IJCTA, 8(5), 2015, pp. 1981-1988

Ranking Features of the Products Based on the Customer Opinion

P. Kanchana Devi*, N. Balaganesh and K. Muneeswaran*****

Abstract: With the rapid development of e-commerce, nowadays more people prefer to buy the products from online. To enhance customer satisfaction and shopping practice, usually online merchants request their customers to express opinions on the products which they have purchased. The reviews will be helpful for the customers to purchase a new product and it will also helps the manufacturers to enhance the quality of the product by improving the features which are all ranked at last. Feature ranking of products aims in identifying and prioritizing the essential features of products based on the online reviews given by the customers. The features which are repeatedly stated by the customers and the opinions on those features are considered to be the measures for ranking. We first extract the product features and then the opinions on those features are identified via sentiment classifier, finally the features are ranked based on their occurrence and opinions.

Index Terms: Product features, feature detection, sentiment classification, feature ranking.

1. INTRODUCTION

The rapidly mounting e-shopping has facilitated consumers to purchase products online. More than 100 million online product sales have been done in the market during the year 2006 by the producers. Nowadays more and more products are newly introduced in market. The people have many choices for buying a particular product which also consists of many categories. Most retail web sites like amazon, flip kart, snap deal encourage patrons to give feedback for their purchasing products. This gives rise to huge collections of online customer reviews on the Web. These comments have become an important resource for both consumers and firms.

In the purchaser feedback, they expresses their opinions on various features of the products. Generally a feature is the characteristic of the product. For example, canon IXUS 265 has more than two hundred features, such as "battery," "picture quality," "cost," "memory card," "size," "display". We dispute that some features are more essential than the others, and have greater impact on the eventual customers decision making as well as firm's product improvement strategies. From the huge amount of feature based reviews, the people has to know what features are more important and which features are important to consider before buying the particular product. With more and more familiar users becoming more contented with the online shopping, an increasing number of people are lettering their reviews. As a result, the number of reviews on almost all products grows rapidly. Some trendy category of the products can get thousands of reviews at some enormous sites. Moreover, many reviews are lengthy but less number of sentences will have the actual focus on the product. This makes it hard for a potential customer to read them and to make knowledgeable decision on purchasing the product. If the customer reads only few reviews, he/she may get influenced by them.

* Research Scholar, Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India, *Email: kanchselvam@gmail.com*

** Research Scholar, Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India, *Email: balaganesh@mepcoeng.ac.in*

*** Senior Professor, Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India, *Email: kmuni@mepcoeng.ac.in*

Also, the large number of reviews available in various online shopping websites makes it hard for the product manufacturers to keep track of customer's opinions for their products. Consumers can conveniently make wise purchasing decision by paying supplementary attentions to the significant features. However, it is impractical for people to manually identify the important features of products from numerous reviews. For these reason, there is a need for an approach which could automatically identify the important features is highly demanded. In this paper, we propose a product feature ranking method to automatically identify the important features of products from online consumer reviews.

Given the online consumer reviews of a specific product, we first identify the features in the reviews using a Stanford parser, and determine consumer's opinion about the features via a sentiment classifier. We then designed a feature ranking approach to identify the top features by concurrently taking into account the feature occurrence and customer's opinion. Specifically, we assume that the overall feature ranking is generated based on the opinions of the customers on multiple features of the specific product, which helps us to assess the degree of importance of the features. Ranked features are also helpful to rank the products by which the consumer make a decision about whether they purchasing it or not. The rest of the paper organized as follows. In section 2 previous works related to our proposed system are discussed, section 3 explains the proposed model and its functionalities, section 4 discuss the results obtained from the proposed work and finally section 5 conclude the proposed work.

2. RELATED WORK

The major research directions in opinion analysis are sentiment classification and opinion mining. Sentiment classification examines ways to classify each review sentence as positive, negative, or neutral. Minqing Hu and Bing Liu[2] generated aspect-based summaries of patron reviews of products sold online. Given a set of online reviews of a particular product, the process identifies the features of the product based on customer opinions and generates a summary using the revealed information.

Ana-Maria Popescu and Oren Etzioni[3] introduces OPINE, an unsupervised information extraction system which extracts features of a product based on the opinions given by reviewers. It uses the relaxation labeling technique to discover the opinions. OPINE solves the opinion mining tasks and outputs a set of product features, each accompanied by a record list of associated opinions[4],[5] which are rated based on strength.

Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu[6] uses a concept of phrase dependency parsing which extends traditional parsing to phrase level. This concept is then implemented in mining relations between product features and expressions of opinions. Jianxing Yu, Zheng-Jun Zha, Meng Wang, and Tat-Seng Chua [7] developed probabilistic feature ranking algorithm to identify the main features based on two observations like occurrence and polarity. In this work, the consumer's opinion on each specific feature in the review influences his/her overall opinion on the product. Xiaowen Ding, Bing Liu, and Philip S. Yu[8] proposed a holistic lexicon-based approach to solve the problem by developing external proofs and linguistic rules of natural language expressions. This approach allows the system to grip opinion words that are context dependent. It also deals with many special words, phrases and language constructs which have impacts on opinions based on their linguistic patterns.

Ohana, B. & Tierney, B [9] uses SentiWordNet, an opinion lexicon derived from the WordNet database where each word is associated with numerical scores representing positive and negative sentiment information. B. Liu [10] approach comprises counting positive and negative expression scores to determine sentiment orientation. H. Wang, Y. Lu, and C. X. Zhai [11] analyzed opinion articulated about the product from the huge amount of online reviews. To find the product rating, first step is to determine the features ranking. For that they considered each individual reviewer's latent opinion on each feature as well as the relative influence on different features. After calculating those weights the overall rating of the entity is decided.

O. Etzioni et al [12] presents the unsupervised method for extracting features effectively from the web not from the online reviews. It presents three ways for extraction. Pattern learning to find out domain-specific extraction rules, which smooth the progress of additional extractions. Subclass Extraction automatically recognizes sub-classes in order to increase recall. List Extraction places lists of class instances and extracts elements of each list. It adopts the pointwise mutual information retrieval which extracts from the web. Tanvir and Mohammad [14] presents a ranking based opinion mining system which uses linguistic and semantic analysis of text to recognize key information components from text documents. It finds the numeric score of all the features in the documents using Senti-WordNet [13] and then calculates the overall orientation of the feature to determine how strong the opinion is for both the positive and negative features. They have used the document parser, which divides the individual documents in creature record size chunks and for those chunks the polarity is identified. For opinion extraction the inverse document occurrence is calculated.

From these observation, some of them leaves to find the implicit opinion of the features. In some work, only adjectives are considered as the opinion, and in some work the frequent features are kept as the important features, and few works are domain dependent. The proposed method overcome such difficulties.

3. PROPOSED SYSTEM

This section provides the detailed description of the proposed system. The system design is shown in Figure1. The dataset contains product reviews which are collected from amazon.in.

3.1. Dataset

Data set consists of nine different categories of electronic products and each category has set of reviews. All the product reviews are collected from online review rife forum website amazon.in [15] and is stored in a structured form. It contains reviewer name, summary of the review, review date, verified purchaser or not, and ratings.

3.2. Product Feature Detection

To acquire more precise identification of features, we here propose three step of feature identification process based on supervised approach. Generally noun and noun phrases are considered to be the features so they must be extracted from the documents. The first step is to identify the features from the reviews, for which each review is parsed using *Stanford Parser* and noun and noun phrases are extracted from the parsing tree as feature candidates. After obtained the noun and noun phrases, some pre-processing steps are done, which includes removal of stop words, punctuation, numbers, convert text into lowercase. The collected noun and noun phrases are available as the separate files, so the step is performed to merge those two files.

All the extracted terms will not be the features and it may contain noises. The names, places and entities also retrieved as the features. But the entity is the general not a feature of a product. To better investigate the reasonability of the ranking results, the proposed approach, considered the second step as the noise elimination from that list. For which, we adopted the dictionary based approach. A set of features(bag of words)are maintained and the product features are extracted by comparing the terms and features there by eliminating the noise. In the feature list, there are some frequent terms also present. In order to obtain the unique features we have to perform the duplicate removal. Then the duplicate features are removed as the final step.

