



ISSN: 0976-3031

Available Online at <http://www.recentscientific.com>

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 8, Issue, 7, pp. 18143-18147, July, 2017

**International Journal of
Recent Scientific
Research**

DOI: 10.24327/IJRSR

Research Article

A METHODOLOGY TO ENHANCE THE ACCURACY OF ASPECT LEVEL SENTIMENT ANALYSIS USING IMPUTATION OF MISSING SENTIMENT

Jenifer Jothi Mary A^{1*} and Arockiam L²

¹Department of Computer Science, St. Joseph's College (Autonomous),
Tiruchirappalli, Tamil Nadu, India

²School of Computing Sciences, St. Joseph's College (Autonomous),
Tiruchirappalli, Tamil Nadu, India

DOI: <http://dx.doi.org/10.24327/ijrsr.2017.0807.0459>

ARTICLE INFO

Article History:

Received 15th April, 2017
Received in revised form 25th
May, 2017
Accepted 23rd June, 2017
Published online 28th July, 2017

Key Words:

Missing Sentiment; Aspect Based
Sentiment Analysis; Subjective Sentences;
Objective Sentences; Polarity

ABSTRACT

Internet has massive collections of consumer reviews expressing various opinions on various aspects of many a product. Sentiment analysis of online documents such as news articles, blogs and micro blogs has received greater than ever attention in recent years. Mostly, these reviews are not organized properly thereby creating problems in information navigation and knowledge acquisition. In order to address this problem, product aspect ranking is explored automatically to identify product features from online consumer reviews. So, this paper proposes a novel method to consider the aspects of the objective sentences while calculating the polarity of a review.

Copyright © Jenifer Jothi Mary A and Arockiam L, 2017, this is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Online reviews play an important role in product sale and production. Many consumers believed that it is significant to read user-generated content before deciding on a product or before going for a purchase. But, the volume and velocity of online reviews keep on increasing because of the use of internet. So, it becomes more difficult for customers to read through all the reviews and reach an intelligent decision. Many customers have their own preferences for choice of a product. So, extraction based on product aspects from the text of online reviews is instrumental for leveraging the online reviews for individual business decision making.

Millions of users share their thoughts and opinions about different events on the blogs or social networking platforms. Sentiment analysis refers to a classification problem where the focus is to predict the polarity of words and then to classify them into positive and negative with the aim of identifying the attitude and sentiments that are expressed in any language.

Aspect Based Sentiment Analysis

The word aspect is used to represent attributes or components of a product or event. For example, "The Iphone camera is bad". The given review is regarding the camera of the iphone. The opinion for this specific product is negative. The sentiment classification task mainly targets on determining the opinions about the product Aspects. The task of generating product aspect-based summaries is different from traditional text summarization. The goal is to obtain aspects summary formed by all the aspects of the products. Customers have opinions whether positive or negative. Opinion mining is carried out at three levels [6]. They are,

- **Document level:** This level provides the opinion as positive, negative or neutral about the entire document.
- **Sentence level:** At this level, the entire document is separated into sentences. Each sentence is categorized as positive, negative or neutral.
- **Aspect level:** At this level the documents entire Aspect is divided as positive, negative or neutral.

*Corresponding author: **Jenifer Jothi Mary**

Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli, Tamil Nadu, India

Related Works

Khan *et al* (2014) proposed a new algorithm for twitter sentiment analysis and based on three way classification algorithm. It resolved the issues and increases the classification accuracy effectively reducing the number of classified neutrals. The results of the proposed framework showed great improvement when comparing with similar work. It achieved an average accuracy, precision and recall respectively 85.7%, 85.3%, 82.2%. A mechanism proposed by Jeyapriya *et al* (2015) was based on phrase-level to examine customer reviews. It was used to extract most important aspects of an item and to predict the orientation of each aspect. The projected system implemented aspect extraction using frequent item set mining in customer product reviews and opinions whether positive or negative opinion. It identified sentiment orientation of each aspect in customer reviews by supervised learning algorithms. The performance of the system was evaluated with the parameters Precision, recall and F-measure.

