ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Product Aspect Ranking in Facilitating Real-World Applications

Vishnu Vardhan Anasuri,
M.Tech Student
Department of CSE ,
Global Institute of Engineering and Technology,
Chilkur,RRDistrict,Telangana

Mrs. M. Jhansi Lakshmi,
Associate professor,
HoD of CSE,
Global Institute of Engineering and Technology,
Chilkur,RRDistrict,Telangana

ABSTRACT:

A product may have hundred of aspects. Some of the product aspects are more important than the others and have strong influence on the eventual consumer's decision making as well as firm's product development strategies. Identification of important product aspects become necessary as both consumers and firms are benefited by this. Consumers can easily make purchasing decision by paying attention to the important aspects as well as firms can focus on improving the quality of these aspects and thus enhance product reputation efficiently. The important product aspects are identified based on two observations: 1) the important aspects are usually commented on by a large number of consumers and 2) consumer opinions on the important aspects greatly influence their overall opinions on the product. We then develop a probabilistic aspect ranking algorithm to infer the importance of aspects by simultaneously considering aspect frequency and the influence of consumer opinions given to each aspect over their overall opinions.This paper provides the description of various techniques for product aspect identification and classification.

KEYWORDS: ***aspect identification; aspect Ranking; Consumer review; Product aspects; Sentiment classification.***

INTRODUCTION

In data mining, we often need to compare samples to see how similar they are to each data's and others using K-nearest neighbor algorithm. For samples whose features have continuous values, it is customary to consider samples to be similar to each other if the distances between them are small. Other than the most popular choice of Euclidean distance, there are of course many other ways to define distance. Results obtained for the ranking of aspects are also encouraging. As mentioned before, it is necessary to propose improved matching methods and new evaluation measures capable of dealing with the inconsistencies that can appear at evaluation step, in order to obtain more reliable results. Our future work will be firstly oriented in this direction. We will also attempt to provide users with the opinion polarity of each identified product aspect and grouping aspects according to the strength of their opinions and their granularity level.

Product aspect ranking is beneficial to a wide range of real-world applications. In this paper, we investigate its usefulness in two applications, i.e. document-level sentiment classification that aims to determine a review document as expressing a positive or negative overall opinion, and extractive review summarization which aims to summarize consumer reviews by selecting informative review sentences. We 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 . Compared to the preliminary conference version , his article has no less than the following improvements:
(a) it elaborates more discussions and analysis on product aspect ranking problem

(b) it performs extensive evaluations on more products in more diverse domains; and
(c) it demonstrates the potential of aspect ranking in more real-world applications.

In summary, the main contributions of this article include:
- We propose a product aspect ranking framework to automatically identify the important aspects of products from numerous consumer reviews.
- We 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 aspect over their overall opinions on the product.
- We demonstrate the potential of aspect ranking in real-world applications. Significant performance improvements are obtained on the applications of document-level sentiment classification and extractive review summarization by making use of aspect ranking.

RELATED WORK

This section briefly survey previous work on product aspect ranking framework starting with the product aspect identification. Existing product aspect identification techniques can be broadly classified into two main approaches: supervised and unsupervised [2]. Supervised learning technique learns an extraction model which is called as aspect extractor, that aspect extractor is then used to identify aspects in new reviews. For this task Hidden Markov Models and Conditional Random Fields [3, 4], Maximum Entropy [13], Class Association Rules and Naive Bayes Classifier [14] approaches have been used. Wong and Lam [3] used a supervised learning technique to train an aspect extractor.

They learned aspect extractor using Hidden Markow Model and conditional random field. Supervised techniques is reasonably effective, but preparation of training examples is time consuming. In contrast, unsupervised approaches automatically extract product aspects from customer reviews without using training examples. Hu and Liu's works [5, 6] focuses on association rule mining based on the Apriori algorithm to mine frequent itemsets as explicit product aspects. In association rule mining, the algorithm does not consider the position of the words in the sentence. In order to remove incorrect frequent aspects, two types of pruning criteria were used: compactness and redundancy pruning.

