

# Web based Quality Assessment of Customer Reviews using Quartile Measure

P.S Hiremath<sup>1</sup>, Siddu P. Algur<sup>2</sup>, Shivashankar S.<sup>3</sup>

<sup>1</sup> Gulbarga University, Department of Computer Science, Gulbarga, Karnataka, India  
Email: hiremathps@hotmail.com

<sup>2</sup> BVB College of Engg and Tech., Department of Information Science and Engg, Hubli Karnataka, India

<sup>3</sup> Karnatak Science College, Department of Computer Science, Dharawad, Karnataka, India  
Email: algursp@bvb.edu, s\_shivashankar@rediffmail.com

**Abstract**—As e-commerce is gaining popularity day by day, the web has become an excellent source for gathering customer reviews / opinions by the market researchers. The number of customer reviews that a product receives is growing at very fast rate (It could be in hundreds or thousands). Customer reviews posted on the websites vary greatly in quality. The potential customer has to read necessarily all the reviews irrespective of their quality to make a decision on whether to purchase the product or not. In this paper, we make an attempt to assess a review based on its quality, to help the customer make a proper buying decision. The quality of customer review is assessed as most significant, more significant, significant and insignificant. A novel and effective web mining technique is proposed for assessing a customer review of a particular product based on the feature distribution. This is performed in three steps : (1) Identify review regions and extract reviews from it. (2) Extract and group the features of reviews by quartile measure and then assign weights to the features belonging to each of the groups. (3) Assess the review by considering the feature weights and group belongingness. Experimental results demonstrate the effectiveness of the proposed technique which measure the quality of review and assess it accordingly. The efficiency of the task of customer review summarization can be enhanced by identifying and eliminating insignificant reviews.

**Index Terms**—Customer reviews, Quartile measure, Summarization, Feature extraction, Feature weight, Web mining.

## I. INTRODUCTION

Of late, the web has become an excellent source for posting customer reviews. The customers can now post reviews of products at merchant sites and express their views on almost everything. In the past few years, there has been an increasing interest in mining and assessing the customer reviews [9, 19, and 20]. However, the customer reviews posted at online shopping sites vary greatly in quality. Thus, it is very essential to have a mechanism which is capable of assessing the quality of reviews for purchase decision or marketing intelligence. Identifying the quality of customer reviews is useful for both potential buyers and product manufacturers. For a potential buyer, it is more convenient and less time consuming to see at a glance feature by feature comparison of customer reviews. For a product manufacturer, it helps to find the strengths and weakness of his/her own products and also that of the competitors.

There are three main review formats commonly found on the web. Different review formats may need different

techniques to identify and assess the quality of the reviews.

Format 1 : Pros and Cons

-The reviewer is asked to describe Pros and Cons separately. e.g., C|net.com uses this format.

Format 2 : Pros, Cons and detailed review

-The reviewer is asked to describe Pros and Cons separately and also write a detailed review. e.g., Epinions.com uses this format.

Format 3 : Free format

-The reviewer can write freely, i.e., no separation of Pros and Cons. e.g., Amazon.com uses this format.

The opinion orientations (positive or negative) of features are known from Format 1 and 2 because pros and cons are separated and thus there is no need to identify them.

In this paper, we propose a novel and an effective web mining technique for assessing the customer review of a particular product based on the feature distribution. Given a product name and a set of URL's of web pages that contain customer reviews on the product, it works in three stages.

Stage1: Identify the review regions and extract reviews from it.

Stage2: Extract and group the features of reviews by quartile measure and assign weights to the features belonging to each of the groups.

Stage3: For any new review, we assess it by considering the feature weights (computed in Stage2) and group belongingness.

This technique also summarizes the review after eliminating insignificant reviews. Experimental results show that the proposed technique can measure the quality of review and assess it accordingly. The efficiency of the task of customer review summarization can be enhanced by identifying and eliminating the insignificant reviews and thus retaining only significant ones.

