ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

A Survey on Product Aspect Ranking Techniques

Ancy. J. S, Nisha. J.R

P.G. Scholar, Dept. of C.S.E., Marian Engineering College, Kerala University, Trivandrum, India.

Asst. Professor, Dept. of C.S.E., Marian Engineering College, Kerala University, Trivandrum, India.

ABSTRACT: Huge collections of consumer reviews are available on the Web expressing various opinions on multiple aspects of products .The important reviews are mostly not organized properly thereby creating problems in information navigation and knowledge acquisition. To address this problem, product aspect ranking framework is explored to automatically identify important product aspects or features from online consumer reviews.The important product aspects are identified based on two observations: i) the important aspects are usually commented on by a large number of consumers and ii) consumer opinions on the important aspects greatly influence their overall opinions on the product. The framework contains three main mechanisms, i.e., aspect sentiment classification, product aspect identification, and probabilistic aspect ranking algorithm. 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. The framework which aims at improving the usability of consumer reviews of a product. This paper provides the description of various techniques for product aspect identification and classification.

KEYWORDS: Product aspects, aspect ranking, aspect identification, sentiment classification, consumer review.

I. INTRODUCTION

Today internet is ruling the world. Everyone has using internet for everything. Increased use of e-commerce application consumer often prefer internet for buying the product. Not only trading the product they post their precious comments over the web. So the main drawback in the reviews is not organized properly.Huge collections of consumer reviews are available on the Web. These reviews have become an important resource for both consumers and firms. Consumers commonly seek quality information from online consumer reviews prior to purchasing a product, while many firms use online consumer reviews as an important resource in their product development, marketing and consumer relationship management.

Most online reviews express consumers overall opinion ratings on the product, and their opinions on multiple aspects of the product. While a product may have hundreds of aspects in which some aspects are more important than the others and have greater influence on consumers purchase decisions as well as firms product development strategies. Here, an aspect is also called feature.Which refers to a component or an attribute of a certain product. A sample review "The batterylife of Nokia Lumia 730 is amazing." reveals positive opinion on the aspect "battery life" of product Nokia Lumia 730. Besides the retail Websites, many forum Websites also provide a platform for consumers to post reviews on millions of products. For example, CNet.com involves more than seven million product reviews; whereas Pricegrabber.com contains millions of reviews on more than 32 million products in 20 distinct categories over 11,000 merchants. Such numerous consumer reviews contain rich and valuable knowledge and have become an important resource for both consumers and firms [2]. Consumers commonly seek quality information from online reviews prior to purchasing a product, while many firms use online reviews as important feedbacks in their product development, marketing, and consumer relationship management

Generally, a product may have hundreds of aspects. For example, iPhone 3GShas more than three hundred aspects (see Fig. 1), such as "usability," "design," "application," "3G network." In which some aspects are more important than the others and have greater impact on the eventual consumers decision making as well as firms' product development strategies. For example, some aspects of iPhone 3GS, e.g., "usability" and "battery," are concerned by most consumers,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

and are more important than the others such as “usb" and “button." Hence, identifying important product aspects will improve the usability of numerous reviews and is applicable to both consumers and firms. Consumers can conveniently make wise purchasing decision by paying more attentions to the important aspects, while firms can focus on improving the quality of these aspects and thus enhance product reputation effectively.
However, it is impractical for people to manually identify the important aspects of productsfrom numerous reviews. Therefore, an approach to automatically identify the important aspects is highly demanded.

In this paper we present the methodology in section II and techniques used for the product aspect identification and product aspect classification in the section no. III and section no. IV respectively and section V illustrates the product aspect ranking.

II. METHODOLOGY

A product aspect ranking framework to identify the important aspects of products from numerous consumer reviews. The framework consist of three main mechanisms, (Fig 1) i.e., product aspect identification,sentiment classification and product aspect ranking algorithm.