3.3. Sentiment classification

In this work, we extract a potential opinion sentence and find the polarity of each opinion in the perspective of its associated feature. If a sentence holds one or more product features and one or more

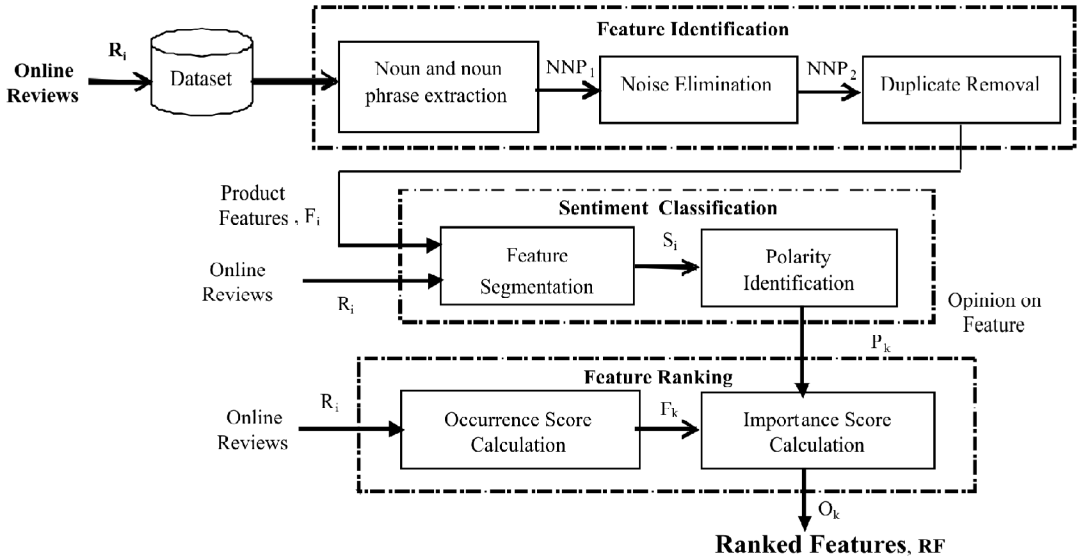

Figure 1: Overall System Design

opinion terms, then the sentence is called an opinion sentence. Our focus is to extract those sentences. Before that some preprocessing steps need to be done in online review dataset, which includes conversion of text into lowercase, removal of punctuation and numbers. This step is needed because the extracted features file is already preprocessed. By which we can perfectly match the feature with the corresponding sentences.

3.3.1. Feature Segmentation

The goal of this first step is to perform the feature segmentation. The process is to match the sentences in a review into corresponding feature. For which, all reviews are splitted into sentences. Match the feature in the sentences with the feature list file as we obtained in feature identification step. Assign all the feature extracted sentences to the respective features. For a particular feature we may have more than one opinion sentences based on their occurrence.

3.3.2. Polarity Identification

Secondly, the polarity is identified for those extracted feature sentences using sentiment classifier. In the reviews, customers express their opinion about the particular feature openly and implicitly. The sentiment classifier uses the supervised method called naïve bayes classifier [1] to find the polarity of the sentence. It considers both explicit and implicit opinions and take very less time for learning the dataset. An explicit opinion on feature f is a subjective sentence that directly utters a positive or negative opinion. An implicit opinion on feature f is an objective sentence that implies an opinion.

Example: Sentence expresses positive opinion explicitly: “The picture quality of this camera is amazing.” Sentence expresses negative opinion implicitly: “The earphone broke in two days.” For each sentence feature probability is calculated. In bayes method, larger probability is more likely to be the actual class label (polarity of the feature). The polarity may be positive, negative, or neutral. According to the probability scores for the particular feature opinion is decided finally.

3.4. Feature ranking

In our feature ranking approach, to improve the effectiveness, we combined the frequency-based method and correlation-based method. In frequency-based method only the feature frequency is considered. In correlation-based only consider the relationship between words.

Based on two characteristics of the features, the ranking is performed. Consumer's opinion on each specific feature in the review influences his/her overall view on the product and the most important features are recurrently mentioned in consumer reviews. So the overall opinion on an feature in a review is an aggregation of the opinions given to precise features in the review as we obtained from polarity identification step, and occurrence of various features have different contributions in the aggregation. To model such aggregation, we formulate the overall rating O_k in each review *r* which *is* generated based on the weighted sum of the opinions on specific features, as in eqn (3).