The rest of the paper is organized as follows. The Section 2 presents the related work. In Section 3, we present the proposed technique of quality assessment of customer review based on a quartile measure for assessing the customer reviews. Section 4 shows the experimental results. Section 5 gives the conclusion.

II. RELATED WORK

Lot of research has been done in text summarization and terminology identification. The authors Dejong [6], Tait [22] and Radev and McKeown [21] propose text summarization using template instantiation. This technique needs to design a template by identifying and extracting primary elements and facts in a document. Paice [18], Kupiec, Pederson and Chen [14], Hovy and Lin [7] have focused on text summarization using text extraction, which is based on representative sentences. Kan and McKeown [12] have proposed a combined approach by merging template instantiation and text extraction. Jacquemin and Bourigault [10], Justeson and Katz [11], Dalle [4] and Church and Hanks [3] have focused on terminology identification using symbolic approach.

Many researchers are working on information extraction from texts. Their main focus is on machine learning and NLP methods for extraction or classification of entities and relations. Extending the same, the other area of research is opinion/review extraction from web pages and opinion summarization based on product features. Dave, Lawrence and Pennock [5] have proposed semantic classifier for product reviews, but it does not mine features of the product.

Liu et. al. [8] have proposed a technique to analyze customer reviews of Format 3. Their focus is on identifying the product reviews and summarizing by determining the orientation of each review. The techniques used are based on unsupervised item set mining. This approach cannot be applied for reviews of Format 2 for obtaining accurate results, because a review contains short and incomplete sentences. Morinaga et al. [17] proposed a system to know the reputation of the product, but it does not focus on analysis of the reviews. Liu et al. [15] proposed a system to (i) compare customer reviews of many competing products, and (ii) identify product features from reviews. The technique is based on NLP and supervised pattern discovery. It identifies product features of reviews of Format 2 consisting of only pros and cons. They also provide a visualization system which can be applied to all review formats.

The major problem with existing studies on assessment of reviews is that they consider all reviews irrespective of the significance of each review. Hence classification of reviews based on significance is an important task. Turney [23] proposed a system that classifies reviews as “thumbs up” for useful or “thumbs down” for unuseful review by using an unsupervised learning algorithm. Pang et al. [19] proposed a supervised learning algorithms for the same problem.

In Kim et al. [13] proposed a system for assessment of quality of reviews, using regression models. They derive ground-truth from user votes for helpfulness and then train the model and test it. Liu et al. [16] proved the biases present in the voting system and proposed a system for classification of quality of reviews by defining standard specification of quality of reviews.

III. PROPOSED TECHNIQUE

We propose a novel and effective technique to extract the customer reviews from the web pages and classify

them into different groups based on their quality. Any method of rating the reviews based on the helpful votes from the customers fails to provide a clear guideline for what a good review consists of [16]. We define four types of review qualities, which are determined by applying quartile measure to ranked frequency distribution of features constituting the reviews. Each quartile contains the features whose frequencies sum to the 25% of the total frequency of all the features. The four types of reviews are :

- (i) **Most Significant Review:** It is the one which contains dominant features with higher rank in the first quartile and sum of these feature weights is maximum.
- (ii) **More Significant Review:** It is the one which contains next dominant features whose rank is less than the first one and lie in the second quartile. The sum of these feature weights is less than the first one.
- (iii) **Significant Review:** It is the one which contains next features whose rank is less than the second one, lie in the third quartile. The sum of these features weight is less than the second.
- (iv) **Insignificant Review:** It is the one which contains all the features whose ranks are less than previous one, lie in the fourth quartile. The sum of these feature weights is least and insignificant.

