

Product Aspect Ranking using Sentiment Analysis: A Survey

Chetan Mate

¹Department of Computer Engineering, MIT College of Engineering, University of Pune, India

Abstract: *In today's world, internet is the main origin of information. There are many e-commerce websites available where people discuss on different issues of product. All e-commerce website provide facility to the consumer to give opinion about their product and services. Consumer reviews contain rich and valuable knowledge for both firms and users. The problem with this information is that these reviews are mostly unorganized therefore creating difficulty for information transfer knowledge accession. We propose a product aspect ranking framework, which automatically determine the important aspects of products from online consumer reviews, aiming at improving the usability of the countless reviews. The important aspects are identified by two observations, a) The important aspects of a product are usually given by a large number of consumers; b) And consumers' opinions on the important aspects influence their overall opinions on the product. However Identifying important product aspects will increase the usability of numerous reviews and is useful to both consumers and firms. It is impractical for people to manually identify the important aspects of products from numerous reviews. Consumers can conveniently make purchasing decision by paying more attention to the important aspects, while enterprise can focus on improving the quality of these aspects and thus enhance product ranking effectively.*

Keywords: *e-commerce, aspect ranking, product aspect, sentiment analysis*

1. INTRODUCTION

Recent years have witnessed that in today's modern day life internet and web application are playing very important role. Web application provides simple way for people to do several task of daily life like browsing, doing online transaction and purchasing product. Due to the evolution and advancement of information technology enabled services many e-commerce websites are available, hence result into more and more number of products are sold on the internet. Thousands of products from various dealers have been offered online. For example, Bing Shopping has indexed more than five million products. Amazon.com registers a total of more than 36 million products. Shopper.com records more than five million products from over 3,000 dealers [1]. Thus this affects more and more number of people for purchasing products online. Over the recent year there is quick growth and emergence of e-commerce technology, inspire customer to buy product online and express opinions on all kind of objects such as product and services. In order to feel customer more pleasant and more secure about online shopping, it has become a conventional practice for online dealers to enable their customers to write reviews on product aspects that they have purchased. Thus people not only buy product online shopping but also they give review about product aspect. As a result, the number of reviews that a product receives grows rapidly. Here, an *aspect*, also called *feature* in literatures, refers to a component or an attribute of a certain product. A sample review "The battery backup of Sony Xperia C mobile phone is amazing." reveals positive opinion on the aspect "battery backup" of product Sony Xperia C mobile phone.

These customer reviews are very important for social influence as well as an economic effect.

These reviews are useful to users for making intelligent decisions about product purchasing and also helpful to merchants for knowing their product's positive and negative attribute. In particular, customer express sentiments (opinion) on various aspects of products. A sentiment can be defined as opinion expressed by consumer. Sentiments represent any viewpoints of consumer such as like or desirable (positive), dislike or undesirable (negative) and may be neutral viewpoint. Sentiment classification can be done at three levels of extraction which are Document level, Sentence level and Aspect level [9]. Such numerous consumer reviews contain rich and precious knowledge and have become an important source for both consumers and firms [1]. Consumers commonly find quality information from online reviews prior to purchasing a product, while many firms use online reviews as important feedbacks in their product evolution, marketing, and consumer relationship management. We argue that 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. Hence, determining important product aspects will improve the usability of countless reviews and is useful to both consumers and firms. Consumers can conveniently make 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. We propose a product aspect ranking framework to automatically identify the important aspects of products from online consumer reviews.

In our proposed work, based on consumer reviews of product first we recognize the important features of product. Then we classify the sentiment on that aspect. And then we develop the aspect ranking algorithm for providing the rating to particular product.

