Indian Institute Of Management Kozhikode

Working Paper

IIMK/WPS/156/MKTG/2014/14

DO EMOTIONS MATTER? EXPLORING THE DISTRIBUTION OF EMOTIONS IN ONLINE PRODUCT REVIEWS

Rahat Ullah
Wonjoon Kim
Naveen C. Amblee
Hyunjong Lee
Alice Oh

IIMK/WPS/156/MKTG/2014/14

DO EMOTIONS MATTER? EXPLORING THE DISTRIBUTION OF EMOTIONS IN ONLINE PRODUCT REVIEWS

Rahat Ullah[1]
Wonjoon Kim[1]
Naveen C. Amblee[2]
Hyunjong Lee[1]
Alice Oh[1]

[1] Korea Advanced Institute of Science and Technology (KAIST)

[2] Assistant Professor, Indian Institute of Management Kozhikode, IIMK Campus PO, Kozhikode– 673570, Email: amblee@iimk.ac.in

DO EMOTIONS MATTER? EXPLORING THE DISTRIBUTION OF EMOTIONS IN ONLINE PRODUCT REVIEWS

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 such as the causal effect on consumer choice, distribution pattern of WOM, and its product type-dependency, 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 needs more attention from researchers. In addition, an important antecedent for the generation of word-of-mouth is a strong emotional response, which in turn triggers the consumer to post a customer review online. However, only a limited number of studies to date have actually examined the content of reviews for their emotional content. To fill this gap, we analyzed the emotional content of a large number of online product reviews using Natural Language Processing (NLP) methods. 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. In addition, we found 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. Investigating further, we did find that there is a difference in the emotional content of reviews between search and experience goods in the early stages of product launch. However, interestingly, we find that these differences disappear over time as the addition of reviews reduces the information asymmetry gap. This provides important evidence to the widely held notion that on the Internet, all goods become search goods. Our findings suggest important managerial implications regarding product development, advertisement, and platform design using WOM content.

I. INTRODUCTION

Word-of-mouth (WOM) has been described as one of the most important means of informal communication among consumers (Sundaram *et al.,* 1998; Dellarocas and Narayan, 2006a; Derbaix and Vanhamme, 2003; Hennig-Thurau *et al.,* 2004). Online customer reviews for products or services can be formally defined as peer-originated product evaluations placed on company or third-party web sites (Mudambi and Schuff, 2010). 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 together 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 (Chevalier and Mayzlin, 2006; Dellarocas and Narayan, 2006a).

Therefore, previous studies of WOM have examined the importance of online WOM in general, and user-generated product reviews in particular. The hypothesis that product reviews have an influence on product sales has received considerable validation in previous empirical studies (Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Liu 2006, Dellarocas et al. 2007, Duan et al. 2008). Two of the most important metrics in the study of WOM are volume and valence. Volume refers to the quantity of WOM communications, while valence refers to the nature of the communications: positive, negative, neutral, or mixed (Liu 2006), and considerable research has focused on the impact of these two metrics on product sales. For example, the volume of product reviews have been shown to have a direct correlation with product sales (Godes and Mayzlin, 2004; Duan *et al.,* 2005; Chevalier and Mayzlin, 2006; Liu, 2006), while research on the valence of the reviews has been less conclusive, with some studies finding a positive relationship between valence and sales (Chevalier and Mayzlin, 2006), and others finding no significant relationship (Liu 2006). The finding that the valence of reviews does not have a consistent and positive relationship with sales has surprised researchers.

Therefore, several studies have attempted to explain this unexpected result, and one of the possible explanations include the popularity factor, whereby products with a large number of reviews (and therefore large prior sales) continue to gain sales, through a cascading effect (Duan et al. 2009). The other two factors are the self-selection bias and the under reporting bias. The self-selection bias occurs when those consumers with a sufficiently high valuation of the product end up purchasing it, and subsequently are more likely to post a very positive review (Hu et al. 2009). The under reporting bias occurs when only those consumers who are extremely satisfied or extremely dissatisfied write reviews, as those consumers who felt that the product was “just okay” may not be motivated to post a review (Hu et al. 2009). These factors taken together can give consumers a highly skewed view of the utility of the product, and while possibly being adjusted for into the decision making, cannot be reliably used in consumer behavior models that incorporate WOM as a factor.

In particular, the chief limitation of previous research is the assumption that simply observing the extent to which consumers share socially, i.e., considering only the frequency or overall rating/valence of WOM, and not its contents, is sufficient for understanding WOM behavior (Rime *et al.,* 1998). Only a few studies recently started to analyze the textual information embedded in online user-generated content in any significant detail. Ghose and Ipeirotis (2010) and Ghose et al.(2012) used multiple aspects of review text, such as lexical, grammatical, semantic, and stylistic levels to identify important text-based features and studied their impact on review helpfulness and product sales, while Rui,Liu and Whinston (2013) used machine learning algorithms and found that that positive Twitter WOM have higher impact on movie sales, whereas negative WOM is associated with lower movie sales.

