A NOVEL APPROACH TO RATE AND SUMMARIZE ONLINE REVIEWS ACCORDING TO USER-SPECIFIED ASPECTS

Hsiao-Wei Hu
School of Big Data Management, Soo Chow University
No.70, Linshi Rd., Shihlin Dist., Taipei City 111, Taiwan
camihu@gmail.com

Yen-Liang Chen
Department of Information Management, National Central University
No. 300, Jhongda Rd., Jhongli District, Taoyuan City 32001, Taiwan
ylchen@mgt.ncu.edu.tw

Po-Tze Hsu
Department of Information Management, National Central University
No. 300, Jhongda Rd., Jhongli District, Taoyuan City 32001, Taiwan
grace60235@hotmail.com

ABSTRACT

As internet use expands, the reviews found on e-commerce websites have greater influence on consumer purchasing decisions. One popular practice of these websites is to provide ratings on predefined aspects of the product, thereby enabling users to obtain summaries of vital information. One limitation of this approach is that rating and summary information is unavailable for aspects of the product that are not predefined by the website. In light of this weakness, this paper proposes a new approach that allows the user to specify the product aspects in which he is interested, whereupon the system automatically classifies and rates all of the online reviews according to those specific aspects. It is worth noting that the proposed method could also assists enterprises to identify the issues of importance to users, which would otherwise be hidden. An understanding of their concerns could be used as a reference in efforts to improve the internal environment and implement service innovations, thereby enhancing customer satisfaction and increasing competitiveness. Analysis of several datasets of hotel reviews made it possible to ascertain the following information for target hotels: (1) the percentages of positive, neutral, and negative comments on various aspects of hotels, as specified by users, (2) average ratings with regard to the aspects specified by users, and (3) categorization of reviews based on specified aspects. Our approach offers the following advantages over current website practices: (1) the functions of our approach are compatible with and can be installed on current e-commerce websites to improve services, (2) users can obtain a summary of information according to their own interests, and (3) our analysis allows users to easily visualize groups of similar opinions.

Keywords: Opinion mining; Sentiment analysis; Normalized google distance; K-means; Online reviews

1. Introduction

The widespread use of network and information technology has led to a wide number of conventional commercial activities being performed online. Many e-commerce systems allow customers to express their opinions regarding the products they have purchased and review the comments posted by previous customers. This option is offered in the hopes of providing reliable, trustworthy information and improving the services they provide. For example, on the hotels.com website, prospective customers can read the reviews written by previous guests about a hotel in which they may be interested. Because these reviews reveal the real experiences of previous customers, they exert a powerful influence on potential customers [Duan et al. 2008; Lu et al. 2014; Zhu et al. 2014; Purnawirawan et al. 2014].

Along with customer reviews, many websites also provide summarized rating information on various predefined aspects of their products and/or services. This helps users to assess review content as quickly as possible [Hu et al. 2012; Gu et al. 2013; Wan & Hakayama 2014]. However, rapid advances in business data analytics have led customers to expect more than just accurate information; they expect better service in the form of information that is both accurate and customized to their needs [Tam & Ho 2006]. Why does customized or personalized information matter? Thirumalai and Sinhab [2011] claim that decision customization that provides choice assistance is positively associated with customer satisfaction. Tam and Ho [2006] also claim that content relevance, self-reference, and goal specificity affect the attention, cognitive processes, and decisions of web users in a variety of ways. In other words, users are receptive to personalized content and find it useful as an aid to decision-making. Although traditional review functions are useful, they fail when the interests of users fall outside the product aspects predefined by the website.

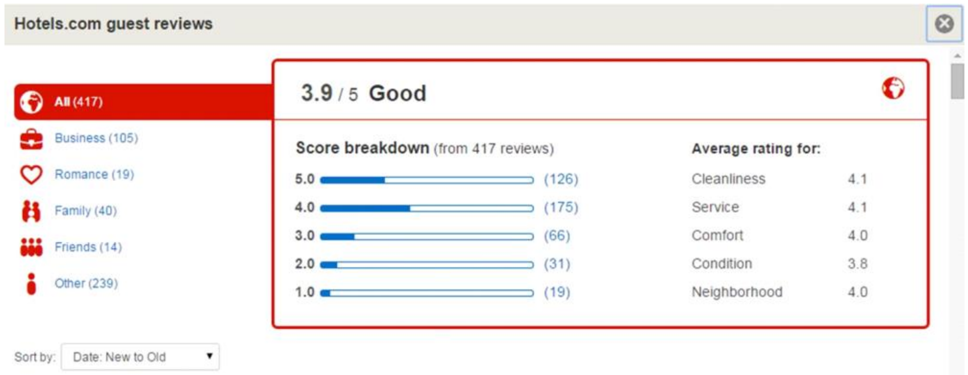
Figure 1 illustrates how online consumer reviews often fail to meet consumer needs. Most existing review websites offer summarized ratings for various aspects of a product, which enables consumers to quickly grasp the content of reviews. In this example, we consider two well-known hotel review websites: hotels.com and booking.com. Hotels.com provides average ratings for each hotel according to the following five predefined aspects: cleanliness, service, comfort, conditions, and neighborhood. Meanwhile, booking.com provides average ratings for each hotel according to the following seven predefined aspects: cleanliness, staff, comfort, facility, location, value for money and free WiFi. It is worth noting that if a customer is interested in the “value for money” or “free WiFi” of a hotel, then hotels.com does not provide the summarized ratings required by the user, because these aspects are not part of their system. A consumer interested in these aspects can then only compare and evaluate the hotels by examining each relevant review one by one, which can be very time-consuming. Or worse yet, a consumer may have an unsatisfactory experience due to the fact that he failed to obtain the required information, leading him to migrate to other websites.

In other words, while information regarding the predefined aspects is helpful in enabling customers to quickly evaluate hotels, it is difficult to acquire an accurate summary from the enormous number of reviews on a website when consumers have unique requirements that are not predefined in the system.

This study therefore proposes a methodology for the rating and clustering of online reviews according to user-specified aspects. For the purposes of testing the proposed methodology, we used hotels.com as a study target; however, our methodology is not specific to this site.

Hotels.com fits the above-mentioned characteristics, in that the system provides an overall rating as well as a summary of average ratings related to cleanliness, service, comfort, conditions, and neighborhood. It also displays many reviews for each hotel, and each review is composed of many sentences.

In this research, we used an opinion mining method to extract implicit opinions; i.e., sentences from reviews are classified according to specific aspects. Analysis is then used to determine the sentiment polarity of these sentences. The use of these methods enabled the formation of a sentiment table for specific aspects of a hotel, showing the numbers of positive, neutral, and negative opinions/sentences in the review. By summarizing the sentiment tables of all reviews for a specific hotel, we obtain an overall sentiment table at the hotel level. Furthermore, this enables the aggregation of values in the sentiment table for a target hotel and makes it possible to obtain ratings related to the performance of the target hotel with regard to each specified aspect. Finally, the sentiment tables of all reviews can be clustered to reveal an aggregate, i.e., an overall opinion with regard to the target hotel.



(a) Five predefined aspects of a hotel in hotels.com

(b) One of the reviews of a hotel in hotels.com

(c) Six predefined aspects of a hotel in booing.com

Hotels.com	Booking.com
Cleanliness	Cleanliness
Service	Staff
Comfort	Comfort
Conditions	Facilities
Neighborhood	Location
	*Value for money
	*Free WiFi

(d) A comparison of predefined aspect between hotels.com and booking.com

Figure 1: Examples of Hotel Websites

The contributions of this study are three-fold. First, regardless of whether a website provides summarized ratings related to pre-defined aspects, the proposed method enables users to obtain the specific information in which they are interested. The proposed method makes it possible to extend the capabilities of review websites so that the needs of users can be met in a more flexible and dynamic manner. Second, using these methods would allow users to quickly evaluate and compare hotels without the need to spend lengthy amounts of time reading through reviews. Third, since consumer perspective is an important reference for enterprises in product innovation and service improvement [Chen & Chen 2015], the proposed method gives enterprises a new channel by which to gain an objective understanding of the perspective of consumers through the collection of user-specified product aspects. The consumer perspectives provide an important reference to understand the internal environment and service innovation of enterprises, and thereby increase their competitiveness [Hennig-Thurau et al. 2004].

The remainder of this paper is organized as follows. Section 2 presents background and related literature. Section 3 details the proposed methodology. Section 4 outlines real datasets collected from hotels.com to demonstrate the validity of our proposed method as well as its effectiveness in analyzing and comparing comments related to actual hotels. Finally, conclusions are drawn and possible future work is discussed in Section 5.

2. Background and Related Literature

This section gives an overview of popular websites containing hotel reviews. We then introduce web personalization as well as existing approaches to opinion mining, comment summarization, and clustering.