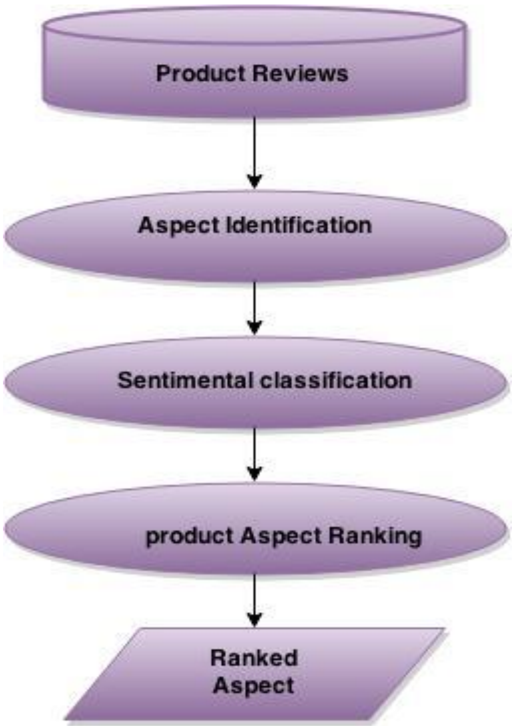

Fig 1: A product Aspect ranking Process

Reviews can be posted on the webs in three different types:
Type (1) - Pros and Cons: The reviewer is asked to describe Pros and Cons separately.
Type (2) - Pros, Cons and detailed review: The reviewer is asked to describe Pros and Cons separately and also write a detailed review.
Type (3) - free format: The reviewer can write freely, i.e., no separation of Pros and Cons.
Different types of reviews may need different techniques to perform the tasks such as product aspect identification, product sentiment classification and product aspect ranking as mentioned in Fig. 1

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

For type (1) and (2), opinion orientations are known because Pros and Cons are separated and thus there is no need to identify them. Only product features need to be identified from the comments of customers. For type (3), we need to identify both product features and opinion orientations.

First, we exploited the *Pros* and *Cons* reviews to increase aspect identification and sentiment classification on free-text reviews In particular, split the free text reviews into sentences, and parse each sentence using parser. After that the frequent noun phrases are extracted from the sentence parsing trees as candidate aspects. Since these candidate aspects may contain noises, further the *Pros* and *Cons* reviews are used to assist them in identification of aspects from the candidates. Then all the frequent noun terms extracted from the *Pros* and *Cons* reviews are collected to form a vocabulary. Each aspect in the *Pros* and *Cons* reviews is represented into a unigram feature, and all the aspects are then used to learn a one-class Support Vector Machine (SVM) classifier [3]. The resultant classifier is used to identify aspects in the candidates extracted from the free text reviews. This task of analysing the sentiments expressed on aspects is called aspect-level sentiment classification [4]. Many techniques are used for sentiment classification which includes the supervised learning approaches and unsupervised approaches such as the lexicon-based approaches. The lexicon-based method uses a sentiment lexicon which contains a list of sentiment words, phrases and idioms, to determine the sentiment orientation on each aspect [5]. On the other hand, the supervised learning methods train a sentiment classifier by using training dataset. The classifier is then used to predict the sentiment on each aspect. Finally a probabilistic aspect ranking algorithm is used to identify the important product aspects from reviews.

III. **ASPECTIDENTIFICATIONTECHNIQUES**

A. *Supervised learning:*

Supervised learning technique use the collection of labeled reviews to learn an extraction model. This extraction model called as extractor is then used for the identification of aspects in ne reviews. Most of the supervised learning techniques are based on the sequential learning. Various literatures show the different technique for the learning of extractor.

Wong and Lam [6] uses the HMM model and conditional random field to learn the extractor.
Li et al [9] uses skip CRF and tree CRF i.e. to integrated CRF variation to learn the extractor. The main disadvantage of this method is that it require Labelled sample for training. These methods are very time consuming to label the samples.

B. *Unsupervised learning:*

In this method the aspects are considered noun or noun phases and occurrence frequency of noun and noun phrases is calculated. The frequent noun or noun phrases are considered as aspects. Hu and Liu [4] use this unsupervised technique for aspect identification. Main disadvantage of this method is that identified aspects candidates may contain noise.
Wu et al [12] uses a phrase dependency parsing. Phrase dependency parsing takes the sentence as input and segments it into phrases. Then these segments are linked with directed arc. Phrase dependency parsing focuses on phrases and not on single word inside phrase. To make sure that identified aspects candidate is to be aspect language model which is based on product reviews is used to predict score of candidates. Model filter out low score candidates. Such model may be biased to frequent terms in the review and cannot sense precisely the related score of aspect terms as a result cannot filter out noise efficiently.
Popesu and Etrioni [10] developed their own systems for aspect identification. They developed OPINE system which is based on KnowIt. All web information extraction system which extract aspects from reviews. Su et al [11] design a reinforcement strategy. This strategy cluster together product aspect and opinion words by iteratively using both content and sentiments

IV. **SENTIMENT CLASSIFICATION TECHNIQUES**

A.*Lexicon Based Approach:*

Opinion words are used in many sentiment classification tasks. Desired states are express with Positive opinion words while, undesired states are expressed with negative opinion words. All opinion words, opinion phrases and idioms are together called as Lexicon.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Hu and Liu [4] use this method. They utilize synonym /antonym relations defined in WordNet to bootstrap the seed word set and finally obtain a sentiment lexicon. Fig 2 shows example of Bipolar adjective structure where synonym relationship is represented by arrow (→) and antonym relationship is represented by dash arrow (--->).

