Original

The impact of emotions on the helpfulness of movie reviews

R. Ullah[a], A. Zeb[a], W. Kim[a,b],*

[a]*Graduate School of Culture Technology, KAIST, Yuseong-gu, Daejeon, South Korea*
[b]*Department of Business and Technology Management, KAIST, Yuseong-gu, Daejeon, South Korea*

Abstract

Online customer reviews have become a significant source of product-related information for consumers. As a result of the growing number of customer reviews, determining which customer reviews are the most helpful is important in reducing information overload. The ways in which reviews can be helpful need to be identified. In this study, we examine the impact of emotional content in online customer reviews on the number of votes those customer reviews receive that indicate they were helpful. We find that content that is more emotional yields more votes. Furthermore, our findings suggest that reviews with positive emotional content have a positive effect on review helpfulness whereas reviews with negative emotional content have no effect on review helpfulness. This study contributes to an understanding of emotional content in word of mouth and has important implications for online retailers and consumers.

1. Introduction

Online word of mouth (WOM), which has been reported as one of the most influential means of natural and informal communication among consumers, is defined as any statement made by a consumer about a product and made available to other consumers on the Internet (Hennig-Thurau et al., 2004). In theory, customers can benefit from having hundreds of online reviews from which to gain information; however, this high number of reviews can create information overload (Jones et al., 2004). Therefore, consumers need to be able to identify a smaller number of reviews that will be helpful, and various online web portals are trying to provide ways to do just that (Mudambi & Schuff, 2010).

Previous studies have examined the significance of online WOM, particularly consumer-generated product reviews. Studies have shown that product reviews have a significant effect on purchase decisions (Chen et al., 2008) and product sales, and was widely confirmed in previous studies (Duan et al., 2008; Chevalier & Mayzlin, 2006; Liu, 2006; Godes & Mayzlin, 2004). Prior research in the study of WOM has shown that volume and valence are the two important metrics (Neelamegham & Chintagunta, 1999), which may affect the number of votes a review receives to indicate its helpfulness (Mudambi & Schuff, 2010). Volume means the total amount of WOM while valence means the character of WOM content that is, whether it is positive or negative. Hence a wide range of research has focused on the effect of volume and valence on product sales. For example, WOM volume is reported as having a significant impact on product sales (Chevalier & Mayzlin, 2006; Chen et al., 2004; Godes & Mayzlin, 2004; Liu, 2006). While some studies have reported that WOM valence is a significant factor in product sales (Godes & Mayzlin, 2004; Chevalier & Mayzlin, 2006), others have found no significant relationship (Liu, 2006).

In particular, previous research considers the volume and valence of WOM without considering its contents, which is not sufficient for understanding WOM behavior (Rime et al., 1998). Only a handful of recent studies have analyzed the textual information embedded in online customer reviews in great detail. For example, Ghose and Ipeirotis (2010) and Ghose et al. (2012) have identified important text-based features such as lexical, semantic, grammatical and stylistic levels and have studied their impact on product sales and review helpfulness, while Rui et al. (2013) used machine-learning algorithms and found that negative WOM is associated with lower sales of movie sales, whereas positive tweets lead to higher movie sales.

However, we are unaware of any previous studies that examine the emotional content of reviews and its relationship to review helpfulness. Research has shown that emotions influence decision making (Raghunathan et al., 2006; Shiv et al., 2005).

*Corresponding author.
E-mail address: wonjoon.kim@kaist.edu (W. Kim).

Moreover, discrete emotions with the same valence affect information processing (Tiedens & Linton, 2001). In particular, we aim to understand why some reviews receive many votes indicating helpfulness (which include both "yes" and "no" votes) whereas others get only a few votes or none. As a result, in this paper, we study which emotional features of online reviews influence the number of votes saying they were helpful.

Thus, we add to the literature on online WOM by analyzing the emotional content of online movie reviews and its impact on the helpfulness of reviews using natural language processing (NLP) techniques that support the human natural linguistic aspects of WOM. In the first stage of our study, we examine the total emotional content and then follow up by looking at positive and negative emotional content and its impact on the helpfulness of reviews. We find that the emotional content of WOM is more influential than the other attributes. This finding distinguishes our approach from that of previous studies, which focus on extracting semantic characteristics from the text of the review and show how many votes indicating helpfulness the reviews received (Cao et al., 2011). In addition, if more helpful reviews can be identified earlier, then they can be acted on sooner by retailers, manufacturers, and third-party providers.