2.1. Hotel Web Sites

Numerous hotel websites offer customer reviews, collect information about hotels from around the world, and provide online booking services. After customers book rooms and avail themselves of hotel services, they can write reviews in order to share their experiences and opinions with others. Based on these reviews, other users decide whether to reserve a room in a particular hotel. Most of these websites have similar online comment mechanisms. Figure 1 presents screenshots from two well-known hotel websites that provide online reviews: hotels.com and booking.com. In these examples, the websites supply customer review classifications based on reviewer profiles. For example, the customer might define him/herself as a “business” or “family” traveler. Customers then give scores based on features predefined by the website. These websites accumulate a large number of written reviews and provide average ratings for each hotel according to predefined aspects such as cleanliness, comfort, location, services, facilities, staff, value for money, free WiFi, condition of the hotel and neighborhood as well as an overall evaluation. After staying in the hotel, customers can assign a score for each aspect within a predefined range, which appears as an average on the website. Thus, customers have access to every individual review and obtain a sense of the average performance of the target hotel when it comes to these predefined aspects.

The classification system used by hotels.com to define the reviewer type is shown in Fig.1(a), where reviewers are classified into six types: all, business, romance, family, friends, and others. Review results for a given hotel can be filtered according to a specified reviewer type. Scores are then broken down into the five predefined aspects of cleanliness, service, comfort, condition, and neighborhood. The score assigned to each aspect is a number between 1 and 5. The site shows the average score of the target hotel for each of the five built-in aspects. In addition, the site also shows the overall average rating for the hotel.

Although these ratings can help users to understand how well each hotel performs with regard to these predefined aspects, it is unable to provide ratings for other aspects. Furthermore, the classification of reviews in these systems is based on the type of reviewer. It would be useful to allow users to (1) specify which aspects of the review content they want to explore and (2) cluster all of the reviews into groups of users with similar perspectives.

2.2. Web personalization

Web personalization is considered the most highly evolved form of automation for the customization of web content according to the needs of users. Recent differentiation strategies to attract and retain users have therefore emphasized web personalization techniques [Ho & Ho 2008]. The immediate objective of personalization technologies is to elucidate user preferences and the context of the search in order to deliver highly-focused relevant content. The long-term objective is to generate business opportunities and increase customer satisfaction [Ho & Ho 2008, Thirumalai & Sinhab 2011]. Tam and Ho [2006] claimed that users are receptive to personalized content and find it useful as an aid in decision-making.

Enterprises employ personalization technologies in a variety of ways, with the aim of generating business opportunities. Some enterprises use personalization technologies as recommenders in the hope of generating selling opportunities [Wang & Benbasat 2005]. Personalization technologies are also used to arrange the index of product pages dynamically, based on click-stream analysis to reduce the search effort required by users. Researchers have also examined the persuasive effects of personalization on user decision-making [Xua et al. 2011, Karimi et al. 2015], to ease business-to-consumer interaction [Ardissono et al. 2002], and to eliminate aimless surfing activities [Shafiq et al. 2015, Hawalah & Faslia 2015, Shahabi and Banaei-Kashani 2003].

Unlike traditional online review services that provide summarized rating related to predefined aspects of products, this study proposes a means of rating and summarizing online reviews according to user-specified aspects, in order to reduce the search effort required by users and to provide information specific to their needs, particularly when the aspects in which they are interested are not predefined in the system.

2.3. Opinion Mining

Opinion mining is the process by which implicit opinions are extracted from comments through sentiment analysis and subjectivity analysis [Pang & Lee 2008]. Opinion mining is used to identify the opinions of users all over the web [Pang & Lee 2008] and is applicable in a variety of domains [Liu & Zhang 2012]. Opinion mining is generally applied in five types of application: product reviews [Bai 2005, Duan et al. 2008, Hu et al. 2011, Jansen et al. 2009, Lee et al. 2008, Li et al. 2010, Pang et al. 2002, Popescu & Etzioni 2005, Scaffidi et al. 2007], business and government intelligence [Archak et al. 2007, Connor et al. 2010, Diakopoulos & Shamma 2010], recommendation systems [Tatemura 2000], stock market prediction[Gu et al. 2006, Bollen et al. 2011] and political inclinations [Larsson & Moe 2011, Papacharissi & de Fatima Oliveira 2012, Tumasjan et al. 2011, Williams & Gulati 2008].

The core of the opinion mining process comprises three steps, involving analysis at the word level, sentence level, and document level [Missen et al. 2013].

First, word-level polarity orientation (determining whether a word is positive or negative) and polarity strength (determining the strength of meaning in a word) are computed. Two approaches have been proposed for word-level processing: the corpus-based approach and the dictionary-based approach. The corpus-based approach exploits inter-word relationships in large corpora. An example of this approach includes the use of language constructs [Hatzivassiloglou & McKeown 1997; Wilson et al. 2005] and evidence of co-occurrence [Baroni & Vegnaduzzo 2004]. The dictionary-based approach uses specific dictionaries, which are custom designed to determine word polarity and strength. An example of this approach involves analyzing the subjectivity, polarity, and strength of words using WordNet [Miller 1995], SentiWordNet [Esuli & Sebastiani 2006], or WordNet-Affect [Valitutt 2004]. In this study, we used SentiWordNet 3.0 for the computation of similarities between adjectives in order to obtain a score with which to rate the sentiment polarity of a word.

Sentiment polarity and polar strength at the sentence-level are based on the results of word-level analysis. A sentiment score at the sentence-level or word-level can be represented by sentiment polarity and polar strength. Two approaches have been used to determine the subjectivity of sentences; testing for the presence of subjective words [Zhang et al. 2009] and identifying similarities among sentences [Yu & Hatzivassiloglou 2003]. Determining sentence polar strength involves obtaining a sentiment score at the sentence-level (sentence score) from the sentiment score at the word-level (word score). Various methods have been devised to compute a sentence score from the word level: assessing the number of polar words [Hu & Liu 2004], assessing word-level polarity scores [Yu & Hatzivassiloglou 2003], and assessing word-level context-aware polarity [Ku et al. 2006], in which the impact of neighboring words and sentiment words are considered. In this study, we used the scores for sentiment words for the computation of sentence scores.

Sentiment polarity and polar strength at the document-level are based on the results of sentence-level assessments. Sentiment analysis at the document-level can be obtained by assessing the sentiment polarity and strength at the sentence-level and word-level. Sentiment analysis at the document level can be divided into three major approaches:

Corpus-based dictionaries: An opinion lexicon is used to identify documents in which opinions are stated, wherein lexicons are prepared using a given test corpus [Gerani et al. 2009; Hui Yan & Si 2006].

Ready-made dictionaries: The use of document-independent ready-made dictionaries, such as General Inquirer [Kennedy & Inkpen 2006] or SentiWordNet [Zhang & Zhang 2006].

Text classification: The problem is treated as a text classification problem [Aue & Gamon 2005, GuangXu et al. 2007], using classification attributes including the number of subjective words/sentences in a document and the number of positive/negative words/sentences in a document. Classification can be conducted according to supervised learning, semi-supervised learning, or unsupervised learning methods.

This study used statistical analysis of reviews to determine the sentiment polarity and strength of reviews of a target hotel with an unsupervised clustering method for the classification of reviews.

Many applications have been developed in the field of opinion mining. The proposed method provides greater flexibility than that of previous solutions by enabling users to discover opinions relevant to their personal interests without the restrictions associated with predefined aspects/perspectives.

2.4. Comment Summarization

Comment summarization is the process of distilling a large amount of textual data within a small but representative package [Hu & Liu 2004]. Opinion mining can help to identify components for use in the expression of opinions, which makes it fundamental to the process of summarization [Ku et al. 2005]. Comment summarization is widely used on E-commerce websites and applications that employ product reviews [Tang et al. 2009]. Opinion mining has been used to summarize the opinions of numerous movie reviews [Zhuang et al. 2006] through text mining techniques that identify similarities between sentences with regard to sentiment polarity. Opinions related to product features predefined by the user can also be extracted from comments in order to create a summarized review [Wang et al. 2013]. Hu and Liu [Hu & Liu 2004] and Wang et al. [Wang et al. 2013] generated summaries for consumer reviews from Amazon; Zhuang et al. [Zhuang et al. 2006] summarized movie reviews from IMDB; and Meng and Wang [Meng & Wang 2009] generated summaries from reviews on ZOL.com, the largest 3C online store in China.

Comment summarization has proven successful in helping users to quickly understand the main points expressed in reviews; however, the methods in this study differ in two fundamental ways: 1) Traditional methods generate summarized comments, while the proposed method generates summaries in the form of sentiment tables and ratings; 2) Traditional methods focus on generating summaries that best represent the original information, while our methods generates summaries that best reflects the demands or interests stipulated by users.

2.5. Cluster analysis

Cluster analysis (i.e., clustering) is the division of data into groups of similar objects. This method can be viewed as a data modeling technique that provides concise summaries of data. Clustering is found in many disciplines and plays an important role in a broad range of applications such as business intelligence [Chen et al. 2012], image pattern recognition [Zhang et al.2012], web searches [Di Marco & Navigli 2013, Maiti & Samanta 2014], and e-commerce [Chen & Wang 2013]. Most applications that use clustering deal with large datasets and/or data with numerous attributes [Han et al. 2011].

