Decision Support Systems 81 (2016) 41–53

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

From valence to emotions: Exploring the distribution of emotions in online product reviews

Rahat Ullah [a], Naveen Amblee [b], Wonjoon Kim [a,c,*], Hyunjong Lee [d]

[a] Graduate School of Culture Technology, KAIST, Yuseong-gu, Daejeon 305-701, Republic of Korea
[b] Indian Institute of Management Kozhikode, India
[c] Department of Business and Technology Management, KAIST, 291 Daehak-ro, Yusoeng-gu, Daejeon 305-701, Republic of Korea
[d] Kakao (Data mining and consulting company), Republic of Korea

ARTICLE INFO

Article history:
Received 27 February 2015
Received in revised form 27 July 2015
Accepted 26 October 2015
Available online 3 November 2015

Keywords:
Online reviews
Word of mouth
Emotional content
Natural Language Processing
Search and experience goods

ABSTRACT

Word-of-mouth (WOM) in the form of online customer reviews has received considerable attention by practitioners and academics. Prior literature has focused more on the understanding of the phenomenon using the frequency or overall rating/valence information of WOM, while questions on how firms can potentially use or design online WOM platforms and benefit from it based on the *content* of WOM are still open, and need more attention from researchers. In addition, an important antecedent for the generation of word-of-mouth is a strong emotional imbalance known as schema discrepancy, which is considered to trigger the consumer to post a customer review online. However, only a limited number of studies to date have actually examined the emotional content of reviews to validate this line of reasoning. To fill this gap, we analyzed the *emotional content* of a large number of online product reviews using Natural Language Processing (NLP) methods. We find that there is a difference in the emotional content of reviews across search and experience goods in the early stages of product launch. However, interestingly, these differences disappear over time as the addition of reviews reduces the information asymmetry gap. This suggests that traditional experience goods are evaluated more like search goods in online environments, because consumers can easily evaluate attributes of products prior to purchase based on the reviews accumulated. In addition, we find that more extreme reviews have a greater proportion of emotional content than less extreme reviews, revealing a bimodal distribution of emotional content, thereby empirically validating a key assumption that underpins much of the extant literature on online WOM. Furthermore, reviews have a greater proportion of positive emotional content within positive extreme ratings as compared to negative emotional content within negative extreme ratings which is a major factor in online WOM generation, and helps explain the commonly observed J-shaped distribution of reviews. Our findings suggest important managerial implications regarding product development, advertisement, and platform design using WOM content.

1. Introduction

Word-of-mouth (WOM) has been described as one of the most important means of informal communication among consumers [16,19,29,59]. Among them, online customer reviews for products or services can be formally defined as peer-originated product evaluations placed on company or third-party web sites [41]. In the case of retail websites (e.g., Amazon.com, Google Play Store, etc.), customer reviews include both the textual product reviews written by customers along with product evaluations in the form of numerical star ratings (e.g., between 1 and 5 stars). Such forms of online product reviews have become one of the key drivers influencing product sales and corresponding marketing strategies because they provide useful information to consumers as well as to product manufacturers and retailers [11,16].

Therefore, previous studies of WOM have examined the importance of online WOM in general, and user-generated product reviews in particular. Specifically, they focused on the volume and valence of online WOM and their effect on product sales [11,15,18,25,40,65]. The findings suggested that while the volume of product reviews have shown a direct correlation with product sales [11,15,25,40,65] through a cascading effect [14], the effect of the valence of the reviews on sales has been less conclusive [11,65]. In other words, these previous studies' scope of inquiry has been limited to WOM frequency or related factors.

However, a key problem with analyzing the volume and valence of WOM to understand consumer's WOM behavior is the under-reporting bias as suggested by Hu et al. [26], whereby only those consumers who are extremely satisfied or extremely dissatisfied write reviews, while

* Corresponding author.
E-mail addresses: rahatkaist@kaist.ac.kr (R. Ullah), amblee@iimk.ac.in (N. Amblee), wonjoon.kim@kaist.edu (W. Kim), lhjgine@kaist.ac.kr (H. Lee).

http://dx.doi.org/10.1016/j.dss.2015.10.007
0167-9236/

those consumers who felt that the product was "just okay" may not be motivated to post a review. This usually leads to a bimodal pattern of review distribution, where the extreme reviews (extreme positive such as 5-star rating and extreme negative such as 1-star rating) tend to dominate in numbers. It has been further suggested that the bimodal distribution is a *J*-shaped distribution [1,11,16,26,36,51] because of the self-selection bias, where consumers who are favorably disposed to a product purchase and then evaluate it [26]. In this paper we attempt to study this skewed bimodal distribution using an emotion-framework known as schema discrepancy [19], where eWOM is viewed as an online form of catharsis (emotional release) [29,50].

In the emotion-framework, it has been shown that emotions play an important role in consumer response, thus firmly establishing the significance of emotion to consumer behavior [46,49,64]. Especially, previous studies consistently emphasized the important role of emotions in social sharing, particularly with respect to consumer WOM behavior. For example, both Berger and Milkman [5] and Stieglitz and Dang-Xuan [57] recently showed how emotional content significantly affects the sharing of online content using psychological field experiments and emotional twitter messages. As such, it is important to examine the emotional content in product reviews to better understand the idiosyncrasies of online WOM communications. However, we are unaware of any previous research that looks more specifically at the emotional makeup of review content, although emotions are the driving force behind online review articulation [1,63]. In this study, we examine the emotional content within reviews based on the schema-discrepancy theory for eWOM.

In addition to acting as a form of catharsis, online WOM communications are very helpful to prospective consumers who use online reviews to help make a decision. Indeed, one of the goals of writing a review (and a means of obtaining catharsis) is to influence decision makers who are seeking information about the product. Consumer read reviews to reduce their purchase risk, and to get advice on whether to purchase the product or not. In other words, reviews can also be a source of product-usage information [29,30]. Therefore, reviews are particularly important for experience goods, which are products dominated by attributes which can only be evaluated after consumption, as opposed to search goods, whose dominant attributes can be evaluated prior to purchase [13,37,47]. Since consumers cannot evaluate experience goods properly prior to purchase, they tend to rely on the past experience of others, which are expressed as online reviews. We expect that as reviews pile up, and consumers have access to more and more past consumption experiences of others, experience goods may be evaluated more like search goods, in that their dominant attributes can now be evaluated prior to purchase and consumption. In a similar vein, it is expected that the schema-discrepancy (expected versus actual) experienced by these later consumers will be less, leading to less emotional reviews.

In this paper we develop an integrated approach toward eWOM, by connecting the related phenomena of customer satisfaction, schema discrepancy, information search, and product type to explain the distribution of reviews. This is important as the extant literature on eWOM is not conclusive and even sometimes contradictory, and this research will help toward understanding the underlying processes at work. Since consumers read reviews to help fill in their information gaps prior to making a decision, it is important that we gain an understanding of the emotional makeup of reviews across product types (search vs. experience) and across time (accumulation of reviews).

To this end, we analyze emotional content and its characteristics in WOM communications by analyzing the emotional content of a large number of online product reviews on Amazon.com using Natural Language Processing (NLP) techniques. By using NLP, we computationally analyze and understand the natural human linguistic aspects of WOM focusing on emotions. More specifically, using NLP, we are able to examine the proportion of emotional content in online product reviews across different star ratings and the distribution of emotional content in positive and negative product reviews as well as over time, which enables us to characterize unexplored characteristics of online WOM content. In addition, we also overcome the limitation of previous studies that focused on only a similar set of products, by examining the differences among customer reviews depending on product types (search vs. experience goods), which were considered to be open questions until now [15,22].

We found that there is a difference in the emotional content of reviews between search and experience goods in the early stages of product launch. However, these differences disappear over time as the addition of reviews reduces the information asymmetry gap. This provides support to the notion that traditional experience goods are evaluated more like search goods in online environments [28], because consumers can easily evaluate products prior to purchase based on the reviews accumulated. In addition, we found that more extreme reviews have a greater proportion of emotional content than less extreme reviews, revealing a bimodal distribution of emotional content, thereby empirically validating a key assumption that underpins much of the extant literature on online WOM. Furthermore, reviews have a greater proportion of positive emotional content within positive extreme ratings as compared to negative emotional content within negative extreme ratings which helps explain the skewed bimodal or *J*-shaped distribution which is commonly found in the frequency distribution of online reviews [10,11,26,36,51].

Our findings not only enable us to understand the fundamental nature of customer reviews, but also suggest several important implications for managers of electronic commerce, especially regarding product development, renovation, and advertisement. For example, product features which occur in 5-star reviews can be emphasized in advertisements because those features are the most emotionally stimulating aspects for consumers. Even the negative emotions can also be used to improve their offerings on the production side and to troubleshoot their inventory and service from a customer service perspective. The longer reviews with less emotional words in mid ratings regions may also reveal valuable information about the strength and weakness of the products that consumers have experienced. Lastly, our approach using NLP methods in analyzing WOM contents introduces a way to utilize large numbers of customer reviews that are usually a rich but unstructured set of consumer data with noise and unusable information in order to extract important marketing insights from them.

To this end, in Section 2, we discuss the theoretical background regarding the role of emotions in consumers' WOM behavior and related WOM distributions. Section 3 presents the pretest we implemented in order to categorize our data into different product types, and the methods we use to test our hypotheses including NLP techniques. Section 4 provides our results regarding the three hypotheses we suggest in Section 2. Finally, in Section 5 we discuss and summarize the main points of this paper and discuss important managerial implications of our findings.