The remaining of the paper proceeds as follows. In the next section, we conduct a review of the literature. We then describe the data and research methodology. Finally, we discuss the results and conclude with some implications.

2. Literature Review

2.1. Motivation to Post Customer Reviews

Consumers write online reviews for various reasons. The most accepted motivation in the literature is that consumers engage in WOM when they are extremely satisfied or extremely dissatisfied (Anderson, 1998). The act of writing the review allows consumers to reduce their dissatisfaction, which is connected with negative emotions (Hennig-Thurau et al., 2004) such as tension and anxiety (Sundaram et al., 1998). Therefore, the involving focusing on negative WOM is sometimes referred to as dissonance reduction (Buttle, 1998). It has also been noted that consumers are involved to engage in negative WOM for reasons related to altruism, seeking vengeance, and seeking advice (Sundaram et al., 1998). In addition, consumers are motivated to engage in positive WOM when their expectations are met or exceeded (Sundaram et al., 1998).

Specifically, in the domain of online WOM, consumers are involved by an ambition for possible economic benefit, social interaction, the potential to enhance their own self-worth, and their concern for other consumers (Hennig-Thurau et al., 2004). By expressing themselves online, they can obtain various kinds of utility (Balasubramanian & Mahajan, 2001). As Dellarocas and Narayan (2006) noted, emotion is an important and common cause for WOM behavior among consumers, whether offline or online, and we find that the reasons for online and offline WOM are similar overall.

2.2. The Role of Emotions

Prior research has examined the relationship between consumer behavior and consumption-related emotions and suggested that emotions play a major role in consumer response (Richins, 1997). In general, emotion has a valence and can be divided into positive and negative emotions. Additionally, emotion enables the processes of communication through which elements of a private experience can be shared with social partners. For example, the surprise emotion has been shown to play a key role in the evocation of consumer WOM. Negative surprise can create negative WOM, while positive surprise can produce positive WOM (Derbaix & Vanhamme, 2003).

However, surprise is a non-valenced emotion that is caused when possibility about a product or service are not met, also known as schema discrepancy (Derbaix & Vanhamme, 2003). In schema discrepancy analysis, a second emotion follows, such as anger (negative) or joy (positive), which causes one to presume that negative or positive surprise was evoked (Hennig-Thurau et al., 2004; Anderson, 1998). Therefore, previous studies consistently emphasize the important role of emotions, especially in consumers' WOM behavior. However, no evidence directly regarding the impact of emotional content in reviews on their helpfulness has been presented.

2.3. Impact of Emotional Content on Votes Indicating Helpfulness

The increasing number of reviews available for various products has created information overload for consumers (Yin et al., 2011). Pinpointing which reviews are most helpful is critical in reducing this information overload. A review diagnosticity theory is defined as the piece of information that is helpful in making informed purchase decisions and is linked to the assumption of information diagnosticity, which includes the question of whether a certain text of information is helpful during the processes of decision-making (Yin et al., 2011). Within the basic model of review diagnosticity, researchers noted that product type, review depth, and the degree of review extremity affect the helpfulness of reviews (Mudambi & Schuff, 2010).

Previous research has shown that emotions play a positive role in decision making (Shiv et al., 2005) and emotions play a key role in driving online conversations (Berger & Milkman, 2012). However, little is known about the significance of emotions in online reviews specifically, how emotions influence consumer judgment in an online environment. Therefore, reviews with emotional content affect perceptions of its helpfulness. Thus, we propose:

H1: Reviews with emotional content will have more votes indicating that they were helpful.

2.4. Impact of Positive and Negative Emotional Content on Votes Indicating Helpfulness

A handful amount of recent studies have attempted to examine the text of reviews for their helpfulness. For example, the

review's readability had a significant effect on the ratio of votes indicating a review's helpfulness, and extremely helpful reviews receive a higher score than do those considered less helpful (Korfiatis et al., 2012). In addition, Cao et al. (2011) used text mining to examine the impact of various aspects of online user reviews —stylistic, basic and semantic characteristics— on the number of votes those reviews receive indicating helpfulness. They suggest that the semantic characteristics are having more effect than other characteristics in affecting reviews' helpfulness, and reviews with extreme opinions are considered more helpful than those with neutral or mixed opinions. Moreover, it has been reported that statements regarding product quality have a positive influence on review helpfulness (Siering & Muntermann, 2013). However, we are unaware of any study that examines positive and negative emotional content in the text of online reviews and their impact on the review's helpfulness. Therefore, we propose that:

H2: Reviews with positive emotional content will have a positive effect on review helpfulness.