However, we are unaware of any previous research that looks more specifically at the emotional makeup of review content. Emotions are the driving force behind online review articulation, since any highly satisfactory or highly dissatisfactory experience elicits strong emotional responses. These emotions create an emotional imbalance, which in turn requires the restoration of balance through the expression of positive emotions or through the venting of negative feelings (Hennig-Thurau et al. 2004). This act of catharsis (Alicke et al. 1992) often takes the form of online product reviews. It is known that shared content such as online WOM plays a crucial role in emotional recovery, relief, and other aspects of social interaction (Maute and Dube, 1999; Derbaix and Vanhamme, 2003; Hennig-Thurau et al., 2004). As such it is important to understand the "emotional structure" of product reviews to better understand the idiosyncrasies of online WOM communications. In other words, the heterogeneity of content among online customer reviews across their distribution has yet to be characterized, as previous studies have limited their scope of inquiry to WOM frequency or related factors rather than emotional content-related characteristics.

Therefore, in this study we have attempted to expand the literature on online WOM by analyzing the emotional content of online product reviews on Amazon.com using Natural Language Processing (NLP) techniques. By using an NLP strategy, we can computationally analyze and understand the natural human linguistic aspects of WOM. More specifically, using NLP, we are able to examine the proportion of emotional contents in online product reviews as well as the role of emotional content in user experiences to identify heterogeneous characteristics of online WOM contents. We are also able to observe changes to the emotional content of reviews over time for search and experience goods.

In the first stage of our study, we examine the distribution of the emotional content of online product reviews across product ratings related with other review characteristics, such as average number of reviews, average length of reviews, etc. This is followed up by an examination of the emotional valence of content across product ratings regardless of whether the distribution is symmetric, because the relationship between emotional valence and WOM distribution remains controversial. Finally, taking into consideration the categories of search and experience goods, we test whether uncertainty related to product quality before consumption is associated with more WOM activity considering the difference product category, i.e. search vs. experience goods. We also take a longitudinal view and make some interesting observations on the evolution of emotional content in online reviews.

Considering the fact that much attention is needed currently to analyze the content of WOM in order to extract and usethis valuable information, our study lays an important ground for other researchers by developing a novel approach to the issue. Our findings not only enable us to understand the fundamental nature of customer reviews including the motivation of WOM generation, but also suggest several important implications for managers of electronic commerce, especially regarding product development, advertisement, and platform design based on the information held within WOM. To this end, in Section 2, we discuss the theoretical background regarding the role of emotions in consumer's 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 technique. Section 4 provides our results and discussion regarding the three hypotheses we suggest in section 2. Finally, in section 5we summarize the main points of this paper and discuss important managerial implications of our findings.

II. LITERATURE REVIEW

Several factors motivate consumers to post online customer reviews. The most widely accepted motivation in the literature is that people share their emotional experiences with others, and that consumers tend to report more frequently at the extreme levels of satisfaction (i.e., highly dissatisfied or highly satisfied; see Anderson, 1998). In addition, Consumers engage in positive word of mouth behavior in order to express their delight with the experience, increase involvement with the product, and the help the company (Hennig-Thurau et al., 2004). On the other hand, by sharing a negative consumption experience through a customer review, consumers can reduce the discontent associated with their negative emotions (Hennig-Thurau *et al.,* 2004), such as anxiety and tension (Sundaram *et al.,* 1998). Therefore, the motivation for engaging in negative WOM is often referred to as dissonance reduction (Buttle, 1998; Dellarocas and Narayan, 2006a, 2006b).[1] Consumers also share useful content for self-enhancement purposes (e.g., to appear knowledgeable; see Wojnicki and Godes, 2008) or for altruistic reasons (e.g., to help others; Sundaram *et al.,* 1998).[2] In addition, the antecedents of offline and online WOM are similar, and expressing emotion is an especially common and important motivation for consumer WOM behavior, whether online or offline (Dellarocas and Narayan, 2006a).

In this respect, previous research has investigated the relationship between consumption-related emotions and consumer behavior (Mano and Oliver, 1993; Westbrook and Oliver, 1991; Richins, 1997). It has been shown that emotions play an important role in consumer response, thus firmly establishing the significance of emotion to consumer behavior (Richins, 1997). Emotion yields important psychological consequences and generates long-lasting mnemonic recurrences as well as acting as motivation for social sharing (Rime *et al.,* 1992). For example, the emotion of surprise has been shown to play a major role in the elicitation of consumer WOM. Positive surprise can generate positive WOM, while negative surprise can lead to negative WOM (Derbaix and Vanhamme, 2003).

---

[1] It has also been suggested that consumers are motivated to engage in negative WOM for altruistic, vengeance-related, and advice-seeking-related reasons (Sundaram, Mitra, and Webster, 1998).

[2] In this case, previous studies have suggested that consumers are motivated for altruistic, product-involvement-related and self-enhancement-related objectives (Sundaram, Mitra, and Webster, 1998). In the more specific domain of online WOM, in addition to this emotional motivation, a desire for social interaction and possible economic incentives have been suggested as additional motivations for consumer WOM behavior (Hennig-Thurau *et al.,* 2004, Balasubramanian and Mahajan, 2001).

Here, a general requirement of emotions is to have a valence (i.e., positive or negative). This emotional valence is caused by schema discrepancy, or when expectations about a product or service are not realized (Derbaix and Vanhamme, 2003). 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 (Anderson, 1998; Hennig-Thurau *et al.,* 2004). Therefore, previous studies consistently emphasized the important role of emotions in social sharing, especially with respect to consumer WOM behavior. For example, Berger and Milkman (2012) recently examined how emotional content affects the sharing of online content using psychological field experiments that showed that positive content is more viral than negative content. Similarly (Stieglitz and Dang Xuan, 2013) found that emotional twitter messages tend to be retweeted more quickly and more often compared to neutral ones. However, clear evidence regarding the effect of emotion on the distribution of online WOM has not been presented but has so far only been assumed to be true. In this paper we attempt to reveal the emotional structure of online WOM.