2. Background and hypothesis

Previous research has investigated the relationship between consumption-related emotions and consumer behavior [46,49,63,64]. They showed that emotions play an important role in consumer response, thus firmly establishing the significance of emotion to consumer behavior [49]. In other words, they showed that emotions yield important psychological consequences and generate long-lasting mnemonic recurrences as well as acting as motivation for social sharing [54]; any highly satisfactory or highly dissatisfactory experience elicits strong emotional responses and these charged emotions create an emotional imbalance [1]. This emotional imbalance is caused by schema discrepancy, whereby expectations about a product or service are not realized, leading to surprise [19]. Then, a second emotion follows, such as joy (positive) or anger (negative), which causes one to assume that either positive or negative surprise was elicited [1,29]. This emotional imbalance and consequent emotion of surprise requires the restoration of balance through the expression of positive emotions or through the venting of negative feelings [29]. In an online environment, this act of catharsis [2] is said to take the form of consumer WOM. Therefore, it is claimed that customer reviews in the form of WOM communications are emotional sharing mechanisms which play a crucial role in emotional recovery, relief, and other aspects of social interaction [19,29,42]. However, there is limited empirical research connecting emotions with online reviews. As such, it is important for us to understand the nature of the emotional content contained within online WOM communications.

Previous studies argued that consumers tend to report more frequently at the extreme levels of satisfaction (i.e., highly dissatisfied or highly satisfied; see [1]). For example, consumers engage in positive word of mouth behavior in order to express their delight with the experience, to increase involvement with the product, and to help the company [29]. On the other hand, by sharing a negative consumption experience through a customer review, consumers can reduce the discontent associated with their negative emotions [29], such as anxiety and tension [59]. Therefore, the motivation for engaging in negative WOM is often referred to as dissonance reduction [7,16,17].[1] In other words, previous research has predicted that stronger emotional responses to a product lead to the posting of customer reviews [19,29,42]. Therefore, we suggest that reviews with more extreme ratings (e.g. 5-Star or 1-Star rated reviews) will have a greater proportion of emotional content than those with less extreme ratings (e.g. 3-Star rated reviews). To provide a solid foundation for previous assumptions regarding emotions and reviews, we propose to validate the following hypothesis:

Hypothesis 1. More extreme reviews will have a greater proportion of emotional content than less extreme reviews.

In addition, the bimodal distribution of reviews, not the emotional contents within reviews, has been identified by previous studies as having a *J*-shaped distribution (instead of a more uniform U-shaped distribution), which shows that online product reviews are overwhelmingly positive, with a dominance of 5-Star ratings, some 1-Star ratings, and very few ratings in between [10,11,26,36,51]. For example, a study of the relationship between book sales and online customer reviews for movies [11] showed a strong *J*-shaped distribution, prompting the authors to concur with previous findings [1,16].

Some studies have attributed this shape to the self-selection bias, where consumers with a preexisting favorable disposition tend to purchase, consume and positively review a product [26]. This bias is attributed to the inherent limitations of WOM frequency information whereby only those consumers with strong emotions post reviews [21,27,29,54], which was discussed in Hypothesis 1. In addition to the self-selection bias stretching the review distribution on the extreme positive side (e.g. 5-star review), another effect stunts the distribution on the extreme negative side (e.g. 1-star review). This is due to the previously described effect where socially unsanctioned emotional experiences (e.g., negative experiences) are less likely to be shared, or they are shared with a delay [1,7,20,42,64]. This may be explained as people may not share the negative experiences associated with the events that trigger reactivation of negative feelings [5,54,55]. In addition, individuals feel reluctant to transmit bad news to avoid feelings of guilt [1].

Indeed, in the case of social media studies have shown that positive content further boosts virality as compared to negative content in social media, except in the case of the news domain [31]. Furthermore, Berger and Milkman [5] also found that people prefer sharing positive rather than negative news when they analyzed a dataset of online New York Times articles. Therefore based on the literature and previous studies on online sharing, we can hypothesize that in addition to extreme reviews having more emotional content, positive extreme reviews will have greater emotional content (due to the self-selection bias) and extreme negative reviews will have less emotional content than their positive counterpart, due to the avoidance of socially unsanctioned experiences. This would lead to the following hypothesis:

Hypothesis 2. In online product reviews, there is more positive emotional content within positive extreme ratings as compared to negative emotional content within negative extreme ratings.

2.1. Information search and the impact on search versus experience products

It is reasonable to expect that the nature of the distribution of emotional content will depend on the type of good involved. Therefore we further examine the distribution of emotional content in WOM based on product types, be they search or experience goods [13,37,47][2]. It is because the likelihood of product attributes eliciting surprise – hence motivating consumers to generate more WOM – is likely to be different for experience and search goods, as experience goods cannot be satisfactorily evaluated before consumption. For a given product, one of these attributes typically dominates, making that product search-dominant or experience-dominant. Thus, product groups are commonly referred to as search products or experience products depending on which attribute is more dominant [37]. Previously we noted that the elicitation of surprise is generally followed by an emotional outburst of joy or anger, which further elicit social sharing via positive or negative WOM [19,53]. Therefore, we can expect that because experience goods have dominant attributes which cannot be evaluated before consumption, the likelihood of these attributes eliciting surprise would be higher than those for search goods. In other words, emotional content of online WOM should be greater in cases of experience goods compared with that of search goods.

On the other hand, if consumers do not have enough preexisting information regarding the product being evaluated, they will engage in information search in order to become informed [41]. According to the theory of information search put forward by Stigler [60], consumers will continue to engage in information search about a product as long as the marginal benefits of this search exceed the marginal costs associated with hunting for information. In this perspective, the Internet has greatly reduced these costs of search, by making information available at the click of a button [34]. In particular, online reviews have transformed word-of-mouth communications, which in the pre-Internet era was hard to obtain and fleeting, into easily available and permanent communications between consumers. Indeed, research on online WOM has confirmed that consumers primarily read online reviews to reduce risks associated with a purchase and to reduce the search costs associated with the decision making process [29].

Therefore, Huang et al. [28] discussed that consumers' perceived ability to judge the quality of a product before purchase changes in the online environment. They showed that “consumers spend similar amounts of time online gathering information for both search and experience” goods (p.55). Essentially, this seems to suggest that consumers are able to adequately evaluate traditional experience goods in an online environment due to the extensive and standardized information that is provided on online retailing websites. In other words, products available online with detailed accompanying information such as online reviews can be evaluated as search goods, even though they may be classified as experience goods in a traditional offline retail environment. This is because the additional product information available in online

[1] It has also been suggested that consumers are motivated to engage in negative WOM for altruistic, vengeance-related, and advice-seeking-related reasons [59].

[2] We use the search vs. experience model as it is among the most widely used and accepted product categorization models in the literature. We excluded credence goods from our WOM analysis due to their limited implications with respect to our research.

Table 1
Mean scores and standard deviations for ability to judge the performance of each product before and after consumption.

Product	Before		After		Classification
	Mean	S.D.	Mean	S.D.	
HP LaserJet P1006 Printer	5.60	2.34	6.59	2.16	S
HP Office jet 6310 All-in-One Printer	5.63	2.14	6.88	2.02	S
Canon Power Shot 8MP Digital Camera	5.55	1.99	6.38	1.82	S
Panasonic Lumix DMC-FZ28K 10MP Digital Camera	5.34	1.93	6.40	1.64	S
Microsoft Office Home and Student 2007	5.14	2.45	6.47	2.39	S
QuickBooks Pro 2009 Software	5.16	2.46	6.32	2.25	S
Adobe Photoshop Elements 7 Software	5.55	2.27	6.71	1.89	S
Microsoft Office Professional 2007 Full Version	5.84	2.40	6.85	2.45	S
Panasonic SRG06FG: 3.3-Cup Automatic Rice Cooker	4.31	1.87	6.09	2.17	E
Sanyo ECJ-D55S : Micro Computerized Rice Cooker/Steamer	4.50	2.08	6.27	2.14	E
Zojirushi NS-KCC05:3Cup Rice Cooker & Warmer	4.28	2.11	5.56	2.17	E
Twilight (The Twilight Saga, Book 1) Paperback	3.88	2.50	5.24	2.18	E
The Shack (Paperback Book)	3.71	2.24	5.12	2.37	E
Quantum of Solace (2008 Movie)	4.32	2.43	5.75	2.64	E
Planet Earth — The Complete BBC Series (2007 TV Show)	4.66	2.20	6.08	2.56	E
Dior Dior Show Mascara	3.15	1.97	5.02	2.49	E
True Blood: The Complete First Season (HBO Series)	3.80	2.42	7.29	8.98	E
Average	4.73	2.22	6.18	2.61	

environments including online reviews enables consumers to make informed purchasing decisions, thereby reducing risk and maximizing utility.

If this assumption on the changes in evaluation of experience attributes of experience products as search attributes in an online environment is correct (note that we only refer to the ability of consumers to evaluate the attribute — the attribute itself does not change), the difference in the distribution of emotional content of reviews between search and experience goods should be small or insignificant if a sufficient volume of informative WOM is available. Put differently, we can expect experience goods with few reviews — either in total or at an early stage of product launch, to exhibit a greater level of emotional content as compared to search goods. However, later consumers who benefit from these early reviews will experience less information asymmetry and therefore less schema discrepancy. In other words, the experience good starts to be evaluated like a search good as reviews pile up. Therefore, when later reviews are considered, the difference in emotional content should cease to exist, as the information asymmetry gap should have been bridged by means of the earlier reviews. However, based on our Hypothesis 2, there will be two caveats. First, the positive emotional content in positive extreme reviews (5-star) should be greater, due to the self-selection bias, including a sense of triumph of having made the "correct" purchasing decision. Second the negative emotional content in negative extreme reviews (1-star) will be suppressed due to the tendency to avoid spreading negative personal experiences. Thus, we state the following hypotheses:

Hypothesis 3. Early online reviews of experience goods have greater proportions of emotional content when compared to search goods, while later online reviews of experience and search goods will have similar levels of emotional content. This effect will be more pronounced for positive reviews than for negative reviews.

