ISSN: 2278-0181
NCRTET-2015 Conference Proceedings

Important Facet of Product Appraisal and Application

Deepika. S[1], P G Student[1], Jennifer Sagaya Rani. A[2],Asst.Professor[2],
Department Of Computer Science Engineering,
[1]PG Student, Parisutham Institute of Technology and Science,
Thanjavur, Tamilnadu, India.

***Abstract*: Numerous consumer reviews of products are now available on Internet. The Consumer reviews contains valuable knowledge for the users. The reviews are often disorganized and leading to difficulties in information navigation. This proposes a product aspect ranking framework ,in which automatically identifies the important aspects of products from online consumer reviews and aiming at improving the usability of the numerous reviews. The important product aspects based on two observations are the important aspects commented on by a large number of consumers and consumer opinions on the important aspects and their overall opinions on the product. The consumer reviews of a product, is to identify product aspects by a shallow dependency parser and then determine the consumer opinions on these aspects of sentiment classifier. This develop a probabilistic aspect ranking algorithm to infer the importance of aspects by simultaneously in the aspect of frequency and the influence of consumer opinions given to each aspect over their overall opinions.**

Keywords*: Sentiment Classifier, Product Review, Consumer Review, Product Aspects.*

I. INTRODUCTION

In Recent years, witnessed the rapidly expanding e-commerce. A recent study from ComScore reports that online retail spending reached $37.5 billion in Q2 2011 U.S. [5]. The Millions of products from various merchants have been offered in online.In particular, BingShopping1 has indexed more than five million products.Amazon.com archives a total of more than 36 million of products.Shopper.com records more than five million of products from over 3,000 merchants. In Most retail Websites it encourage consumers to write reviews to express their own opinions on various aspects of products. Some aspect, also called feature in literatures, which refers to a component or an attribute of a certain product. A sample review is “The battery life of Nokia N95 is amazing" it reveals the positive opinion on aspect “battery life" of the product Nokia N95.

In the retail Websites, there are many forum Websites provide a platform for the consumers to post reviews on millions of products. In Particular, CNet.com involves more than seven million of product reviews; whereas Pricegrabber.com contains millions of reviews on more than 32 million of products in 20 distinct categories over 11,000 merchants. Such large consumer reviews contain valuable knowledge and have to become an important resource for the both consumers and firms [9]. Most of the Consumers commonly seek quality of information from the online reviews to purchasing a product,many firms use online reviews as a important feedback in their product development, marketing, and the consumer relationship management. A product may have hundreds of aspects.

For instance, iPhone 3GS has more than three hundred aspects (Fig. 1), such as “usability," “design," “application," “3G network". Some aspects are more important than the others, and have greater impact on 'decision making as well as firmsproduct' development strategies. Some aspects
of iPhone 3GS, e.g., “usability" and “battery," are concerned by most consumers, and are more important than the others such as “USB" and “Button". It is Motivated by the observations,in this we propose a product aspect ranking framework to automatically identify the important aspects of the products from online by consumer reviews. The assumption is that the important aspects of a product possess has the following characteristics:

Frequently commented in the consumer reviews for product and consumers' opinions on the product. The frequency-based solution is to regard the aspects that are frequently commented in consumer reviews denoted as important. The consumers’ opinions on the frequent aspects may not influence the overall opinions of the product, and not the influence of purchasing decisions.

For instance, most of the consumers frequently criticize bad “signal connection" of iPhone 4, but they still give the high overall ratings to iPhone 4. Some aspects such as “design" and “speed," has not be frequently commented, but they are more important than “signal connection." The frequency-based on the solution is unable to identify the important aspects. Then, a basic method is to influence the consumers’ opinions on specific aspects over the overall ranking on the product is to count the cases in their opinions on specific aspects and the ratings are consistent, and then ranks the aspects according to the number of the consistent cases. This method assumes that an overall ranking was derived from the specific opinions on the different aspects individually, and not precisely characterize the correlation between the specific opinions and the overall ranking. This method propose an effective aspect ranking approach to infer the importance of product aspects on the bases of reviews and consumers' opinion.