Bimodal Distribution of WOM and Emotions

In addition, online WOM studies have shown that online product reviews are overwhelmingly positive, with mostly 5-Star ratings, some 1-Star ratings, and very few ratings in between (Chevalier and Mayzlin, 2006; Kadet, 2007). These findings are contrary to the case of offline reviews, which has shown a distribution skewed toward negative WOM (Anderson, 1998; Chatterjee, 2001), in which extremely dissatisfied customers are responsible for greater WOM than are high-satisfaction customers. Therefore, these contradictory findings of previous studies and the overall limited understanding of online WOM distribution patterns require us to examine this issue differently.

In fact, one of the most widely reported trends in consumer WOM behavior in online environments is its bimodal distribution, which is observed when online customer reviews are ordered by valence (Westbrook and Oliver, 1991; Rime *et al.,* 1992; Dellarocas and Narayan, 2006a; Hu *et al.,* 2009). Researchers have identified a *J*-shaped distribution showing more positive reviews than negative ones, which can be found in most online reviews and other feedback sources (Resnick and Zeckauser, 2002; Cabral and Hostacsu, 2010; Hu *et al.,* 2009). For example, a study of the relationship between book sales and online customer reviews for movies (Chevalier and Mayzlin, 2006) showed a strong *J*-shaped distribution, prompting the authors to concur with previous findings (Anderson, 1998; Dellarocas and Narayan, 2006a).

The J-shaped distribution of customer reviews has been attributed to two types of biases. The first is referred to as the self-selection bias, where consumers with a preexisting favorable disposition tend to purchase, consume and positively review a product (Hu, Pavlou and Zhang 2009). The second bias is referred to as the underreporting bias, which is the truncated posting of reviews by only those consumers who felt a strong emotional response, either positive or negative (Anderson 1998; Hu, Pavlou and Zhang 2009). In contrast with these results, other studies have suggested that dissatisfaction heightens the incidence and magnitude of emotional responses (Buttle, 1998; Westbrook and Oliver, 1991; Derbaix and Pharm, 1991; Maute and Dube, 1999).[3] It has been reported that the influence of emotions on consumer post-purchase responses may be more pronounced for dissatisfactory experiences than for normal consumption experiences, especially in the case of private offline behavior. For example, Anderson (1998) reported that dissatisfied customers tell twice as many people about their experience as satisfied customers. Similarly, Desatnick (1992) asserted that 90% or more of those who are dissatisfied with a service will share their experience with at least 9 other people, and 13% of those unhappy customers will tell their stories to more than 20 people.

However, whether online or offline, previous studies did not consider WOM *emotional content* within the complex consumption environment even though the content shared through WOM seems to play a crucial role in the emotional recovery or relief of consumers, social interaction related to the reviewed products, and concern for others, among other WOM-related reasons (Rime *et al.,* 1998). Therefore, here we attempt to analyze WOM content and its characteristics to understand the phenomena underlying consumer WOM behavior. Specifically, we investigate the role of emotions in consumer WOM behavior under the associated systematic skewedness.

Previous research has predicted that stronger emotional responses to a product lead to the posting of customer reviews (Maute and Dube, 1999; Derbaix and Vanhamme, 2003; Hennig-Thurau *et al.,* 2004). In other words, the more emotionally intense an event is, the more likely it is to be shared, because emotions elicit social sharing. Therefore, we expect 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.

On the other hand, it has previously been suggested that dissatisfaction with a product heightens the magnitude of emotional responses compared to satisfaction with a product (Desatnick, 1992; Derbaix and

---

[3] Buyer responses to dissatisfaction have been classified into a typology of exit, voice (word-of-mouth), and loyalty by Hirschman (1970); the categorization has been used to explain responses to dissatisfaction in a variety of contexts (Allen and Keaveney, 1985; Rosse and HGulin, 1985; Rusbult and Farrell, 1983).

Pharm, 1991; Buttle, 1998; Maute and Dube, 1999; Berger and Milkman, 2012), while others have suggested that negative and positive WOM distributions are balanced (Rime et al., 1992; Anderson, 1998). However, several of these studies did not account for the self-selection bias whereby consumers with a preexisting favorable disposition tend to purchase, consume, and subsequently positively review a product. These issues may be attributed to shortcomings in survey methodologies and to the inherent limitations of WOM frequency information (Rime *et al.,* 1992; Hennig-Thurau *et al.,* 2004; Hu *et al.,* 2006; Dellarocas and Wood, 2008).

Nonetheless, we still need to take into account the previously described notion that socially unsanctioned emotional experiences (e.g., negative experiences) are less likely to be shared, or they are shared with a delay (e.g., require a rewrite before being posted) (Buttle, 1998; Westbrook and Oliver, 1991; Derbaix and Pharm, 1991; Maute and Dube, 1999; Anderson, 1998). Moreover it has been reported that emotional twitter messages tend to be retweeted more quickly and more often as compared to the neutral ones (Stieglitz and Dang Xuan, 2013). Other studies, such as those of (Hansen et al 2011) have reported that positive sentiment further boost virality as compared to the negative sentiments except in the case of the news domain. Therefore, we expect a greater proportion of positive emotional content in positive extreme ratings as compared to the proportion of negative emotional content within negative extreme ratings.