2. LITERATURE SURVEY

In this section, we review existing works related to the Proposed framework. We start with the works on the aspect identification from the consumer review. Existing methods for aspect identification based on supervised learning approaches and the

lexicon-based approaches. 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 [14]. The supervised learning methods learn a sentiment classifier based on training dataset. 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. [7]. Supervised learning is based on the training data and cannot perform well without sufficient training samples. However, labeling training data is tedious and time-consuming. In this work, the *Pros* and *Cons* reviews have explicitly categorized positive and negative opinions on the aspects. Most existing supervised methods are based on the sequential learning technique. For example, Wong and Lam [7] learned aspect extractors using *Hidden Markov 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.* [9] integrated two CRF variations, i.e., Skip-CRF and Tree-CRF. On the other side, 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. The next task is aspect sentiment classification, which determines the orientation of sentiment expressed on each aspect. Two major approaches for aspect sentiment classification include lexicon-based and supervised learning approaches. The lexicon-based methods are typically unsupervised. They rely on a sentiment lexicon containing a list of positive and negative sentiment words. To generate a high-quality lexicon, the bootstrapping strategy is usually employed. For example, Hu and Liu [12] started with a set of adjective seed words for each opinion class (i.e., positive or negative). They utilized synonym/antonym relations defined in *WordNet* to bootstrap the seed word set, and finally obtained a sentiment lexicon.

Document-level sentiment classification aims to classify an opinion document as expressing a positive or negative opinion. Existing works use unsupervised, supervised or semi-supervised learning techniques to build document level sentiment classifiers. Unsupervised method usually relies on a sentiment lexicon containing a collection of positive and negative sentiment words. It determines the overall opinion of a review document based on the number of positive and negative terms in the review. Supervised method applies existing supervised learning models, such as SVM and Maximum entropy (ME) etc. [5], while semi-supervised approach exploits abundant unlabeled reviews together with labeled reviews to improve classification performance. The two widely used methods are the sentence ranking and graph-based methods [10]. In these works, a scoring function was first defined to compute the informativeness of each sentence. Sentence ranking method [9] ranked the sentences according to their informativeness scores and then selected the top ranked sentences to form a summary. Graph-based method [7] represented the sentences in a graph, where each node corresponds to a sentence and each edge characterizes the relation between two sentences. A random walk was then performed over the graph to discover the most informative sentences, which were in turn used to compose a summary.

3. PROPOSED WORK

Following four steps are used in proposed work.

- 1) Reviews extraction and Preprocessing.
- 2) Aspect Identification of the product
- 3) Classify the positive and negative reviews of product by sentiment classifier.
- 4) The probabilistic ranking algorithm used for product ranking.

3.1 Reviews Extraction and Preprocessing

First step of product aspect ranking framework is data preprocessing which is a very important step. Compared to formal text document the reviews are generally less formal and written in particular manner. If the sentiment analysis applied on unstructured review generally achieve very poor performance in most case. Therefore the preprocessing techniques on reviews are important

for obtaining acceptable result on sentiment analysis. Following are the methods of preprocessing.

- 1) Stemming: In stemming we will remove the postfix from each word such as ingtion etc. Eg. Buying will become buy after stemming.
- 2) In tokenization we will tokenize each sentence by space. Means we will remove the spaces. Stop word removal like a, an, the etc.

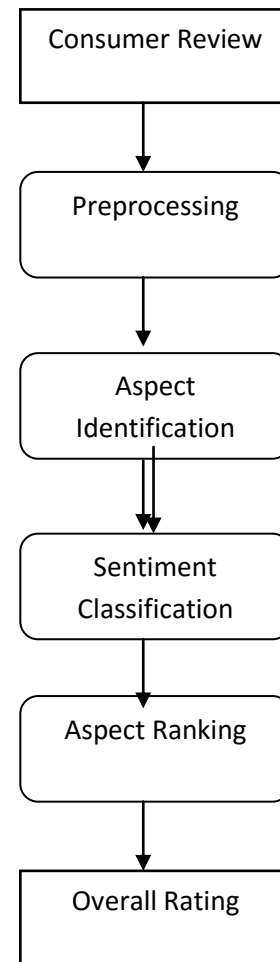


Figure 3.1: Architecture Diagram

3.2 Aspect Identification of the product

Aspect identification is second important step in which we identify aspect from countless consumer reviews. The reviews are available on different ecommerce websites. But the problem with these reviews is that they are composed in different formats on various websites. Consumer review consists of positive and negative reviews. On

