RESEARCH ARTICLE OPEN ACCESS

Ranking of Products on the Basis of Aspects- A Probabilistic Approach

Neha M. Toshniwal [1], Prof. D.V.Gore [2]
Department of Computer Science and Engineering
PESMCOE
Pune - India

ABSTRACT

As the E-commerce is expanding, the online purchasing of the products has also increased. The reviews for the purchased product are given by the customers and there are large numbers of reviews, of which the important aspects of the products are identified by using the product aspect ranking framework. Firstly we identify the important aspects of products which are assumed to be frequently commented in the reviews, and then the sentiment classification for the aspect is done by using the sentiment classifier. Finally, the aspects are ranked by using the probabilistic aspect ranking algorithm.

***Keywords:* -** E-commerce, aspect ranking, product aspect, sentiment classification, consumer reviews..

I. INTRODUCTION

Recent years have witnessed the rapidly expanding e-commerce. For example, Bing Shopping has indexed more than five million products. Amazon.com archives a total of more than 36 million products. Shopper.com records more than five million products from over 3,000 merchants. Most retail Websites encourages consumers to write reviews to express their opinions on various aspects of the products.

Consumer reviews contain rich and valuable knowledge for both firms and users. However, the reviews are often disorganized, leading to difficulties in information navigation and knowledge acquisition.

This article proposes a product aspect ranking framework, which automatically identifies the important aspects of products from online consumer reviews, aiming at improving the usability of the numerous reviews.

In particular, given the consumer reviews of a product, we first identify product aspects by a shallow dependency parser and determine consumer opinions on these aspects via a sentiment classifier. 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 presents the ranking of the aspects of the products. The important aspects are identified from the customer reviews, which are commented frequently in the reviews.

For ranking the aspects of the products, the important aspect identification is done on the customer online reviews which are frequently commented in the reviews. In this paper we use the new approach of assigning the weight to the important aspects and the association between the weight of the overall opinion and the opinion of the aspect by using probabilistic ranking algorithm. Ranking is done on the basis of the importance of the aspect.

II. RELATED WORK

This section explains the work done on the product aspect ranking method, where different methods for the sentiment classification are present. There are various machine learning approaches in which mostly used are the supervised classification methods.

In machine learning technique the training dataset is collected first and classifier is trained on the training data and the feature selection is done on the basis of the term present and the term frequency, once the classification technique is used. Existing aspect identification methods were classified into supervised and unsupervised methods. Supervised learning methods gives a training dataset that is the model should be learned, called as aspect extractor which is used for aspect identification.

The Hidden markov model, Maximum entropy [1], class association rules and Nave Bayes (NB)[2] were used for aspect identification.

Wang and Lam[3] used supervised learning technique were hidden markov model and conditional random field were used as extractor and was effective but the preparation of training data set is time consuming.

The next step after aspect identification is sentiment classification in which the sentiments of the aspects are identified. There are various methods for sentiment classification; the mostly used methods are supervised methods, in which the training dataset is used to train the classifier.

Zha et.al. [5] used SVM as a classifier for sentiment classification, classified the positive and negative aspects on the basis of the sentiments and then compared SVM ,NB and ME and found that SVM performance is better than other. SVM has disadvantage, as its performance reduces for the small dataset.

M. Govindarajan [6] proposed a new approach, which is the ensemble of NB and SVM classifier known as Hybrid Approach, which efficiently integrates the advantages of NB and SVM; the hybrid approach gives higher accuracy than NB and SVM.

The hybrid approach proposed was improved by Nugen et.al. [7] where a two stage system was developed with reject option. Here the document level sentiment classification was done in which the documents was given as input to NB and the documents rejected by first stage were given has input to second stage were SVM classifier was used to classify the documents. A Two stage classifier gives better result with reject option.

The bigram feature of word gives gain in sentiment analysis was shown by Wang and Manning [8], presented a novel approach were SVM classifier was built over NB log count ratio as features, were NBSVM with bigram gives better result.

Ranking of aspects is the next step to sentiment classification, which is used to rank individual aspects according to their importance.

Wang et.al.[10] used latent aspect rating analysis model which reviews the opinion on each aspect and the effect on the different aspects. It focused on opinion level of aspects and customer rating analysis, instead of aspect ranking.
Synder and Barzilay [11] proposed a multiple aspect ranking method in which each aspect was ranked individually.

III. SYSTEM ARCHITECTURE

The Proposed system takes the consumer reviews as input and performs various steps on these reviews to rank the aspects as output.
The proposed product aspect ranking framework, firstly pre-processes the dataset which are consumer reviews of products, and sentiment classification is done and finally applies the aspect ranking algorithm to rank the aspects.

The following fig 1.shows the flow of the system.