H3: Reviews with negative emotional content will have a negative effect on review helpfulness.

3. Methodology

To test our hypotheses, we selected a list of movies from Box Office Mojo[1] released nationwide in the United States from May to September 2011. We automatically retrieved reviews from the Internet Movie Database (IMDb)[2] using a customized software tool. The final dataset comprises 41 movies (16,821 reviews) after excluding the titles for which no reliable information about all variables was available. Using the natural language processing (NLP) technique, we compiled a word list from reviews with the Stanford morpheme analyzer (Toutanova et al., 2003) and extracted words that were adjectives, nouns, and verbs, and other morphemes and special characters were filtered out. We also made a positive/negative bag of words (BOW) and used coefficient correlation (CC) to measure the dependency between a word t and a polarity category c. "CC" is defined in equation 1 and can be viewed as a "one-sided" chi-square metric (Zheng et al., 2004). In equation 1, N is the total number of documents that are used to calculate CC. N is the number of documents for a movie in a week. $P(t, c)$ means the probability that word t occurred in class c. $\bar{t}$ means words that are not t in a BOW, and $\bar{c}$ means categories that are not c. Therefore, CC measures the lack of independence between word t and category c. Finally, we computed the number of words adjacent to each word.

$$CC(t,c_i) = \frac{\sqrt{N}\left[P(t,c_i)P(\bar{t}\,\bar{c_i}) - P(t,\bar{c_i})P(t,\bar{c_i})\right]}{\sqrt{P(t)P(\bar{t})P(c_i)P(\bar{c_i})}} \quad (1)$$

We used seven basic emotions and descriptors (as in Derbaix & Vanhamme, 2003) that could be used to measure each emotion, as presented in Table 1. Examples of emotion-related words are listed in Table 2.

4. Analysis and Results

The dependent variable, helpfulness, is calculated as the percentage of people who found a review helpful (Helpfulness %). This was obtained by dividing the number of people who voted that the review was helpful by the total number of votes, as prescribed by Mudambi and Schuff (2010) and Yin et al. (2014). The IMDB website presents the question "Was the above review useful to you?" along with "Yes" and "No" options. Following the work of Mudambi and Schuff (2010) and Yin et al. (2014), the rating, rating squared, and word count are used as control variables. Rating refers to a review's star rating, which ranges from 1 star to 10 stars. A quadratic term of the star rating is included because of the nonlinear relationship between helpfulness and the star rating while the review length was measured by the total number of words in a review.

4.1. Summary Statistics

The summary statistics of reviews containing emotional content are listed in Table 3. Across all reviews, the total emotional content is about 17%; positive emotional content is about 12%, compared with negative emotional content, which is about 4%.

Table 1
Primary Emotions Identified in Derbaix and Vanhamme (2003).

Fear (surprised, amazed, astonished)
Anger (angry, mad, enraged)
Enjoyment (joyful, delighted, happy)
Distress (sad, downhearted, discouraged)
Disgust (disgusted, feeling of distaste or revulsion)
Surprise (surprised, amazed, astonished)
Contempt (disdainful, contemptuous, scornful)

Table 2
Examples of Some Emotion-related Words.

Words indicating positive motions		Words indicating negative emotions	
Love	Delightful	Disappointment	Sad
Great	Gorgeous	Worse	Irritating
Pleased	Tremendous	Unhappy	Disgust
Wonderful	Phenomenal	Bad	
Happy	Outstanding	Stupid	
Amazed	Superb	Angry	
Excellent		Displeased	
Surprise		Awful	
Fun		Nonsense	

Table 3
Descriptive Statistics for the Full Sample (*n*=16,821).

Variable	Mean	Std. Dev.	Min	Max
Review helpfulness	0.48	0.06	0	1
Total emotional content	0.17	0.07	0.03	0.35
Positive emotional content	0.12	0.05	0.05	0.27
Negative emotional content	0.04	0.04	0	0.16
Rating	5.82	0.84	1	10
Length	70.21	71.67	19	917

We used an approach similar to that of Mudambi and Schuff (2010) and Yin et al. (2014) to analyze our data using a Tobit regression. This approach is appropriate because of the censored nature of the dependent variable: its value is in range and was constructed as a ratio (Greene, 2002). To investigate the effect of the percentage of total emotional content and positive and negative emotional content on the votes indicating helpfulness, we performed a Tobit regression, as shown in Table 4.

