Reputation based online product recommendations

Chiranjeevi Pandi[1,*], Teja Santosh Dandibhotla[2], Vishnu Vardhan Bulusu[3]

*1. ACE Engineering College
Telangana State, India*

*2. Sreenidhi Institute of Science and Technology
Yamnampet, Ghatkesar, Telangana State, India*

3. Jntuhcem, Manthani, Telangana State, India

chiruanurag@gmail.com

ABSTRACT. *Crucial data namely product aspects and opinions are extracted from online product reviews. The obtained opinions are further analyzed for orientations. These orientations that are positive, negative or neutral are counted to determine the sentiment of the aspect. The sentiments are often turned (unforeseen rise or fall) and due to this the quality of recommended products by the recommendation system is less. The purpose of this study is to assess the importance of aspects reputations in the similarity based product recommendations. A simulation model was established through the analysis of product reviews for ranking the aspects and identifying the frequent aspects among them. The case based reasoning of the searched product against the available similar products from the category are finally compared on the basis of aspect reputations. This comparison provides the list of sorted reputed products in the decreasing order of similarity as recommendations. Through this study, it was found that the recall measure calculated on the reputation based recommendations is better than sentiment based recommendations. The findings of this research are promising in terms of product recommendations using reputation.*

RÉSUMÉ. *Les données cruciales, à savoir les aspects des produits et les opinions, sont extraites des critiques de produits en ligne. Les avis obtenus sont ensuite analysés pour des orientations. Ces orientations qui sont positives, négatives ou neutres sont comptés pour déterminer le sentiment de l'aspect. Les sentiments sont souvent tournés (hausse ou chute imprévue) et à cause de cela, la qualité des produits recommandés par le système de recommandation est moindre. Le but de cette étude est d'évaluer l'importance de la réputation des aspects dans les recommandations de produits fondés sur la similarité. Un modèle de simulation a été établi à travers l'analyse des critiques de produits afin de classer les aspects et d'identifier les aspects les plus fréquents parmi eux. Le raisonnement fondé sur le cas du produit recherché contre les produits similaires disponibles de la catégorie est finalement comparé sur la base de la réputation des aspects. Cette comparaison fournit la liste des produits réputés classés dans l'ordre décroissant de similitude indiqué dans les* *recommandations. Cette étude a révélé que la mesure de rappel calculée sur la base des recommandations basées sur la réputation est meilleure que les recommandations basées sur le sentiment. Les résultats de cette recherche sont prometteurs en termes de recommandations de produits utilisant la réputation.*

KEYWORDS: *product aspects, opinions, aspect rank, frequent aspects, aspect reputation, product similarity, product recommendations.*

MOTS-CLÉS: *aspects du produit, opinions, rang d'aspect, aspects fréquents, réputation d'aspect, similitude de produit, recommandations de produit.*

DOI:10.3166/ ISI.23.5.81-103

1. Introduction

Over the last two decades the world has witnessed a rapid growth in various forms of data on the World Wide Web. The online flow of information is getting more and more wobbling. The web content is growing at lightning fast speed with social media as a huge part. E-commerce sites are integral parts of social media which allowed online shopping community for scripting their reviews. The quality of the online product is perceived through online consumer reviews. Online product reviews provide crucial pieces of information namely aspects and opinions in order to carry out the task of opinion mining. These crucial pieces of information from these reviews affect the purchase intentions of the consumers (Wang *et al*., 2017). Often the opinions that are framed on the product aspects facilitate to determine sentiments of the aspects (Peng and Lee, 2008).

The sentiments of the product aspects are often turned (Gârbacea *et al*., 2014) (unforeseen rise or fall) as monitoring the reputation of the product demands continuous analysis of the product aspects sentiments through online tracking of product reviews on daily basis (Silvestro *et al*., 2017). These turnovers in the aspect opinion counts have reduced the quality of aspects sentiments based product recommendations in the current product recommendation environment (Perales *et al*., 2017). Also the provision of equal importance to all product aspects worsens the sentiment based recommendations.

In order to overcome these problems, this paper establishes a model based on the statistical aspects reputations for providing the product recommendations. This task is carried out by first retrieving the base cases for the customer queried product. Then, the ranking of extracted product aspects of both the query case and base cases is performed by using aspect sentiment and aspect gain ratio. The opinion orientations of these ranked aspects are also analyzed. Next, these ranked aspects are empirically analyzed for frequency to learn the most frequent opined aspects among the highly ranked aspects. Further, common aspects across the query case and the base cases are extracted. Furthermore, the statistical reputation values of these common aspects on time based opinion orientation counts are determined. These statistical reputations of the common aspects from the base cases are used in determining the product similarity with the query case.

The findings of the reputation way of product recommendations are finally compared against the aspects sentiments based product recommendations. The results are observed to be promising in this direction.

2. Related works

The E-Commerce platforms allow the consumers of the product to pen their feelings in the form of online reviews using natural language. More often the reviews are written in free flow, unstructured format allowing reviewers to write lengthy reviews. These expressed writings involve in them the knowledge levels of the language of the reviewer in the form of sentences. The automated understanding of human intentions from these review sentences for a fellow human is easy. The same task is very challenging to carry out by the machine. In order to mitigate this problem, a popular tool namely Natural Language Processing (NLP) is used. NLP provides the ability to the machine to analyze the human language (either speech or text) and get the understanding of the language with the maximum accuracy.

In order to carry out this task by the machine, the processing environment depends on various dictionaries and lexicons. A lexicon is a collection of lexemes (basic unit of language with one or several words intended to convey the meaning as a whole) in the alphabetical order. WordNet is a lexical knowledge base for English language. It groups the English words into sets of synonyms called synsets. The major purpose of WordNet is to support automatic text analysis in many of the artificial intelligence applications. Most of the synsets are connected to other synsets in the WordNet through semantic relations. These relations differ based on the type of word. The various semantic relations based on the type of the word are: i) hypernyms, hyponyms, holonyms and meronyms fall under Noun category, ii) hypernym, troponym, entailment, coordinate terms fall under Verb category, iii) related nouns, similar to, participle of verb fall under Adjective category and iv) root adjectives fall under adverbs category. These semantic relations are used for determining similarity between the concepts. SentiWordNet is an enhanced lexical resource which contains sentiment scores for the WordNet word types. The main purpose of SentiWordNet is to support the task of opinion mining.