The more attentions to the important aspects, during firms can focus only on improving the quality of the aspects and enhance the product reputation effectively. It is impractical for the people to manually identify the important aspects of products from large or numerous reviews. Hence, an approach to automatically identify the important aspects of the product is highly demanded.

Fig. 1. Numerous aspects of the product *iPhone 3GS*.

For a product like camera, the aspects such as “lenses" and “picture quality" is greatly influences consumer opinions on the camera, thus, they are more important than the aspects such as “a/v cable" and “wrist strap." In which, it identify important product aspects and will improve the usability of numerous reviews in beneficial to both consumers and firms. Some consumers are conveniently make wise purchasing decision by paying more attention to the important aspects of the product.

II. RELATED WORKS

As shown in Fig. 2, given the consumer reviews of a specific product is, first identify the aspects of reviews by a shallow dependency parser [10] and then analyze consumer opinions on the aspects in the bases of a sentiment classifier. Then develop a probabilistic aspect ranking or rating algorithm, which is effectively exploits the aspect of frequency as well as the influence of a consumers’ opinions given to each aspect over their overall opinions on the product in a unified probabilistic order.

For example, assume the overall opinion in a review is generated based on a weighted aggregation of the opinions on a specific aspects,in which the weights essentially measure the degree of importance of the aspects. Then, a probabilistic regression algorithm is developed to infer the importance weights by the incorporating aspect frequency and associations between the overall opinion and their opinions on specific aspects.In the bases, to evaluate the proposed product aspect ranking the framework, then collect a large collection of product review consisting of 94,560 consumer reviews on 21 products in the eight domains. These reviews are crawled from the multiples' of prevalent forumWebsites,i.e.,CNet.com,Viewpoints.com, Reevoo.com and Pricegrabber.com etc. This corpus is available by the request for future research on aspects of ranking and related topics.

Extensive experimental results on the corpus demonstrate shows the effectiveness of the proposed product aspect ranking the framework. The Product aspect ranking is beneficial to a wide range of real-world applications. It investigate the usefulness in two applications, i.e. document-level sentiment classification which aims to determine a review document as expressing a positive or negative opinion from the consumer, and extractive review summarization that aims to summarize the consumer reviews by selecting the informative review sentences.

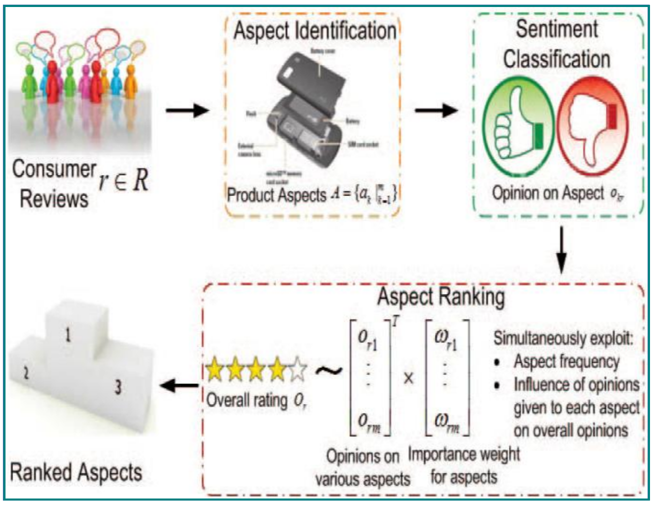

Fig.2. Flowchart of the proposed product aspect ranking framework.

Then perform extensive experiments to evaluate the efficacy of aspect ranking in these two applications and achieve significant performance improvements. Product aspect ranking was first introduced in our previous work [8]. Which is compared to the preliminary conference version, it has no less than the following improvements: First, elaborates more discussions and analysis on the product aspect ranking problem; Second, performs the extensive evaluations on more products in more diverse domains; and Third, demonstrates the potential aspects of ranking in more real-world applications.

In summary, the main contributions of this process which includes:

• Propose a product aspect ranking framework to automatically identify the important aspects of products from numerous consumer reviews.