The k-means algorithm [Han et al. 2011] is a well-known and commonly used clustering algorithm. It takes input parameter k and partitions the objects into k clusters. The algorithm begins by selecting k objects to represent the cluster centers. The remaining objects are assigned to the most similar clusters, as determined by the distance from or similarity to objects associated with the cluster centers. After assigning all of the objects to clusters, the algorithm computes the mean value of objects in each cluster as new cluster centers. This process iterates until the criterion function converges. The k-means algorithm is scalable and efficient at processing large datasets. In this study, we used the k-means algorithm for the clustering of all normalized review sentiment vectors into k groups, in which the reviews in the same cluster present similar sentiments with regard to the specified aspects. The mean vector of a cluster reveals the characteristics of the reviews to which it belongs. This enables the identification of k common types of popular opinions of the target hotel.

3. Research Design

The proposed approach classifies sentences according to an aspect specified by the user, determines the polarity of statements made regarding that aspect, and then rates and summarizes reviews accordingly. Specifying the aspects of a review on which to focus involves having the user provide aspect names along with a number of terms, either nouns or adjectives, which are represented by term set D. In the following, we outline an example to illustrate this approach using the five user-defined aspects of D1(Value), D2(Location), D3(Service), D4(Meals), D5 (Facilities). Table 1 shows a list of terms provided by the user to describe each aspect. Fig. 2 presents a sample review obtained from hotels.com.

Table 1: List of terms for specified aspects D

D_1: Value	D_2: Location	D_3: Service	D_4: Meals	D_5: Facilities
Value	Location	Service	Food	Room
Price	View	Front desk	Drink	Bed
Amount	Station	Staff	Breakfast	Internet
Rate	Store	Check-in	Afternoon tea	Facility
Cheap	Mall	Check-out	Buffet	Pool
Worth	Airport	Parking	Bar	Spa
Low	Distance	Fast	Restaurant	Lobby
Inexpensive	Far	Friendly	Dinner	Wi-Fi
Economical	Close	Helpful	Lunch	Toilet
Reasonable	Convenient		Brunch	Bathroom
Fee	Train		Delicious	Dirty
			Tasty	Broken

Figure 2: Sample Review on hotels.com

This method derives the sentiment of reviews according to user-provided terms (D) related to five user-defined aspects, (D1 to D5). Reviews are classified into groups to reveal the major opinions expressed with regard to the target hotel.

This approach involves four phases, the framework of which is presented in Fig. 3. In the first phase, review data is preprocessed so that term sets can be generated for each review. In the second phase, each sentence in a review is classified by aspect, in accordance with the terms in the sentence. SentiWordnet 3.0 is used to identify the polarity of sentiments expressed in each sentence. A sentiment table is then generated for the review analysis phase, in which the numbers of positive, neutral, and negative sentences related to each specified aspect are listed.

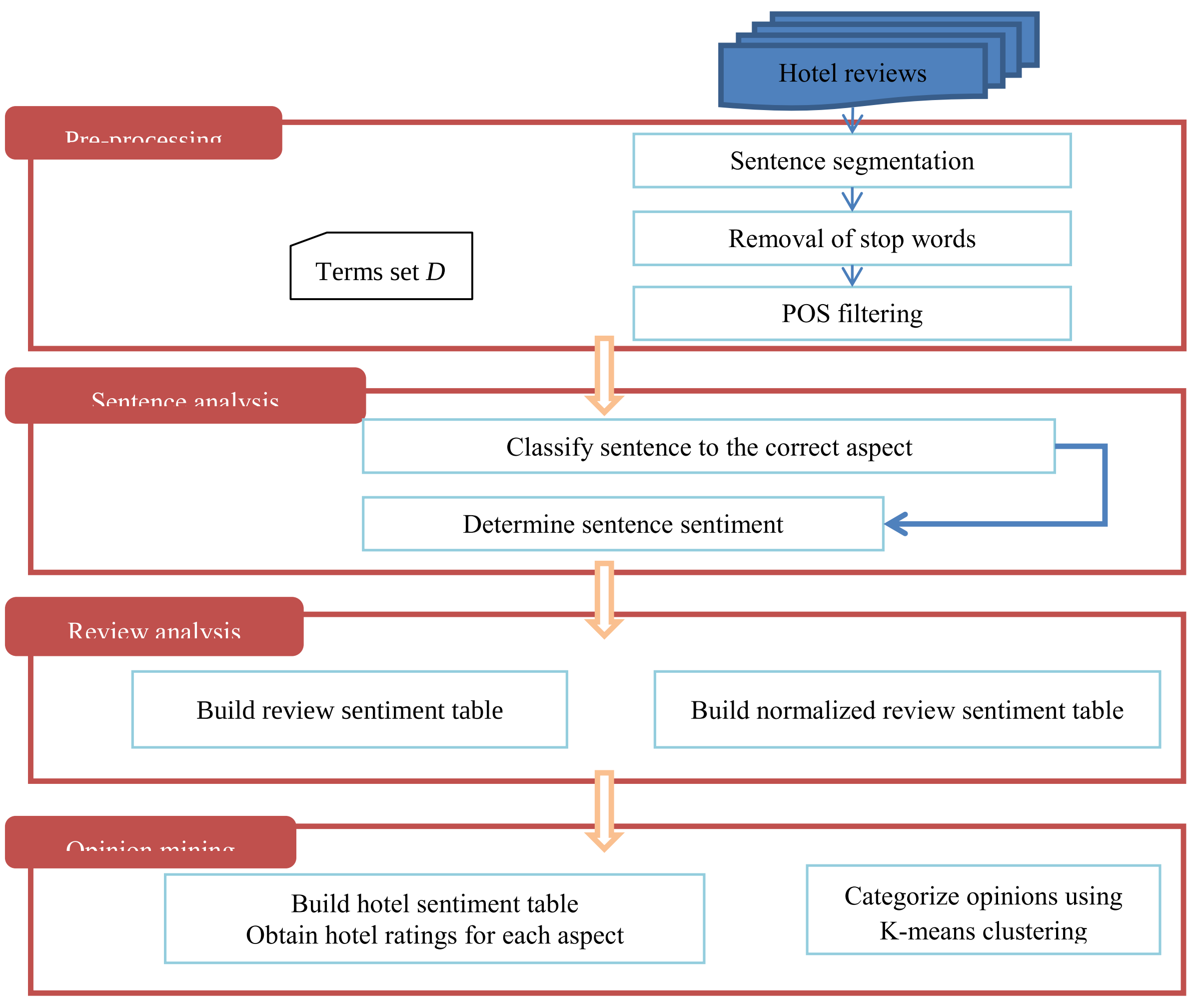


Figure 3: Operational Framework of Proposed Approach

In the opinion mining phase, review sentiment tables containing all of the reviews associated with the target hotel are used to produce a normalized sentiment table showing the percentages of positive, neutral, and negative sentences related to each aspect. The sentiment table is then aggregated to derive ratings for each aspect of the target hotel. To identify the major points in the reviews with regard to target hotels, we began by normalizing the review sentiment table to enable its representation in the form of a normalized vector. The k-means algorithm was then used to cluster all normalized sentiment vectors into k groups, in which the reviews in each group share similar sentiments with regard to the specified aspects. This makes it possible to categorize the opinions related to the target hotel.

In the following, we detail the four phases of preprocessing, sentence analysis, review analysis and opinion mining in Sections 3.1, 3.2, 3.3 and 3.4, respectively.

3.1. Pre-Processing

In this phase, we generate noun and adjective sets from each sentence in the reviews. There are three tasks in this phase: sentence segmentation, removal of stop words, and POS filtering. The job of sentence segmentation is to find distinct terms in sentences. We partitioned sentences according to ending punctuation, including ".", "?", and "!". We then partitioned the words in the sentences according to the spacing between words. Table 2 presents the segmentation results of a sample review in Fig. 2.

Table 2: Segmentation results for the sample review in Fig. 2

(1)	1	this	2	was	3	A	4	very	5	beautiful	6	hotel	7	that	8	was
	9	quiet	10	and	11	luxurious										
(2)	1	the	2	room	3	Was	4	superior	5	with	6	comfortable	7	beds	8	and
	9	very	10	quiet												
(3)	1	i	2	was	3	busy	4	working	5	so	6	not	7	much	8	time
	9	to	10	use	11	the	12	restaurant	13	or	14	other	15	facilities	16	however
	17	the	18	2nd	19	floor	20	lobby	21	bar	22	was	23	very	24	nice

Stop words that are not important are then removed to avoid excess noise in the analysis of text. Table 3 presents the results following the removal of stop words for the sample review in Fig. 2.

Table 3: Results following the removal of stop words in Table 2

(1)	1		2		3		4		5	beautiful	6	hotel	7		8	
	9	quiet	10	and	11	luxurious										
(2)	1		2	room	3		4	superior	5		6	comfortable	7	beds	8	and
	9		10	quiet												
(3)	1		2		3	busy	4	working	5		6	not	7		8	time
	9		10	use	11		12	restaurant	13	or			15	facilities	16	
	17		18	2nd	19	floor	20	lobby	21	bar	22		23		24	nice

We utilized the Stanford Log-linear Part-Of-Speech Tagger [Porter 1980] during POS filtering to assign a POS tag to each word. The fact that all of the words have tags makes it possible to extract noun and adjective terms that could be used for the classification of sentences according to the aspect to which they belong. Adjective terms are then used to determine the sentiment of the sentence. Finally, we extract the negative adverb terms which could make a positive sentiment appear negative or a negative sentiment appears positive.

