EVALUATING PERFORMANCE OF REVIEW SYSTEM BASED ON KERNEL BASED RELEVANCE VECTOR CLASSIFIER TECHNIQUE

B.Dhanalakshmi*
Computer Science Department, Sathyabama University, Old Mamallapuram Road,
Chennai, Tamilnadu, 600119, India
dhana.baskaran@gmail.com

A.Chandrasekar
Computer Science Department, StJoseph's College of Engineering, Old Mamallapuram Road,
Chennai, Tamilnadu, 600119, India
drchandrucse@gmail.com

Abstract - Generally, a product may have a number of aspects. Some aspects are more important than the others, and have a greater effect on the repeated consumers' decision making. A product aspect based ranking framework using kernel based relevance vector classifier technique, which automatically identifies the important aspects of products from online consumers' reviews. When the consumer wants to buy a product, ranking of the aspects for the products are done based on the reviews provided by the consumers in product related sites. A discriminative classifier technique with supervised learning is used to analyze the online reviews from consumers about hotels. Graphical output will be provided to the consumer considering positives and negatives of product aspects. A separate graph will be provided for aspects that exist in both positive and negative. Thus, consumers can use this aspect based ranking system to buy the best product considering all the aspects.

***Keywords*:** Relevance vector classifier; Supervised learning; Aspect based Ranking.

1. Introduction

The usage of online websites for shopping has been increasing every day. The reviews for the product plays an important role in the consumer satisfaction. Opinion mining is the process of tracking the view of consumer about a certain product. Opinion mining is the process of analyzing business or market related data reviews and searches for consistent patterns or relationships between parameters and to check the evaluated patterns by applying the test patterns to new set of data. The consumers' reviews play a major role in deciding the purchasing behaviour for online shopping based on the other customers' reviews by observing their reviews through their blogs and other social networking sites. [4]

The aim of data mining is prediction which directs the business applications. The process of data mining consists of three stages namely the initial exploration of data, model pattern identification and deploying the model to new data to generate predictions.

For the past few years it has been a challenging research to deal with people's opinions and improvise the business intelligence. Opinion mining here deals with sentiment analysis and pattern analysis with market test data. The evaluated reviews of the opinions are used in many applications such as improvising product quality and profit for the business organizations. The customers' comments on the products are analyzed by identifying the aspects using discriminative classifier technique and assigning ranks for the products features. Knowledge discovery involves extraction of human opinions about the product, identifying the aspects and ranking them using relevance ranking technique. Relevancy ranking identifies the association among the aspects based on the correlation factor and thus used in deriving opinions from the customers' online reviews.

2. Related Work

Edison Marrese-Taylor, Juan D. Velasquez and Felipe Bravo-Marquez proposed a deterministic technique [2] which used more complex NLP based rules for sentiment and subjective analysis. This model analyzed the opinions from trip advisor in the context of Tourism department and even identified the aspect oriented expressions from travelers' opinions and thus the tool added valuable information to improvise the business. The infrequent and inbuilt aspect expressions which were not handled in this model ad that inspired our work.

Jun-ze Wang, Zheng Yan, Laurence T.Yang and Ben-xiong Huang identified the correlation among reviews to know the degree of relevance to the article it belong to [3]. Here they have analyzed real data from various sites and based on the pertinence the reviews were ranked for opinion fusions. Rather than classifying the reviews as spam or not spam, this model also analyzes user interests, preferences and taste. The need of improvising the review pertinence estimation has been achieved in the proposed model.

Other related work includes XU Xueke et al., focused on developing the novelistic aspect level opinion mining model for online customer reviews. The model demonstrated the aspect identification and aspect based sentiment tokens. This work lacks more context information to better identify the opinion words and linguistic heuristic rules are not incorporated and hence they are implemented in our model.

On the other hand, its required to discuss about the work of Kaiquan Xu et al., which analyzes customer reviews and discover potential risks and marketing strategies. The analysis fails to recognize the comparative relations entities and thus reducing the errors accumulated in the system evaluation.

3. Problem Definition and Analysis

Aspect identification and ranking for any product is done using Kernel based Relevance Deterministic Vector Classification (KRDVC) technique where the review documents are determined and analyzed and the process is given in Fig. 1. The objective of the proposed system is to enable the identification of aspects from the online reviews and ranking the potential aspects and thereby evaluating the aspects in the diverse product domain. Extractive sentiment classification and summarizing the reviews are done by this ranking technique. Customers rely on these reviews to apprehend consumers' experience, and producers depend on this consumer-generated content to capture user sentiments about their product [1].