• Develop a probabilistic aspect ranking algorithm to infer the importance of various aspects by simultaneously exploiting aspect frequency and the influence of consumers’ opinions given to each of the aspect over their overall opinions on the product.

• Demonstrate the potential of aspect ranking in real-world applications. Then, Significant performance improvements are obtained on the applications of the document-level sentiment classification and extractive review summarization by making use of the aspect ranking.

III. METHODS

Product Aspect Identification:

The consumer reviews are composed in different formats on various forum Websites. These 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 to write a paragraph of detailed review in the free text. Some Websites such as, Viewpoints.com, only ask for an overall rating and a paragraph of free-text review. Some others Websites such as Reevoo.com just require an overall rating and concise positive and negative opinions on certain aspects. Besides an overall rating, a consumer review consists of Pros and Cons reviews, free text review.

In the Pros and Cons reviews, which identify the aspects by extracting the frequent noun terms in the reviews. Previously it is shown that aspects are usually nouns or noun phrases [7], and it can obtain highly accurate aspects by extracting frequent noun terms from the Pros and Cons reviews [6]. To identify aspects in the free text reviews, a straightforward solution is to employ an existing aspect identification approach. The most notable existing approach is that proposed by Hu and Liu [4]. Then it is first identify 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.

Sentiment Classification on Product Aspects:

This task for analyzing the sentiments expressed on aspects is called aspect-level sentiment classification. The Exiting techniques include the supervised learning approaches and the lexicon-based approaches, which is unsupervised. Lexicon-based methods utilize a sentiment lexicon consisting of a list of sentiment words, phrases and idioms, is to determine the sentiment orientation on each aspect [3]. These method are easily to implement, their performance relies heavily on the quality of the sentiment lexicon. Then, supervised learning methods train a sentiment classifier based on training corpus. The classifier is then used to predict the sentiment on each aspect. Many learning-based classification models are applicable, for instance, Support Vector Machine (SVM), Naive Bayes, and Maximum Entropy (ME) model etc. [2]. Supervised learning is dependent on the training data and cannot perform well without sufficient training samples. However, labeling training data is laborintensive and time-consuming. In this work, the Pros and Cons reviews have explicitly categorized positive and negative opinions on the aspects. These reviews are valuable training samples for learning a sentiment classifier.

Probabilistic Aspect Ranking Algorithm:

The proposed of probabilistic aspect ranking algorithm to identify the important aspects of a product from consumer reviews. Generally, important aspects have the following characteristics: Frequently commented in consumer reviews; and Consumers' opinions on these aspects greatly influence their overall opinions on the product. The overall opinion in a review is an aggregation of the opinions given to specific aspects in the review, and various aspects have different contributions in the aggregation. The opinions on the important aspects have strong (weak) impacts on the generation of overall opinion. The learned sentiment classifier is then leveraged to determine the opinion of the opinionated expression, i.e. the opinion on the aspect.

IV. EVALUATION

It shows the extensive experiments to evaluate the effectiveness of the proposed product aspect ranking framework, including product aspect identification, sentiment classification on aspects, and aspect ranking.

Evaluations of Product Aspect Identification on Free Text Reviews:

It is compared to aspect identification approach with the following two methods: (a) the method proposed by Hu and Liu, which extracts nouns and noun phrases as aspect candidates, and identifies aspects by rules learned from association rule mining; and (b) the method proposed by Wu et al, that extracts noun phrases from a dependency parsing tree as aspect candidates, and identifies aspects by a language model built on the reviews.