3.2. Sentence Analysis

Pre-processing is used to reveal pertinent nouns and adjectives in each sentence. Figure 4 presents the nouns and Fig. 5 presents the adjectives extracted from the review in Fig. 2. Some adjectives in a sentence are useful for classifying the sentence according to its aspect, whereas others are not. For example, adjectives such as good, high, low, bad, great, and excellent are useless in classifying aspect because they are general adjectives that can appear when referring to any aspect. On the contrary, specific adjectives such as yummy, expensive, cheap, clean, dirty and tasty are more useful for classifying aspects. Thus, adjectives are removed using a general adjective stop list and adjectives that are not in the stop list are deemed representative. This study used nouns as well as representative adjectives to identify the aspect to which a sentence pertains. We then compare the similarity of terms in a sentence and the term in set D for each respective aspect. In this manner, sentences can be classified according to user-defined aspects.

(1)	hotel						
(2)	room	beds					
(3)	time	use	restaurant	facilities	floor	lobby	bar

Figure 4: Nouns Extracted from Review in Fig. 2

(1)	beautiful	quiet	luxurious
(2)	superior	comfortable	quiet
(3)	busy	2nd	nice

Figure 5: Adjectives Extracted from Review in Fig. 2.

Two main tasks are addressed in this phase: the classification of sentences according to the aspect to which they pertain, and determining the sentiment of sentences. The following two subsections introduce how these two tasks can be accomplished.

3.2.1. Classifying Sentences to Correct Aspects

The term set of a sentence, including nouns and representative adjectives, is used to match noun set D assigned to each aspect. The sentence is then assigned to the most relevant aspect, as follows:

(1) NGD (Normalized Google Distance) [Cilibrasi & Vitanyi 2007] or WordNet::Similarity [Pedersen et al. 2004] is used to compute word-word distances (called semantic similarities).

(2) Each word in a sentence is classified according to the aspect to which it pertains, based on the results in Step 1.
The semantic similarity between a word used in a review and a user-determined word is then used to represent an aspect as the average or the max distance between this word and all words related to a given aspect.

(3) The sentence is classified according to the aspect to which it pertains based on the results in Step 2.
The sentence is then assigned to an aspect classification, according to the number of words in that sentence pertaining to that aspect. In the following, we provide a more detailed description of these three steps.

Step 1 – Compute distances between words

NGD (Normalized Google Distance) was used to compute the semantic similarity between words used in a review sentence and those selected by the user to represent an aspect classification. When the value of NGD(x, y) is relatively small, words can be considered to have greater semantic similarity.

$$NGD(x,y) = \frac{max\{log\, f(x), log\, f(y)\} - log\, f(x,y)}{log\, M - min\{log\, f(x), log f(y)\}} \quad (1)$$

where, M=50,000,000,000 is the total number of webpages that the Google search engine indexes, x and y represent words used to compute similarity. In addition, $f(x)$ and $f(y)$ are the number of pages containing x and y, respectively, while $f(x, y)$ represents the number of pages containing both x and y.

Formula (2) is the reverse of formula (1), in which the distance of x and y is reversed as the similarity between x and y ($Sim_{x,y}$). The semantic similarity between the two words increases with the value of $Sim_{x,y}$. In this study, this value represents the similarity between a term in a review sentence and a user-generated noun representing an aspect classification.

$$Sim_{x,y} = 1 - NGD(x, y) \quad (2)$$

The other word-word distance function we used to reveal semantic similarity is WordNet similarity [Pedersen et al. 2004], as outlined in Formula (3). The semantic similarity between words is proportional to the value of the WordNet Similarity.

$$\mathrm{Sim}_{x,y} = \begin{cases} 1 & ,\text{if } x = y \\ \text{WordNet} :: \mathrm{Sim}(x, y) & ,\text{otherwise} \end{cases} \tag{3}$$

where, *x* and *y* represent the words that are being analyzed for similarity. In this formula, *x* is not equal to *y*. If *x*=*y*, then $Sim_{x,y}$=1.

Step 2 – Classifying a word in sentence to an aspect

In this step, words used in the review allow the classification of sentences pertaining to user-generated aspects. Two methods can be used to accomplish this. The first method involves assigning the term to the aspect classification with the maximum average similarity, which involves assigning the term to the aspect classification with the maximum similarity.

In the first method, we compare term *x* (a term used in the review) to the five aspect terms with the greatest similarity (denoted as *5NN*) in order to compute the average semantic similarity. The similarity between word *x* and aspect D_i is defined as $\sum_{y \in 5NN \cap D_i} simialrity\ (x, y)/|D_i|$. Word *x* is then assigned to aspect D_i if it most closely fits (has the largest semantic similarity) the terms in D_i when compared to all other aspects.

In the second method, the similarity between word x (as used in a review) and aspect D_i is defined as possessing the greatest similarity between x and all words in D_i. Accordingly, word x is assigned to the aspect classification with the greatest semantic similarity.

Both methods classify review sentences to the sixth aspect (D_0, which signifies "others") when the similarity is less than a given threshold. This is done to avoid the problems inherent in forcibly assigning a word to an aspect when designated words are dissimilar. We set the threshold as $0.5/|D_i|$, in which D_i represents the original aspect classification.

Step 3 – Classify sentences to an appropriate aspect

After identifying the aspect classification of each word in Step 2, we can determine the aspect of each sentence according to the classification results of the words in the sentence. A sentence can be assigned to aspect D_i if the highest numbers of terms in the sentence are assigned to D_i. Figure 6 shows how the aspect classification of the three sentences in Fig. 2 is determined.

(1) | hotel (D_0) | ➡ (D_0)

(2) | room (D_5) | beds (D_5) | ➡ (D_5)

(3) | time (D_0) | use (D_0) | restaurant (D_4) | facilities (D_5) | floor (D_5) | lobby (D_5) | bar (D_4) | ➡ (D_5)

Figure 6: Classify Sentences to Aspects

3.2.2. Determining Sentence Sentiment

We used the dictionary SentiWordNet 3.0 [Pedersen et al. 2004] to evaluate the sentiments associated with adjectives used in a review sentence. The sentiment score of adjective a in a sentence is designated as SWN(a). The nearness of negative adverbs such as "not" reverses the sentiment polarity of adjective a.

After obtaining the sentiment scores of all adjectives in the review sentences, an average sentiment score is computed to represent the sentiment score of the sentence as a whole.

We use threshold value γ to identify whether the sentiment of the sentence is positive (+1), neutral (0) or negative (-1). If the score is not less than γ, then it is 1. If it is not greater than -γ, then it is -1. Otherwise, it is 0.

A pretest is used to determine threshold γ. This study tagged 100 sentences for the identification of sentiment polarity of each sentence. We then compared these results to those obtained through human tagging in order to select the γ value with the lowest error rate. In this case, γ =was set to 0.2.

3.3. Review Analysis

Sentence analysis in Section 3.2 was used to identify the aspect and sentiment polarity of each sentence in the reviews. A review sentiment table was then generated for each review in the review analysis step. This table displays the number of positive, neutral, and negative sentences pertaining to the aspects of interest. This is easily accomplished by collecting and summarizing the results from Phase 2. Table 4 presents an example of a review sentiment table.

Table 4: Example of review sentiment table

	D_1	D_2	D_3	D_4	D_5
Neutral	1	0	1	0	1
Positive	3	2	0	0	0
Negative	0	0	5	2	0

Absolute numbers may be misleading when analyzing the frequency of occurrence, because the number of sentences pertaining to a given aspect differs according to aspect. In order to avoid being misled by these numbers, we must normalize the table in order to demonstrate the relative percentages of positive, neutral, and negative sentences pertaining to each aspect. Following this normalization process, the results in Table 4 appear as follows in Table 5.

Table 5: Example of normalized review sentiment table

	D_1	D_2	D_3	D_4	D_5
Neutral	0.25	0	0.167	0	1
Positive	0.75	1	0	0	0
Negative	0	0	0.833	1	0

3.4. Opinion Mining

The review analysis in Section 3.3 enabled the aggregation of reviews for a target hotel into an overall review sentiment table. This table lists the number of positive, neutral, and negative sentences written about this particular hotel, as they pertain to the aspects specified by the user. Figure 7 illustrates this process using a simple example. This makes it possible to derive a normalized table showing the percentages of positive, neutral, and negative sentences, as they pertain to each user-specified aspect.

