PRODUCT ASPECT RANKING USING DOMAIN DEPENDENT AND DOMAIN INDEPENDENT REVIEW

Priti Sole[1], M.C.Kshirsagar[2]

[1]*VACOE, Ahmednagar, Pune University*
[2]*VACOE, Ahmednagar, Pune University*

Abstract

In today's world, internet is the main source of information. There are many blogs and forum sites available where people discuss on different issues and also almost all ecommerce website provide facility to the users to express opinion about their product and services which is important information available on the internet .The problem with this information is that this reviews are mostly not organized therefore creating difficulty for knowledge acquisition. There are many solution exist to resolve this problem but the available existing methods depends on extracting product aspect only considering single domain relevant review corpus. To address this problem, a method is explored to identify product aspect from online review is by taking into account the difference in aspect statistical characteristic across different corpus. This paper shows need of automatically identifying important product aspects from available online customer review and an approach of aspect ranking. This paper also shows the related work on this domain. Our methodology confirmed product aspect which are less nonspecific in domain independent corpus and more domain specific. Then customer opinion expressed on these aspects is determined using sentiment classifier and finally ranking of product aspect is calculated using it's ranking relevance score of each aspect .

Keywords— *Product aspect, aspect ranking, sentiment classification, customer review, opinion mining, aspect identification, product ranking.*

--***--

1. INTRODUCTION

In today's modern day life internet and web application are playing very important role. Web application provides easiest way for people to do several activities of daily life like browsing, doing online transaction and purchasing product. Due to the development and advancement of information technology enabled services a lot of e-commerce websites are available, hence result into more and more number of products are sold on the web. Thus this influence more and more number of people for buying products online. Over the last year there is rapid growth and emergence of e-commerce technology, motivate customer to buy product online and express opinions on all kind of objects such as product and services. In order to feel customer more comfortable and more secure about online shopping, it has become a common practice for online merchants to enable their customers to write reviews on products that they have purchased. Thus people not only buy product online but also they give feedback about product. As a result, the number of reviews that a product receives grows rapidly For example of customer review is: **"The picture quality of Nikon D5000 Digital camera is good".**

These customer review has social impact as well as an economic impact. The customer reviews in shopping web sites are very much helpful for product advertisement in which satisfied customers post their comment to know other user what they like about the product and how much. These reviews are useful to other users for making informed decisions about product purchasing and also helpful to merchants for knowing their product's positive and negative attribute.

Customer write review on various aspects of product. In this paper the word aspect is used to represent any particular part ,attribute or feature of product. In particular, customer express sentiments on various aspects of products A sentiment can be defined as opinion expressed by reviewer. In above example reviewer gives positive opinion on specific aspect picture quality of product digital camera .Sentiments represent any viewpoints of customer such as like or desirable (positive),dislike or undesirable(negative) and may be neutral viewpoint. For instance, sentiment classification looks for emotional expression such as good, bad excellent ,poor etc. Sentiment classification can be done at three levels of extraction which are Document level, Sentence level, Aspect level [9]. This mechanism mainly focus on splitting review document at above specified levels to determine expressed opinion whether it is positive ,negative or neutral. The task of summaries is clearly different from traditional text summarization [1] because it does not summarize the reviews by selecting or rewriting a subset of original sentences from the reviews.

The fig.1 shows example of customer review on cell phone LG G3. This review shows feedback on multiple aspects of the cell phone LG such as display, camera and processor. It means that product may have number of aspects. Figure shows that the reviewer gives overall rating only for cell phone LG without giving any individual rating on each aspect. Hence user would not be able to draw any conclusion about the reviewer's opinion on each aspect. Although different reviewer may give a cell phone LG the same overall rating, its reason of same overall rating may vary according to different reviewer. For example, one reviewer may like display quality of cell phone ,but other user may like the quality of processor because may be he\she is interested in processing power of cell phone. And also some aspects has more importance as compared to other aspects. For instance, for camera product aspect such as "lenses" and "picture quality " are concerned by most customers and are more important than the other aspects such as "a/v cable " and "wrist strap". Therefore it is necessary to tell other users different reviewers individual rating on each important aspect because customer opinion on an important aspects of product influence his or her overall opinion/rating of product. Yet customer provide individual rating on each aspect it is difficult for user to manually determine rating on important aspect because of rapidly increasing large volume of online review.