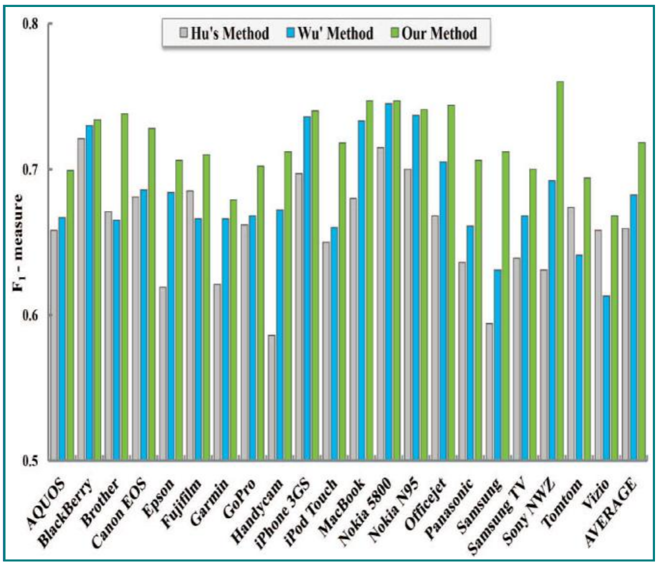

Fig. 3. Performance of product aspect identification

Fig. 3 shows the performance comparison on all the 21 products in terms of F1-measure. From these results, we can see that the proposed approach get the best performance on all the 21 products. It significantly outperforms Hu's and Wu's methods by over 9.0% and 5.3% respectively in terms of average F1-measure. This indicates the effectiveness of Pros and Cons reviews in assisting aspect identification on free text reviews. Hence, by exploiting the Pros and Cons reviews, our approach can boost the performance of aspect identification.

Evaluations of Sentiment Classification on Product Aspects:

To compared the following methods of sentiment classification: one unsupervised method. The opinion on each aspect is determined by referring to the sentiment lexicon. This lexicon contains a list of positive/negative sentiment words. The opinionated expression modifying an aspect is classified as positive (or negative) if it contains a majority of words in the positive (or negative) list; and three supervised methods. It is employed three supervised methods, including Naïve Bayes (NB), Maximum Entropy (ME), and Support Vector Machine (SVM). The sentiment classifiers were trained on the Pros and Cons reviews. In particular, SVM was implemented by using libSVM [1] with linear kernel, NB was implemented with Laplace smoothing, and ME was implemented with L-BFGS parameter estimation.

V. APPLICATIONS

Aspect ranking is beneficial to a wide range of realworld applications. We here investigate its capacity in two applications, i.e. document-level sentiment classification on review documents, and extractive review summarization.

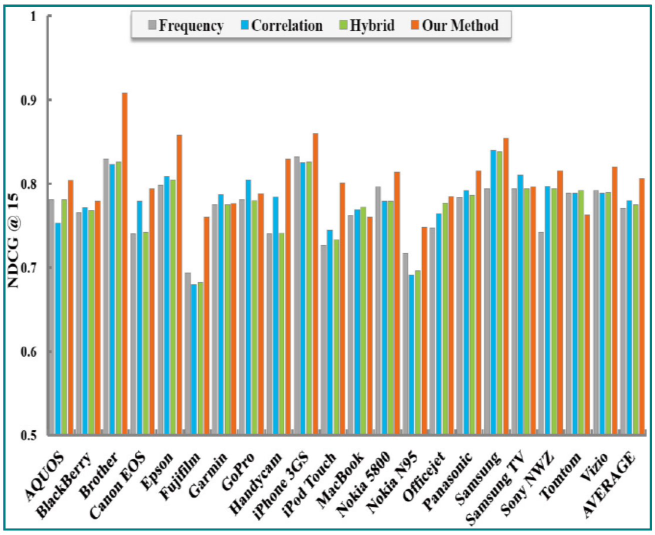

Fig. 4. Performance of aspect ranking in terms of NDCG@*15*

Document-level Sentiment Classification:

