

## 5. Conclusion

The basic objective of this work is to analyze the positive, negative and neural words from real time dataset. Thus our experimental result is obtained by analyzing Amazon product review pages. Thus finally positive, negative, neutral and sentiment value is been calculate. This paper has proposed to rate a product based on the online customer

reviews. The framework consists of collection of reviews, Identification of the important aspects, Classification of aspects as positive, negative and neutral and finally ranking of aspects based on the ranking algorithm. This ranking helps in rating the product. We assume that the aspects are those which are frequently commented by the user and based on this assumption the overall product rating is done. This rating of products enables the user to identify the best product available online.

## 6. References

1. Choi Y, Cardie C. Learning with compositional semantics as structural inference for subsentential sentiment analysis. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP); 2008. p. 793–801.
2. Nakagawa T, Inui K, Kurohashi S. Dependency tree-based sentiment classification using CRFs with hidden variables. Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL); 2010. p. 786–94.
3. Das S, Chen M. Yahoo! for Amazon: Extracting market sentiment from stock message boards. Proceedings of the Asia Pacific Finance Association Annual Conference; 2001.
4. Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP); 2002. p. 79–86.
5. Ikeda D, Takamura H, Ratnoff L, Okumura M. Learning to shift the polarity of words for sentiment classification. Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP); 2008. p. 18.
6. Gowri S. Mala GSA. Efficacious IR system for investigation in digital textual data. Indian Journal of Science and Technology. 2015 Jun; 8(12):140–5. ISSN: 0974-6846.
7. Wong TL, Lam W. Hot item, mining and summarization from multiple auction websites. Proc 5th IEEE ICDM; Washington DC, USA. 2005. p. 797–800.
8. Hu M, Liu B. Mining and summarizing customer reviews. Proc SIGKDD; Seattle, WA, USA. 2004. p. 168–77.
9. Wu Y, Zhang Q, Huang X, Wu L. Phrase dependency parsing for opinion mining. Proc ACL Singapore. 2009; 3(3):1533–41.
10. Hu M, Liu B. Mining and summarizing customer reviews. Proceeding of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04); 2004. p. 168–77.
11. Kim S, Hovy E. Determining the sentiment of opinions. Proceedings of International Conference on Computational Linguistics (COLING '04); USA. 2004.