



## ONLINE PRODUCT RANKING BASED ON REVIEWS

Gnana Kirubahanar V.<sup>1</sup>, Elumalai P. S. G.<sup>1</sup> and Yovan Felix A.<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Sathyabama University, Chennai, India

<sup>2</sup>School of Computing, Sathyabama University, Chennai, India

E-Mail: [kribbs0420@gmail.com](mailto:kribbs0420@gmail.com)

### **ABSTRACT**

Online shopping is the trend nowadays; there has been an increasing demand for easy analysis of the many consumer reviews posted in the internet. The more number of reviews there are about a product, the more difficult it becomes for consumers to navigate through the reviews to arrive at an informed decision about the product. We propose a product feature ranking framework, which automatically identifies the important aspects of products from online consumer reviews, aiming at improving the usability of the various reviews. The important product aspects are the specifications that are frequently commented on by the masses. The more frequent an aspect is mentioned, the more important it is to increase the overall opinion of the product by document level classification and extractive review summarization.

**Keywords:** segregating, stemming, product aspects, features purchase portal.

### **1. INTRODUCTION**

Invention is a method that is used to analyze reviews in order to automatically identify, aspects or features of a reviewed product and the most important aspects or features from among those identified that influence the overall product perception. Basic invention mechanism consists of three steps: Sentiment classification, Aspect ranking algorithm and Aspect identification. Underlying mechanism has been applied to sentiment classification of document level and extractive review summarization. Innovation could be utilized by companies distinguishing the intrigued open view of their products, in the matter of statistical surveying and estimate, and by items amassing audits. A product may have distinguished and unique features or specifications that varies with other products. A phone will be having hundreds of specification such as display, camera, usb, wi-fi, applications and so on. Each consumer has their own desire for specifications that may vary with others. The current trend of online shopping is reviewing the products. Most of the online shopping portal encourages rating the products with star ranking. The consumers are able to comment as well as rank the product as they desire. Based on these comments and rating the consumers are able to have a clear idea of what the product is and how the product's usage and whether it is worth spending money on the product. The main problem arise when a product gets a thousands of reviews and the consumer fails to recognize the details of the product and its specification.

### **2. RELATED WORKS**

To understand the concept of product ranking aspect in online shopping several literatures have been studied and out of which some of the concepts that are related to the product are explained.

With reference to the paper published by the authors Gediminas Adomavicius and Alexander Tuzhilin [10], they give a survey on how a product can be recommended to others based on its usability, content and advance recommendation process. They have also discussed about extending the recommendation system for

more detailed understanding for the users, providing more flexible and less protruding type of recommendation for the benefit of the users.

Jianming He and Wesley W. Chu [9], in their social network based recommendation system has discussed about recommendation of a product through online or social network that would benefit the online consumers. They discussed about how data can be extracted from social media by semantic filtering that analyze the data and is given to a set of people, by which they could give their opinions or reviews and share it with each other to get an collaborative idea, so each user can be benefited by these recommendation for the product they choose with the help of others opinion on the selected product. They have also given a detailed information on how opinions can be made use to others.

Yuanbin Wu, Qi Zhang, Xuanjing Huang, Lide Wu [13], in their phrase dependency parsing for opinion mining, they presented a detailed information about how mining of opinions from numerous number of reviews in an online shopping commerce. The authors have introduced a phrase dependency parsing which helps in extracting the specification of the product and users opinion and provides relation between them.

L. Gong, J. Zeng, and S. Zhang [4], Text stream clustering algorithm based on adaptive feature selection, in which the authors have discussed about text stream analysis which gives the highlight about a particular topic in a survey. Clustering plays a vital role in analyzing text stream. The normal long duration clustering method gives disappointing results. The authors have introduced a new text stream cluster algorithm to focus on the particular information that needs to be extracted. This information provides us a better understanding for extracting all the reviews from online portal.

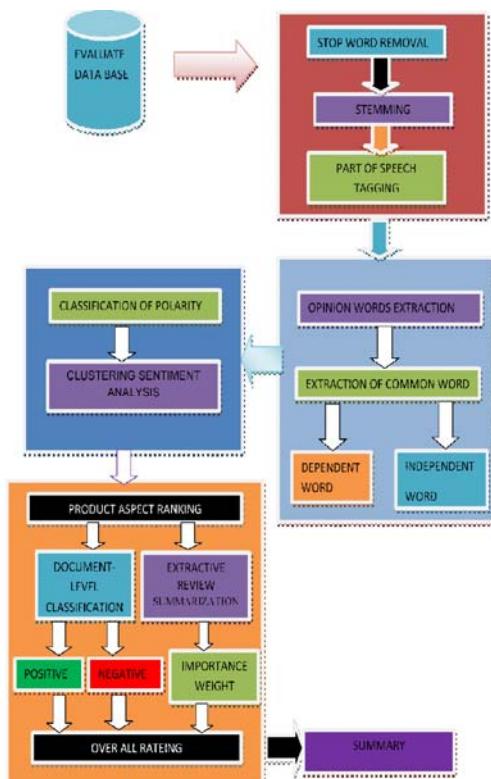
Rutuja Tikait, Ranjana Badre, Mayura Kinikar [12], Product Aspect Ranking Technique, in this survey the authors give a detailed description of how products are reviewed based on the important aspects that benefit both the firm and the consumers. They provide a description of identifying a product and its classification. The important



aspects of a product might differ among the consumers. Based on the pros and cons of the important aspect might affect the overall review on the product despite the good value of other aspects. So product aspect ranking is introduced with the help of sentiment classification.