2.1. Aspects extraction

Aspect level opinion mining aims at obtaining the aspects from the unstructured reviews and finding the opinion orientation of the aspect. This analysis reveals the impressions of the users about the product whether they are enchanted by the product or otherwise.

Research on aspect mining is considered as a major research work for over two decades. Quite a number of researchers have focused their research on this particular subject. The recent works focused on aspect extraction is presented in this section. Samha *et al*. extracted frequent and infrequent product aspects from reviews using Conditional Random Field (CRF) classifier. They have achieved 75% precision in their aspect extraction task. Kumar *et al*. (2017) designed comprehensive feature extraction approach to extract the maximum and accurate product aspects from a huge number of online product reviews. This approach performed superior to anything the specific way for extracting the product aspects in the semantic environment. Santosh *et al*. (2016) derived a hybrid LDA based on Feature Ontology Tree (FOT) in order to extract product aspects from asymmetric collection of product reviews. Qian Liu *et al*. proposed a novel lifelong learning approach to extract aspects from product reviews. This approach uses semantic similarity and aspect associations in order to extract the aspects. Soujanya Poria *et al*. (2016) used 7- layer deep Convolutional Neural Network (CNN) with linguistic patterns to extract product aspects. There observed a recall of 88.32% on cell phone product reviews and minimum of 84.3% recall on DVD product reviews. However, this was achieved under careful consideration of weights and their updates in the CNN. Recently, Chiranjeevi *et al*. extracted aspects from product reviews by using corpus linguistic rules and distant supervision. They have implemented CRF classifier in order to extract product aspects. They have achieved 79.1% recall on cameras product reviews.

2.2. Aspect ranking

The impact of product aspects on consumption intentions is examined by two important ways. They are namely expert suggestion and syntax-based analysis.

Pan and Chiou (2011) proposed an intuitive approach to adopt expert opinions. However, the problem is that the expert's opinion cannot represent the online users widely, especially for credence products. To ensure the quality is put on the first priority for the study of online reviews. Ghose *et al*. (2012) addressed the cognitive inconsistency using crowdsourcing. While crowdsourcing obtains needed services or ideas by seeking contributions from a crowd of people, it is still far from reaching the online community.

Alternative methods are syntax-based analysis, where the context is used to determine aspect ranking. Guo *et al*. (2013) provided guidelines for aspect ranking by exploiting the structure pattern hidden in sentences. Main steps are building and training model and running the model on testing dataset to obtain final results. Wu *et al*. applied the concept of network theory to detect the aspect ranking, in which aspects are treated as nodes, and constitute a huge network along with edges between aspects. An algorithm which is similar to HITS is then employed to compute the node's authority which represents the aspect ranking. It is an easy-to-use method, but difficult to form a unified procedure for network construction. Eirinaki *et al*. (2012) emphasized that the more number of modifier words are there for an aspect, the more important is that aspect. Therefore, first the number of modifier words for the aspect is computed. Then the frequent adjectives are counted to determine aspect ranking. However, errors would arise in certain cases with many modifiers corresponding to some aspects but few for other aspects. Wei *et al*. (2017) incorporated sentiment analysis to rank the product aspects. These ranked product aspects were further analyzed for purchase intentions by the consumer.

2.3. Opinion identification and orientation

2.3.1. Dictionary based approaches

The works on dictionary based approach for opinion word identification and orientation was presented in (Kim and Hovy, 2004). Initially, the set of opinion words are gathered with well known orientations. Then, the size of the set is improved by adding the synonyms and antonyms by searching in the WordNet (Miller *et al*., 1990) or thesaurus (Mohammad *et al*., 2009) The iteration is continued until no new words are found. After the completion of this entire cycle, to correct the errors or to remove the errors manual inspection is carried out.

The major loop hole in dictionary based approach (Qiu *et al*., 2010) is it cannot able to find opinion words with context and domain specific orientations.

2.3.2. Corpus based approaches

The corpus based approaches identify the opinion words by using machine learning technique. Corpus based method depends on syntactic patterns. The large corpus is used to find the opinion words. Mckeown and Hatzivassiloglou presented one of the methods. They manually considered the opinion adjectives as list of seed, opinion words and their orientations. Following are the constraints like OR, AND, EITHER-OR, BUT etc; 'AND 'conjunction represents the same orientation.

Some negating expressions are there, such as 'but', which changes the opinion. So, it is necessary to identify the orientation of adjective. To determine orientation of adjectives of the same or different orientations, learning is applied to a large corpus. Then, generate a graph by linking the adjectives and then perform clustering on the graph to produce positive and negative words.

The Conditional Random Fields (CRF) method is a statistical method used to extract the opinion phrases. This method was presented by Jiaoa and Zhoua (2011) to distinguish sentiment polarities. This method achieved high performance. A two-level CRF model was used by Xu *et al*. (2011). In this work, the complicated dependencies between words, entities and relations, and the unreliable interdependencies among relations were utilized. They made a graphical model from customer reviews and extracted the comparative relations between products.

Cruz *et al*. (2013) proposed a taxonomy-based approach to extract feature-level opinions and map them into feature taxonomy. This structure represents the attributes of an object and opinionated parts. The main goal of this work was domain specific opinion mining. They define how people express their opinions in the document in domain-specific scenario. Some resources will be automatically induced from a set of annotated documents. To improve the performance of the domain, various techniques are used. The domain independent approaches were used to build the accurate opinion extraction systems with different parameters.

Dictionary-based approach is efficient when compared to corpus based approach. Using corpus based approach alone is not efficient because it is not possible to define all English words in a corpus, but corpus based approach has a great advantage that can help to find orientations, and the context of specific opinion words using a domain corpus.