3. Method

We conducted an empirical study using online customer reviews from online retailing giant Amazon.com. We developed a custom software tool to automatically retrieve all customer reviews for a sample of items offered on Amazon.com and subsequently analyzed 15,849 online customer reviews. To analyze such a large number of reviews systematically, we used several NLP techniques, as explained in the Method section. We selected 17 randomly chosen products from the following general product categories on Amazon.com: "Computers and Offices"; "Electronics"; "Home and Garden"; "Books"; "Movies, Music and Games"; "Toys, Kids and Babies"; and "Grocery, Health, and Beauty," We then conducted a pretest to classify the products into search and experience products.

3.1. Pretest

To classify the products as search and experience products, we use the method outlined by Krishnan and Hartline [37], and Hsieh et al. [32]. A sample of 58 undergraduates at a leading university in South Korea was used for the pretest. Students were provided with a list of the selected Amazon.com products, along with a short explanation about how some products could be easily evaluated prior to purchase, while others were more easily evaluated post-consumption. Students were then asked to rate their ability to judge the performance of each product *before* purchase on a 9-point Likert scale. The scale ranged from 1 ("Not at all") to 9 ("Very well"). Next, the students were provided with another explanation about how some products could not be easily evaluated even after consumption. Students were asked to use the same scale to rate their ability to judge their performance of each product *after* purchase and consumption. We randomly presented the products to the students to avoid ordering artifacts.

The mean scores and standard deviations for the ratings are provided in Table 1. Following Krishnan and Hartline [37] and Huang et al. [28], we classify products with a higher score than the scale midpoint of 5 (before $t = 3.61$, $p < .007$, after $t = 22.906$, $p < .001$) on both scales as search products (S), because this rating pattern implies that these products are perceived as being easy to evaluate prior to purchase[3]. Products with a score lower than the scale midpoint of 5 ($t = -6.691$, $p < .001$) on the first evaluation but a higher score than the scale midpoint of 5 ($t = 3.664$, $p < .005$) on the second evaluation were classified as experience products (E) because they are perceived as being more difficult to evaluate prior to purchase but easy to evaluate post-consumption[4].

We also manipulated the classification with stronger criteria, such as the mean and the maximum deviation value of all products. For example, products rated greater than the mean value of all products on

[3] Products with low scores on both scales can be classified as "credence" goods, or goods perceived as being difficult to evaluate even after consumption. However, we excluded such credence goods from our WOM analysis as described in footnote 5.

[4] In the classification evaluations, the description of the product was not present to avoid categorization biases.

both scales were categorized as search products, while products rated less than the mean value on the first scale but greater than the mean value on the second scale were categorized as experience products. Even with this manipulation, however, the results were consistent with the expected previous criteria. Therefore, based on this classification system, we are able to place each product into only one of the two categories.

3.2. NLP and opinion mining

Natural Language Processing (NLP) is a branch of artificial intelligence within computer science that deals with analyzing, understanding, and generating the natural languages (as opposed to computer languages) that humans use. One of the key aspects of NLP is the identification of keywords that carry the semantics of emotional strength of words in terms of their positivity and negativity. Recently, language modeling is being widely used for speech recognition, part-of-speech tagging, syntactic parsing, and information retrieval [61]. One approach of language modeling is N-gram models where an n-gram is defined as a contiguous sequence of *n* items consisting of words, letters, syllables, phonemes or base-pairs from a set of text [6,68]. Among the different n-gram models, we used unigram and bigram for the analysis of opinion words extraction and to figure out the context of each word. Eq. (1) represents the unigram model while Eq. (2) represents the bigram model. Unigram models represent the probability of each word in the document and its assumption is that all words are independent to each other. Bigram models represent a document as the product of probabilities of two word phrases [4]. In this model, a word probability is dependent on its previous word. We built a “bag-of-words” model and a “bag-of-phrases” model: a document could be represented as a vector of word counts, where the dimension of each vector is the number of unique words in the dataset. This approach of analyzing word frequencies is called the “bag-of-words” model in NLP. “Bag-of-phrases” uses unique two consecutive words instead of unique single words [12].

$$\text{Unigram} : \text{P(Document)} = \text{P}(Word_1)\text{P}(Word_2)\cdots\text{P}(Word_n) \tag{1}$$

$$\begin{aligned}\text{Bigram} &: \text{P(Document)}\\ &= \text{P}(Word_1)\text{P}(Word_2|Word_1)\cdots\text{P}(Word_n|Word_{n-1})\end{aligned} \tag{2}$$

Thus we build the language models to extract features as well as to understand the context of each opinion word. Here, we explain several of the important NLP concepts we used: first, we counted each unique word in a review along with how many times it appeared, regardless of word order. Once this was done, the document could be represented as a vector of word counts, where the dimension of each vector is the number of unique words in the dataset, i.e. “bag-of-words” model. Then, we use *SentiWordNet* to extract emotional strength. Here, *SentiWordNet* is a lexical resource that assigns three scores for each word in terms of positives, negatives, and objectives with ternary classifiers based on *WordNet* that is a large lexical database of English built by Miller [23,24,45]. Esuli and Sebastiani [23] state that the method is “based on the quantitative analysis of the glosses associated to synsets that are sets of cognitive synonyms, and on the use of the resulting vectorial term representations for semi-supervised synset classification. The three scores are derived by combining the results produced by a committee of eight ternary classifiers, all characterized by similar accuracy levels but different classification behaviors” (p.417) [24].

Due to the fact that the vector dimension is very high, the comparison of reviews is computationally complex. To reduce this dimensionality, we used one of the NLP “feature selection” algorithms to determine which words were the most salient in reviews. The basic concept behind this algorithm is that words that appear frequently in only a small number of reviews are more salient than words that appear frequently in most reviews. For example, in a camera review data set, most of the reviews will contain the words “picture,” “camera,” and “take.” Thus, with respect to our task of identifying words that carry emotion, such words are not useful. The “feature selection” algorithm is thus used to determine which meaningful words, or words useful to our purpose, are the most salient when comparing reviews.

The well known feature selection methods are term frequency, information gain, mutual information, and chi-square statistics, to name a few. Among the various feature selection algorithms, we used one based on the chi-square test that has been shown to be an effective feature selection method [35], excellent in text categorization [66] as well as being suitable for an imbalanced-dataset [44]. Jiang and Argamon [35] also showed that the chi-square method is analytical and computationally efficient, making feasible the processing of huge volumes of data in real time. Here, we set the two variables as the word and the class, where the class was either negative or positive. Thus, if a word

Table 2
An example of most frequent emotional words for selected products.

Search products				Experience products			
Cannon Camera		HP Laser Jet printer		Quantum of Solace		Planet Earth	
Positive emotional words	Negative emotional words	Positive emotional words	Negative emotional words	Positive emotional words	Negative emotional words	Positive emotional words	Negative emotional words
Love	Disappointment	Pleased	Worse	Excellent	Worse	Amazing	Disappointment
Great	Worse	Happy	Displeased	Love	Bad	Love	Bad
Pleased	Unhappy	Amazing	Angry	Great	Terrible	Awesome	Boring
Wonderful	Bad	Awesome	Worried	Enjoyed	Disappointed	Wonderful	Unhappy
Happy	Stupid	Excellent	Disappointment	Terrific	Horrible	Stunning	Terrible
Amazed	Suck	Great	Awful	Spectacular	Ridiculous	Glad	Offense
Excellent		Love	Hate	Wonderful	Stupid	Pleased	Sad
Surprise		Wonderful	Bad	Superb	Garbage	Great	Deprived
Fun		Perfect	Suck	Fun	Shocked	Enjoy	Horrible
Terrific		Fun	Horrible	Tremendous	Sorry	Excellent	Scared
		Surprise	Bother	Wonderful	Suck	Fantastic	Sorry
			Bore			Incredible	Suck
						Spectacular	Worst
						Superb	
						Happy	
						Phenomenal	
						Thrilling	
						Unbelievable	
						Outstanding	
						Surprise	

occurred as frequently in the positive class as it did in the negative class, then the frequency of that word was considered to be independent of class, and thus its chi-square value would be low. If, on the other hand, a word occurred frequently in the positive class but infrequently in the negative class, then the word would not be considered independent of class, and the chi-square value would thus be high. The computation of the chi-square value was as follows:

$$x^2(D,t,C) = \frac{\left(N_{t,c} - E_{t,c}\right)^2}{E_{t,c}} \quad (3)$$

where D is the data set, t is the random variable for words (terms), c is the random variable for the class (positive or negative), N is the observed frequency count of term t in class c, and E is the expected frequency of term t in class c.

Then, the reviews dataset is divided into positive reviews and negative reviews. For the positive and negative review datasets, bag-of-word models are built to apply the feature selection method. If a word is more relative to positive or negative, it will have a higher chi-square value and if a word is not relative to positive or negative, it will have a value nearer to zero. The processes of our NLP system are as follows:

1. *Stop Word Removal*: We removed stop words such as "the," "as," and "of" that occur very often in English documents but do not carry semantic importance. We used a stop-word list via an Application Programming Interface (API) in the Onix Text Retrieval Toolkit 5 (Lextek International, U.S.A.).
2. *Ranking by Word Frequency*: We ranked words by frequency, which reflects the general topics of the documents.
3. *Feature Selection*: We used the feature selection method of comparing the chi-square distributions of the words based on their appearance in positive and negative reviews.
4. *Valence Scoring*: For each word in the top-ranked word lists, we used *SentiWordNet* to retrieve its positive, negative, and objective scores.
5. *Two-word Phrases*: The words we chose as the top-ranked words were sometimes ambiguous, because most words in natural language have multiple possible meanings. Therefore, we considered which words occurred most often next to those words. We called a word and its immediately following word as "two-word phrase". We counted the frequencies of the words immediately following the top-ranked words, and selected the top five adjacent words in frequency. For example, one of top ranked words, "memory", has multiple meanings i.e. ability to remember, or component of electric devices. The most frequent word appearing to the right of "memory" is "card" in the reviews. It means that the word "memory" is used as "memory card" or "memory cards" in the user reviews of cameras. We show the top frequent emotional words for two search and two experience products as examples in Table 2.
6. *Emotions Content*: Based on the two-word phases extracted, we further retrieve emotional content from two-word phrases using the bigram model. In doing so, we use emotion measures suggested in the psychology literature. Detailed discussion on the measures is described below.