3.1. *Challenges in review mining*

The mining challenges depends on the data extracted and the techniques used. The review documents always contain reviews of mixed views. Consider an article which compares two car brands say Hyundai and Mercedes, contains positive views about Hyundai and negative views about Mercedes. The positive review word may be considered as negative and vice versa at certain situation. The review word "mileage" turns positive or negative depending on the review context it is used in. Hence subjective opinion mining is a big challenge in the general text mining system.

Some opinions may be misleading like certain people may post positive comment about a product which may be considered as negative or neutral for some. Certain product opinions may be unrelated or outdated and there is a need to sort out the domain specific reviews which add to the subjective and sentimental context of the product reviews. Certain reviews need a comparing factor to derive a conclusion about the product. For example, the mileage of Hyundai car is better than mileage of Mercedes but worse than the mileage of Toyota car. Here Hyundai car has positive evaluation but when we compare with Toyota car it has a negative opinion.

3.2. *Aspect Extraction and Summarization*

The review documents need to be evaluated at the document level. The product reviews may have positive or negative view. Before applying any technique for sentiment detection, preprocessing on opinions collected is carried out. [5] For example, we consider hotel reviews, where the customer posts both positive and negative comments. The reviews can be written freely i.e. both positive and negative contexts can be given at the same time. The related aspects had to be identified and the orientation of the identified aspects has to be determined. This framework will analyze and grade the hotels based on certain aspects. These aspects considered should be domain dependent and filter the opinion results accordingly. The aspects are identified from free text reviews using supervised learning technique where the reviews are analyzed by splitting them into sentences and parse each sentence using Stanford Parser2. We determine each aspect in the positive and negative section where the reviews are represented into a unigram feature using Support Vector Machine (SVM) which is used to classify the reviews. Synonym clustering is also done to identify the unique aspects.

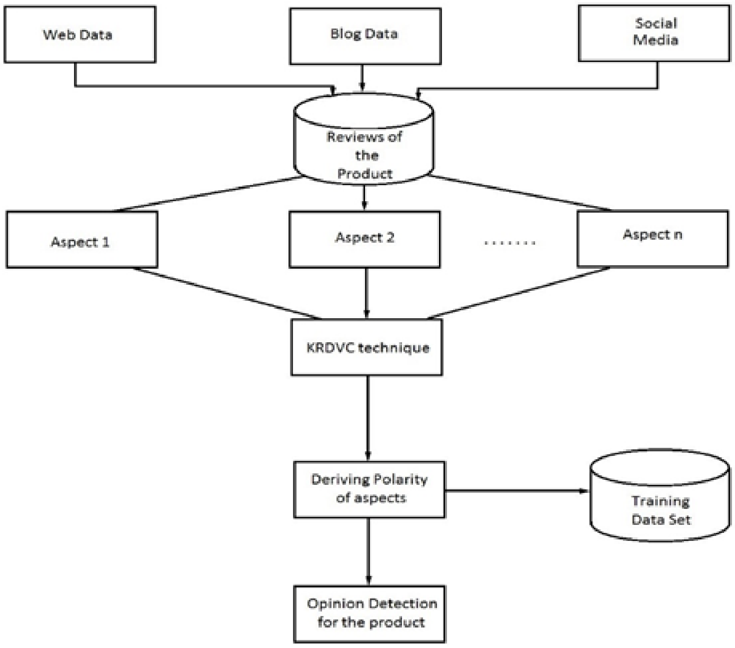

Fig. 1. Kernel Based Vector Classifier Model for mining opinions

3.2.1. *Sentiment classification on aspects of the products*

The aspects and is sentiments are analyzed using supervised learning approach and lexicon based approach. This approach is used to determine the orientation of each aspect based on their sentiment. This learning approach follows the classifier method based on the training sets. The performance of the supervised technique depends on the sufficient training sets otherwise it will be a time consuming process. This technique finds the opinion derived from the analyzed aspect. SVM classifier technique is implemented by kernel based relevance vector classifier algorithm. Input for the algorithm is consumer reviews where each review is associated with different aspects and training sets with set of aspects along with sentiments and opinions. Computing relevance score for the aspect set and thus classifying the aspects based on the score and deriving an opinion as positive or negative.