2.4. Sentiment analysis

2.4.1. Statistical based approaches

The review in statistical based approach is represented as a combination of latent aspects and their respective ratings. It is implicit that aspects and their ratings correspond to the multinomial distributions and clustering the head terms as aspects and sentiments as ratings.

Finding seed opinion words or co-occurrence patterns is carried out using statistical techniques. Fahrni and Klenner proposed that finding seed words can be done by deriving posterior polarities using the co-occurrence of adjectives in a corpus. To construct the corpus, all the document words are included from the dictionary. So that word unavailability problem can be overcome with this approach (Turney, 2002).

Word polarity can be identified by studying the word frequency in a large annotated corpus. The polarity is considered to be positive when the word occurred more frequently. As opposed to this, if the word occurs in negative texts more frequently, then consider its polarity is negative. If the word has equal frequencies, then it is considered as a neutral word. There are certain opinion words which are similar in opinion, in a corpus this type of words appears together frequently. Therefore, if two opinion words appear together frequently within the same context, they are may have the similar polarity. To determine the unknown word polarity, calculate the relative frequency of co-occurrence of the word with other words. This is done using PMI (Turney, 2002). Hu *et al*. (2012) predicted that the writing style of the reviews depend on the interest of the customer. They worked on website amazon.com and identified that approximately 10.3% of the products are subject to online reviews manipulation.

2.5. Recommender systems

The recommender systems (RS) are the information filtering systems which deal with the large amount of information that is dynamically generated based on users preferences, interests and observed behaviours.

Sentiment based product recommendations have gained research importance in the recent times. The knowledge discovered in terms of product features and opinions from online product reviews among the category of products are useful to the customer in personalized recommendations. These feature level sentiments are aggregated to form the product sentiment. Li Chen and Feng Wang proposed (2014) a novel explanation interface that fuses the feature sentiment information into the recommendation content. They also provided the support for multiple products comparison with respect to similarity using the common feature sentiments. Gurini *et al*. proposed friends recommendation technique in Twitter using a novel weighting function which is called Sentiment-Volume-Objectivity (SVO) that considers both the user interests and sentiments. Li *et al*. (2015) proposed a recommender system that recognizes the sentiment expressions from the reviews, quantified with the sentiment strength and appropriately recommend products according to customer needs. Recently, Dong *et al*. (2016) developed a product recommendation strategy that combines both similarity and sentiments to suggest products.

The work of Wang *et al*. (2017) motivated the authors to take up the current work. It is argued that aspects ranking helps to understand the attention of the consumers in terms of highly ranked aspects. The proposed task is ranking the product aspects using aspect gain ratio and aspect sentiments for estimating the turnovers with aspects sentiments. This helps in recommending the reputed products to the customer.

3. Online product recommendations using statistical reputations

A typical workflow of the proposed workflow is presented in figure 1 below. Initially, the incoming product reviews are pre-processed. The steps in pre-processing the reviews are explained below.

This module is used to pre-process the incoming reviews to a standard format. The steps in pre-processing are namely review tokenization, stop words removal and Part-of-Speech (PoS) tagging. The process of review tokenization divides the sentence into individual tokens. Then, the stop words list is applied on the tokens to remove those words which carry no meaning in the analysis. The stop words are compiled from the reviews itself. This compilation is carried out by sorting the terms in the decreasing order of collection frequency and thereby hand-filtering those terms for their semantic content relative to the domain of the product reviews. Finally, PoS tagging is carried out on the list of filtered tokens to associate the unambiguous word categories with each of the token. The Stanford log-linear Part of Speech tagger is used (Kristina *et al*., 2003) for tagging the tokens.