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.

Information Search and the Impact on Search versus Experience Products

We also examine the distribution of emotional content in WOM based on product categories. Generally, all products can be categorized as either search or experience (Darby and Karni, 1973; Krishnan and Hartline, 2001; Nelson, 1970). Here, products with *search* attributes are those that can be evaluated by a consumer prior to purchase and consumption, whereas products with *experience* attributes are those that can only be evaluated post-consumption[5] (Krishnan and Hartline, 2001). 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 (Krishnan and Hartline, 2001). 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 (Rime et al. 1998; Derbaix and Vanhamme 2003). 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.

If consumers do not have enough preexisting information regarding the product being evaluated, they will engage in information search in order to become informed (Mudambi and Schuff 2010). According to the theory of information search put forward by Stiger (1961), 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 mouse (Lynch and Ariely 2000). 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 (Hennig-Thurau et al. 2003).

In fact, studies on information search have also shown that the online environment can change experience products into search products (Klein 1998). This is because the additional product information available in online environments including online reviews enables consumers to make informed purchasing decisions reducing risk and maximizing utility. Concurring with earlier research, Huang *et al.* (2009) discussed the possibility that all products become search products on the Internet because consumers can use the experience of previous consumers to make decisions. They also found that consumers spend similar amounts of time online gathering information for both search and experience products. Essentially this means that consumers are now 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, all products available online with detailed accompanying information such as online reviews are now search goods, even though they may be classified as experience goods in a traditional offline retail environment.

Therefore, considering this online effect on product type evaluation, the difference in the distribution of emotional content between search and experience goods can be small or insignificant if the consumers were able to adequately evaluate the experience goods. 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. 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 take on attributes of 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.

Thus, we state the following hypotheses:

> *Hypothesis 3a*: Early online reviews of experience goods have greater proportions of emotional content when compared with the early reviews of search goods.
>
> *Hypothesis 3b*: Later online reviews of experience and search goods will have similar levels of emotional content.

III. EMPIRICAL STUDY

We conducted an empirical study using online customer reviews from Amazon.com[4]. 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 26 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.

Pretest

To classify the products as search and experience products, we use the method outlined by Krishnan and Hartline (2001) and Hsieh *et al.* (2005). 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 26 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 their performance with each product *before* purchase the products 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

---

[4] http://www.amazon.com

and Hartline (2001) and Huang *et al.* (2009), 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, because this rating pattern implies that these products are perceived as being easy to evaluate prior to purchase[5]. 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 because they are perceived as being more difficult to evaluate prior to purchase but easy to evaluate post-consumption[6].

Based on these criteria, products rated greater than the mean value of all products on both scales were categorized as search products (S), 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 (E). In addition, when the change of the mean value of a product from the first to the second scale was greater than the average change of all products, the product was also considered an experience product (E). We also manipulated the classification with stronger criteria, such as the maximum deviation value of all 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.[7]

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

	Before		After		
Product	Mean	S.D.	Mean	S.D.	Classification
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

[5] Products with low scores on both scales can be classified as "credence" goods, or goods perceived as being difficult to evaluate even after consumption. We excluded such credence goods from our WOM analysis due to their limited implications with respect to our research.

[6] In this case, the mean of the second scale needs to be within the standard deviation of the average score of all products.

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

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	

Measures

Rational/Emotional Index: Previous research, including studies from the psychology literature (Izard, 1977; Resnik and Stem, 1977; Plutchik, 1980; Tse *et al.,* 1989; Liebermann *et al.,* 1996), examined the characterization of emotions and suggested the categories of "rational" and "emotional." Previous studies have also suggested that the importance of such categories in directing analysts to follow a systematic method of evaluation, thus improving correspondence among the categorizations. For example, Richins (1997) analyzed the previous literature in depth to identify a set of consumption emotion descriptors, and Derbaix and Vanhamme (2003) suggested seven primary emotions and descriptors that could be used to represent and measure them, as presented in Table 2.

Table 2. Primary emotions identified in the previous literature.

Izard 1977	Plutchik 1980	Richins 1997	Derbaix and Vanhamme 2003
Fear	Fear	Fear, Worry	Fear (Surprised, Amazed, Astonished)
Anger	Anger	Anger	Anger (Angry, Mad, Enraged)
Enjoyment	Joy	Joy	Enjoyment (Joyful, Delighted, Happy)

Distress/Sadness	Sadness	Sadness, Discontent, Loneliness	Distress (Sad, Downhearted, Discourage)
Disgust	Acceptance	Peacefulness	Disgust (Disgusted, Feeling of Distaste or Revulsion)
Interest	Disgust	Excitement, Optimism	Surprise (Surprised, Amazed, Astonished)
Surprise	Expectancy	Surprise	Contempt (Disdainful, Contemptuous, Scornful)
Shame/Shyness	Surprise	Shame	
Guilt		Guilt	
Contempt		Romantic Love	
		Love	
		Contentment	
		Envy	

Using the information from Table 2, we categorize WOM reviews into rational and emotional content. Because 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 sentiment scoring for each product and each number-of-stars level. In addition, adjacent words were also considered in the analysis to correctly identify the words used in the classification. We show the list of emotional words in our data in Table 3. We then aggregated the frequencies of the emotional words used in each of the reviews according to their star ratings.