Marrese-Taylor *et al* (2014) suggested a new and more complex NLP-based rules for subjective and sentiment classification at the aspect-level in tourism domain. An aspect-based opinion mining tool was used to create a prototype and analyze opinions from TripAdvisor. Results showed that the approach achieved 92% F-measure in determining the sentiment orientation. Proios *et al* (2015) proposed a model for the analysis of tandem product reviews, which considered the aspects. The model was discussed with two different visualizations i.e., one for business and another for end-users. The former uncovered the strong and weak points against the competitors by the aggregated aspect-based opinions provided by all users. And the later was a collaborative filtering approach. It calculated user similarity over a content-based clustering of users and items. It was also demonstrated by the prototype in different scenarios using the Yelp challenge dataset.

Che *et al* (2015) presented a framework for aspect-based sentiment analysis using sentiment sentence compression. Sent_Comp used to remove the unnecessary sentiment and compressing a complicated sentiment sentence into shorter and easier to parse. They also applied a conditional random field model to automatically compress sentiment sentences. It used four product domains of the Chinese corpora. The influence of exclamation marks of the sentiment value of informal text in social media was demonstrated by Teh *et al* (2015). This was experimented on Potts' online tool and others for determining sentiment scores for comments with "!!!!!!" and "!!" to examine the number of exclamation marks for sentiment value. If a comment is positive, then the value assigned manually to an expression that contains more exclamation marks. Vu *et al* (2016) proposed PUMA (Phrase-Based Extraction of User Opinions in Mobile App Reviews) approach to extract user opinions in app reviews. This approach included parts-of-speech (PoS) technique to extract. This was a technique to cluster similar phrases and a technique to monitor phrase clusters with negative sentiments. PUMA used to study two popular apps and found that it could reveal severe problems of those apps reported in there views.

Nair *et al* (2015) proposed a method of extracting the sentiment from Malayalam film review. Method comprising of machine learning, semantic orientation and rule based were used for sentiment analysis of a given text. They compared both statistical methods namely SVM (Support Vector Machine) and CRF (Conditional Random Field) for analyzing the sentiments of movie reviews. From that, they found SVM performed better than CRF. This helped the users to analyze and to assign the rank for the popularity of new arrival films. The system checked the polarity at the sentence level, resulted in an accuracy of 91%. Nair *et al* (2014) used a method for extracting the sentiments from a Malayalam sentence with respect to the film reviews. It also handled smileys which conveyed the exact meaning of a particular situation. The negation rule algorithm was considered in the sentiment extraction. This system produced 85% polarity at the sentence level for the movie reviews.

Statistical sequence leveling technique, discussed by Chatterji *et al* (2015) was used to identify the aspects mobile domain as given by the reviewers. This technique trained 950 manually annotated sentences. The sentiments of a sentence identified using a bag-of-word technique (based on SVM) trained 5450 manually annotated sentences. The association of sentiments with aspects identified 70.08% accuracy. Roy *et al* proposed a method for aspect tagging without using any parsers for sentiment analysis. This method allowed the algorithm to be extended to any language. This might be a viable method for end-to-end aspect extraction for given proper training set. Poria *et al* (2014) proposed a novel rule-based approach to explore common-sense knowledge and sentence dependency trees to detect both aspects. Two popular review datasets were used for evaluating the system against state-of-the-art aspect extraction techniques which obtained the higher detection accuracy for both datasets.

Motivation

In reviews, customer posts one's opinions on a variety of topics, discuss current issues, express complains in positive, negative or neutral sentiment on products that one buys. In fact, companies have started using these reviews to get the public opinions of their product. Companies study these user reviews and try to make possible changes according to the reviews. However, the sheer vastness of these reviews make it very difficult to analyze and generate knowledge out of it. Therefore, it is necessary to find a novel mechanism to classify and to improve the accuracy of the reviews.

Objective

The objective of this paper is to propose a mechanism to find the polarity of the objective sentences where aspects are present. This will improve the accuracy of the aspect polarity to support the customer.

Proposed Work

The proposed mechanism uses customer reviews to extract opinions of the consumers based on aspects and mine the given reviews. In this mechanism, reviews are divided into individual sentences. These separated review sentences are given as input to the SentiWordNet lexicon to find the subjective (opinionated) and objective (non-opinionated) sentences.