The system model of the proposed technique, namely, the web based Quality Assessment of Customer Reviews using Quartile Measure, is shown in the Fig.1. It consists of the following components :

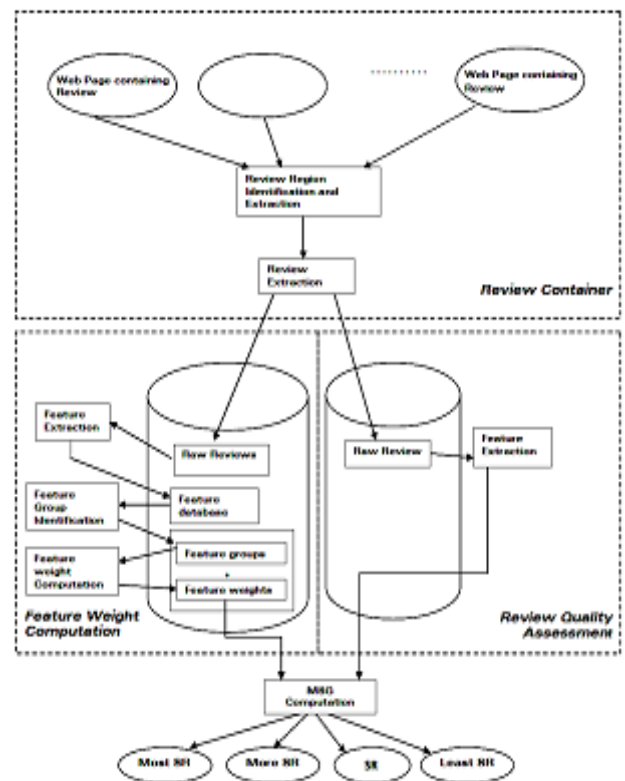


Figure 1 : System model of proposed technique

- Review container
- Feature weight computation
- Review quality assessment

The output of each component is the input for the next component.

**A. Review Container**

Given a URL of the web page containing customer reviews on the product, the review container has two components, namely :

(i) Review region extractor

It identifies and extracts only the relevant review region of a given web page [25], leaving out the other irrelevant information.

(ii) Review Extractor

It extracts the individual reviews from the review region extracted by review region extractor (identifying pros and cons as separate reviews), and stores it in two raw review databases, one for pros and another for cons in the format shown in the Table 1(a) and (b), respectively.

Table 1(a): The format of pros review database with sample records

Review No	Content of the review
R <sub>1</sub> (pros)	Great picture quality, price, great zoom ratio, nice control layout, nice LCD size
R <sub>2</sub> (pros)	Price, ease of use, nice quality photos, LCD screen, small size 2xZoom.

Table 1(b): The format of cons review database with sample records

Review No	Content of the review
R <sub>1</sub> (cons)	Battery usage, included software could be improved, included 16MB memory is stingy, need extended warranty
R <sub>2</sub> (cons)	Unreliable, long delays between pictures. Bad interface which forces you to press OK between pictures.

**B. Feature Weight Computation**

The second component of the proposed system extracts the features from the customer reviews, groups them according to their rank based on quartile measure and assign weights to each feature of every group which are used for assessing and classifying the new reviews in to different groups. In our study, we have used customer reviews on the products, like digital cameras crawled from epinion.com, as our dataset. The data set consists of one thousand two hundred (1200) reviews on five (5) types of digital cameras. Half of these reviews (i.e. 600 reviews) are used for feature weight computations and the remaining half are used for assessing the quality of review.

The steps involved in feature weight computation are

- Feature Extraction
- Feature Grouping
- Assigning the weights to features in a group

**Feature Extraction:**

Output of review extractor stored in raw review data base is the input to the feature extractor. As discussed in section 1, there are three common review formats. In our work, we focus on the reviews of Format 2. Due to the separation of pros and cons, there is no need to decide the orientation for reviews as discussed in [14]. The existing methods of [15] are used to extract the product features from the customer reviews of Format 2. In [15], the authors extract the product features of Format 2, using POS tagger and also check for grouping synonyms. The

method in [16] makes use of edit distance to compare the similarities between the surface strings of two mentions, and uses contextual similarity to reflect the semantic similarity between the two mentions. Thus, the features extracted from the customer reviews are stored in the feature database after checking whether that feature is already existing or not. A sample of extracted features stored in the database is shown in the Table 2.