### 3. PROPOSED WORK

The current trend of online or e-commerce shopping mainly depends on the reviews and ratings given by the users who have already bought the product. Every consumer has their own taste of specification, feature, usability of a product. For some users, they consider only a certain feature of the product. Based on this criterion of the particular feature the users might give negative reviews about the product. This will affect the overall specification of the product. There are some users who like the model of a phone but not its performance but gives a rating for the model [11]. This will be a disadvantage for the new consumers that checks on the rating and comes to a conclusion that the product is worth buying without knowing the full features of the product. To minimize all the confusions in buying a product based on the lot of comments given by the users we introduce a ranking system which helps the consumers to save time and get a precise ranking scale from 1 to 5 based on the extraction of reviews and processing it by sentiment classification where each and every word from the numerous number of review is analyzed and given positive and negative values to each words and then the overall ranking is given to the consumer.



**Figure-1.** System architecture.

### 4. SYSTEM ARCHITECTURE

To acquire the ranking of a product, first the reviews and specification is extracted using stream clustering method which is based on “Tweet Stream Clustering Algorithm”. The extracted reviews are then stored in a database where it is evaluated for the processing of reviews by stop word, stemming and parts of speech tagging. Later the common words are extricated and recognized into area dependent and autonomous words which help in appointing the positive and negative qualities to the words. The Support Vector Machine Training Algorithm is utilized for gathering the positive and negative extremity independently [13]. The Product Aspect Ranking Framework comprises of archive level order and extractive survey rundown. In archive level classification, the parts of the item are isolated as positive and negative in view of the extremity. The extractive review synopsis gives the significance and weight estimation of the perspective specified in the review. Based on the segregation and the weight estimation, the overall raking of an aspect is provided. Figure-1 displays the outline of all the process from extraction of reviews from the database till the overall ranking of the product.

### 5. METHODOLOGY

For implementing this ranking product the entire system is divided into five phase, each phase functions on its own criteria. When a phase is being processed the rest of the phases will be waiting until their functions are called. It functions in a step by step manner.

#### A. Pre-processing of review

The initial phase starts by copying the url of the product page from the website to the product ranking webpage. The whole product its name, reviews and specifications are readily updated in the database as text file. In this phase the unwanted words such as “is, a, the, that...” are removed by stop word removal. Stemming algorithm is used to convert the various forms of a word into normal form i.e. the word ‘connecting’ is changed to its normal form as ‘connect’. After stemming, part of speech tagging (POS tag) is used to assign part of speech such as noun adjective to the words in the reviews. Figure-2 and Figure-3 shows the extraction of reviews and processing words.

#### B. Opinion word feature mining

After the extraction of common words, they are classified into dependent and independent words (i.e) unigram and bigram. The unigrams are single words which are mostly adjective where adjectives express feelings and



Figure-2. Extraction of review.

attitude in the words of consumer's review. The bigrams are dependent words where "good" is a unigram and gives positive polarity but when combined with "not" it will give "not good" which gives negative polarity.

| Keywords   |              |          |                  |             |                   |                |                   |
|------------|--------------|----------|------------------|-------------|-------------------|----------------|-------------------|
| Text       | SentenceText | Mentions | SentencePartType | KeywordType | SentimentResult   | SentimentValue | SentimentPolarity |
| repair     | 3            | keyword  | noun             | neutral     | 0.630689010183783 | -              | -                 |
| warranty   | 3            | keyword  | noun             | neutral     | 0.630689010183783 | -              | -                 |
| slot       | 3            | keyword  | noun             | neutral     | 0.93095025706275  | -              | -                 |
| operating  | 3            | keyword  | noun             | neutral     | 0.93275375233526  | -              | -                 |
| months     | 3            | keyword  | noun             | neutral     | 0.97996426621193  | -              | -                 |
| brand      | 3            | keyword  | noun             | neutral     | 0.980664827454275 | -              | -                 |
| resolution | 3            | keyword  | noun             | neutral     | 0.983467043625158 | -              | -                 |
| price      | 3            | keyword  | noun             | neutral     | 0.98371477624277  | -              | -                 |
| system     | 3            | keyword  | noun             | neutral     | 0.991297478491273 | -              | -                 |
| anything   | 3            | keyword  | noun             | neutral     | 0.996912024108087 | -              | -                 |
| screen     | 3            | keyword  | noun             | neutral     | 0.99826721784173  | -              | -                 |
| phone      | 4            | keyword  | noun             | negative    | 0.92715808536262  | -              | -                 |
| refused    | 1            | keyword  | adjective        | positive    | 0.192177908184795 | -              | -                 |
| Faulty     | 1            | keyword  | adjective        | negative    | 0.220742955119402 | -              | -                 |
| damage     | 1            | keyword  | adjective        | negative    | 0.364701672575452 | -              | -                 |
| full       | 5            | keyword  | adjective        | positive    | 0.474612943847289 | -              | -                 |
| compared   | 3            | keyword  | adjective        | neutral     | 0.989776261128437 | -              | -                 |
| more       | 3            | keyword  | adjective        | neutral     | 0.992399092903131 | -              | -                 |
| other      | 3            | keyword  | adjective        | neutral     | 0.997995323741236 | -              | -                 |