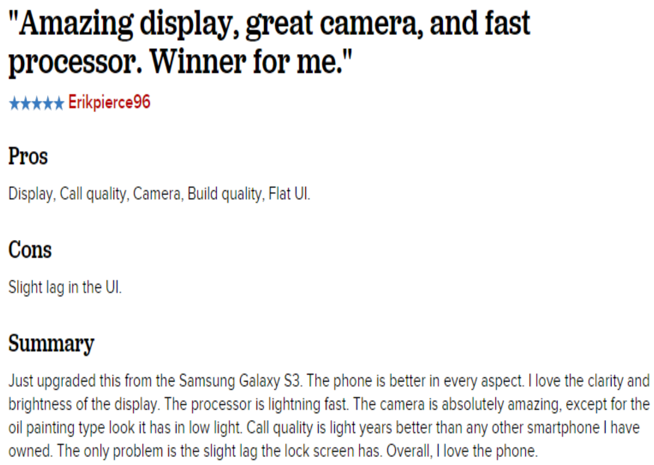

Fig.1. Example of customer review on Cell phone LG G3

Straight we can say that frequently commented aspects are more important but this is not the true. Frequency based solution is not able to determine truly important aspects because sometimes customer opinion on the frequently commented aspect may not affect his/her overall opinion. For example , reviewer frequently criticize on "bad exterior " of HTC but still give high overall rating .So, we can not say that highly commented aspect in review is truly important .We need another kind of solution to find out important aspect. This is the motivation of this paper. This motivate us to present a new approach to automatically determine the important product aspects and rank aspect according to it's importance score so that user can easily understand what are important aspect of product.

2. RELATED WORK

In this section, we present some of the existing research and related work in the aspect based opinion mining :

Existing product aspect extraction approaches can be divided into two categories, called, supervised and unsupervised. Supervised method requires set of labeled reviews as training example. A supervised learning method is then applied to make an extraction model and then it is capable of identifying product aspect. Various technique such as Hidden Markov Model and Conditional Random Field[3,4],Maximum entropy[2] are used for this task. However, supervised methods are time consuming because it requires time for preparing training examples.

M. Hu and B. Liu[5]2004 proposed unsupervised method. In this paper they showed that their proposed summarization task is different from traditional method because they determine only those aspects of product on which opinion is expressed. They consider that product aspect are generally noun or noun phrases. NLProcessor linguistic parser is used to do part of speech tagging to determine syntactic structure of sentence that determines whether a word is noun,verb,adjective etc..Thus they identified noun or noun phrase which identified as aspect and those aspects which are frequently commented by user finally determined.

Y. Wu, Q. Zhang, X. Huang, and L. Wu[7] 2009 Explored phrase dependency parser to extract noun phrases from review as candidate aspect They observed that a lot of product aspects are phrases.Firltly they identify dependency grammar structure of sentence determining relation between head and its dependent. Then phrase dependency tree is constructed from the result obtained in first stage. Finally candidate product aspect are identified opinion expressed on this aspect is also determined.

Sentiment classification task focus on determining semantic orientation on each aspect ie.positive,negative or neutral. Sentiment analysis conducted at one of the three level :Document level, Sentence level and aspect level. Sentiment classification has mainly two approaches that are lexicon based and supervised learning . Lexicon based methods are unsupervised and they depend on sentiment lexicon containing desirable and undesirable words .In contrast supervised method determine the opinion on aspects by using sentiment classifier.

B.Pang,L.Lee,and S.Vaithyanathan [4] 2002 proposed three machine learning methods, naïve Bayes,maximum entropy and support vector machine to classify whole movie reviews into positive or negative opinion. They conclude that standard machine learning methods produced good result as compared to human-generated baseline.They also showed that naïve bayes gives worst result whereas support vector machine gives best result. In most of the comparative studies it is found that support vector machine outperforms other machine learning methods in sentiment classification.

P.D.Turney[8]2002 proposed unsupervised method to classify review documents as recommended(positive) and not recommended (negative) in. In this paper Pointwise Mutual Information (PMI) and Information Retrieval (IR) algorithm is used to measure semantic orientation of word. A semantic numerical score of word is calculated by considering mutual information between the given word and already defined positive word and subtracting mutual information between given word and predefined negative word. Depending on average semantic numerical score the review document get final remark as either positive or negative. But the disadvantage of this method lies in time required to execute queries and accuracy in another application.

F.Li[6] 2010 in this paper, they focus on object feature based review summarization. In this paper they proposed a new methlodology based on Conditional Random Fields.

T. Wilson, J. Wiebe, and P. Hoffmann[13]2005 presented an approach to predicting contextual sentiments at the phrase level by applying machine learning techniques on a variety of feature factors. First they determine whether opinion expression is neutral or not. Then they distinguish sentiment polarity into positive, negative or neutral opinion at phrase level.

Yongyong Zhail, Yanxiang Chenl, Xuegang [14] (2010) attempted to create a novel framework for sentiment classifier learning from unlabeled documents. The process begins with a collection of un-annotated text and a sentiment lexicon. An initial classifier is trained by incorporating prior information from the sentiment lexicon which consists of a list of words marked with their respective polarity. The labeled features use them directly to constrain model‟s predictions on unlabeled instances using generalized expectation criteria. The initially-trained classifier using generalized expectation is then applied on the un-annotated text and the documents labeled with high confidence are fed into the self-learned features extractor to acquire domain-dependent features automatically. Such self-learned features are subsequently used to train another classifier which is then applied on the test set to obtain the final results.