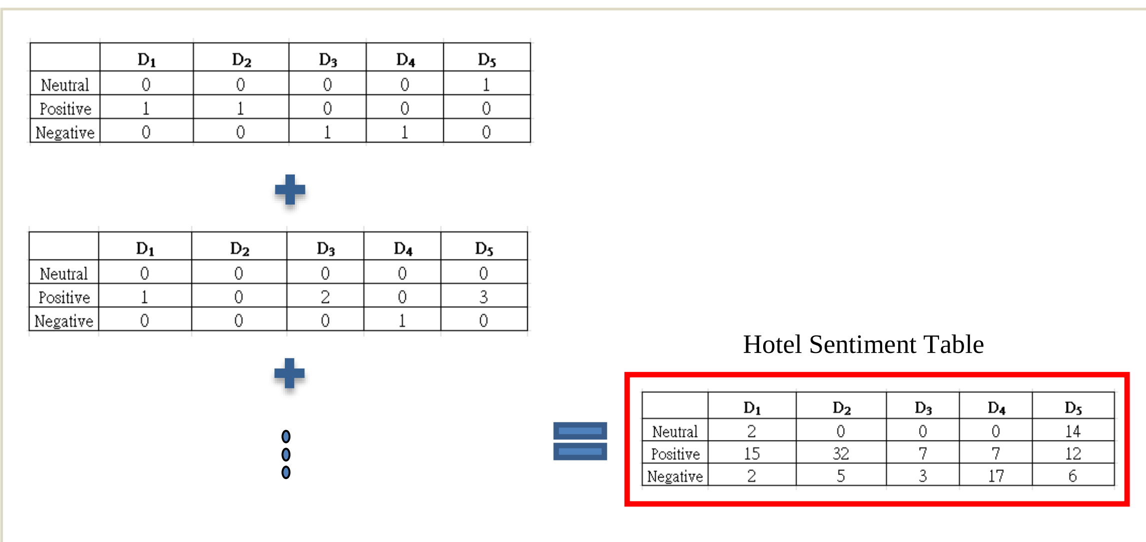


Figure 7: Example of Hotel Sentiment Table

The hotel sentiment table makes it possible to compute a rating for this hotel with regard to multiple user-specified aspects. Let $n_{+,i}$, $n_{o,i}$, $n_{-,i}$ be the numbers of positive, neutral, and negative sentiments related to aspect D_i, respectively. Let $n_{all,i} = n_{+,i} + n_{o,i} + n_{-,i}$. Then the rating of aspect i can be defined as $\frac{n_{+,i}+0.5\times n_{0,i}}{n_{all,i}}$. The use of these equations enables us to obtain an overview of how well the target hotel performs with regard to each aspect. As shown in Fig. 7, the ratings for aspects D1, D2, D3, D4 and D5 are 0.84, 0.86, 0.7, 0.28 and 0.59, respectively.

To isolate important comments in reviews about the target hotel requires that the review sentiment table is transformed into a normalized vector form. The k-means algorithm is useful for clustering normalized review sentiment vectors into k groups, in which the reviews in any given cluster include similar sentiments related to specified aspects. The mean vector of a cluster reveals the characteristics of the reviews in that cluster. This makes it possible to classify popular opinions related to the target hotel into k typical types.

4. Experiments

In this section, we evaluate the classification accuracy of the two methods used in the sentence analysis phase of the proposed method by evaluating whether sentences are correctly classified and determining the accuracy of sentence sentiment classification. We then apply the proposed method to the review data of two real hotels featured on hotels.com. Finally, we apply the Delphi method to examine user satisfaction with the proposed method.

4.1. Experiment 1: Evaluating Sentence Analysis Phase

The two tasks in the sentence analysis phase are sentence-to-aspect classification and sentence-sentiment classification. The accuracy of these phases exerts a strong influence on the summarization of information; therefore, these were evaluated first in Experiment 1.

4.1.1. Data Set and Evaluation Measures

Two hundred and fifty review sentences were selected from hotels.com. These were grouped into one of five aspect classifications and then further assigned to one of the three sentiment classifications. The review sentences were then tagged using their respective aspect classifications and sentiment polarities. Table 6 presents the aspect classification and sentiment polarity distribution of the review sentences. Not all of the aspects and sentiment polarities appeared in the data set with the same frequency; therefore, we allowed samples of different sizes in different cells. Sentence that do not fit any of the five aspect classifications may be classified into the sixth class, referred to as "others".

Table 6: Distribution of sentences according to aspect classification and sentiment polarity

	D_1: Value	D_2: Location	D_3: Service	D_4: Meals	D_5: Facilities	Total
Neutral	4	8	3	4	10	29
Positive	32	38	34	32	23	159
Negative	13	5	15	9	20	62
Total	49	51	52	45	53	250

In traditional classification problems, a confusion matrix is usually used to measure accuracy, precision, recall, and F-measure for the evaluation of classification performance. In the field of machine learning, a confusion matrix, also known as a contingency table or an error matrix [Stehman 1997], is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning method.

Sentence-aspect classification and sentence-sentiment classification are no different from a traditional classification problem; therefore, we are able to use the same measurements to evaluate the performance of our classification results. We first applied the following accuracy formula to evaluate whether aspect analysis and sentiment analysis led to accurate classifications in the overall dataset:

$$Accuracy = \frac{number\ of\ correct\ classified\ sentences}{Total\ number\ of\ sentences} \tag{4}$$

We then built a confusion matrix to test the precision, recall, and F-measure in each class. The confusion matrix is a two aspect matrix with two attributes, predicated class and actual class, as shown in Table 7.

Table 7: Confusion Matrix

Confusion Matrix		Actual	
		Yes	No
Predicted	Yes	TP	FN
	No	FP	TN

This enables the computation of precision, recall, and F-measure for each class using the following formula:

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F-measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

4.1.2. Results of Experiment 1

We began by examining the accuracy of sentence-to-aspect classification results. In the sentence analysis phase, NGD and WordNet were used to define semantic similarity between two words. We then employed two methods (Average and Max) to measure the semantic similarity between the words used in review sentences and the words selected by the user to represent an aspect. The four results presented in Table 8 indicate that the NGD method, when used in conjunction with the Average method can achieve a maximum accuracy of 83.2%.

Table 8: Accuracy of aspect classification using combination of two methods

Methods \ Measures	Average	Max
NGD	83.2%	50.4%
WordNet::Similarity	76.8%	76.8%

Each cell in Table 8 can be further broken down using a confusion matrix. For example, Table 9 presents a confusion matrix for the combination of NGD and Average (accuracy 83.2%). We can see four review sentences classified as “other” as well as the precision, recall, and F-measure for all aspects in Fig. 8.

Table 9: Confusion matrix (accuracy 83.2%)

Result		Actual class					
		D_0	D_1	D_2	D_3	D_4	D_5
Predicted class	D_0	0	1	1	0	0	2
	D_1	0	40	3	1	2	5
	D_2	0	3	45	4	1	3
	D_3	0	0	0	45	0	5
	D_4	0	2	1	1	41	1
	D_5	0	3	1	1	1	37

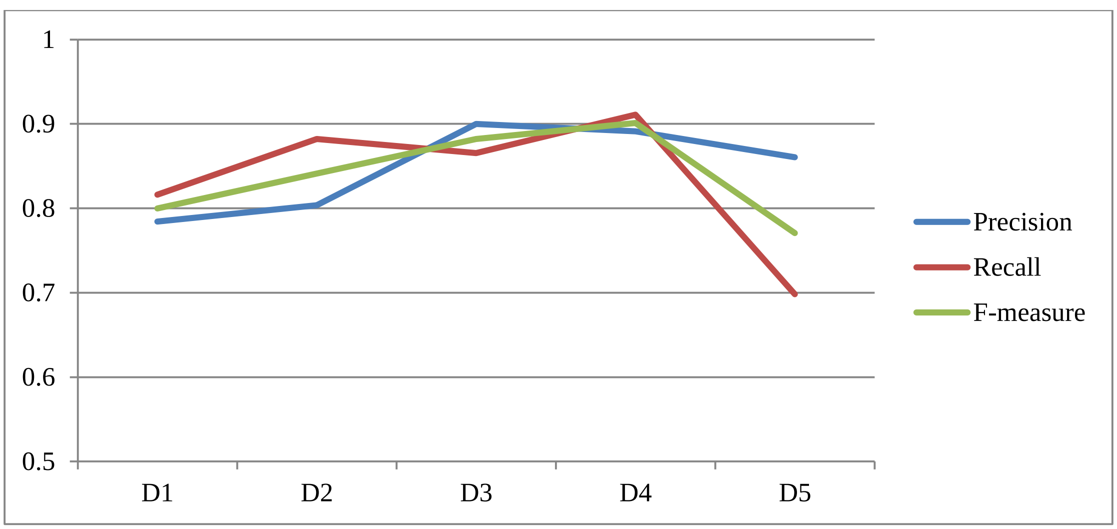


Figure 8: Precision, Recall, and F-Measure of each Aspect Classification

Table 10 presents the results of sentence-to-sentiment classification.

According to Table 10, sentiment polarity analysis achieved accuracy of 75.6% ((20+131+38)/250). Figure 9 shows the precision, recall, and F-measure of neutral, positive, and negative sentiments.