Table 3. 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	

Method

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 key words that carry the semantics of sentiment. Here we explain several of the important NLP concepts we used to identify and categorize sentiment-expressing words.

Review Representation and Opinion Mining

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. This approach to analyzing word frequencies is called the "bag-of-words" model in NLP. We used SentiWordNet to extract emotional strength of words in terms of their positivity and negativity. SentiWordNet uses WordNet synset with ternary classifiers (A. Esuli and F. Sebastiani, 2006). Within SeniWordNet, we extracted minimum, maximum, and average scores of each word in terms of positives and negatives. For example, the word "beautiful" is appeared in two times in SentiWordNet because of its usage cases. Its minimum positive value is 0.625, maximum is 0.75, and average is 0.6875; while its minimum, maximum, and average scores of negatives are zero.

Feature Selection. Because 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 sentiment, 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.

Among various feature selection algorithms, we used one based on the chi-square test that has been shown to be an effective feature selection method (Jiang and Argamon, 2008). This 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 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:

$$\chi^2(D,t,C) = \frac{\left(N_{t,c} - E_{t,c}\right)^2}{E_{t,c}} \tag{1}$$

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 of term t in class c, and E is the expected frequency of term t in class c. Based on equation (1), we identified and extracted from the Amazon.com product review data the words that were the best candidates for representing sentiment, and then processed them using the following five steps:

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) [8] 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.

[8] An Application Programming Interface (API) is an interface implemented by a software program that enables it to interact with other software.

4. *Sentiment Scoring*: For each word in the top-ranked word lists, we used SentiWordNet[9] 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-word "memory" has multiple meanings – ability to remember this, or component of electric devices. The most frequent word to the next 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 top frequent emotional words for two search and two experience products as examples in table4.

Table 4. 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	

[9] SentiWordNet is a lexical resource for opinion mining based on quantitative analyses of the glosses associated to synsets and vectorial term representations for semi-supervised synset classification. The scores are produced by a committee of eight ternary classifiers.

IV. RESULTS AND DISCUSSION

We first examined the distribution of the volume of product reviews based on the number of stars awarded. As shown in Figure 1, 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 (1998) and Dellarocas and Narayan (2006a), and confirms that consumers are more likely to engage in interpersonal communication 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. (2009). 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.

Figure 1. J-shaped distribution of total number reviews.

Distribution of Emotional Content

The summary statistics of reviews containing emotional content for 17 products in two product categories are shown in Table 5. The distribution of non-normalized or absolute emotional content is similar to that of the total number of reviews. After normalization (adjusting for review length), 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.

Table 5. 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

To validate H1 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 6, "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-Stars pair, the 2-3-Stars pair, and the 3-4-Stars pair 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 (H_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 in the next section.

We gain additional insight by examining the length of the reviews associated with the overlapping of the distributions of emotional words, as shown in Figure 2. 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 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.

Table 6. 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

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

Figure 2. J-shaped distribution of the percentage of emotional words per review compared with the upper-shape distribution of the number of words per review.

Distribution of Positive and Negative Emotional Content

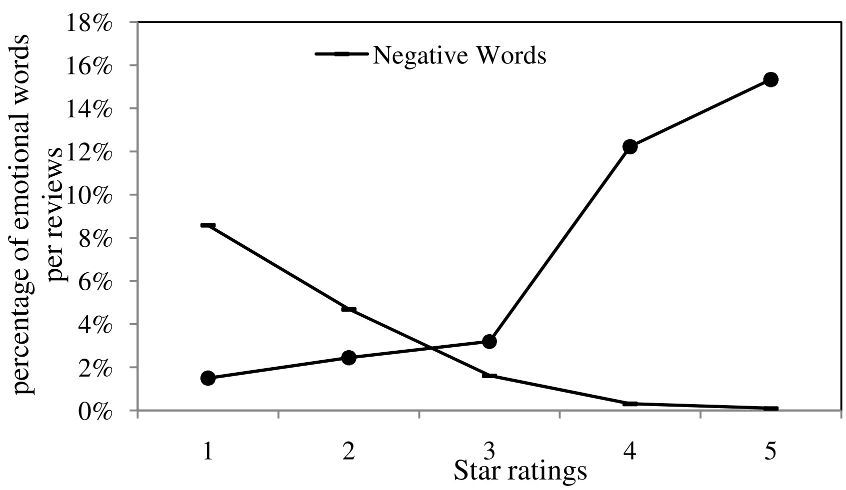
Our second hypothesis stated that there is a greater proportion of positive emotional content within positive extreme ratings as compared to the proportion of negative emotional content within negative extreme ratings. To validate this hypothesis, we perform several ANOVA tests on our dataset. As mentioned earlier the results (Table 6, “Overall” column) which validated H1 also help validate H2, as they show that the positive reviews (5-Star and less extreme 4-Star) both have on average a significantly greater proportion of emotional content than the negative reviews (1-Star and less extreme 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 content more specifically, we analyzed our results further by segmenting the emotional content into positive and negative content. 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 1-Star reviews ($\mu_{4starpositive}$=12.92%, $\mu_{2tarnegative}$=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 Figure 3). 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 6, "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 H2 and also help reconfirm H1.