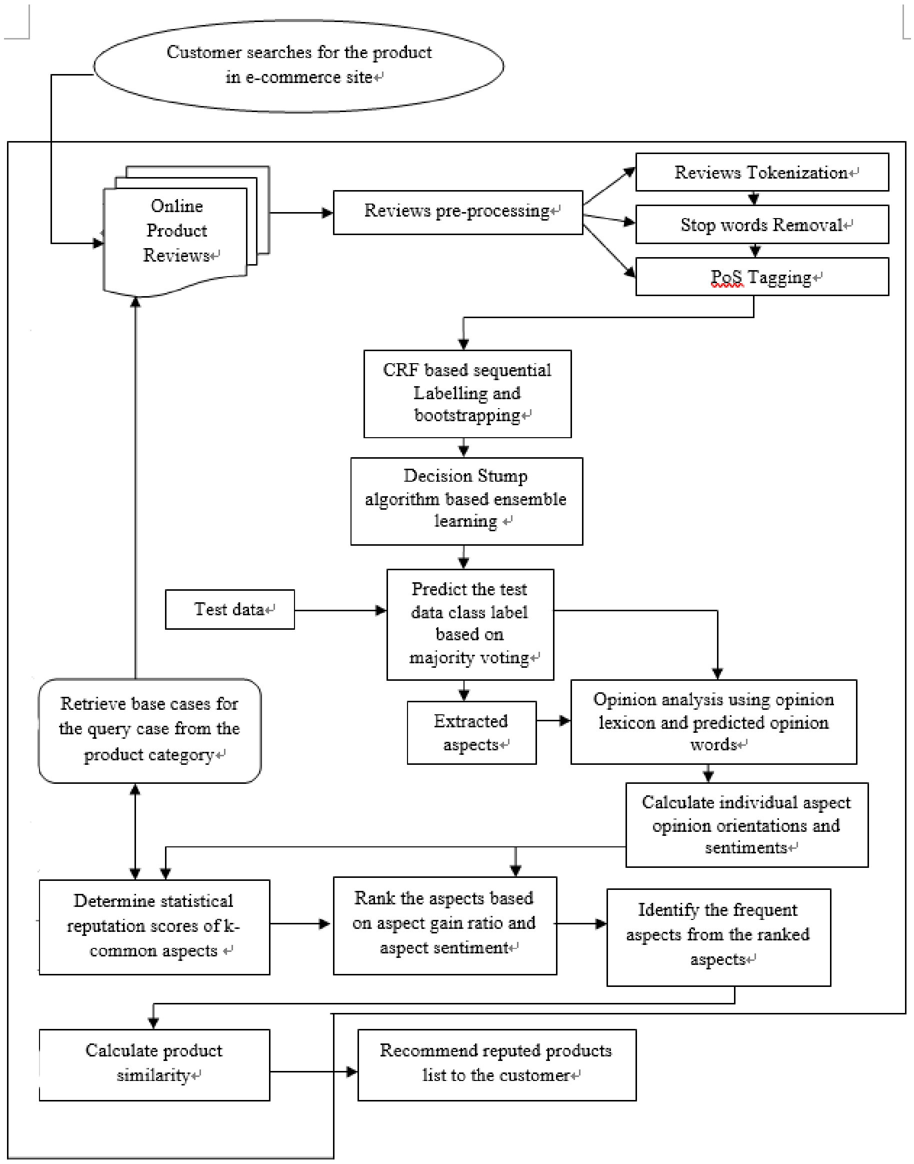

Figure 1. Proposed model

3.1. Product aspects extraction

The pre-processed reviews are labelled sequentially with 'A' for nouns and noun phrases in order to extract them as aspects, 'OW' for adjective terms in order to extract them as opinion words and 'O' for all other PoS tagged terms. The sequential labelling method used is Conditional Random Field (CRF).

Given the sequence of terms in the review sentence, a list of features for each term is encoded for CRF training. These features are as follows:

(1) Review Token: This indicates which word type is the actual instance to be labelled.

(2) PoS tag: The PoS tag of the word.

(3) Class label: The labels are A for aspect terms (nouns), OW for adjectives and O for other PoS terms. Following is the encoding approach of the review terms in the sequence for the review "keyboard and sound are awful".

Review Term	**PoS tag**	**Class label**
keyboard	NN	A
and	CC	O
sound	NN	A
are	VBP	O
awful	JJ	OW

The training of the labelled review tokens is carried out using decision stump classification algorithm. The decision stump classifier in this work is termed as base classifier. The technique used in the training is bagging. It is an ensemble learning technique used to improve the classification performances of various learned weak classifiers. The basic principle of bagging is to manipulate data using bootstrap techniques to produce new derived training sets known as bootstrap replicates (datasets). Bootstrapping is the process of taking a random sample with replacement of the same size (N instances) from the training dataset. By using the process of bootstrapping, approximately 2/3 of the instances in the original training dataset are used. These instances are referred to as in-the-bag instances, while the other 1/3 of instances are called out-of-bag instances. Then a base classifier is trained on each bootstrap dataset and this process is repeated multiple times resulting in multiple classification models.

The performance of a bagging ensemble technique depends on the performance of each individual classifier participating in the ensemble. There are three key parameters for forming an effective bagging ensemble classifier: diversity among classifier members, accuracy of each base classifier, and ensemble size or number of iterations that constitute the ensemble (Fazelpour *et al*., 2016).

The test review tokens are applied against each weak classifier in order to predict the aspect class label. Each weak classifier returned prediction is counted as one vote. The bagged classifier finally counts the votes. When the number of votes for the label 'A' are more than the other labels 'O' and 'OW', then label 'A' is assigned to the test review token. This completes the aspect extraction process.

In the similar manner, when the number of votes for the label 'OW' are more than the other labels 'O' and 'A', then label 'OW' is assigned to the test review token. This completes the opinion word extraction process. In order to verify whether the extracted tokens are the opinion words a comparison is made with the expanded opinion lexicon.

3.2. Opinion orientation using contextual clues and sentiwordnet scores and aspect sentiment calculation

The extracted opinion words are analysed for orientations in the following steps.

(i) The opinion word and the seed terms are assigned with the numerical scores available under adjective category from Sentiwordnet. This is carried out by finding the contextual clues surrounding the opinion word. These contextual clues will help to disambiguate the sense of the opinion word. The contextual clues are finalized based on the typed dependency grammatical relations.

(ii) The distance between the opinion word and the seed term and the distance between the seed terms is calculated as given below.

$$\text{distance}(w_i,w_j) = \text{sentiwordnetscore}(w_i) - \text{sentiwordnetscore}(w_j) \qquad (1)$$

where w_i is either the opinion word or the seed term and w_j is the seed term. The distance measure is modified as the application of distance is carried on non-hierarchical semantic network [40] i.e., on adjectives.

(iii) The semantic orientation (SO) of the opinion word is determined as given below.

$$\frac{SO(\text{opinion word}) \;=\; distance(\text{opinion word},\ bad) \;-\; distance(\text{opinion word},\ good)}{distance(good,\ bad)} \qquad (2)$$

(iv) The opinion word is deemed to be positive when the orientation measurement is greater than zero, and negative otherwise.

The extracted aspects and the corresponding opinion orientations are generated as a pair. Now the positive, negative and neutral opinion orientations on the aspects are counted separately. These counts are used to calculate the sentiment of each feature. The sentiment of a feature is calculated as;

$$Sent(Fi,P) = \frac{Pos\_Opinion\_Count(Fi,P) - Neg\_Opinion\_Count(Fi,P)}{Pos\_Opinion\_Count(Fi,P) + Neg\_Opinion\_Count(Fi,P) + Neu\_Opinion\_Count(Fi,P)} \qquad (3)$$

3.3. Retrieving experiential base cases on the basis of query case using aspect reputation score

From each review R_i the above approaches as described in sections 3.1, 3.2 generate a set of valid aspects A_1, ..., A_{mi}, the opinion orientations with their counts, the associated sentiment values on the aspects. The base cases and the query case are constructed in a straightforward fashion, as a set of product aspects paired with corresponding reputation scores as;

$$\mathrm{Case}(B_i) = \{(A_j, \mathrm{AREP}(A_j, B_i)) : A_j \in \mathrm{Aspects}(B_i)\} \quad (4)$$

$$\mathrm{Case}(Q) = \{(A_j, \mathrm{AREP}(A_j, Q)) : A_j \in \mathrm{Aspects}(Q)\} \quad (5)$$

$$AREP(A,P) = [TPOOC \,?\, TNOOC]*100\% \big/ [TPOOC + TNOOC] \quad (6)$$

where i = 1,2,3,... and j = k-common aspects between base cases and the query case. B is the similar product for the customer searched product (base case) and Q is the customer searched product (query case). The base case aspects (Aspects(B)) for a product B are the union of the valid aspects extracted from its reviews. Each of these aspects is present in a number of B's reviews and with different sentiment scores across the similar products. The set {B,Q} belongs to P.