3.3. Emotion measures

Previous research, including studies from the psychology literature [33,39,48,56,62,63], examined the categorizations of emotions. Prior studies have also suggested the importance of such categories in directing analysts to follow a systematic method of evaluation, thus improving correspondence among the categorizations. For example, Richins [49] analyzed the previous literature in depth to identify a set of emotion descriptors, and Derbaix and Vanhamme [19] suggested seven primary emotions and descriptors that could be used to represent and measure them, as presented in Table 3.

Table 3
Primary emotions identified in the previous literature.

[33], (p.7)	[48], (p.249)	[49], (p.144)	[19], (p.111)
Fear	Fear	Fear, Worry	Fear (Fearful, Afraid, Scared)
Anger	Anger	Anger	Anger (Angry, Mad, Enraged)
Enjoyment	Joy	Joy	Enjoyment (Joyful, Delighted, Happy)
Distress/Sadness	Sadness	Sadness, Discontent, Loneliness	Distress (Sad, Downhearted, Discouraged)
Disgust	Acceptance	Peacefulness	Disgust (Disgusted, Feeling of Distaste or Revulsion)
Interest	Disgust	Excitement, Optimism	
Surprise	Expectancy	Surprise	Surprise (Surprised, Amazed, Astonished)
Shame	Surprise	Shame	
Humiliation		Guilt	
Contempt		Romantic Love	Contempt (Contemptuous, Disdainful, Scornful)
		Love	
		Contentment	
		Envy	

Sentiments are generally confused with emotion. A sentiment is built when a certain object is constantly thought of or perceived by a person and, over time, the person creates a frame of mind toward the object such as good/bad, positive/negative, favorable /unfavorable, etc. [9]. Therefore, sentiments are differentiated from emotions by the duration in which they are experienced [43]. Sentiments have been found to be held and formed for a longer period and more stable than emotions, while emotions are complex psychological phenomenon [43] as shown in Fig. 1. Moreover, sentiments are more stable and dispositional than emotions [8] and are formed and directed toward an object[5], whereas emotions are not always targeted toward an object [43]. Therefore, sentiment words are different from emotional words, and we focus only on emotional words as shown in Table 3. We show an example of sentiment and emotional words in Table 4.

Using the information from Table 3, we identified emotional content in WOM reviews. Since it was not possible to analyze all 15,849 reviews manually, we used NLP-based computational techniques, which include ranking by word frequencies, 'stop word' removal, feature selection, and valence scoring for each product and each number-of-stars level. In addition, adjacent words were also considered in the analysis to correctly identify the emotional words used in the classification. We also manually checked emotional content in order to confirm that they are actually matched one of the seven categories of emotions listed in Table 3. Examples of emotional content extracted from the WOM reviews are shown in Table 5. We then aggregated the frequencies of the emotional words used in each of the reviews according to their star ratings.

3.4. Overall NLP algorithm

Fig. 2 shows the overview of the NLP algorithm of our study. After we retrieved the review data from Amazon.com, we first ran the unigram model of NLP which included stop word removal and ranking word frequency. Based on the results of the unigram model analysis, we ran a feature selection process based on the chi-square distributions of the words and valence scoring process. Using the results of feature selection and valence scoring processes, we ran a bigram model to analyze two-word phrases in product reviews. Using the processed two-word phrases, we further analyzed the distribution of emotional content in

[5] An object refers to a person, a thing, a condition, a place or an event at which a mental state is directed [52].

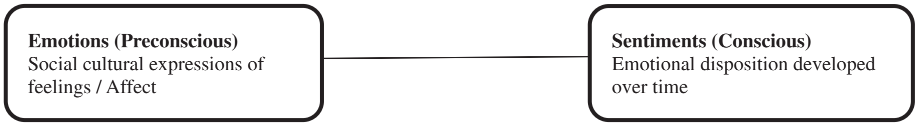

Fig. 1. Difference between sentiments and emotions.

product reviews of those selected products from Amazon.com. This algorithm is unique in the point that we have emotional content analysis in the process of bigram model analysis [3,12]. In addition, this algorithm is more advanced than previous approaches because we combined domain-aware sentiment strength and domain-free sentiment strength [67]. Domain-aware sentiment strength of a word is calculated using the feature selection method and the domain-free sentiment length is extracted from SentiWordNet. We revealed context of each emotional keywords using bigram model.

4. Results

We first examined the distribution of the volume of product reviews based on the number of stars awarded. As shown in Fig. 3, this resulted in a *J*-shaped distribution with the highest volume of reviews associated with 5-Star ratings, followed by 4-Star, 1-Star, 3-Star, and 2-Star ratings, in that order, regardless of product type.

This result is in agreement with the findings of Anderson [1] and Dellarocas and Narayan [16], and confirms that consumers are more likely to communicate their experiences with others when they have very positive/satisfactory or very negative/dissatisfactory experiences, versus more neutral ones. As predicted, we find evidence for the self-selection bias based on the overwhelming dominance of reviews with 5-Star ratings, in agreement with the findings of Hu et al. [26]. In addition, the average number of reviews is much larger for experience products than for search products. This lends some initial support to our theorizing that experience goods are harder to evaluate and as such will more often mismatch with consumers' pre-consumption expectations, thereby generating more WOM online.

The summary statistics of reviews containing emotional content are shown in Table 6. The distribution of non-normalized or absolute emotional content is similar to that of the total number of reviews. However, after we normalized the emotional content, i.e., the percentage value of emotional content in reviews within each star rating for each product, we find that the 5-Star rating is associated with the largest percentage of emotional words (14.76%), followed by 4-Star rating (12.92%), 1-Star rating (9.07%), 2-Star rating (6.54%) and finally the 3-Star rating, which has the smallest percentage of emotional words (4.21%). This pattern seems to conform to the prediction of our first hypothesis.

To validate Hypothesis 1 which stated that more extreme reviews will have a greater proportion of emotional content than less extreme reviews, we performed an ANOVA on the emotional content associated with each star rating for the products in our dataset. The ANOVA results show significant differences in the overall emotional content of extreme and non-extreme reviews (see Table 7, "Overall" column). The percentage of emotional content for positive/negative extreme valences (1-Star and 2-Star vs. 4-Star and 5-Star) is significantly different compared with the middle valence (3-Stars). The means of the percentages of emotional content for the 1–2-Star pairs, the 2–3-Star pairs, and the 3–4-Star pairs are shown to be statistically different. The mean difference for the 5-Star rating was higher than that for the 4-Star rating. These findings confirm our Hypothesis 1 that more extreme reviews will have a greater proportion of emotional content than do less extreme reviews. We also find that the 5-Star and 4-Star ratings both contain significantly more emotional content than the 1-Star rating, which lends preliminary support to our second hypothesis, and is discussed in detail later.

We gain additional insight by examining the length of the reviews associated with the overlapping of the distributions of emotional words, as shown in Fig. 4. The average length of reviews (average number of words per review) peaks at the 2-Star level, showing a reversed *J*-shaped distribution. In other words, as reviews become more emotional, and include a greater number of emotional words, their lengths become shorter. Thus, extreme ratings tend to have fewer numbers of words but higher frequencies of emotional words, while less extreme ratings tended to have greater numbers of words but lower frequencies of emotional words. This is noteworthy because our results show that the ratings of 2 and 3 stars include more non-emotional information, which may be more valuable to both consumers and vendors. This could be interpreted to mean that extreme reviews are mostly

Table 4
A sample of sentiment and emotional words in our data set.

Sentiment words		Emotional words	
Like	Solid	Lovely	Anger
Easy	Better	Amazing	Disgust
Favorable	Unfavorable	Happy	Sorry
Reliable		Wonderful	Boring
Impressed		Pleased	Upset
Simple		Great	Scared
Decent		Excellent	

Table 5
List of all emotional words.