Table 2: A sample of extracted features

f <sub>1</sub>	f <sub>2</sub>	f <sub>3</sub>	f <sub>4</sub>	f <sub>5</sub>	f <sub>6</sub>	-	f <sub>63</sub>
Price	Picture Quality	Zoom	Speed	Battery life	Memory Card	--	warranty

**Feature Grouping:**

Here we present the most effective method of identifying to which group each of the extracted feature belongs to. The input for this component is the raw reviews and featured set extracted by the earlier step. Consider that there are a total of m customer reviews for a particular product. From these reviews, we extract n features as explained in the Section 2. We construct a review matrix of order of m X n. The algorithm to develop the review matrix is given in the Algorithm 1.

Algorithm 1 : Algorithm for review matrix construction.

For each review R<sub>i</sub> in the raw review database

```

{
  For each feature fj in the review
  {
    If fj is present in Ri then Mij = 1
    else Mij = 0
  }
}
    
```

A sample review matrix constructed for few reviews (illustrated in the Tables 1 and 2) using the Algorithm 1 is given in the Table 3.

Table 3: Sample review matrix

	Price	Picture Quality	Zoom	Speed	Battery life	Memory Card	-	warranty
Review no	f <sub>1</sub>	f <sub>2</sub>	f <sub>3</sub>	f <sub>4</sub>	f <sub>5</sub>	f <sub>6</sub>		f <sub>n</sub>
R <sub>1</sub>	1	0	0	0	0	0		
R <sub>2</sub>	0	0	0	0	1	0		
R <sub>3</sub>	0	0	0	0	0	0		
R <sub>4</sub>	0	0	0	0	0	0		
R <sub>5</sub>	1	0	0	0	0	0		
:								
R <sub>m</sub>	-	-	-	-	-	-	-	-
Total	∑Mi <sub>1</sub>	∑Mi <sub>2</sub>	∑Mi <sub>3</sub>	∑Mi <sub>4</sub>	∑Mi <sub>5</sub>	∑Mi <sub>6</sub>	-	∑Mi <sub>n</sub>

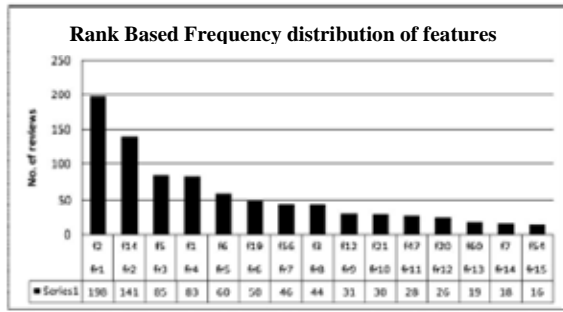


Figure 2 : Rank based frequency distribution of features

**Assigning Ranks to Features:**

The assignment of rank to each feature in the review matrix is based on the frequency of occurrence of each feature in the review set  $R_m$ . The frequency of occurrence is calculated by column sum  $\sum_i M_{ij}$  for each feature  $f_i$  as shown in the Table 3. Next the features are arranged in the decreasing order of their frequencies. The feature with maximum frequency, is assigned rank one, that with the second maximum is assigned rank two and so on. The rank based frequency distribution of features is shown in the Fig. 2, in which only the 15 dominant features of the reviews considered in our experimentation are represented. Let  $fr_i$  denote the feature with rank  $i$ , e.g., if  $fr_1$  is picture quality then picture quality has rank one,  $fr_2$  is ease of use then it has rank two and so on.

