Survey of Product Reviews using Sentiment Analysis

R. Jenitha Sri* and P. Ajitha

Sathyabama University, Chennai-600119, Tamil Nadu, India;
ravikumar.jenithasri7@gmail.com, hannahgracelyne@gmail.com

Abstract

Background/Objective: Online shopping encompasses large variety of products and reviews which gives rich and valuable source of information for both enterprise and user. The review available on the internet is often disorganized that makes the user a difficulty in navigating the information and gaining knowledge. The aim of this study is to predict the users' opinion of a product based on their online reviews. **Methods:** We propose a technique called semantic orientation, which automatically finds the frequently used terms for an aspect of a product from online customer reviews and other important context considered here is a dynamic dataset. Firstly the product aspect is identified, and then sentiment classification is done. Additionally, other techniques like stop words removal, context based mining and stemming is employed. **Findings:** It provides an efficient way of predicting the user's opinion and thereby suggesting them. **Applications/Improvements:** The proposed system has been tested in various products and is analysed. This system can be implemented in real world application like predicting opinions of the user. Hence improving the user's accessibility.

Keywords: Aspect Ranking, Context Mining, Product Aspect, Sentiment Classification, Stemming, Stop Words

1. Introduction

The user-generated content on the web in the form of reviews, blogs, social networks, tweets etc. for various products that are purchased is of great increase. Now a day's people prefer purchasing online. In order to enhance customer shopping experience, it has become a common practice for online merchants to enable their customers to write reviews on products that they have purchased. This helps the customer to know more about the product that they are going to buy. This information is not only useful to the customers, but also for the institutions and companies, providing them with ways to research their consumers, manages their reputations and identifies new opportunities. In the past few years, many researchers studied the problem, which is called opinion mining or sentiment analysis. In order to overcome these problems, some of the main tasks carried out are: to identify the product feature that has been commented and to classify the reviews based on positive and negative. This is quite challenging. Reviews in thousands or hundreds will take more time and critical to study. So each opinion should be taken as an input and identify the opinion in form of POS

(part-of-speech). All those reviews are present in the form of sentence and each is stored in the dataset. These are also used by companies to know what people think about their product. This study helps to improve the drawbacks in their upcoming products. This enables the companies to track the product details like feedback. Amazon is one of the leading online shops which mainly depend on the customers who purchase online. So they have to concentrate more on reviews that are mentioned online by the customer because online review is a place where people may express their opinion freely. People who prefer online shopping will cite the reviews and based upon the reviews commented they go for shopping. In order to classify the reviews based on sentiment, lot of mining algorithms are studied, examined and implemented for opinion mining. In this paper various mining algorithms are studied, discussed to implement the best way to predict the opinion of the product.

2. Related Work

Amazon represents the persistent labelling concept and it identifies the process information and the name of that

product process. To detect the best features of a product, the languages of each pattern are identified by using the model known as a supervised pattern mining method. The relation between opinions and product features helps to improve the product review rating. Also, it improves the strength of user opinions, which means that feedback on the product should be in text format. In multi-document summarization, proposed two new approaches to summarize the word classification. One is sentence extraction and the second approach is summarizer to produce summaries[1].

In lexicon based approach, the major task is to classify the product opinion based on positive and negative. This work is said to be sentiment orientation. Here come the problems like confection (i.e.) multiple ways to arrive at final solution and another drawback is to deal with the context based words[2].

The term document frequency is proposed by document classification to calculate the frequency of word that is occurring in the document[3]. The other researches have been carried out for the sentiment classification. In Sentiment classification, classifying terms from a document into its grammatical rules, or parts of speech have also been explored. In part of speech information is used as part of a feature set for performing sentiment classification on a data set of reviews[4].

The user opinions divided by opinion mining also named as sentiment analysis. Measuring the opinion related text and hidden that text is done by using the probability distribution estimation technique. To detect the best features of a product, the languages of each pattern are identified by using the model known as a supervised pattern mining method. But it is not necessary to identify the opinion orientations[5].