A. *Aspect Identification*

Aspect identification is the important and difficult phase in product aspect ranking.

For identification of aspects the online consumer reviews are taken as input. Consumer reviews consists of positive and negative reviews. Some websites have the overall rating of the product, while some consist of reviews in paragraph in free text form. CNet.com, Viewpoint.com, Revoo.com are some websites for reviews and in different format.

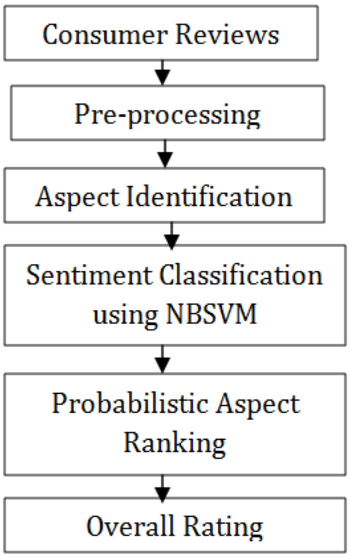

Fig :1 Product Aspect Ranking Framework.

Then parsing is done by using the Stanford parser. The parser generates the noun phrase candidates. We collect the noun terms from the parsing output of the pros and cons reviews and free text reviews.

As the identified aspects consists of some synonym terms, so the synonym clustering is done by using synonym dictionary and the aspects are identified uniquely.

B. *Sentiment Classification*

The task of identifying the sentiments which are expressed on aspects is called as aspect level sentiment classification.

The reviews are classified as positive or negative reviews based on polarity of the aspects in the reviews.

Here for sentiment classification the method is the combination of two methods NB (Naive Bayes) and SVM (Support Vector Machine) is been used.

The SVM sentiment classifier is used for sentiment classification, which is good for longer datasets, The Lib linear SVM is used for classification which is built over NB Log-count ratios as feature values to which the pros and cons reviews are given as the training samples, which are used to determine the customer opinion on the aspects in free text reviews.

c. ***Product Aspect Ranking***

In ranking algorithm we identify the important aspects of the products from the consumer reviews. The ranking is done by using the Probabilistic Aspect Ranking Algorithm.

The consumer opinion of the product aspect has a great influence on the overall opinion of the product. The important aspects are frequently commented by the consumer in the reviews.

Let S be the Overall system which consists of-

- S $=\{\{R\},\{A\},\{S\},\{O_r\}\}$
- R $=\{r_1, r_2 \ldots r_m\}$ = Set of consumer reviews.
  R $=\{a_k\}$ (k=1 to m aspects).
- A$=\{a_1, a_2 \ldots a_m\}$ = set of m aspects in the review.
- S $=\{P, N\}$ = Sentiment classification of aspects.
- P $=\{P_1, P_2, \ldots P_m\}$ = Pros or Positive aspects.
- N $=\{N_1, N_2, \ldots N_m\}$ = Cons or Negative aspects.
- O_{rk} = Opinion of aspect k.
- $w_r=\{w_1, w_2, \ldots w_m\}$= weight of the aspects.
- W_{rk}= Importance of aspect a_k in review r.
- O_r = Overall rate of aspect.
- $O_r = \sum_{K=1}^{m} W_{rk}\, O_{rk}$

IV. EXPERIMENTAL RESULTS

The Data-set reviews are taken from the Cnet.com website. the reviews are present in the form of the Pros and Cons and the free text reviews.

The reviews of the products iPhone, Blackberry, Samsung and Lenovo mobile phones are used for the ranking of the aspects.

The performance parameter used are shown below. F1-measure was used as the evaluation metric for aspect identification and aspect sentiment classification.

F1-measure = 2 * (precision*recall/(precision+ recall))

To evaluate the performance of the aspect ranking the widely used Normalized Discounted Cumulative Gain at top k (NDCG@k)as the evaluation metric.

$$NDCG@k = \frac{1}{Z}\sum_{i=1}^{k}\frac{2^{t(i)}-1}{\log(1+i)},$$

Where, t(i)is the degree of important aspect at i position, Z is derived from top k aspects, which is normalization term for perfect ranking.

Here we compared the result of sentiment classification using SVM and NBSVM and found that NBSVM gives better result in terms of SVM and the Ranking result also improves by using NBSVM classifier as shown in fig 2 and fig 3.