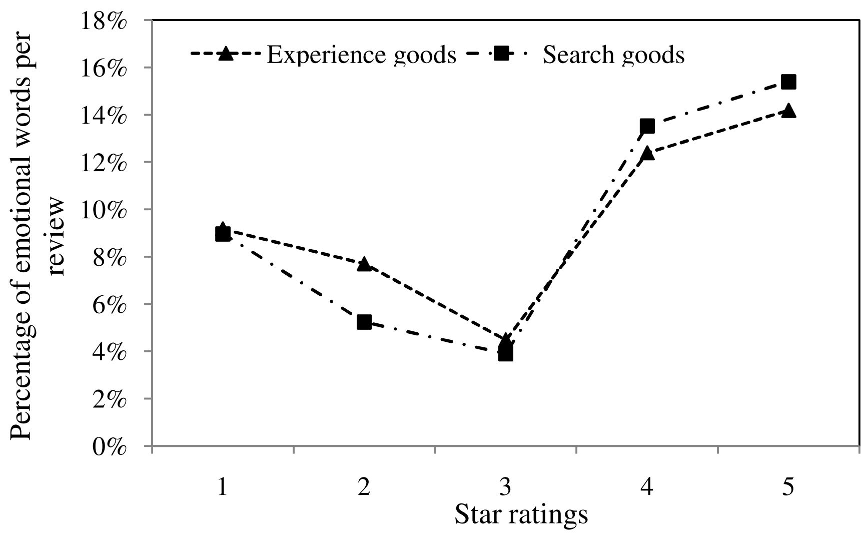
Figure 3. Percentage of positive and negative emotional words per review.



Emotional Content Distribution between search and experience Products

For further analysis of the adjusted numbers of 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 6 ("Search" and "Experience" columns). Figure 4 shows the distribution of emotional content across Star-ratings for Search and Experience goods.

Figure 4. Percentage of emotional words associated with customer reviews for search and experience products.



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 H1 for both product types. We separated the positive and negative emotional content, and found that the results become more pronounced, similar to the results for "Positive" and "Negative"[10].

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 7, "All Emotional Words" column) suggest that there are no significant differences in the emotional content between the search and experience products. Once again we segment the emotional content into positive and negative, and now find some interesting results (Table 7, "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.

---

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

Table 7. 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

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

Evolution of Emotional content distribution in Search and Experience products over time

In order to understand the *evolution* of emotional WOM content over time, we first 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 data collection began. We used three months as the cutoff point as this criterion has been used previously by Li and Hitt (2008) while studying the longitudinal effects of WOM[11]. 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 7, "First Quarter Emotional" column). 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. This lends support to H3a.

Interestingly however, this difference is only found for the 5-Star rating. Since this discrepancy does not show up for the other Star-rating levels, we state that hypothesis 3a is only partially confirmed by this finding. One possible explanation is that the early 5-Star reviews are written by enthusiasts and tend to be far more emotional than later 5-Star reviews. Another explanation is that the information asymmetry still present at this early stage creates greater surprise and subsequent highly positive emotional reaction from the consumers in the form of excessively emotional 5-Star reviews. When we look at the results for the

[11] Li and Hitt (2008) looked at the evolution of the average customer rating of online WOM communications, and not the emotional content within WOM.

reviews posted after the first three months (Table 7, "Remaining Emotional" column), we find that there is no longer a significant difference in the emotional content of search vs. experience products across Star-ratings. This confirms Hypothesis 3b.

Figure 5. Percentage of emotional words associated with customer reviews for search and experience products in first quarter of period.

Figure 6. Percentage of emotional words associated with customer reviews for search and experience products in remaining period except first quarter of time.

Robustness Check of Results

To check the robustness of our results, we ran OLS[12] 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 8. 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.

Table 8. OLS estimation for the distribution of emotional words

	Model 1	Model 2	Model 3
Dependent Variable	**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

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

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***

***, **,* represent significant at the ≤.01, ≤.05, <.1 level, respectively.

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 H1. The standardized coefficient is the highest for the 5-Star rating, confirming that the positive extreme reviews benefits from the self-selection bias of consumers. The dummy variable for product type is not significant which confirms H3b 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.

V. 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 regards 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 role of emotions in customer reviews. In the current work we used NLP techniques, including the feature-selection process, 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 rating reviews between the 2-Star and 3-Star 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. However we only found this for very positive 5-Star reviews, thereby only partially validating our hypothesis. We did not find any difference between the other Star-rated reviews (1-Star, 2-Star, 3-Star, 4-Star), implying that schema discrepancy is not larger in the case of negative experiences.

We then hypothesized that all goods will become experience goods on the Internet as the reviews pile up, 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 for H3b validated this hypothesis broadly, 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. We conclude that in the long run, the information asymmetry gap is bridged, thereby turning experience goods into search goods.

Our findings have important managerial implications. First, with regards 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 cannon digital camera, the words “zoom”, “flash”, and “size” have been found more often in the mid-ratings and they actually are the important 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 plays a positive role in decision making (Baba Shiv et al, 2005).

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 eminent and appropriate emotions to bid in advertising auctions and highlight in their advertisements and in-store displays. For example, HP DeskJet 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.

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 know 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 do not know the impact of these emotions on decision making. That is, while we know that there is more positive emotional content in positive reviews as compared to negative content in negative reviews, we do not know if these emotions have a proportionate or disproportionate impact on decision making. Future research should examine and model the impact of positive vs. negative emotional words in reviews 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. Fourth, studies could consider a two-stage review model to investigate the underlying causes of positive and negative reviews, allowing researchers to split the current bimodal distribution of reviews into two unimodal distributions, which could then be used to inform management and marketing decisions. By overcoming such limitations, we can more completely understand consumer behavior represented by user-generated content, including within the domain of WOM.