**Grouping of features based on Quartile measure:**

Now, we describe the quartile measure which is used to group the features. Quartile measure is a statistical term describing a division of observations in to four intervals. Each quartile contains one-fourth of the total observations. In our study, the features are arranged on the basis of their ranks and they are grouped in to four types, namely, Most significant, More significant, Significant, and Insignificant. The algorithm for grouping the features into quartiles is given in the Algorithm 2.

Algorithm 2 : Grouping features into quartiles.  
 Step1: Compute the sum of the frequencies of all the features

$$\text{Total} = \sum_i \sum_j M_{ij}$$

Step2: Compute Quartile Value  
 $QV = \text{Total} / 4$ ; ( 25% of the total)  
 Step3: The first few features with higher ranks, say upto rank  $a$ , belong to first quartile, whose frequencies sum to the  $QV$  and these are grouped into the first group.  
 $G_1 = \{fr_1, fr_2, \dots, fr_a\}$   
 Step4: The next few features with lesser ranks, say upto rank  $b$ , belong to the second quartile, whose frequencies sum is to  $QV$ , and these are grouped into the second group.  
 $G_2 = \{fr_{a+1}, fr_{a+2}, \dots, fr_b\}$   
 Step5: Next few features belong to the third quartile with ranks lesser than the features belonging to the second quartile, say upto rank  $c$ , whose frequencies sum to  $QV$ , and these are grouped into the third group.  
 $G_3 = \{fr_{b+1}, fr_{b+2}, \dots, fr_c\}$

Step6: The remaining features, with ranks less than  $c$  and their frequencies summing to  $QV$ , belong to fourth quartile and these are grouped into the fourth group.

$G_4 = \{fr_{c+1}, fr_{c+2}, \dots, fr_d\}$   
 where  $d$  is the last rank. Further these Groups ( $G_1, G_2, G_3$  and  $G_4$ ) are used for assigning the weights to each feature and the procedure is described in the next section

**Assigning the weights to features in a group:**

The calculation and assignment of the weights for each of the features belonging to a group is described below:

We compute the weight  $Wr_i$  for the feature  $fr_i$  with rank  $i$  by using the formula

$$Wr_i = (\text{frequency of } fr_i) / QV$$

where  $i = 1, 2, \dots, a$  for group  $G_1$ ,  
 $i = a+1, a+2, \dots, b$ , for group  $G_2$ ,  
 $i = b+1, b+2, \dots, c$ , for group  $G_3$ ,  
 and  $i = c+1, c+2, \dots, d$ , for group  $G_4$ .

The weights computed for each feature are used to assess the quality of reviews, which is discussed in the next section

**C. Review Quality Assessment**

The third step of the proposed technique is to find out the group to which a given review belongs to based on its quality. Review from the raw review database, feature groups and feature weights are inputs for rating the review quality assessment. The algorithm for review quality assessment is given in the Algorithm 3.

- Algorithm 3 : Review quality assessment.
1. Identify and extract the features appearing in the given review.
  2. Find the groups to which the given review belongs to and also get the weight of each feature from feature group and weight database.
  3. Form the set  $WG_i$  of all the weights of features belonging to group  $G_i$
  4. Compute the sum  $SWG_i$  of weights for each set  $WG_i$ ,  $i = 1, 2, 3, 4$ .
  5. Let  $MSG$  be maximum of  $SWG_1, SWG_2, SWG_3, SWG_4$
- The review quality assessment is determined by the following criteria :
- If  $MSG = SWG_1$ , then Review is Most Significant.
  - If  $MSG = SWG_2$ , then Review is More Significant.
  - If  $MSG = SWG_3$ , then Review is Significant.
  - If  $MSG = SWG_4$ , then Review is Least Significant.

**IV. EXPERIMENTAL RESULTS**

For the purpose of experimentation, we apply the proposed technique to see how effective it is in assessing the quality of review from pros and cons in reviews of Format 2. We also show its effectiveness on the task of customer review summarization. We conduct experiments by taking reviews from the web pages and assessing them as most significant, more significant, significant, and insignificant review.