3.4. Product aspects ranking of both query case and base cases

The extracted aspects are ranked in order to determine the frequent aspects that affect the reputation of the product. This is carried out by measuring the gain ratio measure on each aspect. The reviews collection has both usefulness and uselessness contexts. The product aspect gain ratio for a single product is tabulated in Table 1 below.

Table 1. The product aspect gain ratio for a single product

Aspects	Usefulness	Uselessness	Number of reviews	Gain Ratio
Battery	720	101	1895	0.0341
Performance	480	156	1754	0.1406
Os	404	54	944	0.0327
Brand	187	73	263	0.0017
network connectivity	44	84	701	0.0760
Camera	40	27	210	0.0071
Price	21	10	38	0.0034
Touch	8	5	61	0.0012
Battery life	6	5	53	0.6251

The aspect gain ratio is defined as;

$$GainRatio(Aspect) = \frac{InformationGain(Aspect)}{SplitInformation(Aspect)} \quad (7)$$

where $InformationGain(Aspect)$ is;

$$InformationGain(Aspect) = Entropy(\text{Re}views) - Entropy(Aspect) \quad (8)$$

And $Entropy(\text{Re}views)$ is;

$$Entropy(\text{Re}views) = -\sum_{i=1}^{n} p_i * \log_2(p_i) \quad (9)$$

Where p_i is the probability of any review from Reviews belong to either usefulness context or uselessness context.

And $Entropy(Aspect)$ is;

$$Entropy(Aspect) = Entropy(U \mid Aspect) = -\sum_{j=1}^{2} p_r(u_j \mid Aspect) * \log_2 p_r(u_j \mid Aspect) \quad (10)$$

Where $U=\{u_1,u_2\}$ is the target concept labels for the Reviews and $p_r()$ is the probability function.

And finally $SplitInformation(Aspect)$ is;

$$SplitInformation(Aspect) = -\sum_{k=1}^{m} \frac{count\_of\_aspects\_in\_k\_reviews}{Total\_number\_of\_reviews} * \log_2(\frac{count\_of\_aspects\_in\_k\_reviews}{Total\_number\_of\_reviews})$$

(11)

In order to strengthen the rank of the aspect, the GainRatio of the aspect is weighed with the sentiment value of each aspect. The Rank of the aspect is calculated as;

$$Rank(Aspect) = GainRatio(Aspect) * Sentiment\_score(Aspect) \quad (12)$$

After calculating the rank of each aspect, all the aspects are sorted in the decreasing order.

3.5. Identification of frequent aspects from the ranked aspects

In order to find out the frequent aspects from the ranked and sorted aspects an empirical evaluation is carried out on the reviews collection. A careful analysis of the corpus revealed that an aspect is considered to be frequent if its occurrence in the reviews is greater than or equal to three percent from the set of aspects that are available. The precision (in %) in the number of frequent aspects extracted is tabulated for the datasets in Table 2 below.

Table 2. Percentage of Product aspects extracted as frequent aspects at different sizes

Size	C1 Precision (%)	C2 Precision (%)	C3 Precision (%)	C4 Precision (%)	C5 Precision (%)
Greater than or equal to 1	61.3	57.3	59.1	60.3	57.8
Greater than or equal to 2	75.8	70.9	72.4	71.3	73.4
Greater than or equal to 3	86.3	87.8	85	86.4	87.3
Greater than or equal to 4	67.2	59.8	50.7	57.6	62.3

The results of the empirical evaluation for extracting frequent aspects carried out on five datasets at four different percentages are shown in Figure 2 below.

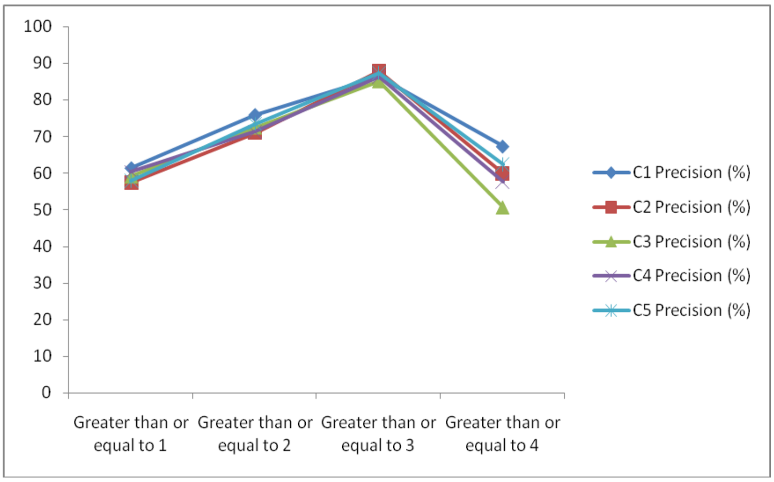

Figure 2. Empirical Evaluation for frequent aspects threshold

The observation from Figure 2 is that at greater than or equal to 4% size, the precision in the number of extracted frequent aspects started to decrease.

3.6. Determination of product similarity between query case and base case using statistical reputations

The case bases of experiential cases are generated earlier to provide aspects reputations based product recommendations. First it is important to understand that these experiential cases do not have fixed set of shared static aspects. Instead each case is represented by its own set of aspects extracted from its own product reviews. There must be a guarantee that some minimal set of shared aspects are available between cases to serve as the basis for comparison.