The goal of document-level sentiment classification is to determine the overall opinion of a given review document. A review document often expresses various opinions on multiple aspects of a certain product. The opinions on different aspects might be in contrast to each other, and have different degree of impacts on the overall opinion of the review document. For example, a sample review document of iPhone 4. It expresses positive opinions on some aspects such as "reliability," "easy to use," and simultaneously criticizes some other aspects such as "touch screen," "quirk," "music play." Finally, it assigns an high overall rating (i.e., positive opinion) on iPhone 4 due to that the important aspects are with positive opinions. Hence, identifying important aspects can naturally facilitate the estimation of the overall opinions on review documents. This observation motivates us to utilize the aspect ranking results to assist document-level sentiment classification. We conducted evaluations of document-level sentiment classification over the product reviews. Specifically, we randomly sampled 100 reviews of each product as testing samples and used the remaining reviews for training. Each review contains an overall rating, which is normalized to [0,1]. We treated the reviews with high overall rating (>0.5) as positive samples, and those with low rating (<0.5) as negative samples. The reviews with ratings of 0.5 were considered as neutral and not used in our experiments. We collected noun terms, aspects, and sentiment terms from the training reviews as features. Note that sentiment terms are defined as those appear in the sentiment lexicon provided by MPQA project.

All the training and testing reviews were then represented into feature vectors. In the representation, we gave more emphasis on the important aspects, and the sentiment terms modifying them. Technically, the feature dimensions corresponding to aspect ak and its corresponding sentiment
terms were weighted which is the importance score is a tradeoff parameter and was empirically set to 100 in the experiments. Based on the weighted features, a SVM classifier was learned from the training reviews and used to determine the overall opinions on the testing reviews. We compared our approach with two existing methods, i.e., Boolean weighting and term frequency (TF) weighting. Boolean weighting represents each review into a feature vector of Boolean values, each of which indicates the presence or absence of the corresponding feature in the review. Term frequency (TF) weighting, weights the Boolean feature by the frequency of each feature on the corpus.

An abstractive summarization attempts to develop an understanding of the main topics in the source reviews and then express those topics in clear natural language. It uses linguistic techniques to examine and interpret the text and then to find the new concepts and expressions to best describe it by generating a new shorter text that conveys the most important information from the original text document. An extractive method summarization method consists of selecting important sentences and paragraphs from the original reviews and concatenating them into shorter from.

VI. CONCLUSION

In this process, it is proposed a product aspect ranking framework to identify the important aspects of products from numerous consumer reviews. The framework contains three main components, i.e., product aspect identification, aspect sentiment classification, and aspect ranking. First, we
exploited 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 infer the importance of various aspects of a product from numerous reviews. The algorithm simultaneously explores aspect frequency and the influence of consumer opinions given to each aspect over the overall opinions. 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.

ACKNOWLEDGEMENT

Sincere thanks to my guide Prof.A.Jennifer Sagaya Rani Asst.Professor Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur for her help and guidance.

REFERENCES

1. J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *J. Neural Parallel Scientific Comput.*, vol. 11, no. 4, pp. 351–368, 2003.
2. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. EMNLP*, Philadelphia, PA, USA, 2002, pp. 79–86.
3. B. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," in *Proc. IT&T Conf.*, Dublin, Ireland, 2009.
4. M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. SIGKDD*, Seattle, WA, USA, 2004, pp. 168–177.
5. ComScore Reports [Online]. Available: http://www.comscore.com/Press_events/Press_releases, 2011.
6. B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proc. 14th Int. Conf. WWW*, Chiba, Japan, 2005, pp. 342–351.
7. B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing*, New York, NY, USA: Marcel Dekker, Inc., 2009.
8. J. Yu, Z.-J. Zha, M. Wang, and T. S. Chua, "Aspect ranking: Identifying important product aspects from online consumer reviews," in *Proc. ACL*, Portland, OR, USA, 2011, pp. 1496–1505.
9. A. Ghose and P. G. Ipeirotis,"Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 10, pp. 1498–1512. Sept. 2010.
10. Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in *Proc. ACL*, Singapore, 2009, pp. 1533–1541.

AUTHOR DETAILS

S.Deepika received B.E(Computer Science and Engineering) from Dhanalakshmi Srinivasan Engineering college, Perambalur in 2013. Now pursuing M.E(Computer Science and Engineering) in Parisutham Institute of Technology and Science, Thanjavur.

A.Jennifer Sagaya Rani received B.E(Computer Science and Engineering) from Sudharsan Engineering College, Pudukkottai, M.E(Computer Science and Engineering) from Anna University, Trichy and currently working as an Assistant Professor in Parisutham Institute of Technology and Science, Thanjavur.