We consider 1200 customer reviews of digital camera, out of which, 600 reviews are used to evaluate the feature weights by grouping them in to four types based on the quartile measure, and the remaining 600 are used to assess the review quality.

The experimental results given in Table 4 show that, for 600 reviews, it could identify 63 features such as picture quality, price, battery etc. The second and third column of the table show the frequency of occurrence of each feature and its corresponding rank, respectively. The frequency of occurrences of features f1, f3,...f63 are 83, 44, ...4, respectively, and their ranks are 4,8,12,...,23, respectively. The fourth column shows to which group each feature belongs to, e.g. G<sub>1</sub> consists of { f<sub>2</sub>, f<sub>14</sub>}, G<sub>2</sub>

consists of {f<sub>5</sub>,f<sub>11</sub>,f<sub>6</sub>,f<sub>19</sub>}, G<sub>3</sub> consists of {f<sub>56</sub>,f<sub>3</sub>,f<sub>12</sub>,f<sub>21</sub>,f<sub>47</sub>,f<sub>20</sub>,f<sub>60</sub>,f<sub>7</sub>,f<sub>54</sub>,f<sub>9</sub>}, and the remaining features in G<sub>4</sub>. The last column of the table consists of values of feature weights, e.g. the weights of features f<sub>2</sub> and f<sub>14</sub> belonging to group G<sub>1</sub> are 0.584 and 0.415, and weights of features f<sub>5</sub>,f<sub>11</sub>,f<sub>6</sub>,f<sub>19</sub> belonging to group G<sub>2</sub> are 0.305,0.298,0.215,0.179, respectively. The values in Table 4 are used to assess the quality of the remaining 600 reviews.

Table 4: Summary of feature sum, rank, feature weight and group information

Feature	Feature sum	Rank	Group	Weight
f1	83	4	G <sub>2</sub>	0.2986
f2	198	1	G <sub>1</sub>	0.584
f3	44	8	G <sub>3</sub>	0.1611
f4	15	16	G <sub>4</sub>	0.0682
f5	85	3	G <sub>2</sub>	0.3058
f6	60	5	G <sub>2</sub>	0.2158
f7	18	14	G <sub>3</sub>	0.0659
f8	15	16	G <sub>4</sub>	0.0682
f9	19	13	G <sub>3</sub>	0.0685
f10	1	26	G <sub>4</sub>	0.0045
f11	13	17	G <sub>4</sub>	0.0591
f12	31	9	G <sub>3</sub>	0.1135
f13	13	17	G <sub>4</sub>	0.0591
f14	141	1	G <sub>1</sub>	0.4159
f15	9	19	G <sub>4</sub>	0.0409
f16	13	17	G <sub>4</sub>	0.0591
f17	1	26	G <sub>4</sub>	0.0045
f18	3	24	G <sub>4</sub>	0.0136
f19	50	6	G <sub>2</sub>	0.1798
f20	26	12	G <sub>3</sub>	0.0952
f21	30	10	G <sub>3</sub>	0.1098
f22	7	21	G <sub>4</sub>	0.0318
f23	2	25	G <sub>4</sub>	0.0091
f24	0	27	G <sub>4</sub>	0.0000
f25	11	18	G <sub>4</sub>	0.0500
f26	0	27	G <sub>4</sub>	0.0000
f27	8	20	G <sub>4</sub>	0.0364
f28	4	23	G <sub>4</sub>	0.0182
f29	7	21	G <sub>4</sub>	0.0318
f30	5	23	G <sub>4</sub>	0.0227
f31	5	23	G <sub>4</sub>	0.0227
f32	0	27	G <sub>4</sub>	0.0000