Three approaches used to collect this opinion word list are manual, dictionary based and corpus based approaches. Manual approach is very time consuming and it is not used alone. It is usually combined with the other two automated approaches to avoid mistakes that resulted from automated methods. The two automated approaches are presented in the following subsections.

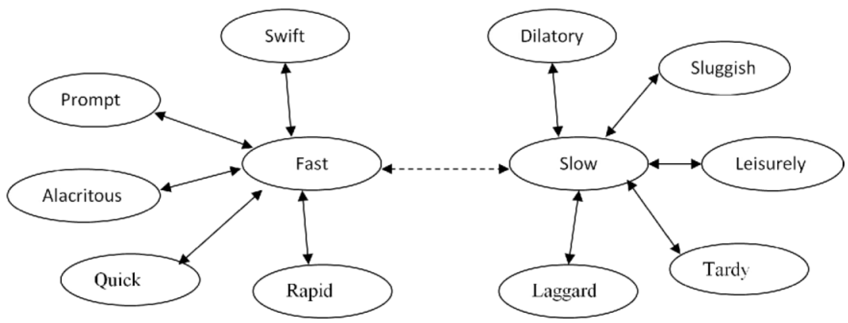

Fig. 2 Bipolar adjective structure,
(→ = similarity; ---> = antonymy)

a) Dictionary-based approach:
[8, 9]presented the main strategy of the dictionary-based approach. A small set of opinion words is collected manually with known orientations. Then, synonyms and antonyms of these words is added to this set which is grown by searching the words in the well-known corpora WordNet [7] or thesaurus [8]. The newly found words are added to the seed list then the next iteration starts. This iterative process stops when no new words are found. After completion of process list is checked manually to remove or correct errors.

This method is unable to find opinion words with domain and context specific orientation which is the main drawback of this method.

b) Corpus-based approach:
The drawback of dictionary based approach is overcome in Corpus- based approach which helps to solve the problem of finding opinion words with context specific orientations. Its methods depend on syntactic patterns.

B. *Holistic Lexicon Based Approach:*
Ding et al [13] presented a Holistic lexicon based method Holistic lexicon-based approachimproves the lexiconbased method in [14] by addressing two issues that the opinion of sentiment words would be content sensitive and conflict in the review.

This method does not look at the current sentence alone rather it uses the external information and evidences in other sentence and other reviews. Some linguistic conventions in natural language expression are used to find the orientation of opinion word. This method required prior domain knowledge or user inputs are needed. This approach is highly effective when sentence contain multiple contradictory opinion words.

C. *Supervised Learning Techniques:*
a) Naïve Bayes Classifier (NB):

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Naïve Bayesian networks are composed of acyclic graph with only one parent andseveral children. There is a very strong assumption of independence with child nodes in the context of their parents. Independence model can be represented with:

$$R = \frac{P(i/X)}{P(j/X)} = \frac{P(i)P(X/i)}{P(j)P(X/j)} = \frac{P(i)nP(X/i)}{P(j)nP(X/j)}$$

When these two probabilities are compared, larger probability is more likely to be the actual class label. Advantage of naive bayes classifier is its short computational time for learning the dataset. Bayes classifiers are usually less accurate than that of other learning algorithms.

b) Maximum Entropy Classifier (ME):
Another classifier is Maximum Entropy classifier. The name Maximum Entropy comes from the fact that the classifier finds the probabilistic model which is the simplest and least constrained. Yet it has some specific constraints. The idea behind maximum entropy is that one should prefer the most uniform models that also satisfy any given constraints. This Classifier is used to converts labeled feature sets to vectors using encoding. This encoded vector is then used to compute weights for each feature that can then be combined to determine the most likely label for a feature set. This classifier is controlled by a set of X{weights}, which is used to combine the joint features that are produced from a feature-set by an X{encoding}. Each C{(featureset, label)} pair is mapped to vector with encoding scheme. The probability of each label is calculated using the following equation:

$$P\left(\frac{\text{fs}}{\text{label}}\right) = \frac{\text{dotprod}(\text{weights}, \text{encode}(\text{fs}, \text{label}))}{\text{sum}(\text{dotprod}(\text{weights}, \text{encode}(\text{fs}, \text{l}))\text{forlinlabels})}$$