Positive Emotional words		Negative Emotional words	
Love	Beautiful	Disappointment	Uninteresting
Great	Delightful	Worse	Irritating
Pleased	comical	Unhappy	Disgusted
Wonderful	Insightful	Bad	Dirty
Happy	Gorgeous	Stupid	Pitiful
Amazed	Tremendous	Suck	Nonsense
Excellent	Fantasized	Dislike	garbage
Surprise	Cool	Angry	Regret
Fun	Phenomenal	Displeased	Allergic
Terrific	Thrilling	Awful	Sad
awesome	Outstanding	Worried	Punished
Stunning	Unbelievable	Hate	unforgettable
Fantastic	Superb	Horrible	Woeful
Perfect	Laugh	Bother	Deprived
Enjoy	Humorous	Annoying	Ruthless
Pretty	Sexy	Unfortunate	Torture
Entertaining	Excited	Ridiculous	Insult
Terrific	Fascinating	Ugly	Weird
Glad		Shocked	Fear
Wow		Trouble	Terrible
Happy		Frustrating	Embarrassed
Romantic		Scared	Horrifying
Incredible		Pathetic	
Fabulous		Sorry	
Spectacular		Painful	
Nice		Upset	

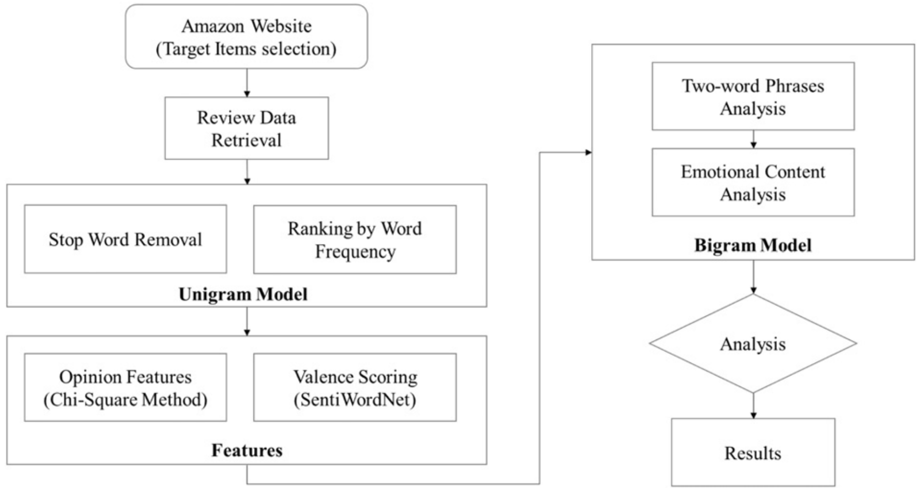

Fig. 2. Overview of NLP algorithm of this study.

emotional/cathartic expressions (either positive or negative) while midrange reviews are more deliberate and thoughtful. As we discuss later in the conclusion section, both (emotional expressions vs. deliberate and thoughtful) can have useful managerial implications.

Our second hypothesis stated that there is a greater proportion of positive emotional content within positive extreme ratings as compared to the proportion of negative emotional content within negative extreme ratings. To validate this hypothesis, we perform several ANOVA tests on our dataset. As mentioned earlier, the results (Table 7, "Overall" column) also validate Hypothesis 2 since the positive reviews (5-Star and 4-Star) both have on average a significantly greater proportion of emotional content than the negative reviews (1-Star and 2-Star). When only the two extreme reviews are considered, the 5-Star review has on average 57.4% more emotional content than the 1-Star review. To check the effects of positive and negative emotional contents more specifically, we analyzed our results further by segmenting the emotional content into positive and negative contents. We do this to account for the fact that even the most glowing review may contain some dissent in the form of negative emotions and vice versa, and even though this will be quite small proportionally, it could obfuscate the results.

We compared the proportion of positive content in 5-Star reviews with the proportion of negative content in 1-Star reviews, and found that on average the 5-Star reviews contained almost twice as much positive emotional content as compared to the negative emotional content in 1-Star reviews ($\mu_{5starpositive} = 15.92\%$, $\mu_{1starnegative} = 7.95\%$, $p < 0.01$). We repeated the analysis to compare 4-Star review to 2-Star reviews, and found that the 4-Star reviews contained nearly two and a half times as much positive emotional content as compared to the negative emotional content in 2-Star reviews ($\mu_{4starpositive} = 12.92\%$, $\mu_{2starnegative} = 5.19\%$, $p < 0.01$). We segmented the products into search and experience goods, and found that the results continued to hold for both categories. These results confirm that positive emotions are more frequently shared than negative emotions. This could indicate that highly satisfactory experiences increase product engagement by consumers, which leads to a greater amount of positive WOM.

When we plot the positive and negative emotional content for each Star-rating level, we find that they intersect between the 2-Star and 3-Star ratings, indicating that this is the area of a neutral review, as the proportion of positive and negative emotional content in the review is about the same (see Fig. 5). Closer scrutiny shows that there is no difference in the amount of positive emotional content in 2-Star and 3-Star reviews, but the 2-Star review contains significantly more negative emotional content (see Table 7, "Positive" and "Negative" columns). Indeed, the 3-Star rating does not contain any more negative emotional content compared to the 4-Star and 5-Star review, indicating that it is very balanced and rational. These results confirm Hypothesis 2.

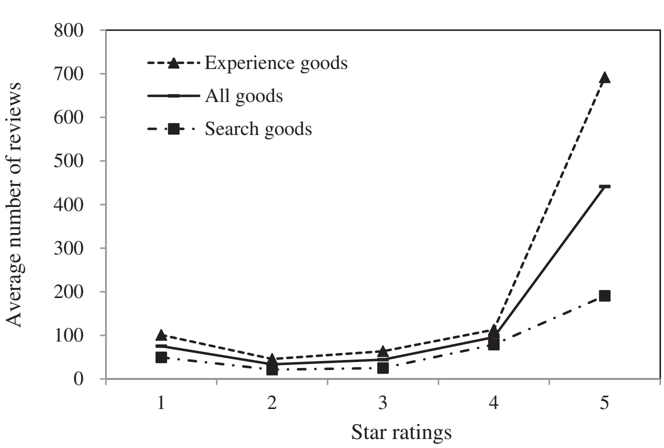

Fig. 3. *J*-shaped distribution of total number reviews.

4.1. Emotional content distribution between search and experience products

For further analysis of the emotional content in product reviews, we segmented the percentages of emotional content into their respective product type categories (i.e., search/experience products). We examined the segmented distributions and confirmed the *J*-shaped distribution of emotional words for both product category types. The results are shown in Table 7 ("Search" and "Experience" columns). Fig. 6 shows the distribution of emotional content across Star-ratings for Search and Experience goods.

For both search and experience products, there were significant differences in the percentage of emotional content between the extreme and mid valences. The extreme valences (1–2-Star ratings and 4–5-Star ratings) show significant differences with the mid valence (3-Star rating) content. Therefore, these results suggest that for both search and experience products, extreme ratings have more emotional content than mid-valence ratings, thus confirming our Hypothesis 1 for both product types. We separated the positive and negative emotional

Table 6
Summary statistics for emotional words in customer reviews.

Search products			Experience products		All products			
Star ratings	Vol.	Average %	Vol.	Average %	Vol.	Average %	Max	Min
1	44 (36.71)	8.96 (4.49)	110.55 (180.31)	9.17 (5.52)	79 (134.23)	9.07 (4.91)	16.57	1.09
2	18 (13.4)	5.24 (2.47)	50.55 (71.94)	7.70 (3.89)	35 (54.29)	6.54 (3.44)	13.56	1.67
3	19.75 (20.81)	3.90 (3.19)	68.66 (93.49)	4.48 (2.42)	46 (72.06)	4.21 (2.73)	9.65	0.46
4	56.87 (46.41)	13.53 (8.25)	117.22 (138.96)	12.39 (4.84)	89 (107.52)	12.92 (6.47)	31.40	1.28
5	142.37 (157.52)	15.40 (3.45)	749.88 (1049.37)	14.18 (5.62)	464 (811.87)	14.76 (4.63)	21.62	0.80
Total	56.2 (85.54)	9.40 (4.37)	219.37 (533.63)	9.58 (4.46)	142.58 (399.09)	9.50 (4.43)	18.56	1.06

content, and found that the results become more pronounced, similar to the results for "Positive" and "Negative"[6].

To further investigate these differences in the emotional content between search and experience products for each star rating, we ran multiple T-tests to compare the emotional content of equivalent Star-ratings. The results (Table 8, "All Emotional Words" column) suggest that there are no significant differences in the emotional content between the search and experience products. However, when we segment the emotional content into positive and negative (Table 8, "Positive Emotional" and "Negative Emotional") we find that experience goods have a small but significantly larger positive emotional content in its negative reviews (1-Star and 2-Star) as compared to search goods.

Then, in order to understand the *evolution* of emotional WOM content over time (and validate our third hypothesis), we arranged reviews for individual products chronologically, and then separated the reviews posted in the first three months of product launch from the reviews posted over the remaining time period, i.e., until our data collection ended. We used three months as the cutoff point as this criterion has been used previously by Li and Hitt [38] while studying the longitudinal effects of WOM[7]. We once again compared the emotional content of various star ratings for search vs. experience goods by means of T-tests. We performed the test for both subsets of reviews — those posted in first 3 months and remaining reviews.

The results show that when only the first three months of reviews are considered, there is a significant difference in the emotional content of 5-Star reviews for search vs. experience goods (results in Table 8, "First quarter emotional" column, Fig. 7). This implies that when reviews are few (as in the first quarter period) there will be information asymmetry regarding the quality of experience goods, leading to greater occurrences of schema discrepancy, which in turn will lead to more emotional reviews being posted. As hypothesized, this difference is only found to be significant for the extreme positive rating (5-Star rating), and not for the other star ratings, including for the 1-star rating, where the effect of the avoidance of sharing of socially unsanctioned experiences is in effect [1,5,54,55].

When we look at the results for the reviews posted after the first three months (Table 8, "Remaining emotional" column, Fig. 8), we find that there is no longer a significant difference in the emotional content of search vs. experience products across Star-ratings. These results confirm Hypothesis 3 suggesting that products which are traditionally considered to be experience products, when made available online with detailed accompanying information such as online reviews, can be evaluated as search products. Note that this does not mean that the product has changed, but that most of its attributes can be adequately evaluated by the consumer prior to purchase, hence making it a search-dominant or search good.

4.2. Robustness check of results

To check the robustness of our results, we ran OLS[8] estimation for three different models. The dependent variables were the proportion of emotional words (Model 1), proportion of positive emotional words (Model 2) and negative emotional words (Model 3). The results of our OLS regression are shown in Table 9. We use the 3 Star-rating and Search Goods as our base dummy variables, as these represent the lowest expected emotional responses, based on our theoretical framework.