The technique is efficient which does not require the use of training examples or predefined sets of domain-independent extraction patterns. However, it suffers from two main shortcomings. First, frequent aspects discovered by the mining algorithm might not be product aspects. The compactness and redundancy pruning rules are not able to eliminate these false aspects. Second, even if a frequent aspect is a product aspect, customers may not be expressing any subjective opinion about it in their reviews.

Wu et al [7] also used the unsupervised method. They used the phrase dependency parser to extract noun and noun phrases and then they used a language model to filter out the unwanted aspects. This language model was used to predict the related score of candidate aspects and was built on product reviews. Candidate having low score were filtered out. However this language model might be biased to frequent terms in the reviews and cannot predict the aspect score exactly as a result cannot filter out noise very efficiently. Subsequently, Popescu and Etzioni [17]developed the OPINE system, which extracts aspects based on the KnowItAllWeb information extraction system [18]. After identification the important aspects next step is sentiment classification which is used to determine the orientation of sentiment on each aspects. Aspect sentiment classification can be done by using two approaches first Lexicon based approach and second supervise learning approach. Lexicon based approach is typically unsupervised. Lexicon consists of list of sentiment words, which may be positive or negative. This method usually employs a bootstrap strategy to generate high quality Lexicon.

Hu and Liu [5] have used this lexicon based method. They obtained the sentimental lexicon by using synonym/antonym relation defined in WordNet to bootstrap the seed word set. Hu's method is improved by Ding et al [8] by addressing two issues: opinion of sentiment word would be content sensitive and conflict in review. They derived the lexicon by using some constraints. Second approach is supervised learning approach which classifies opinions on aspects by using sentiment classifier. Sentiment classifier is learned from training corpus which is used to classify the new aspects opinions. Many learning models are applicable for this purpose [9].

Bopong and Lee [10] used 3 machine learning techniques SVM, Naïve Bayes and Maximum Entropy for determining whether the review is positive or negative. The product aspect ranking is to predict the ratings on individual aspects. Wang et al. [15] developed a latent aspect rating analysis model, which aims to infer reviewer's latent opinions on each aspect and the relative emphasis on different aspects. This work concentrates on aspect-level opinion estimation and reviewer rating behavior analysis, rather than on aspect ranking. Snyder and Barzilay [16] formulated a multiple aspect ranking problem. Justin Martineau and Tim Finin [19] present Delta TFIDF, a general purpose technique to efficiently weight word scores. This technique calculate the value of aspect in document but does not take into account the frequency of words associated with aspect with it.

The two evaluated real-world applications. We start with the works on aspect identification. Existing techniques for aspect identification include supervised and unsupervised methods. Supervised method learns an extraction model from a collection of labeled reviews. The extraction model, or called extractor, is used to identify aspects in new reviews. Most existing supervised methods are based on the sequential learning (or sequential labeling) technique [18].

For example, Wong and Lam [36] learned aspect extractors using Hidden Marköv Models and Conditional Random Fields, respectively. Jin and Ho [11] learned a lexicalized HMM model to extract aspects and opinion expressions, while Li et al. [16] integrated two CRF variations, i.e., Skip-CRF and Tree-CRF. All these methods require sufficient labeled samples for training. However, it is time-consuming and labor-intensive to label samples. On the other hand, unsupervised methods have emerged recently. The most notable unsupervised approach was proposed by Hu and Liu [12]. They assumed that product aspects are nouns and noun phrases. The approach first extracts nouns and noun phrases as candidate aspects. The occurrence frequencies of the nouns and noun phrases are counted, and only the frequent ones are kept as aspects.