REFERENCES

[1] A. Kadet, Rah-Rah Ratings Online, Smart Money Magazine,23 (February), 116, (2007).
[2] A. Esuli, F. Sebastiani, SentiWordNet: A publicly available lexical resource for opinion mining," in Proceedings of Language Resources and Evaluation (2006).
[3] A. Ghose, P.G. Ipeirotis, Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer, IEEE Trans. Knowledge Data Engrg., IEEE Computer Society, Washington, DC, (2010).
[4] A. Ghose, P.G. Ipeirotis, B. Li, Designing ranking systems for hotels on travel search engines by mining user-generated and crowd-sourced content, Marketing Science, 31(3) (2012) 453-520.
[5] A.C. Wojnicki, D. Godes, Word-of-Mouth as Self-Enhancement, HBS Marketing Research Paper No. 06-01. Available at SSRN, (2008).
[6] A. Resnik, B. L.Stem, An Analysis of Information Content in Television Advertising, The journal of Marketing,41(1) (1977) 50-53.
[7] A.O. Hirschman, Exit, Voice and Loyality: Responses to Decline in Firms Organizations and States, Harvard University Press, Cambridge, M.A, (1970).
[8] B.C. Krishnan, M.D. Hartline, Brand Equity: Is it More Important in Services? Journal of Services Marketing, 15(5) (2001) 328-342.
[9] 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.
[10] 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.
[11] 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 Researc, 23(2005) 85-92.
[12] C.E. Izard, Human Emotion, New York, NY: Plenum Press, (1977).
[13] C.E. Rusbult, D. Farrell, A Longitudinal Test of the Investment Model: The Impact on Job Satisfaction, Job Commitment, and Turnover of Variations in Rewards, Costs, Alternatives, and Investments, Journal of Applied Psychology, 68(3) (1983) 429-438.
[14] C. Dellarocas, The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms, Management Science, 49(10) (2003) 1407-1424.
[15] C. Dellarocas, N.F. Awad, X.M. Zhang, Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning, Proceedings of the 24th International Conference on Information Systems, Washington, D.C, (2004).
[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) (2006a) 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, (2006b).
[18] 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.
[19] 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.
[20] C. Derbaix, J. Vanhamme, Inducing Word of Mouth by Eliciting Surprise - A Pilot Investigation, Journal of Economic Psychology, 24(1) (2003) 99-116.
[21] C. Derbaix, M. T.Pharm, Affective Reactions to Consumption Situations: A Pilot Investigation, Journal of Economic Psychology, 12(2) (1991) 325-355.
[22] D. Godes, D. Mayzlin, Using Online Conversations to Study Word-of-Mouth Communication, Marketing Science, 23(4) (2004) 545-560.

[23] D. Godes, D. Mayzlin, Y. Chen, S. Das, C. Dellarocas, B. Pfeiffer, B. Libai, S. Sen, M. Shi, P. Verlegh, The Firm's Management of Social Interactions, Marketing Letters, 16(3-4) (2005) 415-428.
[24] D.S. Sundaram, K. Mitra, C. Webster, Word-of-Mouth Communications: A Motivational Analysis, Advances in Consumer Research, 25(1) (1998) 527-531.
[25] 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.
[26] E.W. Anderson, Customer Satisfaction and Word of Mouth, Journal of Service Research, 1(1) (1998) 5-17.
[27] F.A. Buttle, Word of Mouth: Understanding and Managing Referral Marketing, Journal of Strategic Marketing, 6(3) (1998) 241-254.
[28] G. J.Stigler, The economics of information, The Journal of Political Economy 69(3) (1961) 213-225
[29] G.H. Bower, Mood and Memory, American psychologist, 36(2) (1981) 129-148.
[30] H. Leventhal, A Perceptual Motor Theory of Emotion, Advances in Experimental Social Psychology, 17(1984) 117-182.
[31] 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.
[32] H. Rui, Y. Liu, A. Whinston, Whose and what chatter matters? The effect of tweets on movie sales, Decision support systems, 55(4) (2013) 863-870.
[33] 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.
[34] J. Berger, K.L. Milkman, What Makes Online Content Viral?, Journal of Marketing Research, 49(2) (2012) 192-205.
[35] J. John G. Lynch, D. Ariely, Wine online: Search costs affect competition on price, quality, and distribution, Marketing Science, 19(1) (2000) 83-103.
[36] K. Dave, S. Lawrence, D.M. Pennock, Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews, Proceedings of the 12th international Conference on World Wide Web Budapest, Hungary, (2003).
[37] L.R. Klein, Evaluating the Potential of Interactive Media through a New Lens: Search versus Experience Goods, Journal of Business Research, 41(3) (1998) 195-203.
[38] L. Cabral, A. Hortacsu, The Dynamics of Seller Reputation: Evidence from e bay, Journal of Industrial Economics, 58(1) (2010) 54-78.
[39] 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.
[40] M. Jiang, S. Argamon, Finding Political Blogs and Their Political Leanings, Proceedings of SIAM Text Mining Workshop 12, Atlanta, GA, USA, (2008b).
[41] M. F.Maute, L. Dube, Patterns of Emotional Responses and Behavioral Consequences of Dissatisfaction, Applied Psychology, 48(3) (1999) 349-366.
[42] M. L.Richins, Measuring Emotions in the Consumption Experience, Journal of Consumer Research, 24(2) (1997) 127-146.
[43] M.R. Darby, E. Karni, Free Competition and the Optimal Amount of Fraud, The Journal of Law and Economics, 16(1) (1973) 67-88.
[44] 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.
[45] 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).
[46] N. Hu, J. Zhang, P. A.Pavlou, Overcoming The J-Shaped Distribution of Product Reviews, communications of the ACM, 52(10) (2009) 144-147.