Sentiment orientation is used to identify whether the given sentence is positive or negative. After separating the sentences, it extracts aspects in the subjective sentences using Aspect word List. Then, it counts the number of positive and negative opinions on each aspect. Finally the total score of each aspect will be calculated using the positive and negative scoring. But, before generating required business decision, this mechanism considers the objective sentences as a factor to improve the accuracy of the result. Though sentiment words are not present in the objective sentences, the proposed mechanism checks the aspect words that are present in these sentences. This will be imputed by the sentiment words using educated guessing to improve the accuracy of the aspect summarization.

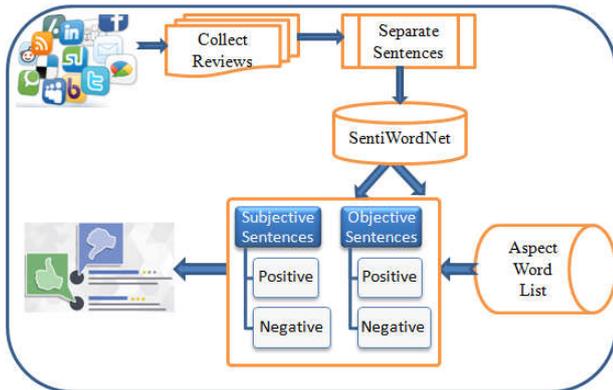


Figure 1 Proposed Framework for Aspect Level Sentiment Analysis

The proposed mechanism aims to improve the accuracy of each aspect by considering the objective sentences along with the subjective sentences. The proposed mechanism for Aspect-based Sentiment Analysis is given in Figure 1.

The aspect summarization of the reviews is done in five phases. They are explained below:

- **Phase I - Review Collection:** A review crawler is used to extract the reviews from the websites. It parses the HTML webpage opinions and extracts the online product reviews. The extracted reviews will be stored in a database that is used as an input for the proposed mechanism.
- **Phase II - Sentence separation:** In this phase, the collected reviews are separated by sentences. Each sentence is counted and again separated as individual words.
- **Phase III - Subjectivity / Objectivity Classification:** A sentence of the review is analysed by the sentiment words or opinions, given in the Sentiwordnet lexicon. This Sentiwordnet lexicon has the dictionary of opinion words. If any sentiment is present in a particular sentence, it is considered as a Subjective Sentence (Sub_Sen). Otherwise, it would be an Objective Sentence (Obj_Sen). These categorized Obj_Sen are removed to avoid further processing overhead.
- **Phase IV -Aspect Extraction:** A reviewer can post reviews on many aspects of a single product. Aspects are nothing but the important features of a product. These aspects may be a single word or a phrase. To extract the aspects, it is necessary identify the aspect words in each sentence of the review. After extracting the aspects, the proposed mechanism finds the sentences' aspect

frequency with its sentiment orientation such as positive or negative.

- **Phase V - Aspect Opinion summarization:** Aspect-wise opinions are summarized not only in the number of Sub_Sen and but also from the Obj_Sen.

METHODOLOGY

The working methodology of the proposed mechanism for Aspect-based Sentiment Analysis is explained in Figure 3.

```

Step 1: Read a review
Step 2: Separate the review by sentences
Step 3: Calculate the number of sentences in the review
Step 4: Read each sentence
Step 5: Separate the sentences by Words
Step 6: Calculate the number of Words in the given sentences
Step 7: Seed sentiment words through SENTIWORDNET
Step 8: if Word is sentiment word
    Assign subjective
    Calculate the number of Subjective sentence
    seed aspects
    if aspect present then
        get the sentiment
        Calculate aspect wise summarization
    end if
else
    Assign objective
    Calculate the number of Objective sentence
    seed aspects
    if aspect present then
        assign the sentiment
        Calculate aspect wise summarization
    end if
end if
Step 9: aspect summarization
    
```

Figure 3 Pseudocode of the Proposed Mechanism

To obtain a best business decision, the proposed mechanism considers the aspects present in the objective sentences Obj_Sen by seeding the aspects to the Obj_Sen. Finally, it performs the aspect wise summarization by considering the subjective and objective sentences. This research paper has used online customer reviews on Samsung galaxy note 7 from Amazon.com. The reviews were collected from 10.10.2016 – 15.10.2016 from the above mentioned website. A model review is depicted in Figure 4.