Table 10: Results of sentiment polarity analysis (γ=1)

The result		Actual class		
		Neutral	Positive	Negative
Predicted class	Neutral	20	20	11
	Positive	7	131	13
	Negative	2	8	38

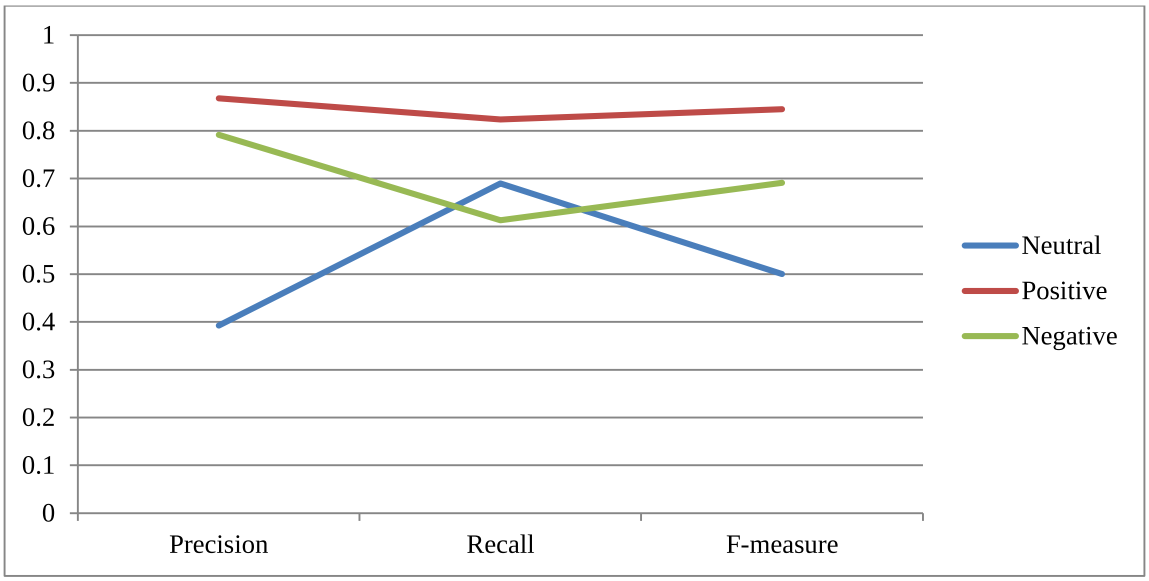


Figure 9: Results of Sentiment Polarity Analysis (γ=1)

4.2. Experiment 2: Actual Data

In the following, we present analysis results obtained using the proposed method, using real data collected from hotels.com. The data was extracted from reviews of two hotels in New York City written between Jan, 2013 and April, 2013.

(1) Hotel 1: “The New York Palace”

At the time of this research, the overall rating of this hotel on hotels.com was 4.6, and 95 reviews were randomly extracted for analysis.

(2) Hotel 2: “Hotel Pennsylvania”

At the time of this research, the overall rating of this hotel on hotels.com was 2.8 and 153 reviews were randomly extracted for analysis.

Following phases 1 - 4 of the analysis, we were able to obtain the hotel sentiment tables for the two hotels. The normalized hotel sentiment tables are presented in Tables 11 and 12. Clearly, Table 11 has a greater number of positive sentiments than does Table 12. These results match the higher overall rating of hotel 1, compared to hotel 2. A shown in the two tables, the ratings of hotel 1 with regard to user-defined aspects were 0.56, 0.665, 0.74, 0.77 and 0.615, respectively, whereas the ratings of hotel 2 were 0.325, 0.27, 0.26, 0.315 and 0.27, respectively.

We then extracted the major opinions related to these hotels, as expressed by reviewers. The review sentiment tables were then clustered into k clusters. Setting k=4 enabled us to identify the four major opinions related to these two hotels, as shown in Tables 13 and 14, respectively.

Table 11: Normalized hotel sentiment table for Hotel 1: "The New York Palace"

	D_1: Value	D_2: Location	D_3: Service	D_4: Meals	D_5: Facilities
Neutral	0.48	0.45	0.3	0.32	0.41
Positive	0.32	0.44	0.59	0.61	0.41
Negative	0.2	0.11	0.11	0.07	0.18

Table 12: Normalized hotel sentiment table for Hotel 2: "Hotel Pennsylvania"

	D_1: Value	D_2: Location	D_3: Service	D_4: Meals	D_5: Facilities
Neutral	0.55	0.18	0.34	0.53	0.36
Positive	0.05	0.18	0.09	0.05	0.09
Negative	0.4	0.64	057	0.42	0.55

In Table 13, we see that the comments in cluster A and cluster D are comparatively positive towards D3 and D4; however, cluster A has a neutral attitude toward D2 and D5, while cluster D has a neutral attitude toward D1. As for cluster B, it is clear that the comments regarding cluster B focus mainly on D4 (Meals), revealing considerable satisfaction with regard to the meals provided by the hotel. Finally, the comments of cluster C are more negative, particularly with regard to D1, D2 and D5.

We cluster opinions into four major types. Cluster A makes positive comments about D1(Value), D3(Service) and D4(Meal) (e.g., they may indicate that this hotel is a good place to go); Cluster B has strong positive attitudes toward D4(Meal) (e.g., they may comment "You must eat here!"); Cluster C makes negative comments about D1(Value), D2(Location) and D5(Facilities) (e.g., they do not like this hotel); Cluster D are positive in regard to every dimension (e.g., they feel that the hotel is wonderful).

Table 13: Cluster centers of hotel 1 "The New York Palace" (k=4)

	Cluster A				
	D_1	D_2	D_3	D_4	D_5
Neutral	0.38	0.88	0.38	0.4	0.78
Positive	0.63	0.12	0.62	0.6	0.11
Negative	0	0	0	0	0.11
	Cluster B				
	D_1	D_2	D_3	D_4	D_5
Neutral	0.63	0.45	0.47	0	0.8
Positive	0.38	0.41	0.27	0.9	0.13
Negative	0	0.14	0.27	0.1	0.07
	Cluster C				
	D_1	D_2	D_3	D_4	D_5
Neutral	0.29	0.39	0.23	0.55	0.28
Positive	0.15	0.35	0.54	0.36	0.5
Negative	0.56	0.26	0.23	0.09	0.22
	Cluster D				
	D_1	D_2	D_3	D_4	D_5
Neutral	0.79	0.18	0.26	0.35	0.05
Positive	0.21	0.73	0.65	0.59	0.75
Negative	0	0.09	0.09	0.06	0.2

As shown in Table 14, the comments in cluster A and cluster D are generally more negative. Negative attitudes are expressed toward D3 (Service); however, cluster A shows particularly negative ratings for D2 and D5, while cluster D reveals negative sentiments toward D4. Clusters B and C both reveal satisfaction with D2 (location). Cluster B reveals dissatisfaction with D5 (facilities), while cluster C does reveals dissatisfaction with D1 (value).

Table 14: Cluster centers of hotel 2 “Hotel Pennsylvania” (k=4)

	Cluster A				
	D_1	D_2	D_3	D_4	D_5
Neutral	0.91	0.11	0	0.9	0.3
Positive	0	0.11	0.05	0	0
Negative	0.09	0.79	0.95	0.1	0.7
	Cluster B				
	D_1	D_2	D_3	D_4	D_5
Neutral	0.88	0	0.86	0.6	0.24
Positive	0	0.4	0.05	0.2	0.08
Negative	0.13	0.6	0.1	0.2	0.68
	Cluster C				
	D_1	D_2	D_3	D_4	D_5
Neutral	0.38	0.11	0	0.5	0.4
Positive	0.06	0.56	0.33	0	0.2
Negative	0.56	0.33	0.67	0.5	0.4
	Cluster D				
	D_1	D_2	D_3	D_4	D_5
Neutral	0.86	0.57	0	0.11	0.31
Positive	0	0	0	0.17	0.54
Negative	0.14	0.43	1	0.72	0.15

4.3. Experiment 3: Evaluation to determine user satisfaction

In the measurement of user satisfaction [Chen & Kumar 2008, Herlocker et al. 2004], the Delphi method [Hsu & Sandford 2007] is widely used to acquire consensus-based opinions from a panel of experts. In this study, we applied the Delphi method to evaluate the proposed method in terms of user satisfaction. For this experiment, we recruited 20 participants with experience booking hotel rooms via hotels.com and subsequently submitting reviews of their experience. The data was extracted from reviews of the Hotel Pennsylvania in New York City written between Jan 2013 and April 2013.

We began by presenting the original data to the participants and then extended the capabilities of the original review website by allowing participants to suggest five aspects in which they were interested. The aspects provided by participants are listed in Table 15.

Table 15: Aspects specified by the twenty participants

Participant ID	Personalized aspects	Participant ID	Personalized aspects
01	Location, Comfort, Room, Service, Value	11	Parking, Price, Value, WiFi, Location
02	WiFi, Amenities, Vibe, Room, Value	12	Business, Cleanliness, Value, Comfort, Quiet
03	Service, Location, Bar, Cleanliness, WiFi	13	Pet, Service, Park, Cleanliness, Room
04	Location, Neighborhood, Landmarks, Childcare, Gym	14	Cleanliness, Value, Airport transfers, Service, Meal
05	Free breakfast, Business facilities, WiFi, Room, Bathtub	15	Convenience, Location, Facilities, Condition, Value
06	Value, Service, Convenience, Meal, Bar	16	Smoking, Service, Value, Business, Price
07	Value, Service, WiFi, Meal, Vibe	17	WiFi, Meal, Coffee, Service, Price, Location
08	Luxury, Meal, WiFi, Bar, Location	18	Room size, Clean, Service, Near central park, Breakfast
09	Location, Vibe, Service, Value, Cheap	19	Comfort, Quiet, Cleanliness, Service, Staff
10	Free WiFi, Free breakfast, Value, Service, Location	20	Carpet, Bathroom, Recommend, Staff, Bed

Our findings obtained using the questionnaires in Figure 10 indicate that over eighty-five percent of the participants were satisfied with the results obtained using this novel approach for the rating and summarizing of online reviews according to user-specified aspects, as shown in Table 16. We compared the result of Q1 and Q2 with the average satisfaction achieved using the proposed method, the results of which indicate that the proposed method produced a high level of user satisfaction.