The results in Table 4 suggest that coefficient for the percentage of emotional content is 0.06 (P<.05), which is statistically significant. Thus the results support H1. The results in Table 4 show the effect of the percentage of positive and negative emotional content on votes indicating helpfulness. The results suggest that the coefficient of the percentage of positive emotional content is 0.04 (P<.003), which is positive and significant. Thus, the results support H2, perhaps because positive emotions are more likely to spread than negative ones (Berger & Milkman, 2012). Another explanation might be that positive valence boosts strength more than negative valence, except in the case of news (Hansen et al., 2011).

In addition, the results in Table 4 also show that the coefficient for the percentage of negative emotional content is –0.002, which is not significant. Thus, the results do not support H3. The coefficient of the percentage of positive emotional content suggests that positive emotions in movie reviews have a positive impact on review helpfulness, while negative emotions in movie reviews have no impact on review helpfulness. Furthermore, the word count is a significant predictor and has a significant and positive effect on a review's helpfulness. The coefficient for the word count is 2.58 (P<.00). This finding is in line with results in Mudambi and Schuff (2010).

5. Discussion and Conclusions

Studies on WOM have examined the motivations behind posting customer reviews (Hennig-Thurau et al., 2004) and showed that emotions are important in that activity (Derbaix & Vanhamme, 2003). However, those studies focused mainly on the distribution of reviews, but the role of emotions in the reviews' content, and its impact on review helpfulness have not been studied before, although the review content and its mixed characteristics have important, useful, and rich information for understanding online consumer behavior. Therefore, we try to examine movie reviews' emotional content and its impact on review helpfulness.

We have reached the following conclusions by examining the emotional content of reviews, finding that more emotional content in online reviews will be considered more helpful. More interestingly, we find that reviews with positive emotional content have a positive effect on review helpfulness whereas reviews with negative emotional content have no effect on review helpfulness. This result emphasizes the importance of total and positive emotions in online reviews.

Our findings have several managerial implications by highlighting the helpfulness of reviews and emotions beyond simple ratings and volume. First, we believe that the content of reviews is very rich and full of information that enables us to understand how consumers evaluate emotional product information in an online marketplace, including how one can quantify emotions and how emotional communication works. Second, our findings suggest that reviews with more emotional content are more helpful, therefore online retailers need to display emotional reviews first. Third, from an advertising perspective, the NLP techniques can help retailers decide which prominent and appropriate emotions to promote and highlight in advertisements and in-store displays. For example, in the case of the film *Pirates of the Caribbean,* phrases such as "lovely story," "amazing video," and "great action," if they occur frequently, could be highlighted in advertisements for the movie's DVD release or sequel. Fourth, reviews are usually rich but unstructured consumer data and are primarily qualitative in nature, which makes them noisy and seemingly impractical to convert into usable information. To address these issues, we employ the NLP methodology, which should be viewed as a beneficial tool for analyzing huge numbers of customer reviews to extract important marketing insights. Managers can understand customers' needs and desires for the products and services better by implementing those techniques to all forms of user-generated content.

This study has several limitations. Specifically, the computational NLP techniques are not as intelligent as humans. However, as NLP techniques become more advanced over time, we can gain greater understanding of consumer behavior in user-generated content. Future research should also examine the objective and subjective nature of reviews and product-type reviews, which can provide more insights regarding customer reviews.

Table 4
Tobit Analysis Results for Final Review Pool (Dependent Variable: Review Helpfulness; n=16,821).

Variable	Coefficient	Standard Error	t-value
Constant	0.18*	0.05	6.19
Total emotional content	0.06*	0.03	0.56
Positive emotional content	0.04*	0.002	1.83
Negative emotional content	–0.002	0.003	–0.074
Rating	0.12*	0.02	5.03
Rating Square	–0.01*	0.002	–4.75
Word count	2.58*	1.24	2.08

*Significant at the ≤.01 level.
Likelihood ratio=–114.98 (P=.000). Pseudo R^2=.26.