Another research has been carried out to understand the reviews that are not mentioned by the keywords. Thus the sentiment needs more understanding which can obtain by labelling the sentence and apply machine learning technique to extract the result[7, 8].

The opining mining task can be solved Opine, a review mining system which is used to find the semantic orientations, this helps to solve the problem like product identification, identify the opinion, and rank the opinion[9].

Document-Word Co-Regularization for Semi-supervised Sentiment Analysis proposes two approaches: first approach is to incorporate sentiment laden terms and the second approach is to adapt new domain[10].

Some other related works proposed few different techniques like domain identification, features of the domain and sentiment analysis. It also deals with the neutral comments that are given. These opinion words can be gathered from opinion lexicon dictionary[11]. Some survey also states that retrieval of comments from online includes various knowledge based techniques like crawling the web, indexing, ranking the information and knowledge representation of a model[12].

3. Proposed Work

online user gives a lot of comments which plays a major role in sentiment classification. In order to overcome the previously discussed drawback and to identify the various comments mentioned, the sentiment orientation algorithm is used in the proposed work. Sentiment orientation algorithm includes two major approaches.

3.1 Corpus based Approach

It is a linguistic based approach. This approach mainly identifies the emotional similarity of the word. This may also call as 3A perspective.

- Annotation: - this may include parsing, POS (parts of speech)
- Abstraction: - mapping of terms
- Analysis: - statistical and generalizing from the dataset.

A corpus driven method is used to find the emotional similarity of the word. This emotional similarity includes words like happy, sad and also include the frequency of the occurrence in the sentence.

3.2 Dictionary based Approach

This approach includes resources like wordnet to automatically identify emotion related words for classification. Word net is a lexical based approach where nouns, verbs, adjective, adverb is grouped into a set of cognitive synonymy (sysnet). Meaningful related words can be navigated with the browser. This helps to identify the large set of text from the comments that are retrieved.

This approach gives lots of advantages, like ease of use, take less time to analysis large set of data, removes unwanted, duplicate content in the dataset. This approach

helps in easy classification of text mentioned in the comments. All the sentiments, words are grouped together, and the best one among the various products are listed separately and stored in the database. These may be useful for the users and the company people to know more about their product which is being highly rated and the best. This approach gives the best outcome of the products among the various listed products from the database for mining.

Figure 1 shows the flow chart explanation for sentiment orientation. When an input is given, it compares the word from the dictionary, if it is found, then the word is added to the wordlist, else the word is compared with the name dictionary and then added to the word list. This helps to find the word easily and makes the process more efficient.

Figure 2 gives the time efficiency comparison between the existing product aspect ranking model and the proposed semantic orientation model.
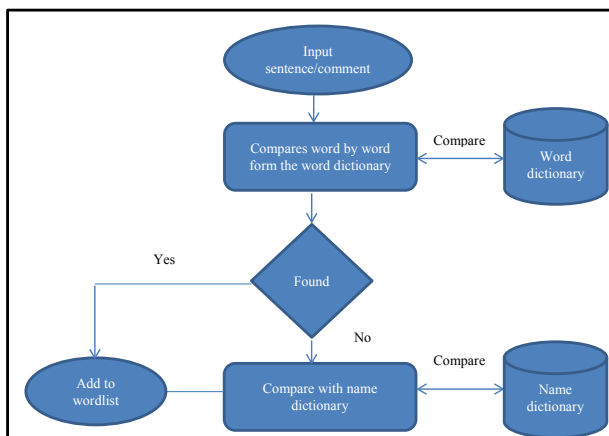

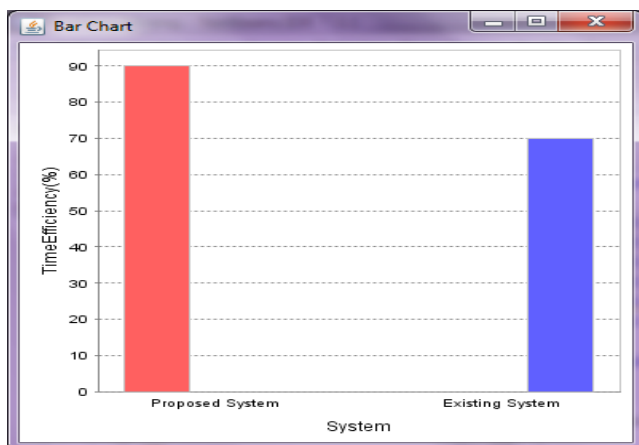**Figure 1.** Flow chart for sentiment orientation.