Here we maintain the different weights w_p for the positive, negative and neutral opinions. After obtaining the importance weights w_p for each review $r \in R$, we compute the overall importance score O_k of each feature a_k by integrating its importance scores P_k over the opinion o_{rk} and the occurrence score of the feature $F_{k.}$ The opinion score of the feature a_k is calculated as in eqn (1) by considering the polarity count along with their important weights. And the occurrence score F_k of a specific feature a_k is calculated as a sub step using (2)

$$\sum_{k=1}^{n} w_p o_{rk} \tag{1}$$

3.4.1. Occurrence Score Calculation

The important features are frequently commented in consumer reviews. For ranking along with the opinion the occurrence of the feature in the online customer reviews also considered as a measure.

$$Occurence(a_k) / \sum_{i=1}^{n} Occurence(a_i) \tag{2}$$

Where a_i is the feature occurrence count for all features *i*. Finally we integrate the opinion score in eqn (1) and occurrence score in eqn (2) to get important feature score in eqn (3).

$$O_k = F_k + P_k \tag{3}$$

According to O_k, the important product features can be identified. The opinions on important features have strong impacts on the generation of overall opinion.

Algorithm rankMain in Figure 2 describes the overall process of our work. We first identified the noun and noun phrases (i) from the review dataset. It may consist of some non feature terms, so those terms are filtered out correctly by using the noise elimination step (ii) as second one. For which we have used the feature dictionary as reference. Followed by, the redundant features are removed as the next step (iii). Then

Algorithm-rankMain(R, FD)

Inputs: R_i-A set of Customer Online reviews
FD-Set of features (Bag of words)

Output: RF-Ranked features

(i)	NNP_1 = extract NNP (R_i)	// Extracted Noun and Nounphrases
(ii)	NNP_2 = eliminate Noise (NNP_1, FD)	//Extracted Features with duplication
(iii)	F_i = remove Duplicate(NNP_2)	// Product Features
(iv)	S_i = segment Feature (R_i, $_{Fi}$)	//Set of sentences {s_1, $_{s2}$,…s_n}
(v)	P_k = detect Opinion (S_i)	//by eqn (1) // Opinion of the features
(vi)	F_k = calculat Occurrence (R_i)//by eqn(2)	//Occurrence of the features
(vii)	RF = rank Feature (P_k, $_{Fk}$)	//by eqn(3)

Figure 2: Overall Algorithm

the consumer's opinion about the feature for segmented sentences (iv) is identified via a sentiment classifier (v). We then designed a feature ranking approach to identify the important features (vii) by concurrently taking into account the feature occurrence (vi) and customer's opinion.

4. RESULTS AND DISCUSSION

This section discusses about the intermediate results of the each component of the system and provides the performance evaluation of the implemented methodologies.

4.1. Experimental Setup

The electronic product categories taken for our work are Apple Macbook Pro MD101HNA, Lenovo 59-442243, Dell Inspiron 3542, Sony DSC-HX400V, Canon IXUS 265 HS 16, Nikon Coolpix L30 20.1, Apple iPhone 6, Sony Xperia M2, Samsung Galaxy S6. The number of electronic product reviews is one of the largest, so the reviews of mobile phone, digital camera and laptop are chosen to construct the data set. This course of action can also be used for miscellaneous products domains and datasets.

Table 1
Ranked features results for Canon IXUS 265 HS 16

Features	*Importance Score*
Battery	42.30238
picture quality	41.52381
Price	39.142858
memory card	39.071428
Usability	37.071429

Table 1 shows the results for the ranked features along with their overall scores of Canon IXUS 265 HS 16. Top 5 features of the product are listed out in the table. Overall score is the aggregation of the opinion score and occurence score of the features. According to that battery is ranked first, because it has high score, and configuration is attained the lowest score, so it is in bottom level.

4.2. Performance Analysis

To evaluate the performance of feature ranking, we take on the widely used *Normalized Discounted Cumulative Gain* at top *k*. Given a ranking list of features, NDCG@*k* is calculated as in eqn (4).

$$NDCG@k = \frac{1}{Z}\sum_{i=1}^{k}\frac{2^{t(i)}-1}{\log(1+i)} \tag{4}$$

where *t(i)* is the importance degree of the feature at position *i*, and *Z* is a normalization term derived from the top-*k* features of a perfect ranking. Normally NDCG is used to measure effectiveness of algorithms. The efficacy of ranking top 5 features is NDCG@5 and for top 10 features is NDCG@10.

$$Ranking\ \ Loss = \sum_{i=1}^{n}\frac{y_i - \overline{y}_i}{m \times n} \tag{5}$$

Table 3 shows sample results by three methods such as frequency-based, correlation-based and our proposed approach. Top 5 features of the product *Apple iPhone6* are listed. From these three ranking lists, we can see that the proposed feature ranking method generates more reasonable ranking than the other