3.2.2. *Opinion Detection graph for representing polarity of the aspects*

Based on the analysis done on the relevant aspects given in the customer reviews, the positive, negative and neutral classification of the opinions determined using the kernel based relevancy vector classifier technique are represented using the detection graph. This graphical representation will clearly explain about the view of the product either positive, negative or neutral. The input relevance aspects represent x axis and the output target vector classified values as positive, negative or neutral which represent y axis.

4. Experiments and results

This section explains about our experimental evaluation where we considered the hotel online review database. Then we compare the performance of our Kernel based relevance vector classifier technique with Support vector classifier technique.

4.1. *Data Set*

The online reviews about a hotel at Chennai are collected from various web sources such as Twitter, Blog sites, Facebook, family friends and makemytrip.com as statistics of data sources given in Fig.2. The data are from different domains such as general hotel reservation web site like makemytrip.com and social medias like Facebook, etc., private blog sites. Generally, the data source from hotel reservation websites will have separate area for positive, negative and neutral comments. To calculate the sentiment of each review comment first the data need to be pre-processed where stemming and stop word removal process are done.

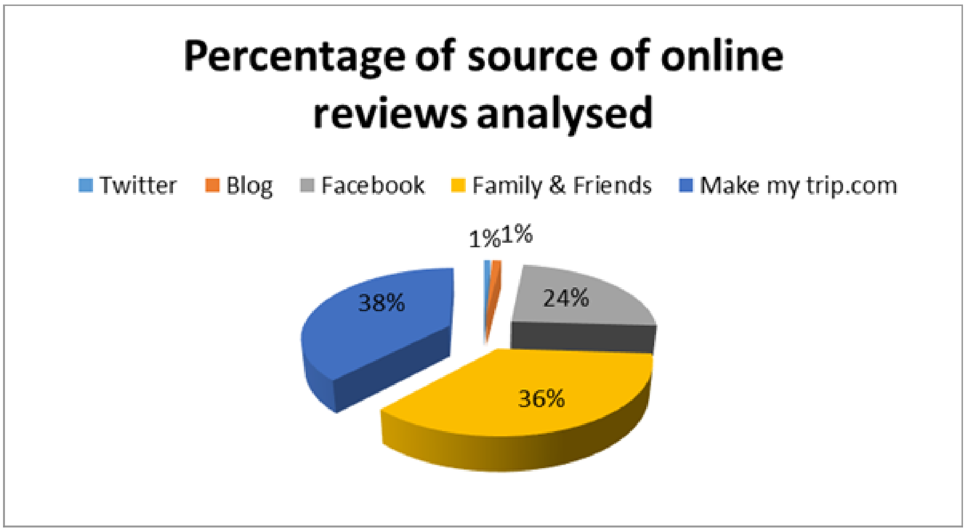

Fig. 2. Statistics of online reviews analyzed

4.2. Aspect Weighting

Each word then need to be analyzed and its weight is calculated with the help of TF-IDF as defined as below-

$$w_i = n_{w,i} * \log(|T| n_{w,T}) \tag{1}$$

where T is the number of all documents containing review, w denotes words, i denotes individual document belonging to T, |T| denotes size of all the documents, $n_{w,i}$ denotes number of times the word w appears in i, $n_{w,T}$ denotes number of documents in which the word w appears in T. Aspect selection step will refine the relevant aspects using kernel based relevant vector classifier technique which are input to a classification process.

4.3. Kernel based relevant vector classifier technique

We now use the relevancy classifier technique where it is desired to find the posterior probability of member class given the input aspect x. Generalizing the linear model the likelihood function can be defined as given in Eq. (2) where r denotes the target classification result, w denotes the weight factor assigned for any aspect, x denotes input vectors denoting the set of aspects, y denotes positive definite kernel which is some real valued function and t denotes the definite time.

$$P(r/w) = \prod_{j=1}^{N} \left\{ \frac{1}{1+\exp(-y(x_j))} \right\}^{t_i} \left\{ 1 - \frac{1}{1+\exp(-y(x_j))} \right\}^{1-t_i} \tag{2}$$

Overall the data reviews collected from various sources consists of user generated content with a short amount of text and an arbitrarily determined scores. Some reviews may be full sentence text while some may be short summaries and keywords. Fig. 3 shows the classification of hotel reviews and the score distribution as positive, negative and neutral.