some website there is a overall grading on the product, some website the reviews are in paragraph in formal text. e.g. CNet.com, Viewpoints.com, Reevo.com are the various websites for reviews of different product and has different formats. From these reviews the aspect are determine as a frequent Noun term and noun phrases. We can get precise aspect by extracting persistent noun from the positive and negative reviews[3]. Hu and Liu proposed most remarkable approach for aspect identification. In this approach it first identifies the noun and noun phrases in the document. The occurrence frequency of noun and noun phrase are counted then only the consistent noun terms are kept as the aspect [4]. Phrase dependency parser used to extract the noun phrase by Wu et al. Some aspect may contain synonym term such as "headphone" and "earphone." In such situation we perform synonym clustering to obtain unique aspect. These synonym terms are obtained from synonym dictionary Websites.

3.3. Sentiment Classifier

Sentiment analysis or Opinion mining is a type of natural language processing used for tracking the mood or polarity of people about product aspect. Sentiment classification used to classify the given text to one or more predefined sentiment categories. Such as Positive, Negative, Neutral. There is various classification techniques are available. There are two types of learning supervised learning and another is unsupervised learning. The lexicon-based approaches are mostly unsupervised. The lexicon based methods make use of a sentiment lexicon consist of list of sentiment words, phrases and idioms, to obtain sentiment polarity on each aspect[9]. the performance of supervised learning dependent on training data. It not able performs well without sufficient data. Supervised learning method train a sentiment classifier based on training dataset. The classifier is used to determine the sentiment on each aspect. There are many learning based classification models are available. Support Vector Machine (SVM), Naive Bayes, and Maximum Entropy (ME) model these are the learning based classification model[10]. Among all these learning based classification methods Support Vector Machine (SVM) having significant accuracy.

3.4 Aspect Ranking Algorithm

In proposed framework important aspects are commented frequently in consumer review and the consumer's opinions on the important aspects are greatly affects 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 sentiment on (un)important aspects have strong (weak) impacts on the generation of overall opinion. In proposed framework we use AFINN dictionary approach. AFINN is a list of English words rated valence with an integer between minus five (negative) and plus five (positive).

4. APPLICATIONS

Aspect ranking framework is useful to a huge number of real world applications. Businesses and organizations are interested in opinions for use them for new product perception, brand perception, reputation management, product and service benchmarking, market intelligence. Consumers are interested in other's opinions when purchasing a product, using a service, tracking political topics etc.

CONCLUSION

In this paper, we have surveyed the reference papers related to Aspect identification, Sentiment classification. We have planned to identify the important aspects of a product from online consumer reviews. Our assumption is that the important aspects of a product should be the aspects that are frequently commented by consumers and consumers' opinions on the important aspects greatly pressure their overall opinions on the product. Based on this assumption, we will try to develop an aspect ranking algorithm which will identify the important aspects by concurrently considering the aspect frequency and the pressure of consumers' opinions given to each aspect on their overall opinions.

REFERENCES

- [1] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, Tat-Seng Chua, "Product Aspect Ranking and Its Applications".IEEE2014.
- [2] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing*, New York, NY, USA: Marcel Dekker, Inc., 2009.
- [3] 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.
- [4] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. SIGKDD*, Seattle, WA, USA, 2004, pp. 168-177.
- [5] Finn, A., Kushmerick, N., and Smyth, B. 2002. Genre Classification and Domain Transfer for Information Filtering. In *Proc. of European Colloquium on Information Retrieval Research*, pages 353-362.
- [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] 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.
- [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] Bing Liu, "Sentiment Analysis and Opinion Mining" pp.7-140,2012.
- [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] B. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," in *Proc. IT&T Conf.*, Dublin, Ireland, 2009.
- [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.