ME classifier was used by Kaufmann [15]to detect parallel sentences between any language pairs with small amount of training data. Other tools were developed to automatically extract parallel data from non-parallel corpora use language specific techniques or require large amounts of training data. Their results showed that ME classifiers can generate useful results for almost any language pair. This can allow the formation of parallel corpora for many new languages.

c) Support Vector Machines Classifiers (SVM):
Support Vector Machines (SVMs) are the newest supervised machine learning technique. SVM uses the notion of a "margin"- a hyperplane that divide two data classes.. An upper bound on the expected generalization error can be reduced by maximizing the margin and thereby largest possible distance between separating hyperplane instances on either side of it. In Fig. 3, X, O are 2 classes and A, B and C are the three hyperplanes. Hyperplane A provides the best separation between the classes, because the normal distance of any of the data points is the largest, so it represents the maximum margin of separation. In the case of linearly separable data, once the optimum separating hyperplane is found, data points that lie on its margin are known as support vector points whose solution is represented as a linear combination of only these points. Other data points are ignored. Therefore, the model complexity of an SVM is unaffected by the number of features encountered in the training data.

For this reason, SVMs are well suited for learning tasks where the number of features is large with respect to the number of training instances.

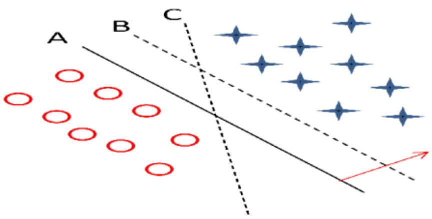

Fig. 3 Using support vector machine on a classification problem

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

V. PRODUCT ASPECT RANKING

After the aspect sentiment classification, classified aspects are then ranked.The probabilistic aspect ranking algorithm to infer the importance of various aspects of a product from numerous reviews. Important aspects have the following characteristics:

a) They are frequently commented in consumer reviews
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 overallopinion.The product aspects are finally ranked according to their importance scores. Zheng et al [1] described the probabilistic aspect ranking algorithm in their literature.

VI. CONCLUSION

This survey paper presented an overview on the product aspect ranking techniques to identify important aspects of products from numerous consumer reviews. Product aspect ranking process contains three main steps i.e. product aspect identification, aspect sentiment classification and aspect ranking. This survey illustrates various product aspect ranking techniques.

REFERENCES

1. Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, and Tat-Seng Chua," Product Aspect Ranking and Its Applications", IEEE transactions on knowledge ans data engineering, Vol. 26, No.5, May 2014.
2. 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.
3. .L. M. Manevitz and M. Yousef, "One-class SVMs for document classification," J. Mach. Learn., vol. 2, pp. 139–154, Dec. 2011.
4. M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. SIGKDD, Seattle, WA, USA, 2004, pp. 168–177.
5. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," in Proc. IT&T Conf., Dublin, Ireland, 2009.
6. T. L. Wong and W. Lam, "Hot item mining and summarization from multiple auction web sites," in Proc. 5th IEEE ICDM, Washington, DC, USA, 2005, pp. 797–800.
7. Miller G, Beckwith R, Fellbaum C, Gross D, Miller K. WordNet: an on-line lexical database. Oxford Univ. Press; 1990.
8. Mohammad S, Dunne C, Dorr B. Generating high-coverage semantic orientation lexicons from overly marked words and a thesaurus. In:
9. Proceedings of the conference on Empirical Methods in Natural Language Processing(EMNLP'09); 2009.
10. F. Li et al., "Structure-aware review mining and summarization,"in Proc. 23rd Int. Conf. COLING, Beijing, China, 2010, pp. 653–661.
11. A. M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Proc. HLT/EMNLP, Vancouver, BC,Canada, 2005, pp. 339–346.
12. Q. Su et al., "Hidden sentiment association in chinese web opinion mining," in Proc. 17th Int. Conf. WWW, Beijing, China,2008, pp. 959–968.
13. Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proc. ACL, Singapore, 2009, pp. 1533–1541.
14. 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.
15. A. Liu, Sentiment Analysis and Opinion Mining. Mogarn& Claypool Publishers, San Rafael, CA, USA, 2012.
16. Kaufmann JM. JMaxAlign: A Maximum Entropy Parallel Sentence Alignment Tool. In: Proceedings of COLING'12: Demonstration Papers, Mumbai; 2012. p. 277–88