In order to do this, a k-comparability is defined. It is a boolean property of two cases Bu and Bv which is true if and only if Bu and Bv share at least k aspects. During retrieval, those base cases that are at least k-comparable (have at least k aspects in common) with the target query case Q are considered, as seen in Equation 13. CB denotes the case base of all product cases.

$$\text{Retrieve}_k(Q) = \{B \in CB: k - \text{comparable}(Q, B)\} \tag{13}$$

The product similarity is calculated by using the traditional cosine similarity measure which is given as follows.

$$Cos(Q, B_i) = \frac{\sum_{i=1}^{n} Q.B_i}{\sqrt{\sum_{i=1}^{n} (Q*Q)*(Bi*Bi)}} \tag{14}$$

Finally, the recommender system provides to the customer with the list of reputed products in the decreasing order of similarity.

4. Results and discussion

The datasets used for the task of reputation analysis are the collection of five categories of product reviews from Amazon. GPS devices (C1), Tablets (C2), Laptops (C3), Smart phones (C4) and cameras (C5) are the product categories for which the reviews are considered for analysis. In each product category, 100 products are considered from the E-commerce application. Table 3 presents the details of the datasets used for this experiment.

The pre-processing of data is carried out by removing stop words and non English words. PoS tagging is performed on the obtained set of words.

In order to carry out aspect extraction from the pre-processed reviews the bagging ensemble technique is utilized. In the model training process, the review tokens are labelled manually for the respective class label in CRF way. Then, this labelled data is random sampled with replacement. This caused the problem of class imbalance. In order to alleviate this problem and also to improve classification performance, the effects that ensemble size (number of iterations) has on the classification performance of imbalanced dataset are investigated.

Table 3. Dataset details

Document attributes	Values
Number of review documents	67458
Minimum sentences per review	2
Maximum sentences per review	43
Average number of reviews written by customers	3.78
Average number of reviews written on the product	28.47

To accomplish this, an extensive experimental study using four ensemble sizes (10, 20, 50, and 100 iterations) within the bagging process across 15 sampled imbalanced datasets is carried out. The results are shown in table 4 below.

Table 4. Classification results

Ensemble Size	Bagging with Decision Stump
10	73.4%
20	75.8%
50	79.9%
100	77.6%

It is observed from the above table results demonstrate is that the ensemble size 50 is a better choice for all scenarios. The precision, recall and F1-scores on the extracted aspects by using Decision Stump based Ensemble classifier on the five product categories is tabulated in table 5 below.

Table 5. Accuracy of the decision stump based Ensemble on the extracted aspects

Category	Precision (%)	Recall (%)	F1-Score (%)
GPS Devices	86.9	70.5	77.8
Tablets	87.6	74.3	80.4
Laptops	88.5	72.9	79.9
Smart phones	89.2	75.8	81.9
Cameras	88.3	79.1	83.4

The precision, recall and F1-scores on the extracted opinion words by using Decision Stump based Ensemble classifier on the five product categories is tabulated in table 6 below.

Table 6. Accuracy of the decision stump based Ensemble on the extracted opinion words

Category	Precision (%)	Recall (%)	F1-Score (%)
GPS Devices	87.6	72.4	79.2
Tablets	88.4	76.6	82.1
Laptops	87.7	71.2	78.6
Smart phones	89.1	78.2	83.3
Cameras	88.9	77.9	83.0

The results from the above two tables specify that the precision acquired on aspects extraction is 88.1% and 88.34% of precision on opinion words extraction. It is observed that a better increase of 4% is obtained when compared with the precisions of both Naive Bayes classifier based ensemble and Support Vector Machine (SVM) classifier based ensemble in both the extraction processes. This shows that the decision stump classifier based ensemble for product aspects extraction and their opinion words extraction has provided better performance than Naive Bayes and SVM classifiers based ensemble in terms of precision and recall.

Table 7. Sentiments of three smart phones

	Iphone 6s plus		**Oppo f1 plus**		**Nokia Lumia 525**	
Product Aspect	**Sentiment Score**	**GainRatio**	**Sentiment Score**	**GainRatio**	**Sentiment Score**	**GainRatio**
Battery	1	0.1356	0.2	0.0627	1	0.0341
Performance	0.6	0.4928	1	0.2671	0.5	0.1406
Os	1	0.0864	0.14	0.0579	-1	0.0327
Brand	1	0.0142	0.6	0.0052	0.42	0.0071
network connectivity	-1	0.0047	1	0.0145	1	0.0060
Camera	1	0.5894	0.23	0.0234	0.77	0.0017
Price	0.57	0.0237	-0.33	0.0178	0.07	0.0034
Touch	1	0.9741	1	0.3687	-1	0.5386
battery life	1	0.8529	-1	0.4328	1	0.6251

The sentiments and gain ratios of the smart phones products are tabulated in table 7 below. The further analysis for product reputation classification is carried out on smart phones. This is because more number of customers are showing interest in purchasing various smart phones in the current scenario.

The ranks of these aspects across the three smart phones are tabulated in table 8 below.

Table 8. Aspect Ranks of three smart phones

	Iphone 6s plus	**Oppo f1 plus**	**Nokia Lumia 525**
Product Aspect	**Rank**	**Rank**	**Rank**
Battery life	0.8529	-0.4328	0.6251
Performance	0.29568	0.2671	0.0703
Battery	0.1356	0.01254	0.0341
Network connectivity	-0.0047	0.0145	0.006
Brand	0.0142	0.00312	0.002982
Camera	0.5894	0.005382	0.001309
Price	0.013509	-0.00587	0.000238
Os	0.0864	0.008106	-0.0327
Touch	0.9741	0.3687	-0.5386

It is observed from the above table that the ranked aspects in the decreasing order for Iphone 6s plus are Touch, Battery life, Camera, Performance, Battery, Os, Brand, Price and Network Connectivity. The ranked aspects in the decreasing order for Oppo f1 plus are Touch, Performance, Network Connectivity, Battery, Os, Camera, Brand, Price and Battery Life. The ranked aspects in the decreasing order for Nokia Lumia 525 are Battery Life, Performance, Battery, Network Connectivity, Brand, Camera, Price, Os and Touch.

It is observed from the above list of ranked aspects that the same aspects from different products of the smartphone category have altogether different positions in terms of ranks. Also it is understood from this observation is that the e-commerce customers tend to pay more attentions on the aspects with higher ranking.