H. Wang, Y. Lu, and C. X. Zhai.[12] 2010 developed a latent rating regression analysis model.Advanatge of this method is that we are able to find out latent rating of each aspect from given text review and overall rating of product. First they find out major aspects by using bootstrapping based algorithm. They assume that overall rating is weighted aggregation of underlying rating on each aspect and it’s weight. Weight is nothing but the importance placed by customer on each aspect. Then latent regression analysis model is used to find out individual reviewer’s underlying ranking on each major aspect and the relative important weight on different aspects. Limitation of this model is that they concentrate on reviewer rating behavior analysis rather than on aspect ranking.

3. WORKING OF SCHEME

Customer review are in the form of unstructured text format. Customer review contains different kind of opinion on different aspects of products. Opinion mining is used to find out aspects from given review corpus and deriving opinion on it. In this paper we present a method which automatically identify important aspects. There are many techniques exist for opinion based mining. But each technique has it’s own consequences as shown in related work. Generally existing techniques depends on mining patterns only from single review corpus. The framework contains three main mechanism ie. Product aspect identification, aspect sentiment classification and product aspect ranking.

For the task of aspect extraction unsupervised natural language processing method is used. This method defines domain independent specific template. We will not only focus on domain dependent corpus but also domain independent corpus. A set of candidate aspects are extracted. Then for each aspect we find out it’s DDR(Domain Dependent Relevance)and DIR(Domain Independent Relevance) score is calculated.DDR represent statistical relation of the candidate aspect to given review corpus and on the other hand DIR represent statistical relation of the candidate aspect to the domain irrelevant corpus. Aspects that has less DIR score means less non specific in given domain independent corpus and more DDR score means more domain specific are confirmed as candidate aspect.

Then next task is to determine semantic expressed on extracted candidate aspect. Sentiment classifier is used to perform this task which classifies semantic orientation on each aspect.

Then probabilistic ranking algorithm is used to find out the ranking score of various aspect of product from numerous review. The algorithm consider aspect frequency and take into account relation between the overall opinion and the opinions on specific aspects. The opinions on important aspects have strong impacts on the generation of overall opinion and on the other hand opinions on unimportant aspects have weak impacts on the generation of overall opinion. By taking into consideration above fact ranking score of each aspect is calculated and then product aspect are finally ranked according to it.

4. CONCLUSION

In this paper ,we have shown the need of determining important product aspect from available online numerous reviews and an approach of product aspect ranking. Also we have shown related work and existing approaches to this domain. In this paper, a method for identifying product aspects from customer reviews has been presented. First of all, the candidate product aspects are identified which are specific to the given review domain and yet not general(domain independent). Customer opinion on these aspects are determined using sentiment classifier. Finally, the identified product aspects are ranked according to their relevance score.

ACKNOWLEDGMENTS

This paper work is completed successfully only because support from each and every one including teachers, colleague, parents and friends. Especially, I am very thankful to those who provide me guidance and make this work reachable. This paper work is supported by my senior, my teachers and some experienced personalities.My acknowledgment of gratitude toward the everyone who make this work reachable.

REFERENCES

[1] Aurangzeb Khan,Baharum Baharudin and Khairullah Khan, "Sentiment Classification Using Sentence-level Lexical Based Semantic Orientation of Online Reviews." Trends in Applied Sciences Research, *6:* 1141-1157, 2011.

[2] blogs," in International Conference on Weblogs and Social Media (ICWSM), pp. 219-222.

[3] M. Hu and B. Liu, "Mining and summarizing customer reviews," presented at the Proceedings of the tenth ACM

[4] B.Pang,L.Lee,and S.Vaithyanathan," Thumbs up? Sentiment classification using machine learning techniques,"in Proc EMNLP, Philadelphia, PA, USA, pp. 79–86,2002.

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

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

[7] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proc. ACL, Singapore, pp. 1533–1541,2009.

[8] P.D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orienta- tion Applied to Unsupervised Classification of Reviews," Proc. 40th Ann. Meeting on Assoc. for Computational Linguistics, pp. 417-424, 2002.

[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] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," J. Machine Learning Research, vol. 3, pp. 993-1022, Mar. 2003.

[12] H. Wang, Y. Lu, and C. X. Zhai, "Latent aspect rating analysis on review text data: A rating regression approach," in Proc. 16th ACM SIGKDD, San Diego, CA, USA, pp. 168–176,2010

[13] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347-354, 2005.

[14] Yongyong Zhail, Yanxiang Chenl, Xuegang Hu, "Extracting Opinion Features in Sentiment Patterns" , International Conference on Information, Networking and Automation (ICINA),2010 .