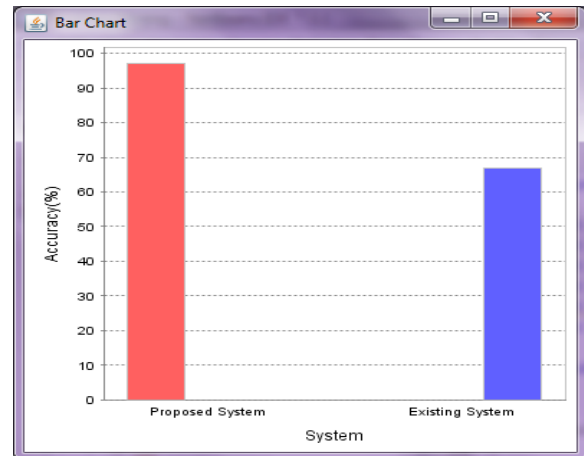**Figure 2.** Comparison of time efficiency.

Figure 3. Comparison of accuracy.

Figure 3 gives the accuracy comparison in percentage for both product aspect ranking and proposed semantic orientation. Hence we have increased 30 % of accuracy in the proposed system.

4. Conclusion

Any product needs an assessment based on product review. In this research work we propose the selection of review or opinion of every user. It estimates the performance of the product after checking the product performance whether good or bad based on reviews. It should be maintained using the proposed new algorithm called semantic orientation. It gives good and complete results based on product review.

5. Acknowledgement

6. References

1. Carenini G, Ng RT, Zwart E. Multi-document summa- rization of evaluative text. Proceedings of ACL, Sydney, NSW: Australia; 2006.
2. Jin W, Ho HH. A novel lexicalized HMM-based learning framework for web opinion mining. Proceedings of 26th Annual ICML, Montreal, QC: Canada; 2009. p. 465–72.
3. Manevitz LM, Yousef M. One-class SVMs for document classification. Journal of Machine Learning Research. 2011 Dec; 2:139–54.

4. Ohana B, Tierney B. Sentiment classification of reviews using SentiWordNet. Proceedings of IT&T Conference, Dublin: Ireland; 2009.

5. Paltoglou G, Thelwall M. A study of information retrieval weighting schemes for sentiment analysis. Proceedings of 48th Annual Meeting ACL, Uppsala: Sweden; 2010. p. 1386–95.

6. Pang B, Lee L. Opinion mining and sentiment analysis in Found. Trends Information Retrieval. 2008; 2(1–2):1–135.

7. Lee, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning technique. Proceedings of EMNLP, Philadelphia, PA: USA; 2002. p. 79–86.

8. Pang B, Lee L, Vaithyanathan S. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts techniques. Proceedings of ACL, Barcelona: Spain; 2004. p. 271–78.

9. Popescu AM, Etzioni O. Extracting product features and opinions from reviews. Proceedings of HLT/EMNLP, Vancouver, BC: Canada; 2005. p. 339–46.

10. Sindhwani V, Melville P. Document-word co-regularization for semi-supervised sentiment analysis. Proceedings of 8th IEEE ICDM, Pisa: Italy; 2008. p. 1025–30.

11. Ruba KV, Venkatesan D. Building a custom sentiment analysis tool based on an ontology for twitter posts. Indian Journal of Science and Technology. 2015 Jul; 8(13):1–9. DOI: 10.17485/ijst/2015/v8i13/61464.

12. Kumar V, Singhal N, Dixit A, Sharma AK. A novel architecture of perception oriented web search engine based on decision theory. Indian Journal of Science and Technology. 2015 Apr; 8(7):635–41. DOI: 10.17485/ijst/2015/v8i7/65156.