The results for Model 1 confirm our finding that more extreme reviews have greater emotional content than less extreme reviews, thereby lending further support to Hypothesis 1. The dummy variable for product type is not significant which confirms Hypothesis 3: that in the long run it is irrelevant if a product is a search or experience good from the viewpoint of emotional content.

Model 2 looks at just the positive emotional content within reviews, and we find that positive 4-Star and 5-Star reviews contain significantly more positive emotional content compared to a less emotional 3-Star review. We also confirm that a 1-Star review contains less positive emotional content than a 3-Star review. There is no difference in the positive emotional content of 3-Star vs. 2-Star reviews, confirming our earlier findings.

Model 3 looks exclusively at the negative emotional content within reviews, and our results confirm that 1-Star and 2-Star rating reviews contain significantly more negative emotional content than 3-Star reviews, with 1-Star having the most negative emotions. There is no difference between the midrange 3-Star review and the positive 4-Star and 5-Star reviews in terms of negative content. By putting the results of the 3 models together, we can conclude that the 3-Star rating is very balanced in terms of its emotional content. The results of OLS estimation for our models provide additional confirmation to our earlier analysis.

5. Discussion and conclusion

The recent growth of online communities and portals has provided ample channels for WOM generation and increased its influence in various markets. Online customer reviews have gained importance due to the richness of their content, their role in the diffusion of products, their impact on purchase quantity and timing, and their effects on consumer learning and decision making. Although many prior studies have examined the phenomenon, they mainly focused on the distributions of reviews at a macro-level and not on the emotional makeup of review content; moreover, their findings are mixed with regard to the actual distribution of WOM content within such reviews. Consequently, understanding the distributions of WOM content and the motivations behind such distributions remains a challenge.

Therefore, in our study we sought to examine the actual content of reviews focusing specifically on the emotional content in customer

[6] See Table 6, "Search Positive", "Search Negative", "Experience Positive", "Experience Negative" columns.

[7] Li and Hitt [38] looked at the evolution of the average customer rating of online WOM communications, and not the emotional content within WOM. We varied the time window of three months but found consistent results.

[8] Because the dependent variables are continuous variables with respect to the emotional words, OLS is appropriate as a tool of analysis here.

Table 7
ANOVA for star-rating pairs for total, positive, negative, search and experience products emotional content.

		Overall	Positive	Negative	Search	Experience	Search positive	Search negative	Experience positive	Experience negative
Star rating (I)	Star rating (J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)
1	2	2.43*	−.932	2.80**	3.73	1.47	−.49	7.39**	−.71	1.87
	3	5.06***	−1.68**	5.88***	5.07**	4.68**	−2.04	9.30*	−1.51	4.34***
	4	−3.14***	−10.70***	7.18***	−4.56*	−3.21	−11.52***	10.16***	−11.07***	6.47***
	5	−5.21***	−13.81***	7.39***	−6.43***	−5.01**	−14.39***	10.31***	−14.36***	6.68***
2	3	2.63***	−.751	3.08**	1.34	3.21	−1.54	1.91	−.79	2.46
	4	−5.57***	−9.77***	4.37***	−8.29***	−4.68**	−11.03***	2.77	−10.35***	4.59***
	5	−8.20***	−12.88***	4.58***	−10.16***	−6.48***	−13.89***	2.92	−13.64***	4.80***
3	4	−8.20***	−9.018***	1.29	−9.63***	−7.90***	−9.48***	.86	−9.55***	2.12
	5	−10.2***	−12.13***	1.50	−11.50***	−9.70***	−12.35***	1.01	−12.84***	2.33
4	5	−2.07	−3.11***	.210	−1.86	−1.79	−2.86	.14	−3.28***	.21
N		17	17	17	8	7	8	8	9	9

***, **, and * Mean difference is significant at the ≤.01, ≤.05, and <.1 levels, respectively.

reviews. In the current work we used NLP techniques to analyze 15,849 product reviews. Our results confirm our hypotheses, including the prediction that extreme reviews would be associated with higher numbers of emotional words, as evidenced by a bimodal distribution. This finding provides for the first time empirical evidence for the widely held assumption of such a distribution of emotional content. More interestingly, we found that the average length of reviews peaks at the 2-Star level, showing a reversed *J*-shaped distribution. In other words, as reviews become more emotional, and include a greater number of emotional words, their lengths become shorter. Therefore, this result suggests that more rational review information can be found around rating 2-Star or 3-Star ratings, which can be highly useful for product development managers and retailers, as well as consumers.

Second, we found that a greater proportion of positive emotional content exists at the positive end of the rating spectrum when compared with negative emotional content toward the negative end of the rating spectrum, confirming a positive skew in online product reviews. Additionally, we found that these distributions overlap between the 2-Star and 3-Star ratings and that the proportion of positive and negative emotions balances out somewhere between the 2-Star and 3-Star ratings. Together with the findings about the reversed J-shaped review length, this result emphasizes the importance of mid ratings. Third, we found evidence to support the notion that when new products are launched, there will be information asymmetry between search and experience goods, owing to the difficulty in properly evaluating the latter prior to purchase, which in turn would lead to more emotional reviews for experience goods. We then hypothesized that experience goods will become more like search goods in an online environment as the reviews pile up (again, by this we refer to changes in the ability to evaluate product attributes, not changes in the attributes themselves), which in turn will reduce information asymmetry, and that eventually there will be little difference in the emotional makeup of reviews for search and experience goods. Our results validated this hypothesis, as we find no difference between the proportions of emotional content for search vs. experience goods across all star-ratings in the later part of the data. This may suggest that in the long run, the information asymmetry gap is bridged, thereby, experience goods become more alike search goods in the online environments as the reviews pile up.

Our findings have important managerial implications. First, with regard to our findings on the longer reviews with less emotional words in mid ratings regions, managers may find more useful information which can be used for product development, product renovation, and advertisement emphasis. These less emotional words with longer length may deliver valuable information about the strength and weakness of the products consumer experienced. For example, in case of canon digital camera, the words “zoom”, “flash”, and “size” have been found more often in the mid-ratings and they actually are the important

Fig. 4. *J*-shaped distribution of the percentage of emotional words per review compared with the upper-shape distribution of the number of words per review.

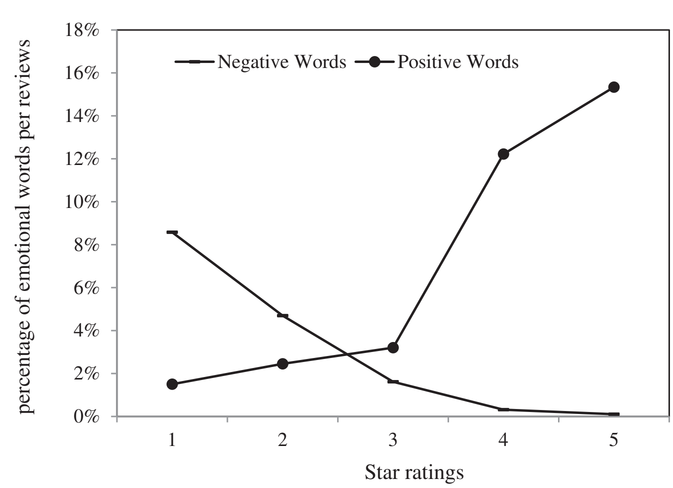

Fig. 5. Percentage of positive and negative emotional words per review.

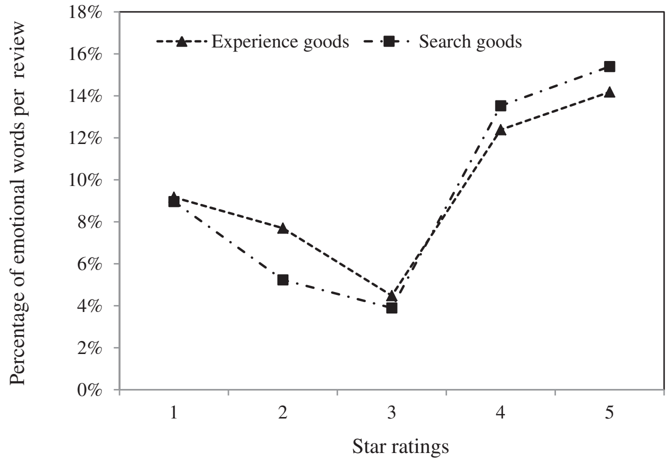

Fig. 6. Percentage of emotional words associated with customer reviews for search and experience products.

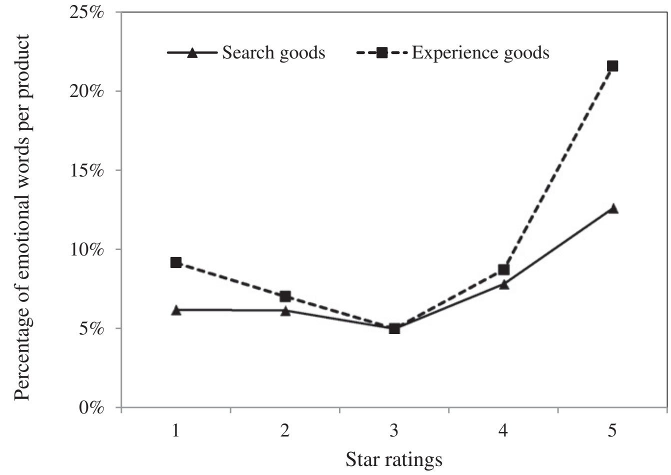

Fig. 7. Percentage of emotional words associated with customer reviews for search and experience products in first quarter of period.

features of the product. In addition, the manufacturer's claims on product features which need to be affirmed by the consumer can be also confirmed from a careful study of such mid ratings of online WOM. Moreover, the effectiveness of informative and persuasive role of product marketing can be also assessed from the analysis of online WOM. In other words, our approaches will be able to provide ideas for product development, make marketing decisions easier, and raise the confidence in the marketing planning processes.