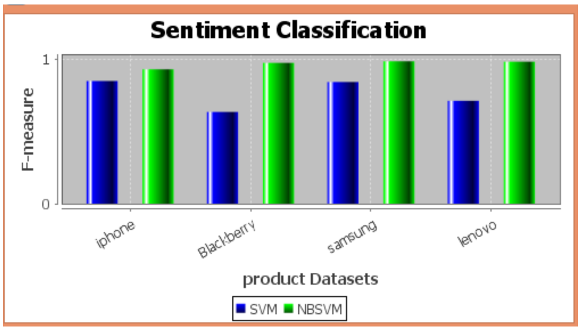

Fig: 2 Sentiment Classification Results

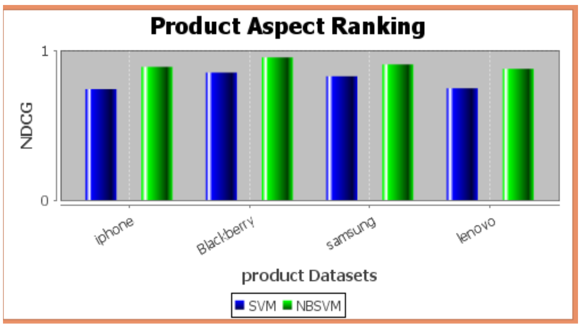

Fig: 3 Product Aspect Ranking Results

V. CONCLUSIONS

We identified the aspects from the pros and cons reviews and the free text reviews, and the sentiment classification for these reviews of the product is done by using the NBSVM classifier and these identified aspects are ranked by using the probabilistic aspect ranking algorithm.

Thus we identified the top most important aspects of the products and ranked them on the basis of their occurrence in the reviews.

This Proposed system gives the good results in terms of F-measure, when compared with SVM classifier as most of the sentiments for the aspects are identified correctly. The Ranking results are also improved by the proposed system.

This System has the disadvantage of error propagation as it considers only uni-grams for sentiment classification. A typical kind of error is caused by the polarity inconsistency between a phrase and the words it contains.

In future the results can be improved and the error propagation can be removed by using a more improved system.

ACKNOWLEDGMENT

I express true sense of gratitude towards my project guide **Prof. D.V. Gore**, of computer department for her invaluable co-operation and guidance that she gave me throughout my research. I specially thank our **P.G coordinator Prof. D.V. Gore** for inspiring me and providing me all the lab facilities, which made this research work very convenient and easy. I would also like to express my appreciation and thanks to our all my friends who knowingly or unknowingly have assisted me throughout my hard work.

REFERENCES

[1] Somprasertsri, G., Lalitrojwong, P, Automatic product feature extraction from online product reviews using maximum entropy with Conference on Information Reuse and Integration... pp. 250 255. IEEE Systems, Man, and Cybernetics Society (2008).

[2] Yang, C.C., Wong, Y.C., Wei, C.P, Learning to extract and summarize hot item features from multiple auction web sites Knowlegde... Inf. Syst. 14(2), lexical and syntactic features. In: Proc. of the IEEE International 143160 (2008)

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.

[4] Y. Wu, Q. Zhang, X. Huang, and L. Wu, Phrase dependency parsing for Opinion miningin Proc. ACL, Singapore, 2009, pp. 15331541

[5] Z-J-Zha, J-Yu, J-tang, M-wang, T-S Chua , Product aspect ranking and its applications IEEE Transaction of Knowledge and Data Engineering, Vol.26, no.5, may-2014.

[6] M.GOVINDARAJAN, Sentiment Classification of Movie Reviews Using Hybrid Method, International Journal of Advances in Science Engineering and Technology, ISSN: 2321- 9009 Volume- 1, Issue-3, Jan.-2014.

[7] Dai Quoc Nguyen and Dat Quoc Nguyen and Son Bao Pham, A Two-Stage Classifier for Sentiment Analysis, International Joint Conference on Natural Language Processing, pages 897901, Nagoya, Japan, 14-18 October 2013.

[8] Sida Wang and Christopher D.Manning, Baselines and Bigrams: Simple, Good Sentiment and Topic Classification, Department of Computer Science, Stanford University, Stanford, CA 94305.

[9] Wouter Bancken, Daniele Alfarone and Jesse Davis, Automatically Detecting and Rating Product Aspects from Textual Customer Reviews, Proceedings of DMNLP, Workshop at ECML/PKDD, Nancy, France, 2014.

[10] H. Wang, Y. Lu, and C. X. Zhai, Latent aspect rating analysis on review text data: A rating regression approach, in Proc. 16^{th} ACM SIGKDD, San Diego, CA, USA, 2010, pp. 168176.

[11] B. Snyder and R. Barzilay, Multiple aspect ranking using the good grief algorithm, in Proc. HLT-NAACL, New York, NY, USA, 2007, pp. 300307.

[12] B. Liu, Sentiment Analysis and Opinion Mining .Morgan Claypool Publishers, San Rafael, CA, USA, 2012.

[13] M. Hui and B. Liu, Mining and Summarizing customer reviews, in Proc. SIGKDD, Seattle, WA, USA, 2004, pp. 168177.