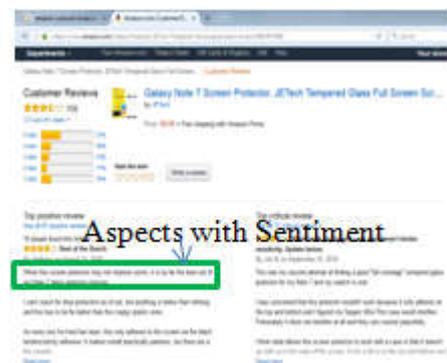


Figure 4 A Model Review

The working methodology of the proposed mechanism is presented in Figure 5. If the sentiment word is present in a sentence, then it is tagged as Sub_Sen. Otherwise, it is tagged

as Obj_Sen by the proposed mechanism. Aspect words will be sent as seed words to the subjective sentences to count the positive and negative sentiments for the score. This process is repeated for the objective sentences after eliminating the sentences without any sentiments.

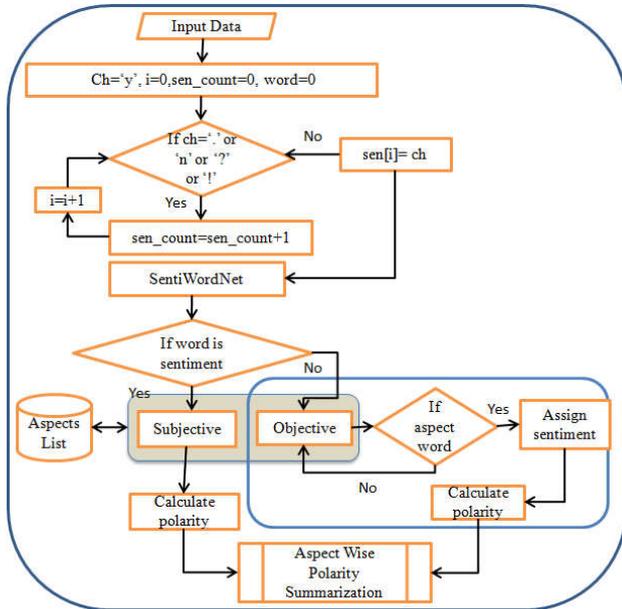


Figure 4 Workflow of the proposed mechanism

Performance Evaluation

The proposed mechanism is evaluated using the following parameters namely, accuracy, precision and recall. Accuracy means how much the given value is correct. Let the list of aspects be a₁, a₂, a₃, a₄ and a₅. The Positive and Negative opinions can be denoted aspve_op and nve_op respectively. Likewise, Subjective Sentence, Objective Sentence and Aspect Summarization are denoted as Sub_Sen, Obj_Sen and Asp_Sum respectively. Each aspect polarity will be calculated using the equation (1). Aspect wise summary of the subjective and objective sentences in terms of positive and negative will be calculated using the equations (2) and (3) respectively.

$$a_{1ij} = \sum pve_op + \sum nve_op \quad (1)$$

$$Sub_Sen = \sum a_i \quad (2)$$

$$Obj_Sen = \sum a_i \quad (3)$$

where, i denote number of positive opinion, j represents the number of negative opinion. Aspect-wise summarization is being calculated using the equation - (4).

$$Asp_Sum_{ij} = \sum (Sub_Sen + Obj_Sen) \quad (4)$$

Finally, both the positive and negative scores are added together to get the accuracy of the sentiment of the given reviews. The abstract of the reviews collected is given in Table I.

Table I Abstract of the Sample Review Collected

Corpus	Samsung galaxy note 7 from amazon.com
No. of Reviews	158
Total number of Sentences	421
Number of Subjective Sentences	293
Number of Objective Sentences	128

Figure 5. presents the sum of positive and negative aspects of the Samsung galaxy note 7 and the mean value of the sum value.