Q1. The degree to which the **original** function “summary of average rating related to five <u>predefined</u> aspects” assists consumers in acquiring hotel information.
☐(5)Very Useful ☐(4)Useful ☐(3)No Comment ☐(2) Useless ☐(1)Very Useless

Q2. The degree to which the **proposed** function “summary of average rating related to five <u>personalized</u> aspects” assists consumer in acquiring hotel information.
☐(5)Very Useful ☐(4)Useful ☐(3)No Comment ☐(2) Useless ☐(1)Very Useless

Q3. The degree to which the **proposed** function “overall opinion related to five <u>personalized</u> aspects” assists consumer in acquiring hotel information.
☐(5)Very Useful ☐(4)Useful ☐(3)No Comment ☐(2) Useless ☐(1)Very Useless

Q4. The degree to which the **proposed** system indicates the availability and quality of “supporting information”.
☐(5)Very Useful ☐(4)Useful ☐(3)No Comment ☐(2) Useless ☐(1)Very Useless

Q5. The degree to which the **proposed** system reduces the length of time required of users to read through reviews in order to obtain specific information
☐(5)Very Useful ☐(4)Useful ☐(3)No Comment ☐(2) Useless ☐(1)Very Useless

Figure 10: Questionnaire for assessing user satisfaction

Table 16: Results of user satisfaction

Question Item	Very Useful	Useful	No Comment	Useless	Very Useless	Total	Satisfaction
Q1	10	7	3	0	0	20	83.33%
Q2	15	5	0	0	0	20	100.00%
Q3	12	4	4	0	0	20	80.00%
Q4	12	5	3	0	0	20	85.00%
Q5	12	4	4	0	0	20	80.00%
Average of Q2 to Q5	12.8	4.5	2.8	0.0	0.0	20.0	86.25%

5. Conclusions and Future Research

Generally, websites that include customer reviews of products or services only give users the ability to search for summarized information from online reviews according to aspects predefined by the web site. When users are interested in particular aspects of a product or service that has not been predefined by the website, they may have difficulty obtaining the information they need. The aim of this research was to bridge the gap between the needs of users and the search features currently found on websites by enabling users to extract a summary of information from the reviews of products and services reviews by specifying the search parameters according to their needs. NGD or WordNet were used to compute similarity between terms in existing review sentences and user-generated terms related to aspects of interest. SentiWordNet 3.0 is then used to analyze the sentiment polarity of every sentence. This results in the generation of a review sentiment table for each review, showing the distribution and sentiments of review sentences with regard to the aspects in which users are interested. The resulting hotel sentiment table provides an overview of all review sentiment tables pertaining to the hotel in question. Finally, each review is represented in vector form and an algorithm is used to cluster the opinions expressed in reviews in order to gain a better understanding of the major types of opinions reported for a given hotel. This study used customer generated reviews from hotels.com as the target data set to test these methods. The methods can be divided into two parts. The first part involves the evaluation of results related to aspect classification and sentiment analysis. The second part involves the clustering of actual reviews in order to obtain an overview of the opinions forwarded in these reviews. Our results demonstrate the efficacy of the proposed method in summarizing (according to the interests of users) the information found on websites in which products or services are reviewed.

The achievements of this study make three particular contributions to the domain:

The proposed method makes it possible to extend the capabilities of review websites by enabling users to obtain information specific to their needs in a flexible and dynamic manner. This can help to enhance user satisfaction and thereby increase the competitiveness of the firm. For example, using these methods in online reviews would make it possible for users to evaluate and compare hotels quickly, according criteria they establish, without the need to spend lengthy amounts of time combing through reviews.

The proposed method gives enterprises a new channel by which to gain an objective understanding of the perspective of consumers through the collection of user-specified product aspects. These selections also provide a valuable reference for enterprises seeking avenues for innovation and service improvement.

The functions of the proposed methodology are compatible with current e-commerce websites to improve services. Furthermore, our analysis allows users to easily visualize groups of similar opinions.

The proposed approach could be improved in the following ways. First, in sentence-to-aspect classification, each sentence was classified with regard to the single aspects in which users are interested. Nonetheless, it is possible that single sentences in customer reviews may actually pertain to several aspects simultaneously, or even pertain to different aspects of varying degrees. Therefore, future researchers could create a methodology allowing for multiple classifications or fuzzy classification. Second, the proposed method defines aspect classification according to a set of user-generated terms, which requires time and effort on the part of users, and renders the system dependent on the expertise of users. Therefore, future developments could be aimed at reducing the reliance of the system on user involvement. Thirdly, the sentence-to-sentiment classification task uses only adjectives; however, many terms that express sentiments are not adjectives. Future approaches might benefit from considering all evaluative terms, such as “hate,” “love,” “like,” “please,” “content”.

Acknowledgements

This work has been supported by the Ministry of Science and Technology in Taiwan within the project number: NSC 101-2410-H-030 -021 -MY3.

REFERENCES

Archak, N., A. Ghose, and P. Ipeirotis, “Show Me the Money!: deriving the pricing power of product features by mining consumer reviews,” *ACM SIGKDD Conference on Knowledge Discovery and Data Mining ACM*, 56-65, New York, USA, 2007.

Ardissono, L., A. Goy, G. Petrone, and M. Segnan, “Personalization in business-to-customer interaction,” *Communications of the ACM*, Vol.45, No. 5: 52-53, 2002.

Aue, A. and M. Gamon, “Customizing sentiment classifiers to new domains: a case study,” Proceedings of international conference on recent advances in natural language processing, 2005.

Bai, X., “Predicting consumer sentiments from online text,” *Decision Support Systems*, Vol.50, No.4: 732–742, 2005.

Baroni, M. and S. Vegnaduzzo, “Identifying subjective adjectives through web-based mutual information,” Proceedings of the 7th Konferenz zur Verarbeitung Natrlicher Sprache, pp. 613-619, 2004.

Bollen, J., H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, Vol.2, No.1:1–8, 2011.

Chen, Y. J. and Y. M. Chen, “Acquiring Consumer Perspectives in Chinese eWOM,” *Journal of Information Management*, Accepted, 2015.

Chen, L. and F. Wang, “Preference-based clustering reviews for augmenting e-commerce recommendation,” *Knowledge-Based Systems*, Vol.50, 44-59, 2013

Cilibrasi, R.L. and P.M.B. Vitanyi, “The google similarity distance,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19:370-383, 2007.

Connor, B., R. Balasubramanyan, B. Routledge, and N. Smith, “From tweets to polls: linking text sentiment to public opinion time series,” *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2010.

Di Marco, A. and R. Navigli, “Clustering and diversifying web search results with graph-based word sense induction,” *Computational Linguistics*, Vol.39, No.4:1--43, 2013.

Diakopoulos, N., and D. Shamma, “Characterizing debate performance via aggregated twitter sentiment,” *Proceedings of the 28th international conference on Human factors in computing systems*, Atlanta, Georgia, 2010.

Duan, W., B. Gu, and A.B. Whinston, “Do online reviews matter? — An empirical investigation of panel data,” *Decision Support Systems*, Vol. 45:1007-1016, 2008.

Esuli, A. and F. Sebastiani, “SentiWordNet: a publicly available lexical resource for opinion mining,” Proceedings of the 5th conference on language resources and evaluation, pp. 417-422, 2006.

Gerani, S., M.J. Carman, and F. Crestani, “Investigating learning approaches for blog post opinion retrieval,” Proceedings of the 31th European conference on IR research on advances in information retrieval, pp. 313-324, 2009.

Gu, B., P. Konana, A. Liu, B. Rajagopalan, and J. Ghosh, “Predictive value of stock message board sentiments,” *McCombs Research Paper No. IROM-11-06*, 2006.

Gu, B., Q. Tang, and A. B. Whinston, “The influence of online word-of-mouth on long tail formation,” *Decision Support Systems*, Vol. 56:474-481, 2013.

GuangXu, Z., H. Joshi, and C. Bayrak, “Topic categorization for relevancy and opinion detection,” Proceedings of the text retrieval conference, USA, 2007.

Han, J., M. Kamber, and J. Pei, “Data Mining: Concepts and Techniques,” Third Edition, The Morgan Kaufmann Series in Data Management Systems, 2011.

Hatzivassiloglou, V. and K.R. McKeown, “Prediting The Semantic Orientation of Adjective,” Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, pp. 174-181, 1997.

Hawalah, A. and M. Faslia, “Dynamic user profiles for web personalization,” *Expert Systems with Applications*, Vol.42, No.5:2547-2569, 2015.

Hennig-Thurau, T., K. P. Gwinner, G. Walsh and D. D. Gremler, “Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet?” Journal of Interactive Marketing, Vol. 18, No. 1, pp. 38-52, 2004.