References

Anderson, E.W. (1998). Customer satisfaction and word of mouth. *Journal of Service Research, 1,* 5-17.

Balasubramanian, S., & Mahajan, V. (2001). The Economic Leverage of the Virtual Community. *International Journal of Electronic Commerce, 5,* 103-138.

Berger, J., & Milkman, K.L. (2012). What makes online content viral? *Journal of Marketing Research, 49,* 192-205.

Buttle, F.A. (1998). Word of mouth: understanding and managing referral marketing. *Journal of Strategic Marketing, 6,* 241-254.

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems, 50,* 511-521.

Chen, P.Y., Dhanasobhon, S., & Smith, M.D. (2008). *All reviews are not equal: The disaggregate impact of reviews and reviewers at Amazon.com.* Available at http://ssrn.com/abstract=918083

Chen, P.Y., Wu, S.Y., & Yoon, J. (2004). *The Impact of Online Recommendations and Consumer Feedback on Sales.* AIS Electronic Library (AISeL).

Chevalier, J.A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43,* 345-354.

Dellarocas, C., & Narayan, R. (2006). A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science, 21,* 277-285.

Derbaix, C., & Vanhamme, J. (2003). Inducing word-of-mouth by eliciting surprise—a pilot investigation. *Journal of Economic Psychology, 24,* 99-116.

Duan, W., Gu, B., & Whinston, A.B. (2008). Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems, 45,* 1007-1016.

Ghose, A., & Ipeirotis, P.G. (2010). *Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer.* Washington, DC: IEEE Trans. Knowledge Data Engrg., IEEE Computer Society.

Ghose, A., Ipeirotis, P.G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science, 31,* 493-520.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science, 23,* 545-560.

Greene, W.H. (2002). *Econometric Analysis.* Uppler Saddle, New Jersey: Prentice Hall.

Hansen, L.K., Arvidsson, A., Nielsen, F.Å., Colleoni, E., & Etter, M. (2011). Good friends, bad news-affect and virality in twitter. In: *Future information technology* (pp. 34-43). Berlin-Heidelberg: Springer.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G., & Gremler, D.D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing, 18,* 38-52.

Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research, 15,* 194-210.

Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications, 11,* 205-217.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing, 70,* 74-89.

Mudambi, S.M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34,* 185-200.

Neelamegham, R., & Chintagunta, P. (1999). A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Science, 18,* 115-136.

Raghunathan, R., Pham, M.T., & Corfman, K.P. (2006). Informational properties of anxiety and sadness, and displaced coping. *Journal of Consumer Research, 32,* 596-601.

Richins, M.L. (1997). Measuring emotions in the consumption experience. *Journal of Consumer Research, 24,* 127-146.

Rimé, B., Finkenauer, C., Luminet, O., Zech, E., & Philippot, P. (1998). Social sharing of emotion: New evidence and new questions. *European Review of Social Psychology, 9,* 145-189.

Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? The effect of tweets on movie sales. *Decision Support Systems, 55,* 863-870.

Shiv, B., Loewenstein, G., & Bechara, A. (2005). The dark side of emotion in decision-making: When individuals with decreased emotional reactions make more advantageous decisions. *Cognitive Brain Research, 23,* 85-92.

Siering, M., & Muntermann, J. (2013). *What Drives the Helpfulness of Online Product Reviews? From Stars to Facts and Emotions.* 11th International Conference on Wirtschaftsinformatik, Leipzig, Germany.

Sundaram, D.S., Mitra, K., & Webster, C. (1998). Word-of-mouth communications: A motivational analysis. *Advances in Consumer Research, 25,* 527-531.

Tiedens, L.Z., & Linton, S. (2001). Judgment under emotional certainty and uncertainty: the effects of specific emotions on information processing. *Journal of Personality and Social Psychology, 81,* 973-998.

Toutanova, K., Klein, D., Manning, C.D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 173-180). Association for Computational Linguistics.

Yin, D., Bond, S., & Zhang, H. (2011). Dreading and Ranting: The Distinct Effects of Anxiety and Anger. In *Online Seller Reviews, ICIS 2011 Proceedings.*

Yin, D., Bond, S., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly, 38,* 539-560.

Zheng, Z., Wu, X., & Srihari, R. (2004). Feature selection for text categorization on imbalanced data. *ACM SIGKDD Explorations Newsletter, 6,* 80-89.