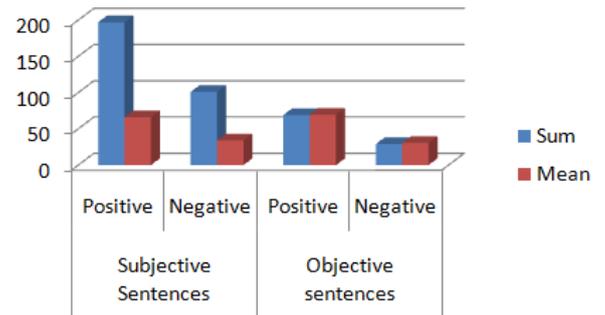


Figure 5 Sum and Mean of the Sample Review collected

While presenting the aspect summary, the proposed mechanism considers the aspects in the objective sentences along with the aspects of the subjective sentences. Though the total number of positive and negative opinions derived in this proposed method is decreased, this mechanism provides the accurate consumer opinion of a product on the basis of aspects.

CONCLUSION

Competing in the business world is a never-ending challenge that one continually faces. This paper presents a generic design of an opinion mining system that aims at improving the products in industrial perspective. The authors have proposed a system which recommends the product aspects that are highly valued by business world. This mechanism can be utilized to analyse the businesses reviews, to know the pros and cons and compare the product with the competitors. It further drills-down the opinions in order to bring a new light on the hidden gems.

References

- Khan, Farhan Hassan, Saba Bashir, and Usman Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme", Decision Support Systems, Vol. 57, 2014, pp. 245-257.
- Jeyapriya, A., and CS Kanimozhi Selvi, "Extracting aspects and mining opinions in product reviews using supervised learning algorithm", Proc. of the International Conference Electronics and Communication Systems (ICECS), IEEE, 2015, pp. 548-552.
- Marrese-Taylor, Edison, Juan D. Velásquez, and Felipe Bravo-Marquez, "A novel deterministic approach for aspect-based opinion mining in tourism products reviews", Expert Systems with Applications, Vol. 41, No. 17, 2014, pp.7764-7775.
- Proios, Dimitris, Magdalini Eirinaki, and Iraklis Varlamis, "TipMe: Personalized advertising and aspect-based opinion mining for users and businesses", Proceedings of the IEEE / ACM International Conference on Advances in Social Networks Analysis and Mining, 2015, pp. 1489-1494.
- Che, Wanxiang, Yanyan Zhao, Honglei Guo, Zhong Su, and Ting Liu, "Sentence compression for aspect-based sentiment analysis", IEEE / ACM Transactions on

- Audio, Speech, and Language Processing, Vol. 23, No. 12, 2015, pp. 2111-2124.
- Bing Liu, "Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers", 2012, pp.7-140.
- Teh, Phoey Lee, Paul Rayson, Irina Pak, and Scott Piao, "Sentiment analysis tools should take account of the number of exclamation marks!!!", Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services, ACM, 2015, p. 35.
- Vu, Phong Minh, Hung Viet Pham, and Tung Thanh Nguyen, "Phrase-based extraction of user opinions in mobile app reviews", Proceedings of the 31st IEEE / ACM International Conference on Automated Software Engineering, 2016, pp. 726-731.
- Nair, Deepu S., Jisha P. Jayan, and Elizabeth Sherly, "Sentiment Analysis of Malayalam film review using machine learning techniques", International Conference on In Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2015, pp. 2381-2384.
- Nair, Deepu S., Jisha P. Jayan, and Elizabeth Sherly, "SentiMa-sentiment extraction for Malayalam", International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2014, pp. 1719-1723.
- Chatterji, Sanjay, Ranjan Kumar Rahul, and Anoop Arora, "A hybrid approach for identifying sentiments around aspects", 2nd International Conference on In Recent Trends in Information Systems (ReTIS), IEEE, 2015, pp. 33-37.
- Roy, Narendra, and Samik Some "Product Aspect Extraction for Sentiment Analysis without using Parsers".
- Poria, Soujanya, Erik Cambria, Lun-Wei Ku, Chen Gui, and Alexander Gelbukh, "A rule-based approach to aspect extraction from product reviews", Proceedings of the second workshop on natural language processing for social media (SocialNLP), 2014, pp. 28-37.

How to cite this article:

Jenifer Jothi Mary A and Arockiam L., A Methodology to Enhance the Accuracy of Aspect Level Sentiment Analysis Using Imputation of Missing Sentiment. *Int J Recent Sci Res.* 8(7), pp. 18143-18147.
DOI: <http://dx.doi.org/10.24327/ijrsr.2017.0807.0459>