Herlocker, J. L., J. A. Konstan, L. G. Terveen, and J. T. Riedl, “Evaluating collaborative filtering recommender systems,” *ACM Transactions on Information Systems*, Vol. 22, No. 1:5-53, 2004.

Hsu, C.C. and B.A. Sandford, “The Delphi technique: making sense of consensus,” *Practical Assessment, Research and Evaluation*, Vol. 12, No. 10, 2007. http://pareonline.net/pdf/v12n10.pdf. Accessed Jan 11, 2016.

Hu, M. and B. Liu, “Mining and summarizing customer reviews,” Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining , pp. 168-177, 2004.

Hu, N., I. Bose, N.S. Koh, and L. Liu, “Manipulation of online reviews: An analysis of ratings, readability, and sentiments,” *Decision Support Systems*, Vol. 52:674-684, 2012.

Hu, N., I. Bose, Y. Gao, and L. Liu, “Manipulation in digital word-of-mouth: A reality check for book reviews,” *Decision Support Systems*, Vol.50, No.322-30, 2011

Hui Yan, J.C. and L. Si, “Knowledge transfer and opinion detection in the trec2006 blog track,” Proceedings of the text retrieval conference, USA, 2006.

Ho, S.Y. and K.K.W. Ho, “The Effects of web personalization on influencing users’ switching decisions to a new website,” *PACIS Proceedings*, Paper 67, 2008

Jansen, B.J., M. Zhang, and K. Sobel, A. Chowdury, “Twitter power: tweets as electronic word of mouth,” *Journal of the American Society for Information Science and Technology*, Vol.60:2169–2188, 2009.

Karimi, S., K. N. Papamichail, and C. P. Holland, The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behavior, *Decision Support Systems*, Accepted Manuscript 2015.

Kennedy, A. and D. Inkpen, “Sentiment classification of movie reviews using contextual valence shifters,” *Computing Intelligence*, Vol. 22, No.2: 110-125, 2006.

Ku, L.W., Y.T. Liang, and H.H. Chen, “Opinion extraction, summarization and tracking in news and blog corpora,” Proceedings of AAAI-2006 Spring symposium on computational approaches to analyzing weblogs, 2006.

Ku, L.-W., L.-Y. Li, T.-H. Wu and H.-H. Chen, “Major topic detection and its application to opinion summarization,” *SIGIR*, 627-628, 2005.

Larsson, A.O. and H. Moe, “Studying political microblogging: Twitter users in the 2010 Swedish election campaign,” *New Media and Society*, Vol.14, No.5:729– 747, 2012.

Lee, D., O. Jeong, and S. Lee, “Opinion mining of customer feedback data on the web,” *2nd international conference on Ubiquitous information management and communication ACM*. 230–235, 2008.

Li, Q., J. Wang, Y.P. Chen, and Z. Lin, “User comments for news recommendation in forum-based social media,” *Information Sciences*, Vol.180, No. 24:4929–4939, 2010.

Liu, B. and L. Zhang, “A survey of opinion mining and sentiment analysis,” *Mining Text Data*, Springer US, 2012. ISBN 978-1-4614-3222-7

Lu, Q., Q. Ye, and R. Law, “Moderating Effects of Product Heterogeneity Between Online Word-of-Mouth and Hotel Sales,” *Journal of Electronic Commerce Research*, Vol. 15, No. 1:1-12, 2014.

Maiti, S. and D. Samanta, “Clustering Web Search Results to Identify Information Domain,” *Emerging Trends in Computing and Communication, Lecture Notes in Electrical Engineering*, Vol.298, 291-303, 2014.

Meng, X. and H. Wang, “Mining user reviews: from specification to summarization,” Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pp. 177-180, 2009.

Miller, G.A., “Wordnet: a lexical database for english,” *Commun ACM*, Vol. 38, No. 11:39-41, 1995.

Missen, M., M. Boughanem, and G. Cabanac, “Opinion mining: reviewed from word to document level,” *Social Network Analysis and Mining*, Vol. 3, No. 1:107-125, 2013.

Pang, B. and L. Lee, “Opinion Mining and Sentiment Analysis,” *Foundations and Trends in Information Retrieval*, Vol. 2:1-135, 2008.

Pang, B., L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques,” *Proceedings of the ACL-02 conference on Empirical methods in natural language processing 2002 Association for Computational Linguistics*, 2002.

Papacharissi, Z., and M. de Fatime Oliveira, “Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt,” *Journal of Communication*, Vol.62, No.2:266-282, 2012.

Pedersen, T., S. Patwardhan, and J. Michelizzi, “WordNet::Similarity: measuring the relatedness of concepts,” in Demonstration Papers at HLT-NAACL, pp. 38-41, 2004.

Popescu, A., and O. Etzioni, “Extracting product features and opinions from reviews,” *Natural Language Processing and Text Mining*, 9–28, 2005.

Porter, M. F., “An algorithm for suffix stripping,” *Program*, Vol. 14, No. 3:130-137, 1980.

Pu, P., L. Chen, and P. Kumar, “Evaluating product search and recommender systems for E-commerce environments,” *Electronic Commerce Research*, Vol. 8, No. 1:1-27, 2008.

Purnawirawan, N., N. Dens, P. D. Pelsmacker, “Expert Reviewers Beware! The Effects Of Review Set Balance, Review Source And Review Content On Consumer Responses To Online Reviews,” *Journal of Electronic Commerce Research*, Vol. 15, No. 3:162-178, 2014.

Shahabi, C. and F. Banaei-Kashani, “Efficient and anonymous web-usage mining for web personalization,” *Journal of Computing*, Vol.15, No. 2:123-147, 2003.

Shafiq, O., R. Alhajj and J. G. Rokne, “On personalizing Web search using social network analysis,” *Information Sciences*, Vol.314, No.1:55-67, 2015

Scaffidi, C., K. Bierhoff, E. Chang, M. Felker, H. Ng, and C. Jin, “Red opal: product-feature scoring from reviews,” 8th ACM Conference on Electronic Commerce, 182–191, NY, USA, 2007.

Stehman, S. V., "Selecting and interpreting measures of thematic classification accuracy". *Remote Sensing of Environment*, Vol.62, No. 1: 77–89, 1997.

Tam, K. Y., and, S. Y. Ho, “Understanding the impact of web personalization on user information processing and decision outcomes,” *Mis Quarterly*, Vol.30, No. 4:865–890, 2006.

Tang, H., S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," *Expert Systems with Applications: An International Journal*, Vol. 36, No. 7:10760-10773, 2009.

Tatemura, J., "Virtual reviewers for collaborative exploration of movie reviews," *Proceedings of the 5th International Conference on Intelligent User Interfaces*, New Orleans, Louisiana, United States ACM, 2000.

Thirumalai, S. and K. K. Sinhab, "Customization of the online purchase process in electronic retailing and customer satisfaction: An online field study," *Journal of Operations Management*, Vol. 29, No. 5:477-487, 2011.

Tumasjan, A., T. Sprenger, P. Sandner, and L. Welpe, "Election forecasts with Twitter: How 140 characters reflect the political landscape," *Social Science Computer Review*, Vol.29, No.4:402-418, 2011.

Valitutt, R., "Wordnet-affect: an affective extension of wordnet," Proceedings of the 4th international conference on language resources and evaluation, pp. 1083-1086, 2004.

Wan, Y. and M. Hakayama, "The Reliability of Online Review Helpfulness," *Journal of Electronic Commerce Research*, Vol. 15, No. 3:179-189, 2014.

Wang, D., S. Zhu, and T. Li, "SumView: a web-based engine for summarizing product reviews and customer opinions," *Expert System with Applications*, Vol. 40, No. 1:27-33, 2013.

Wang, W., and I. Benbasat, "Trust In and Adoption of Online Recommendation Agents," *Journal of the Association for Information Systems*, Vol.6, No.3, Article 4, 2005

Williams, C., and G. Gulati, "What is a social network worth? Facebook and vote share in the 2008 presidential primaries," *In Annual meeting of the American Political Science Association*, 1–17, Boston, MA: APSA, 2008.

Wilson, T., J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level ssentiment analysis," Proceedings of the conference on human language technology and emprical methods in natural language processing, pp. 347-354, 2005.

Xua, H., X. (Robert) Luob, and J. M. Carrolla, and M. B. Rossona, "The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing," *Decision Support Systems*, Vol.51, No.1:42-52, 2011

Yu, H. and V. Hatzivassiloglou, "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences," Proceedings of EMNLP-03, pp. 129-136, 2003.

Zhang, E. and Y. Zhang, "UCSC on TREC 2006 blog opinion mining." Proceedings of the text retrieval conference (TREC-2006), USA, 2006.

Zhang, Z., Q. Ye, R. Law, and Y. Li, "Automatic detection of subjective sentences based on chinese subjective patterns," *Communications in computer and information science*, Vol. 35: 29-36, 2009.

Zhuang, L., F. Jing, and X.Y. Zhu, "Movie review mining and summarization," Proceedings of the 15th ACM international conference on Information and knowledge management, pp. 43-50, 2006.

Zhu, L., G. Yin, and W. He, "Is This Opinion Leader's Review Useful? Peripheral Cues for Online Review Helpfulness," *Journal of Electronic Commerce Research*, Vol. 15, No. 4:267-280, 2014.