Table 2
Performance Analysis

Categories	NDCG@5	NDCG@10	Categories	NDCG@5	NDCG@10
Canon IXUS 265 HS 16	0.899	0.920	Lenovo59-442243	0.930	0.906
Sony DSC-HX400V	0.901	0.910	Dell Inspiron3542	0.862	0.883
Nikon Coolpix L30 20.1	0.852	0.874	AppleiPhone 6	0.900	0.899
Apple Macbook Pro MD101HNA	0.921	0.910	SonyXperia M2	0.899	0.907

Table 3
Top 5 features ranked by three methods for Apple iphone6

S. No	*Frequency based method*	*Correlation based method*	*Our method*
1	Phone	Phone	battery life
2	battery life	battery life	display
3	3G	looking	internet connectivity
4	appearance	3G	screen size
5	feature	feature	weight

methods. For example, the feature “*phone*” is ranked at the top by the other methods. However, “*phone*” is generally the entity of the product but not important feature. As we removed numbers during preprocessing step, internet connectivity is added as an alternative to 3G. From which we can conclude that our approach is able to automatically identify the important and more reasonable features from numerous customer reviews as shown in last column of the Table III. Therefore, we can conclude that our approach is able to automatically identify the important features from numerous customer reviews and rank them effectively.

5. CONCLUSION AND FUTURE WORK

In this work, we have introduced the feature ranking of products which identifies the essential features of products from customer online reviews. The framework encloses three main components, i.e., feature detection, sentiment classification, and feature ranking approach based on occurrence and opinion calculation. It helps us to extract the significant features of the products to be considered before buying that particular product. And bottom level features are helpful for manufactures to improve those feature quality. In this work we considered only explicit features, future work will be based on implicit features and will try to improve performance by using different algorithms.

Acknowledgement

The authors wish to thank the Management and Principal of Mepco Schlenk Engineering College, for the support in carrying out this research work.

References

[1] Zheng-Jun Zha; Jianxing Yu; Jinhui Tang; Meng Wang; Tat-Seng Chua, “Product Aspect Ranking and Its Applications,” IEEE Transaction onKnowledge and Data Engineering, Vol. 26, No. 5, May 2014, pp. 1211-1224

[2] Minqing Hu and Bing Liu, “Mining and summarizing customer reviews,” in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, 2004, pp. 168-177.

[3] Ana-Maria Popescu and Oren Etzioni, “Extracting product aspects and opinions from reviews,” in Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for

Computational Linguistics, Stroudsburg, PA, USA, 2005, pp. 339-346.

[4] Kang Liu; Liheng Xu; Jun Zhao, "Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model," In Proceedings of IEEE Transactions on Knowledge and Data Engineering,, vol. 27, no. 3, 2015, pp. 636-650.

[5] Maryam K. Jawadwala, Seema Kolkur, "Feature Ranking in Sentiment Analysis," In Proceedings of International Journal of Computer Applications on Knowledge and Data Engineering, Stroudsburg, Volume 94 – No 13, 2014, pp. 339-346.

[6] Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu, "Phrase dependency parsing for opinion mining," in Proceedings of the conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Stroudsburg, PA, USA, Vol. 3, 2009, pp. 1533-1541.

[7] Jianxing Yu, Zheng-Jun Zha, Meng Wang, and Tat-Seng Chua, "Aspect ranking: identifying important product features from online consumer reviews," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Stroudsburg, PA, USA, Vol 1,1496-1505.

[8] Xiaowen Ding, Bing Liu, and Philip S. Yu, "A holistic lexicon-based approach to opinion mining," in Proceedings of the 2008 International Conference on Web Search and Data Mining, ACM, New York, NY, USA, 231-240.

[9] Ohana, B. & Tierney, B, "Sentiment classification of reviews using SentiWordNet," in Proc. 9th. IT&T Conference, Dublin Institute of Technology, Dublin, Ireland, October 2009, pp. 22-23

[10] B. Liu, "Sentiment Analysis and Opinion Mining," Mogarn & Claypool Publishers, San Rafael, CA, USA, 2012.

[11] 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. 168–176.

[12] 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.

[13] B. Pang and L. Lee, "Opinion mining and sentiment analysis," in Found. Trends Inform. Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[14] Tanvir Ahmad and Mohammad Najmud Doja, "Ranking System for Opinion Mining of Features from Review Documents," in International Journal of Computer Science Issues, vol. 9, No. 1, July 2012, pp, 440-446.

[15] The Dataset for this work was collected from Online purchasing website amazon.in