Second, managers can also use extreme ratings, where more emotional information may be available, particularly in high-extreme ratings (5-Star), to get a better idea of the emotions associated with their products. Platform designers can develop ecommerce portals which organize and present the emotionally attractive reviews prominently. In addition, product features which occur in 5-star reviews can be emphasized in marketing because those features are the most emotionally stimulating information for consumers. Since we extracted these key emotional words using NLP extraction algorithms, the entire process can be fully automated, with the highest scoring positive emotional words being automatically displayed. This can significantly assist the consumers in their decision making processes, since previous studies in neuroscience and psychology have shown that emotions play a positive role in decision making [58].

Third, even the negative emotions conveyed by the users can be useful to managers. On the retailing side, managers can use the negative emotions to troubleshoot their inventory and service (e.g. if the word “horrible” is related to customer service), while on the production side, negative emotions can be used to improve their offerings (e.g. the words “ridiculous” is used to describe the storyline). While a certain amount of negative emotional content will always be present, using this information as a feedback mechanism which in turn plugs into a continuous quality improvement process can be of great benefit to firms situated at various locations in the value chain.

Fourth, from an advertising perspective, our study also can be of interest to the marketing manager. Using our findings and approaches, firms can generate more effective marketing campaigns by using the positive emotional words to help create advertisements and other promotional materials that emphasize these positive emotions that are associated with the product. The data on emotions in reviews can help retailers decide on which prominent and appropriate emotions (via keywords) to bid for in advertising auctions and highlight in their advertisements and in-store displays. For example, HP Desk Jet printer can be advertised with the emotional word “excellent,” while the video planet earth can be with the word “amazing” based on the customers WOM frequency (see Table 3). Since these emotions have previously been evoked within consumers by the products, they are likely to resonate better with potential customers, thereby increasing sales. In addition, it is more likely that expectations based on these promotions will sync with post consumption experiences (since these are the most commonly evoked emotions), leading to greater satisfaction, which in turn can help create greater brand loyalty.

Finally, reviews are usually a rich but unstructured set of consumer data with noise and unusable knowledge and information. To overcome these issues, NLP methods can be used as valuable tools for analyzing large numbers of customer reviews to extract important marketing insights. In addition to the techniques outlined in our study, there are also more advanced NLP techniques that can be used for more complex syntactic and semantic analyses, including word sense disambiguation and discourse analysis. By applying those techniques to user-generated-content including customer reviews, managers can understand customers' needs and wants more deeply. This new approach

Table 8
T-test for the mean differences in emotional words.

		All emotional	Positive emotional	Negative emotional	First quarter emotional	Remaining emotional
Search (I)	Experience (J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)	Mean difference (I − J)
1	1	−.202	−1.26**	3.65	−2.95	0.15
2	2	−2.46	−1.48***	−1.86	−.88	−1.73
3	3	−.58	−.736	−1.30	−.005	−1.75
4	4	1.14	−.812	−.04	−.91	−2.00
5	5	1.21	−1.23	.02	−8.96***	0.78

***and ** Mean difference is significant at the ≤.01 and ≤.05 levels, respectively.

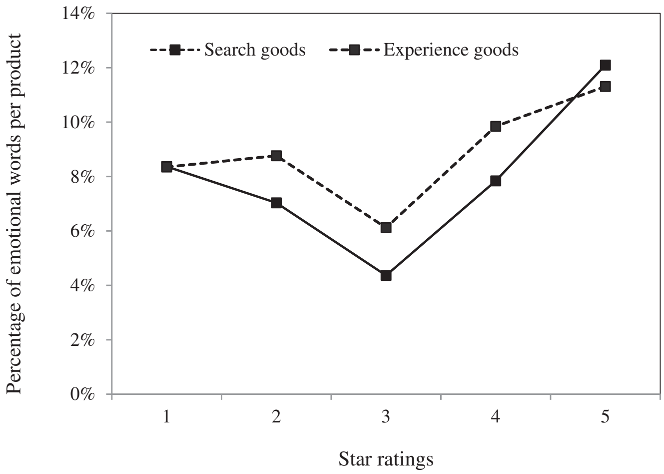

Fig. 8. Percentage of emotional words associated with customer reviews for search and experience products in remaining period except first quarter of time.

will eventually become important for managers, because firm's strategic use of this large amount of user-generated-content may affect firm's market performance as more and more this large amount of user-generated-content are available [69,70].

While our findings lend themselves to various possibilities for extension, they are also constrained by a few specific limitations. An important limitation to acknowledge is that while we understand the emotional makeup of product reviews and in the course of this research developed techniques to identify the top positive and negative emotional words, we are as yet unaware of the impact of these emotions on decision making. Second, future work could expand the focus of analysis from positive/negative review quality to more diverse categories such as the subjective/objective nature of reviews and product function-related features, etc., which would yield a wider range of insights regarding customer reviews. Third, research could extend our analysis to other user-generated-content, such as online communities and blogs, again yielding richer insights. By overcoming such limitations, we can more completely understand consumer behavior represented by user-generated content, including within the domain of WOM.

Acknowledgment

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012-S1A3A-2033860).

Table 9
OLS estimation for the distribution of emotional words.

Dependent Variable	Model 1	Model 2	Model 3
	Emotional words	Positive emotional words	Negative emotional words
	Coefficients	Coefficients	Coefficients
Constant	4.11***	2.72	0.962
Experience products	0.179	1.11*	0.433
Rating 1	4.87***	−1.76*	6.76***
Rating 2	2.33	−1.15	4.00***
Rating 4	8.72***	9.52***	−0.27***
Rating 5	10.55***	12.61***	−1.09***

*** and * represent significant at the ≤.01 and <.1 levels, respectively.

References

[1] E.W. Anderson, Customer satisfaction and word of mouth, Journal of Service Research 1 (1) (1998) 5–17.
[2] M.D. Alicke, J.C. Braun, J.E. Glor, M.L. Klotz, J. Magee, H. Sederhoim, R. Siegel, Complaining behavior in social interaction, Personality and Social Psychology Bulletin 19 (3) (1992) 286–295.
[3] M. Araújo, P. Gonçalves, M. Cha, F. Benevenuto, iFeel: a system that compares and combines sentiment analysis methods, Proceedings of the 23rd International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, Seoul, Korea 2014, pp. 75–78.
[4] I. Arnon, N. Snider, More than words: frequency effects for multi-word phrases, Journal of Memory and Language 62 (1) (2010) 67–82.
[5] J. Berger, K.L. Milkman, What makes online content viral? Journal of Marketing Research 49 (2) (2012) 192–205.
[6] A.Z. Broder, S.C. Glassman, M.S. Manasse, Z. Geoffrey, Syntactic clustering of the web, Computer Networks and ISDN Systems 29 (8) (1997) 1157–1166.
[7] F.A. Buttle, Word of mouth: understanding and managing referral marketing, Journal of Strategic Marketing 6 (3) (1998) 241–254.
[8] A. Ben-Ze'ev, The Subtlety of Emotions, MIT Press, Cambridge, MA,USA, 2000 12.
[9] C.D. Broad, Emotion and sentiment, The Journal of Aesthetics and Art Criticism 13 (2) (1954) 203–214.
[10] L. Cabral, A. Hortacsu, The dynamics of seller reputation: evidence from E Bay, Journal of Industrial Economics 58 (1) (2010) 54–78.
[11] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, Journal of Marketing Research 43 (3) (2006) 345–354.
[12] F. Coenen, P. Leng, R. Sanderson, Y.J. Wang, Statistical identification of key phrases for text classification, Machine Learning and Data Mining in Pattern Recognition Lecture Notes in Computer Science2007 838–853.
[13] M.R. Darby, E. Karni, Free competition and the optimal amount of fraud, The Journal of Law and Economics 16 (1) (1973) 67–88.
[14] W. Duan, B. Gu, A.B. Whinston, Informational cascades and software adoption on the internet: an empirical investigation, MIS Quarterly 33 (1) (2009) 23–48.
[15] W. Duan, B. Gu, A.B. Whinston, The dynamics of online word-of-mouth and product sales—an empirical investigation of the movie industry, Journal of Retailing 84 (2) (2008) 233–242.
[16] C. Dellarocas, R. Narayan, A statistical measure of a population's propensity to engage in post-purchase online word of mouth, Statistical Science 21 (2) (2006) 277–285.
[17] C. Dellarocas, R. Narayan, What motivates consumers to review a product online? A study of the product specific antecedents of online movie reviews, Proceedings of the Workshop on Information Systems and Economics, Evanston, USA, 2006.
[18] C. Dellarocas, X.M. Zhang, N.F. Awad, Exploring the value of online product reviews in forecasting sales: the case of motion pictures, Journal of Interactive Marketing 21 (4) (2007) 23–45.
[19] C. Derbaix, J. Vanhamme, Inducing word of mouth by eliciting surprise — a pilot investigation, Journal of Economic Psychology 24 (1) (2003) 99–116.
[20] C. Derbaix, M.T. Pharm, Affective reactions to consumption situations: a pilot investigation, Journal of Economic Psychology 12 (2) (1991) 325–355.
[21] C. Dellarocas, C.A. Wood, The sound of silence in online feedback: estimating trading risks in the presence of reporting bias, Management Science 54 (3) (2008) 460–476.
[22] C. Dellarocas, G.G. Gao, R. Narayan, Are consumers more likely to contribute online reviews for hit or niche products? Journal of Management Information Systems 27 (2) (2010) 127–157.
[23] A. Esuli, F. Sebastiani, Determining the semantic orientation of terms through gloss classification, Proceedings of the 14th ACM international conference on Information and knowledge management, ACM, Bremen, Germany 2005, pp. 617–624.
[24] Andrea Esuli, Fabrizio Sebastiani, Sentiwordnet: a publicly available lexical resource for opinion mining, Proceedings of Language Resources and Evaluation, 5th Conference on Language Resources and EvaluationGenova, IT, European Language Resources Association (ELRA), Paris, FR 2006, pp. 417–422 (ISBN 2-9517408-2-4).
[25] D. Godes, D. Mayzlin, Using online conversations to study word of mouth communications, Marketing Science 23 (4) (2004) 545–560.
[26] N. Hu, J. Zhang, P. A.Pavlou, Overcoming the J-shaped distribution of product reviews, Communications of the ACM 52 (10) (2009) 144–147.
[27] N. Hu, P.A. Pavlou, J. Zhang, Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word of mouth communication, Proceedings of the 7th ACM conference on Electronic commerce, New York, USA, 2006.
[28] P. Huang, N.H. Lurie, S. Mitra, Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods, Journal of Marketing Research 73 (1) (2009) 55–69.
[29] T. Hennig-Thurau, K.P. Gwinner, G. Walsh, D.D. Gremler, Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? Journal of Interactive Marketing 18 (1) (2004) 38–52.
[30] T. Hennig-Thurau, G. Walsh, Electronic word-of-mouth: motives for and consequences of reading customer articulations on the internet, International Journal of Electronic Commerce 8 (2) (2003) 51–74.
[31] L.K. Hansen, A. Arvidsson, F.A. Nielsen, E. Colleoni, M. Etter, Good friends, bad news — affect and virality in twitter, Future Information Technology,Communications in Computer and Information Science 185 (2011) 34–43.
[32] Y.-C. Hsieh, H.-C. Chiu, M.-Y. Chiang, Maintaining a committed online customer: a study across search-experience-credence products, Journal of Retailing 81 (1) (2005) 75–82.
[33] C.E. Izard, Human Emotions, Plenum Press, New York, NY, 1977.
[34] J.G. Lynch, D. Ariely, Wine online: search costs affect competition on price, quality, and distribution, Marketing Science 19 (1) (2000) 83–103.