4.4. Optimizing the result using Lagrangian formulation

In this proposed method the vector classifier result can be optimized using Lagrangian formulation.

$f(x)$: $\frac{1}{2}\|w\|^2$

$g(x)$: $y_i (w.x_i + b) - 1 = 0$

so Lagrangian $L = f(x) + \Sigma\alpha_i g_i(x)$

Max $L = \frac{1}{2}\|w\|^2 - \Sigma\alpha_i [y_i (w.x_i + b) - 1]$

When Lagrangian method applied on training data, the constraints will be replaced by constraints on the Lagrangian multipliers. The maximum value will give the optimized classifier result.

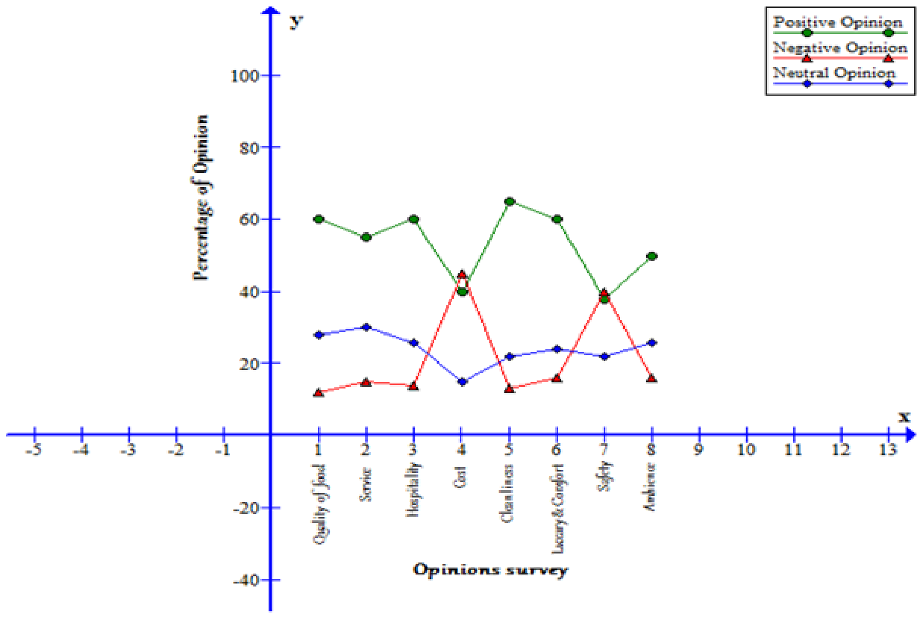

Fig. 3. Classification of reviews based on kernel based relevance vector classifier technique

Conclusion

The countenance of reviews of consumers in specialized product sites, blogs, social networking sites for analyzing the quality of products and services has become the main way of communication, due to remarkable growth of online web environment in recent years [6]. In this work, opinion mining is applied on online data reviews about hotels from various web sources. The aspect weightings are calculated which give the priorities mathematically. The vector classifier technique is intended for estimating consumer sentiments regarding hotels. The prototype can help find the hotel aspects from online data sources and classify the positive, negative and neutral reviews. The aspects are analyzed and their review classification are shown in graphical output. This graph guides the consumers to know about the product/domain before affording it. In future, this work can be implemented in health clubs, retail industries, hospital management and course management.

References

[1] Akshay Kangale, et al. (2015): Mining consumer reviews to generate ratings of different product attributes while producing feature-based review-summary, pp. 3272–3286.
[2] Edison Marrese-Taylor, Juan D.Velasquez, Felipe Bravo-Marquez (2014): A novel deterministic approach for aspect-based opinion mining in tourism products reviews, pp. 7764–7775.
[3] Jun-ze Wang, Zheng Yan, Laurence T.Yang and Ben-xiong Huang. (2015): An approach to rank reviews by fusing and mining opinions based on review pertinence, pp. 3-15.
[4] Shahab Saquib Sohail, et al. (2016): Feature extraction analysis of online reviews for the recommendation of books using opinion mining technique, pp. 754–756.
[5] Pradhan, V.; Vala, J.; Balani, P. (2016): A survey on sentiment analysis algorithms for opinion mining, 133 (9), pp. 7–11.
[6] Dhanalakshmi, B; Chandrasekar, A (2013): Fast and efficient opinion mining technique for online reviews using ranking and classification algorithms, 8(7), pp. 1587-1595.