[47] P. Chatterjee, Online Reviews-Do Consumers Use Them?, Advances in Consumer Research, 28(1) (2001) 129-133.
[48] P. Gupta, J. Harris, How e-WOM recommendations influence product consideration and quality of choice: A motivation to process information perspective, Journal of Business Research, 63(9-10) (2010) 1041-1049.
[49] 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.
[50] 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.
[51] P. Nelson, Information and Consumer Behavior, The Journal of Political Economy, 78(2) (1970) 311-329.
[52] R. Plutchik, Emotion, A Psycho evolutionary Synthesis, New York, Academic, (1980).
[53] R.A. Westbrook, R.L. Oliver, The Dimensionality of Consumption Emotion Patterns and Consumer Satisfaction, Journal of Consumer Research, 18(1) (1991) 84-91.
[54] R.L. Desatnick, Managing to keep the customer, Strategic Change, 1(5) (2006) 273-285.
[55] R.E. Allen, T.J. Keaveney, Factors Differentiating Grievants and Nongrievant, Human Relations, 38(6) (1985) 519-534.
[56] R. N.Bolton, J.H. Drew, Mitigating the Effect of Service Encounters, Marketing Letters, 3(1) (1992) 57-70.
[57] S. Balasubramanian, V. Mahajan, The Economic Leverage of the Virtual Community, International Journal of Electronic Commerce, 5(3) (2001) 103-138.
[58] 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.
[59] S. Stieglitz, L. Dang-Xuan, Emotions and Information Diffusion in Social Media: Sentiment of Micro blogs and Sharing Behavior Journal of Management Information Systems 29(4) (2013) 217-248.
[60] 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.
[61] W. Duan, B. Gu, A.B. Whinston, Do online reviews matter? An empirical investigation of panel data, Decision Support Systems, 45(4) (2008) 1007-1016.
[62] 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.
[63] X. Li, L.M. Hitt, Self-Selection and Information Role of Online Product Reviews, Information Systems Research, 19(4) (2008) 456-474.
[64] Yi-Ching 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.
[65] Y. Liu, Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue, Journal of Marketing Research, 70(3) (2006) 74-89.
[66] 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.

Indian Institute of Management Kozhikode

Type of Document: (**Working Paper/Case/ Teaching Note, etc.)** **WORKING PAPER**	*Ref. No.: (to be filled by RCP office)* ***IIMK/WPS/156/MKTG/2014/14***
Title: **DO EMOTIONS MATTER? EXPLORING THE DISTRIBUTION OF EMOTIONS IN ONLINE PRODUCT REVIEWS**	
Author(s):	***Institution(s)***
Rahat Ullah	Korea Advanced Institute of Science and Technology (KAIST)
Wonjoon Kim	Korea Advanced Institute of Science and Technology (KAIST)
Naveen C. Amblee	Assistant Professor Department of Marketing Management Indian Institute of Management Kozhikode Kerala, India 673 570 Email: amblee@iimk.ac.in
Hyunjong Lee	Researcher at WISEnut, South Korea
Alice Oh	Korea Advanced Institute of Science and Technology (KAIST)
Subject Areas : **Marketing, Ecommerce**	*Subject Classification Codes, if any:*
Supporting Agencies, if any:	*Research Grant/Project No.(s):*
Supplementary Information, if any:	*Date of Issue: (to be filled by RCP office)* ***April 2014***
Full text or only abstract to be uploaded on website: (please choose one) ***ABSTRACT***	*Number of Pages:* ***34***

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 such as the causal effect on consumer choice, distribution pattern of WOM, and its product type-dependency, 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 needs more attention from researchers. In addition, an important antecedent for the generation of word-of-mouth is a strong emotional response, which in turn triggers the consumer to post a customer review online. However, only a limited number of studies to date have actually examined the content of reviews for their emotional content. To fill this gap, we analyzed the emotional content of a large number of online product reviews using Natural Language Processing (NLP) methods. 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. In addition, we found 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. Investigating further, we did find that there is a difference in the emotional content of reviews between search and experience goods in the early stages of product launch. However, interestingly, we find that these differences disappear over time as the addition of reviews reduces the information asymmetry gap. This provides important evidence to the widely held notion that on the Internet, all goods become search goods. Our findings suggest important managerial implications regarding product development, advertisement, and platform design using WOM content.

Key Words/Phrases: Ecommerce, online word-of-mouth, emotional content, reviews, NLP

Referencing Style Followed: ***APA***

Research, Conference And Publication Office
Indian Institute Of Management Kozhikode
IIMK Campus P.O., Kozhikode 673 570
Kerala, India
Telephone +91 495 2809 238
E-mail rcp@iimk.ac.in
website www.iimk.ac.in