[35] M. Jiang, S. Argamon, Finding political blogs and their political leanings, Proceedings of SIAM Text Mining Workshop 12, Atlanta, GA, USA, 2008.
[36] A. Kadet, Rah-rah ratings online, Smart Money Magazine 23 (2007) 116 (February).
[37] B.C. Krishnan, M.D. Hartline, Brand equity: is it more important in services? Journal of Services Marketing 15 (5) (2001) 328–342.
[38] X. Li, L.M. Hitt, Self-selection and information role of online product reviews, Information Systems Research 19 (4) (2008) 456–474.
[39] Y. Liebermann, A. Flint-Goor, Message strategy by product-class type: a matching model, International Journal of Research in Marketing 13 (3) (1996) 237–249.
[40] Y. Liu, Word of mouth for movies: its dynamics and impact on box office revenue, Journal of Marketing Research 70 (3) (2006) 74–89.
[41] S.M. Mudambi, D. Schuff, What makes a helpful online review? A study of customer reviews on Amazon. com, MIS Quarterly 34 (1) (2010) 185–200.
[42] F.M. Maute, L. Dube, Patterns of emotional responses and behavioral consequences of dissatisfaction, Applied Psychology 48 (3) (1999) 349–366.
[43] M. Munezero, C.S. Montero, E. Sutinen, J. Pajunen, Are they different? affect, feeling, emotion, sentiment, and opinion detection in text, IEEE Transactions on Affective Computing 5 (2) (2014).
[44] D. Mladenic, M. Grobelnik, Feature selection for unbalanced class distribution and naive Bayes, Proceedings of the Sixteenth International Conference on Machine Learning 1999, pp. 258–267.
[45] G.A. Miller, WordNet: a lexical database for English, Communications of the ACM 38 (11) (1995) 39–41.
[46] H. Mano, R.L. Oliver, Assessing the dimensionality and structure of consumption experience: evaluation, feeling, and satisfaction, Journal of Consumer Research 20 (3) (1993) 451–466.
[47] P. Nelson, Information and consumer behavior, The Journal of Political Economy 78 (2) (1970) 311–329.
[48] R. Plutchik, Emotion, a Psycho Evolutionary Synthesis, Academic, New York, 1980.
[49] L.M. Richins, Measuring emotions in the consumption experience, Journal of Consumer Research 24 (2) (1997) 127–146.
[50] L.M. Richins, Word of mouth communication negative information, Advances in Consumer Research 11 (1) (1984).
[51] P. Resnick, R. Zeckauser, Trust among strangers in internet transactions: empirical analysis of Ebay's reputation system, The Economics of the Internet and E-Commerce, Advances in Applied Microeconomics 11 (2002) 127–157.
[52] J.A. Russell, L.F. Barrett, Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant, Journal of Personality and Social Psychology 76 (5) (1999) 805–819.
[53] B. Rime, C. Finkenauer, O. Luminet, E. Zech, P. Philippot, Social sharing of emotion: new evidence and new questions, European Review of Social Psychology 9 (1998) 145–189.
[54] B. Rime, P. Philippot, S. Boca, B. Mesquita, Long-lasting cognitive and social consequences of emotion social sharing and rumination, European Review of Social Psychology 3 (1992) 225–258.
[55] B. Rime, C. Finkenauer, O. Luminet, E. Zech, P. Philippot, Social sharing of emotion: new evidence and new question, European Review of Social Psychology (2011) 145–189.
[56] A. Resnik, B.L. Stem, An analysis of information content in television advertising, The Journal of Marketing 41 (1) (1977) 50–53.
[57] S. Stieglitz, L. Dang-Xuan, Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior, Journal of Management Information Systems 29 (4) (2013) 217–248.
[58] B. Shiv, G. Loewenstein, A. Bechara, The dark side of emotion in decision-making: when individuals with decreased emotional reactions make more advantageous decisions, Cognitive Brain Research 23 (2005) 85–92.
[59] D.S. Sundaram, K. Mitra, C. Webster, Word-of-mouth communications: a motivational analysis, Advances in Consumer Research 25 (1) (1998) 527–531.
[60] J.G. Stigler, The economics of information, The Journal of Political Economy 69 (3) (1961) 213–225.
[61] F. Song, W.B. Croft, A general language model for information retrieval, Proceedings of the eighth international conference on Information and knowledge management, ACM, Kansas City, Missouri, USA 1999, pp. 316–321.
[62] D.K. Tse, R.W. Belk, N. Zhou, Becoming a consumer society: a longitudinal and cross cultural content analysis of print ads from Hong Kong, the People Republic of China, and Taiwan, Journal of Consumer Research 15 (4) (1989) 457–472.
[63] R. Ullah, A. Zeb, W. Kim, The impact of emotions on the helpfulness of movie reviews, Journal of Applied Research and Technology 13 (3) (2015) 359–363.
[64] R.A. Westbrook, R.L. Oliver, The dimensionality of consumption emotion patterns and consumer satisfaction, Journal of Consumer Research 18 (1) (1991) 84–91.
[65] J. Yang, W. Kim, N. Amblee, J. Jeong, The heterogeneous effect of WOM on product sales: why the effect of WOM valence is mixed? European Journal of Marketing 46 (11/12) (2012) 1523–1538.
[66] Y. Yang, J.O. Pedersen, A comparative study on feature selection in text categorization, Proceedings of the Fourteenth International Conference on Machine Learning 1997, pp. 412–420.
[67] B. Yang, C. Cardie, Context-aware learning for sentence-level sentiment analysis with posterior regularization, Proceedings of 52nd Annual Meeting of Association for Computational Linguistics (ACL), Association for Computational Linguistics, Baltimore Maryland 2014, pp. 325–335.
[68] J.M. Torres-Moreno, J.M. Torres-Moreno, Evaluating Document Summaries, In Automatic Text Summarization, John Wiley & Sons, Inc. 2014, pp. 243–273.
[69] C. Oh, Y. Cho, W. Kim, The effect of a firm's strategic innovation decisions on its market performance, Technology Analysis & Strategic Management 27 (1) (2015) 39–53.
[70] B. Brown, M. Chui, J. Manyika, Are you ready for the era of 'big data', 4McKinsey Quarterly, 2011 24–35.

Rahat Ullah received Ph.D. in Cultural Management from the Graduate School of Culture Technology at KAIST. His research interests include electronic commerce, economics of information systems, social media, and online word of mouth. His current research focuses on the issues in online WOM and online product reviews.

Naveen C. Amblee is Assistant Professor of Marketing Management at the Indian Institute of Management Kozhikode. He holds a Ph.D. in International Management from the University of Hawaii at Manoa. His research interests are in the field of Internet Marketing with an emphasis on online word-of-mouth, digital products and social media. His research articles have been published in the International Journal of Electronic Commerce and the European Journal of Marketing.

Wonjoon Kim is an Associate professor at the Graduate School of Innovation and Technology Management and the Department of Business and Technology Management at KAIST. His research focuses on strategic management of innovation in high-tech industry, economics of information systems, social media, online word of mouth. He received several best paper awards in innovation area. His research can be found at the following journals: Research Policy, Energy Policy, Journal of Applied Economics, etc. He holds a Ph.D. in Economics and M.S. in Material Science and Engineering from Seoul National University.

Hyunjong Lee is a researcher at Daumkakao which was formed from a merger of Daum Communications and Kakao to Daum Kakao in 2014. His research focuses on Statistical natural language processing, Computational linguistics, Opinion mining, and Machine learning. He holds a M.S. in Computer Science from KAIST.