Figure-3. Stop word, stemming, part of speech tagging

The classification of words goes through domain dependent words such as blurry, compact, sharp. This word requires other words to accompany it to give a meaning. Whereas the domain independent words such as nice, good, beauty do not need other words to give a meaning to it.



Figure-4. Overall ranking.

#### C. Clustering of opinion word extremity

In this progression, we amass the information based on the positive, neutral and negative extremity. The gathering is done in light of the k-closest neighbors, where we embed edges between a hub and its neighbors. In the wake of bunching, a component based synopsis and chart is created which gives a simple and reasonable comprehension of the audits.



Figure-5. Segregation of polarity.

#### D. Polarity segregation and ranking

Once the data has been grouped based on the polarity, we segregate the words and display them as positive, neutral and negative aspects of the product. The polarity against each feature is pre-defined by the admin i.e., good as positive polarity, battery under neutral, worse under negative polarity. The user can view the specification of the product in a segregated manner and the rating is calculated for each feature based on the number of times a feature is mentioned in the survey of the



product. Figure-4 and Figure-5 displays the final ranking of a product and weightage to every single word.

## 6. CONCLUSIONS

In this era, where shopping is done at the click of a mouse, this paper work offers a fast and efficient way of forming a decision by going through all the reviews for a product. Reviews offered by buyers are reliable as they tend to be genuine and give a hands-on experience feedback. Skimming through all the reviews of a given product can be exhausting and time consuming at times, this paper addresses in addition to making sure all the features of a product addressed in a review is considered in making the decision about a product.

## REFERENCES

- [1] C. C. Aggarwal, J. Han, J. Wang and P. S. Yu. 2003. A framework for clustering evolving data streams. In: Proc. 29<sup>th</sup> Int. Conf. Very Large Data Bases. p. 8192.
- [2] T. Zhang, R. Ramakrishnan and M. Livny. 1996. BIRCH: An efficient data clustering method for very large databases. In: Proc. ACM SIGMOD Int. Conf. Manage. Data. pp. 103-114.
- [3] P. S. Bradley, U. M. Fayyad and C. Reina. 1998. Scaling clustering algorithms to large databases. In: Proc. Knowl. Discovery Data Mining. p. 915.
- [4] L. Gong, J. Zeng, and S. Zhang. 2011. Text stream clustering algorithm based on adaptive feature selection, Expert Syst. Appl. 38(3): 13931399.
- [5] Q. He, K. Chang, E.-P. Lim and J. Zhang. 2007. Bursty feature representation for clustering text streams. In: Proc. SIAM Int. Conf. Data Mining. pp. 491-496.
- [6] J. Zhang, Z. Ghahramani, and Y. Yang. 2004. A probabilistic model for online document clustering with application to novelty detection. In: Proc. Adv. Neural Inf. Process. Syst. pp. 16171624.
- [7] S. Zhong. 2005. Efficient streaming text clustering. Neural Netw. 18(5/6): 790798.
- [8] C. C. Aggarwal and P. S. Yu. 2010. On clustering massive text and categorical data streams, Knowl. Inf. Syst. 24(2): 171196.
- [9] Jianming He and Wesley W. Chu. 2010. A Social Network-Based Recommender System (SNRS). Volume 12 of the series Annals of Information Systems. pp. 47-74.
- [10] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE transactions on knowledge and data engineering. 17(6).
- [11] Albert Mayan, J. Sharmila Latha T, Kislay Sinha. 2015. Security Analysis of Three Factor Authentication Schemes for Banking. ARPN Journal of Engineering and Applied Sciences. 10(8): 3504-3509, ISSN: 1819-6608.
- [12] Yuanbin Wu, Qi Zhang, Xuanjing Huang, Lide Wu. 2009. Phrase Dependency Parsing for Opinion Mining, Yuanbin Fudan University School of Computer Science. Vol. 3.
- [13] Yovan Felix A., Sasipraba, T. 2016. Incident mapping and EAS using decision support system. ARPN Journal of Engineering and Applied Sciences.
- [14] Rutuja Tikait, Ranjana Badre, Mayura Kinikar. 2014. Product Aspect Ranking Techniques: A Survey. 2(11).