For a given review, e.g., “Great picture quality, price, great zoom ratio, nice control layout, nice lcd size”, the extracted features are Picture quality (f2), price (f1), Zoom (f3), control (f30) and lcd (f12). We identify the groups to which each of the extracted features belong, namely, f2 belongs to G1, f1 to G2, f3 and f12 to G3, and f30 to G4. Then we get the weights of each feature from the feature weight database. Next, we obtain the sum of weights of features of each group namely, 0.584,0.298,0.274,0.029. Further, we assess the quality of the review based on the maximum value and categorize it

Feature	Feature sum	Rank	Group	Weight
f <sub>33</sub>	3	24	G <sub>4</sub>	0.0136
f <sub>34</sub>	6	22	G <sub>4</sub>	0.0273
f <sub>35</sub>	0	27	G <sub>4</sub>	0.0000

f <sub>36</sub>	8	20	G <sub>4</sub>	0.0364
f <sub>37</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>38</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>39</sub>	8	20	G <sub>4</sub>	0.0364
f <sub>40</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>41</sub>	3	24	G <sub>4</sub>	0.0136
f <sub>42</sub>	1	26	G <sub>4</sub>	0.0045
f <sub>43</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>44</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>45</sub>	1	26	G <sub>4</sub>	0.0045
f <sub>46</sub>	5	23	G <sub>4</sub>	0.0227
f <sub>47</sub>	28	11	G <sub>3</sub>	0.1026
f <sub>48</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>49</sub>	11	18	G <sub>4</sub>	0.0500
f <sub>50</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>51</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>52</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>53</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>54</sub>	16	15	G <sub>3</sub>	0.0586
f <sub>55</sub>	0	27	G <sub>4</sub>	0.0000
f <sub>56</sub>	46	7	G <sub>3</sub>	0.1685
f <sub>57</sub>	8	20	G <sub>4</sub>	0.0364
f <sub>58</sub>	8	20	G <sub>4</sub>	0.0364
f <sub>59</sub>	15	16	G <sub>4</sub>	0.0682
f <sub>60</sub>	19	13	G <sub>3</sub>	0.0696
f <sub>61</sub>	4	23	G <sub>4</sub>	0.0182
f <sub>62</sub>	11	18	G <sub>4</sub>	0.0500
f <sub>63</sub>	4	23	G <sub>4</sub>	0.019512

Table 5 : Summary of experimental results

Group name	No. of features in each group	No. of reviews in each group
Most SR (G <sub>1</sub> )	2	244
More SR (G <sub>2</sub> )	4	128
SR (G <sub>3</sub> )	10	76
ISR (G <sub>4</sub> )	47	152

accordingly. Since, the maximum value 0.584 corresponds to G1, the given review is assessed as most significant review.

The table 5 shows the summary of the assessed 600 reviews. There are 2,4,10 and 47 features in Most Significant Review, More Significant Review, Significant Review and Insignificant Review, respectively. It also shows 244 reviews in Most Significant Review, 128 reviews in More Significant Review, 76 reviews in Significant Review and 152 in Insignificant Review.

The experimental results show that, there are large number of features belonging to insignificant group. They do not influence a buying decision significantly. Hence, such reviews may be ignored while summarization of customer reviews. The summarization of customer review gives scope for the future work.

#### V. CONCLUSION

In this paper, we proposed a novel and effective web mining technique for assessing the customer review for a particular product based on feature distribution. The quality assessment of a customer review is categorized as most significant review, more significant review, significant review or insignificant review.

This is performed in three steps : (1) Identify review regions and extract reviews from it. (2) Extract and group the features of reviews by quartile measure and then assign weights to the features belonging to each of the groups. (3) Assess the review by considering the feature weights and group belongingness. Further these weights and group information is used to assess a given review. The experimental results show that there are significant number of features belonging to insignificant group. The reviews containing these features may be ignored while review summarization, and thus optimizing the process of quality assessment.

#### ACKNOWLEDGMENT

The authors are grateful to the reviewers for their helpful comments which improved the quality of the paper considerably.