The frequent aspects identified from these three products aspects list after careful analysis for the occurrence in the reviews to be greater than or equal to three percent are Battery life, Touch Performance, Battery, Brand and Camera respectively.

It is observed from the above list of frequent aspects that the rank of the aspects has no relation with the frequency of the aspect. This is because the incorporation of gain ratio measure into the aspect ranking.

In order to compare the reputations of the k-common aspects of the three products for providing recommendations, Cosine Similarity is considered. The k-common aspects identified from the single category products and their reputation scores after the customer searched for Iphone 6s plus are tabulated below in table 9. The value of k is found is 6.

Table 9. List of k-common aspects and their reputation scores

k-common aspects	AREP(Q)	AREP(B1)	AREP(B2)
Battery life	0.494	0.507	0.612
Touch	0.851	0.666	-0.772
Performance	0.454	0.684	0.526
Battery	0.877	0.843	0.90
Brand	0.688	0.371	0.731
Camera	0.93	0.444	0.191

The value of 'k' is confined to 6 as these are the at least aspects that are shared among the considered single category products for evaluation. The product recommendations are based on the reputations of these 6 common aspects.

The variations in the number of k-common aspects on the similar products using reputations and cosine similarities are tabulated in table 10 below.

Table 10. Variations in k and cosine similarities with the queried product

k	Cosine(Q,B1)	Cosine(Q,B2)
2	0.87	0.79
3	0.45	0.38
4	0.54	0.51
6	0.29	0.48

The product similarity with the queried product using sentiments of the k-common aspects is displayed in Figure 3 below.

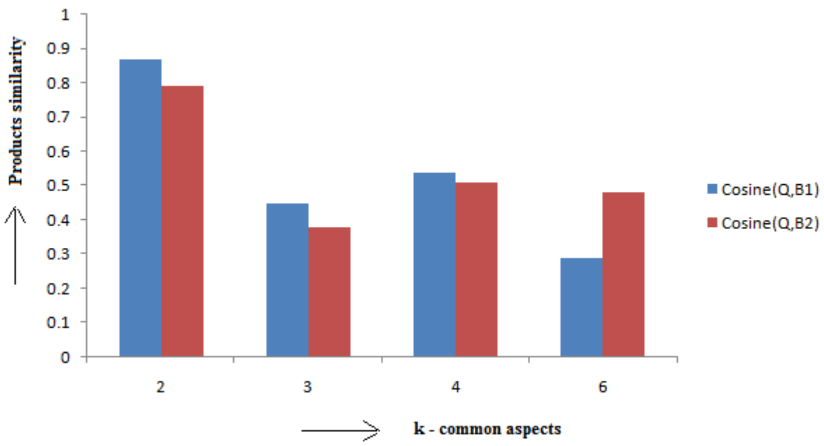

Figure 3. Product similarity with the queried product using reputations of the k-common aspects

From the results presented in above table it is observed that for different values of 'k' (2,3,4,6) the cosine similarity returned the similar products as recommendations in the same order (product B1 comes first in the list and then the product B2) by using the reputations on k-common aspects. The product with higher cosine value between two similar products is shown as first product in the recommendations list. But for k value of 6, the order in the product recommendations was changed. This is because the product P2 has higher cosine value and P1 has lower cosine value when compared with the searched product.

In order to evaluate the utility of the recommendations produced by the recommender system, Precision, Recall and F1-score metrics are used. The formulae for precision, recall and F1-score are given below.

$$\frac{Precision = |good\ products\ recommended|}{|all\ recommendations|} \tag{15}$$

$$\frac{Recall = |good\ products\ recommended|}{|all\ good\ recommendations|} \tag{16}$$

$$\frac{F1-score\ =\ 2\ *Precision\ *\ Recall}{Precision\ +\ Recall} \tag{17}$$

The parameters provided in table 11 below compares the information retrieval metrics on the product recommendations between the works carried out by Wang and Wang (2015) with the results obtained from the current work. They used user opinions that are written in online reviews as preferences to recommend the products through sentiment analysis. They used Collaborative Filtering based recommender system to provide recommendations. The ‘k’ value is the number, where similar product preferences are given by the number of users. In the current work, the aspect level reputations were calculated in terms of opinion orientation counts. The recommendation system implemented was case based recommender model. The ‘k’ value is the number of common product aspects considered for calculating the similar product recommendations.

Table 11. Comparison of information retrieval measures on the product recommendations

RS type		‘k’ type		‘k’ value		Precision (%)		Recall (%)		F1-Score (%)	
Opinion-enhanced CF based RS model (Wang Work)	Reputation based RS model (Present work)	No. of users with similar product preferences	No. of common aspects among the similar products	20	6	10	50	6	100	75	67

The recall value from the Table 8 specify that the recommender system was able to provide better recommendations when compared with the recommendations produced by Wang and Wang (2015) in their work on single category recommendations. This shows that opinion orientation counts used in calculating the reputations of the product aspects improves the product recommendations.

5. Conclusion and future work

The online product recommendations using the reputations of the product aspects were carried out successfully. The objectives are to provide the product recommendations using AREP measure and to support the customer with better purchase decisions. Decision stump based Ensemble was developed for aspects and their opinions extraction. In the process of this work it is found that the ensemble size 50 is a better choice for the machine learning process. It was shown that the decision stump classifier based ensemble for product aspects extraction and their opinions extraction has provided better performance than Naive Bayes and SVM classifiers in terms of precision and recall. Also, the experimental results in terms reputation score based product recommendations indicate that the proposed model is effective.

In future, the product cases retrieval is further improved by working on different aspect weighting approaches.