Subsequently, Popescu and Etzioni [28] developed the OPINE system, which extracts aspects based on the KnowItAll Web information extraction system [8]. Mei et al. [22] utilized a probabilistic topic model to capture the mixture of aspects and sentiments simultaneously. Su et al. [32] designed a mutual reinforcement strategy to simultaneously cluster product aspects and opinion words by iteratively fusing both content and sentiment link information. Recently, Wu et al. [37] utilized a phrase dependency parser to extract noun phrases from reviews as aspect candidates

EXISTING SYSTEM

Existing techniques for aspect identification include supervised and unsupervised methods. Supervised method learns an extraction model from a collection of labeled reviews. The extraction model, or called extractor, is used to identify aspects in new reviews. Most existing supervised methods are based on the sequential learning (or sequential labeling) technique. On the other hand, unsupervised methods have emerged recently. They assumed that product aspects are nouns and noun phrases. The approach first extracts nouns and noun phrases as candidate aspects. The occurrence frequencies of the nouns and noun phrases are counted, and only the frequent ones are kept as aspects.

DISADVANTAGES:

- The reviews are disorganized, leading to difficulties in information navigation and knowledge acquisition.

- The frequency-based solution is not able to identify the truly important aspects of products which may lead to decrease in efficiency of the review.

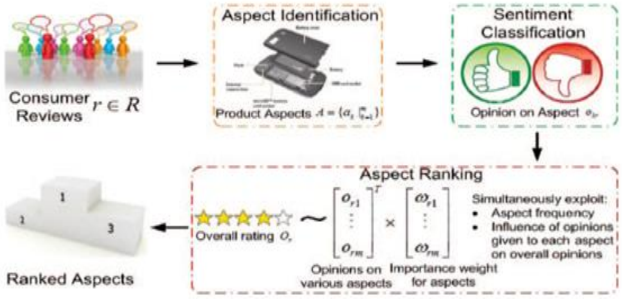
PROPOSED SYSTEM

- We propose a product aspect ranking framework to automatically identify the important aspects of products from numerous consumer reviews.
- We 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 aspect over their overall opinions on the product.

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

ADVANTAGES:

- Identifies important aspects based on the product, which increases the efficiency of the reviews.
- The proposed framework and its components are domain-independent


Consumer Reviews $r \in R$
Aspect Identification
Product Aspects $A = \{a_k\}_{k=1}^{m}$
Sentiment Classification
Opinion on Aspect o_{rk}
Aspect Ranking
Overall rating O_r
$\sim \begin{bmatrix} o_{r1} \\ \vdots \\ o_{rm} \end{bmatrix}^T \times \begin{bmatrix} \omega_{r1} \\ \vdots \\ \omega_{rm} \end{bmatrix}$
Opinions on various aspects
Importance weight for aspects
Simultaneously exploit:
• Aspect frequency
• Influence of opinions given to each aspect on overall opinions
1
2
3
Ranked Aspects

IMPLEMENTATION

Sentiment Classification on Product Aspects:-

The task of analyzing the sentiments expressed on aspects is called aspect-level sentiment classification in literature Exiting techniques include the supervised learning approaches and the lexicon-based approaches, which are typically unsupervised. The lexicon-based methods utilize a sentiment lexicon consisting of a list of sentiment words, phrases and idioms, to determine the sentiment orientation on each aspect. While these method are easily to implement, their performance relies heavily on the quality of the sentiment lexicon. On the other hand, the 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 example, Support Vector Machine (SVM), Naive Bayes, and Maximum Entropy (ME) model etc.. 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. We thus exploit *Pros* and *Cons* reviews to train a sentiment classifier, which is in turn used to determine consumer opinions (positive or negative) on the aspects in free text reviews. Specifically, we first collect the sentiment terms in *Pros* and *Cons* reviews based on the sentiment lexicon provided by MPQA project. These terms are used as features, and each review is represented as a feature vector.

Probabilistic Aspect Ranking Algorithm:-

In this section, a probabilistic aspect ranking algorithm to identify the important aspects of a product from consumer reviews. Generally, important aspects have the following characteristics:

(a) they are frequently commented in consumer reviews; and

(b) 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. That is, the opinions on (un)important aspects have strong (weak) impacts on the generation of overall opinion. To model such aggregation, we formulate that the overall rating Or in each review r is generated based on the weighted sum of the opinions on specific aspects.