#### REFERENCES

- [1.] Bo, Pang, and Lillian Lee. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." *ACL*. 2004.
- [2.] Bo, Pang, and Lillian Lee. "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales." *ACL*. 2005.
- [3.] Church, K, and P Hanks. "Word association norms, mutual information and lexicography." *Computational Linguistics* 16, no. 1 (1990): 22-29.
- [4.] Daille. "Study and Implementation of Combined Techniques for Automatic Extraction of Terminology." In *The Balancing Act: Combining Symbolic and Statistical Approaches to Language Processing*. Cambridge: MIT Press.
- [5.] Dave, K, S Lawrence, and D Pennock. "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews." *WWW-2003*. 2003.
- [6.] Dejong, G. "An Overview of the FRUMP System." In *Strategies for Natural Language Parsing*, 149-176. 1982.
- [7.] Hovy, E, and C Y Lin. "Automated Text Summarization in SUMMARIST." *ACL Workshop on Intelligent, Scalable Text Summarization*. 1997.
- [8.] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." *KDD*. 2004.
- [9.] Hu, Minqing, and Bing Liu. "Mining opinion features in customer reviews." *AAAI*. 2004.
- [10.] Jacquemin, C, and D Bourigault. "Team extracting and automatic indexing." In *Handbook of Computational Linguistics*, edited by R Mitkov. Oxford University Press, 2001.
- [11.] Justeson, J S, and S M Katz. "Technical Terminology: some linguistic properties and an algorithm for identification in text." *Natural Language Engineering* 1, no. 1 (1995): 9-27.
- [12.] Kan, M, and K McKeown. "Information Extraction and Summarization: Domain Independence through Focus Types." Columbia University Technical Report CUCS-030-99, 1999.
- [13.] Kim, Soo-Min, Patrick Pantel, Tim Chklovski, and Marco Pennacchiotti. "Automatically Assessing Review Helpfulness." *EMNLP-2006*. 2006.
- [14.] Kupiec, J, J Pedersen, and F Chen. "A Trainable Document Summarizer." *SIGIR-1995*. 1995.
- [15.] Liu, Bing, Hu Minqing, and Cheng Junsheng. "Opinion observer: Analyzing and Comparing Opinions on the Web." *WWW-2005*. Chiba, 2005.
- [16.] Liu, Jingjing, Yunbo Cao, Chin-Yew Lin, Yalou Huang, and Ming Zhou. "Low-Quality Product Review Detection in Opinion Summarization." *Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Prague, 2007. 334-342.
- [17.] Morinaga, S, K Yamanishi, K Tateishi, and T Fukushima. "Mining Product Reputation on the Web." *KDD-2002*. 2002.
- [18.] Paice, C D. "Constructing Literature Abstracts by Computer: techniques and prospects." *Information Processing and Management* 26 (1990): 171-186.
- [19.] Pang, B, L Lee, and S Vaithyanathan. "Thumbs up? Sentiment Classification Using Machine Learning Techniques." *EMNLP-2002*. 2002.
- [20.] Popescu, Ana-Maria, and O Etzioni. "Extracting product features and opinions from reviews." *HLT-EMNLP*. 2005.
- [21.] Radev, D, and K McKeown. "Generating natural language summaries from multiple on-line sources." *Cisticscomputational Lingu* 24, no. 3 (September 1998): 469-500.
- [22.] Tait, J. *Automatic summarising of english text*. PhD Dissertation, Cambridge: University of Cambridge, 1983.
- [23.] Turney, P. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews." *ACL-2002*. 2002.
- [24.] Zhang, Zhu, and Balaji Varadarajan. "Utility Scoring of Product Reviews." *CIKM'06*, 2006.
- [25.] Hiremath, P.S Benchalli, S.S., Algur Siddu P., and Renuka Udupudi. "Mining Data Regions from Web Pages" *COMAD-2005*. 2005