References

Chen G. L., Chen L. (2014). Recommendation based on contextual opinions. *International Conference on User Modeling, Adaptation, and Personalization*, pp. 61-73. https://doi.org/10.1007/978-3-319-08786-3_6

Cruz F. L., Troyano J. A., Enríquez F., Ortega F. J., Vallejo J. A. (2013). 'Long autonomy or long delay?' The importance of domain in opinion mining. *Expert Systems with Applications*, Vol. 40, No. 8, pp. 3174-3184. https://doi.org/10.1016/j.eswa.2012.12.031

Dong R. H., O'Mahony M. P., Schaal M., McCarthy M., Smyth B. (2016). Combining similarity and sentiment in opinion mining for product recommendation. *Journal of Intelligent Information Systems*, Vol. 46, No. 2, pp. 285-312. https://doi.org/10.1007/s10844-015-0379-y

Eirinaki M., Pisal S., Singh J. (2012). Feature-based opinion mining and ranking. *Journal of Computer and System Sciences*, Vol. 78, No. 4, pp. 1175-1184 https://doi.org/10.1016/j.jcss.2011.10.007

Fazelpour A., Khoshgoftaar T. M., Dittman D. J., Naplitano A. (2016). Investigating the variation of ensemble size on bagging-based classifier performance in imbalanced bioinformatics datasets. *Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on. IEEE*, pp. 377-383. https://doi.org/10.1109/IRI.2016.57

Gârbacea C., Tsagkias M., de Rijke M. (2014). Detecting the reputation polarity of microblog posts. *Frontiers in Artificial Intelligence and Applications*, Vol. 263, pp. 339-344. http://dx.doi.org/10.3233/978-1-61499-419-0-339

Ghose A., Ipeirotis P. G., Li B. B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, Vol. 31, No. 3, pp. 493-520. https://doi.org/10.1287/mksc.1110.0700

Guo J. L., Peng J. E., Wang H. C. (2013). An opinion feature extraction approach based on a multidimensional sentence analysis model. *Cybernetics and Systems*, Vol. 44, No. 5, pp. 379-401. https://doi.org/10.1080/01969722.2013.789649

Hu N., Boseb I., Kohc N. S., Liua L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, Vol. 52, No. 3, pp. 674-684. https://doi.org/10.1016/j.dss.2011.11.002

Jiao J., Zhou Y. Q. (2011). Sentiment polarity analysis based multi-dictionary. *Physics Procedia*, Vol. 22, pp. 590-596. https://doi.org/10.1016/j.phpro.2011.11.091

Kim S. M., Hovy E. (2004). Determining the sentiment of opinions. *Proceedings of the 20th international conference on Computational Linguistics*, pp. 1-7. https://doi.org/10.3115/1220355.1220555

Kristina T., Klein D., Manning C., Singer Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. I*n Proceedings of HLT-NAACL*, pp. 252-259. https://doi.org/10.3115/1073445.1073478

Li X., Wang H. M., Yan X. W. (2015). Accurate recommendation based on opinion mining. *Genetic and Evolutionary Computing*, pp. 399-408. https://doi.org/10.1007/978-3-319-12286-1_41

Miller G. A., Beckwith R., Fellbaum C., Gross D., Miller K. J. (1990). Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, Vol. 3, No. 4, pp. 235-244. https://doi.org/10.1093/ijl/3.4.235

Mohammad S., Dunne C., Dorr B. (2009). Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Vol. 2, pp. 599-608. https://doi.org/10.3115/1699571.1699591

Pan L. Y., Chiou J. S. (2011). How much can you trust online information? Cues for perceived trustworthiness of consumer-generated online information. *Journal of Interactive Marketing*, Vol. 25, No. 2, pp. 67-74. https://doi.org/10.1016/j.intmar.2011.01.002

Pang B., Lee L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, Vol. 2, No. 1-2, pp. 1-135. http://dx.doi.org/10.1561/1500000011

Perales C., Hugo M., Rodriguez-Oliveros R. (2017). Sentiment analysis as reputational risk indicator. *SSRN Electronic Journal*, pp. 1-15. http://dx.doi.org/10.2139/ssrn.3051870

Poria S., Cambria E., Gelbukh A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, Vol. 108 pp. 42-49. https://doi.org/10.1016/j.knosys.2016.06.009

Qiu G., He X. F., Zhang F., Shi Y., Bu J. J., Chen C. (2010). DASA: Dissatisfaction-oriented advertising based on sentiment analysis. *Expert Systems with Applications*, Vol. 37, No. 9, pp. 6182-6191. https://doi.org/10.1016/j.eswa.2010.02.109

Rada R., Mili H., Bicknell E., Blettner M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 1, pp. 17-30. https://doi.org/10.1109/21.24528

Ravi Kumar K. C., Teja Santosh D., Vishnu Vardhan B. (2017). Extracting opinion targets from product reviews using comprehensive feature extraction model in opinion mining. *Indian Journal of Science and Technology*, Vol. 10, No. 21, pp. 1-6. https://doi.org/10.17485/ijst/2017/v10i21/105733

Silvestro F., Bagnasco A., Lanza I., Massucco S., Vinci A. (2017). Energy efficient policy and real time energy monitoring in a large hospital facility: A case study. *International Journal of Heat and Technology*, Vol. 35, No. S1, pp. S221-S227. https://doi.org/10.18280/ijht.35Sp0131

Teja Santosh D., Vishnu Vardhan B., Ramesh D. (2016). Extracting product features from reviews using Feature Ontology Tree applied on LDA topic clusters. *Advanced Computing (IACC), 2016 IEEE 6th International Conference on. IEEE*, 2016 https://doi.org/10.1109/IACC.2016.39

Turney P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417-424. https://doi.org/10.3115/1073083.1073153

Wang W., Wang H. W. (2015). Opinion-enhanced collaborative filtering for recommender systems through sentiment analysis. *New Review of Hypermedia and Multimedia*, Vol. 21, No. 3-4, pp. 278-300. https://doi.org/10.1080/13614568.2015.1074726

Wang W., Wang H. W., Song Y. (2017). Ranking product aspects through sentiment analysis of online reviews. *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 29, No. 2, pp. 227-246. https://doi.org/10.1080/0952813X.2015.1132270

Xu K. Q., Liao, S. Y., Li J. X., Song Y. X. (2011). Mining comparative opinions from customer reviews for competitive intelligence. *Decision Support Systems*, Vol. 50, No. 4, pp. 743-754. https://doi.org/10.1016/j.dss.2010.08.021