Experimental Data and Settings:-

The details of our product review corpus, which is publicly available by request. This dataset contains consumer reviews on 21 popular products in eight domains. There are 94,560 reviews in total and around 4,503 reviews for each product on average. These reviews were crawled from multiple prevalent forum Websites, including cnet.com, viewpoints.com, reevoo.com, gsmarena.com and pricegrabber.com. The reviews were posted between June 2009 and July 2011. Eight annotators were invited to annotate the ground truth on these reviews.

Evaluations of Product Aspect Identification on Free Text Reviews:-

Our aspect identification approach with the following two methods: (a) the method proposed by Hu and Liu in [12], 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. in [37], that extracts noun phrases from a dependency parsing tree as aspect candidates, and identifies aspects by a language model built on the reviews.

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.

Extractive Review Summarization:-

As aforementioned, for a particular product, there is an abundance of consumer reviews available on the internet. However, the reviews are disorganized. It is impractical for user to grasp the overview of consumer reviews and opinions on various aspects of a product from such enormous reviews. On the other hand, the Internet provides more information than is needed. Hence, there is a compelling need for automatic review summarization, which aims to condense the source reviews into a shorter version preserving its information content and overall meaning. Existing review summarization methods can be classified into abstractive and extractive summarization. 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 etc. from the original reviews and concatenating them into shorter from.

CONCLUSION

The 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. 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. Moreover applied product aspect ranking to facilitate two real-world applications, i.e., document-level sentiment classification and extractive review summarization. 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.

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] C. C. Chang and C. J. Lin. (2004). Libsvm: A library for support vector machines [Online]. Available: http://www.csie.ntu.edu.tw/∼cjlin/libsvm/

[3] G. Carenini, R. T. Ng, and E. Zwart, “Multi-document summarization of evaluative text,” in Proc. ACL, Sydney, NSW, Australia, 2006, pp. 3–7.

[4] China Unicom 100 Customers iPhone User Feedback Report, 2009.

[5] ComScore Reports [Online]. Available: http://www.comscore.com/Press_events/Press_releases , 2011.

[6] X. Ding, B. Liu, and P. S. Yu, “A holistic lexicon-based approach to opinion mining,” in Proc. WSDM, New York, NY, USA, 2008, pp. 231–240.

[7] G. Erkan and D. R. Radev,“LexRank: Graph-based lexical centrality as salience in text summarization,” J. Artif. Intell. Res., vol. 22, no. 1, pp. 457–479, Jul. 2004.

[8] O. Etzioni et al., “Unsupervised named-entity extraction from the web: An experimental study,” J. Artif. Intell., vol. 165, no. 1, pp. 91–134. Jun. 2005.

[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] V. Gupta and G. S. Lehal, “A survey of text summarization extractive techniques,” J. Emerg. Technol. Web Intell., vol. 2, no. 3, pp. 258–268, 2010.

[11] W. Jin and H. H. Ho, “A novel lexicalized HMM-based learning framework for web opinion mining,” in Proc. 26th Annu. ICML, Montreal, QC, Canada, 2009, pp. 465–472.

[12] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in Proc. SIGKDD, Seattle, WA, USA, 2004, pp. 168–177.

[13] K. Jarvelin and J. Kekalainen, “Cumulated gain-based evaluation of IR techniques,” ACM Trans. Inform. Syst., vol. 20, no. 4, pp. 422–446, Oct. 2002.

[14] J. R. Jensen, “Thematic information extraction: Image classification,” in Introductory Digit. Image Process., pp. 236–238.

[15] K. Lerman, S. Blair-Goldensohn, and R. McDonald, “Sentiment summarization: Evaluating and learning user preferences,” in Proc. 12th Conf. EACL, Athens, Greece, 2009, pp. 514–522.

[16] F. Li et al., “Structure-aware review mining and summarization,” in Proc. 23rd Int. Conf. COLING, Beijing, China, 2010, pp. 653–661.

[17] C. Y. Lin, “ROUGE: A package for automatic evaluation of summaries,” in Proc. Workshop Text Summarization Branches Out, Barcelona, Spain, 2004, pp. 74–81.