The Effects of Sentiment and Readability on Useful Votes for Customer Reviews with Count Type Review Usefulness Index

Ruth Angelie Cruz
The Catholic University of Korea
(ruthcruz@catholic.ac.kr)

Hong Joo Lee
The Catholic University of Korea
(hongjoo@catholic.ac.kr)

....

Customer reviews help potential customers make purchasing decisions. However, the prevalence of reviews on websites push the customer to sift through them and change the focus from a mere search to identifying which of the available reviews are valuable and useful for the purchasing decision at hand. To identify useful reviews, websites have developed different mechanisms to give customers options when evaluating existing reviews. Websites allow users to rate the usefulness of a customer review as helpful or not. Amazon.com uses a ratio-type helpfulness, while Yelp.com uses a count-type usefulness index. This usefulness index provides helpful reviews to future potential purchasers. This study investigated the effects of sentiment and readability on useful votes for customer reviews. Similar studies on the relationship between sentiment and readability have focused on the ratio-type usefulness index utilized by websites such as Amazon.com. In this study, Yelp.com's count-type usefulness index for restaurant reviews was used to investigate the relationship between sentiment/readability and usefulness votes. Yelp.com's online customer reviews for stores in the beverage and food categories were used for the analysis. In total, 170,294 reviews containing information on a store's reputation and popularity were used. The control variables were the review length, store reputation, and popularity; the independent variables were the sentiment and readability, while the dependent variable was the number of helpful votes. The review rating is the moderating variable for the review sentiment and readability. The length is the number of characters in a review. The popularity is the number of reviews for a store, and the reputation is the general average rating of all reviews for a store. The readability of a review was calculated with the Coleman‐Liau index. The sentiment is a positivity score for the review as calculated by SentiWordNet. The review rating is a preference score selected from 1 to 5 (stars) by the review author.

The dependent variable (i.e., usefulness votes) used in this study is a count variable. Therefore, the Poisson regression model, which is commonly used to account for the discrete and nonnegative nature of count data, was applied in the analyses. The increase in helpful votes was assumed to follow a Poisson distribution. Because the Poisson model assumes an equal mean and variance and the data were over-dispersed, a negative binomial distribution model that allows for over-dispersion of the count variable was used for the estimation. Zero-inflated negative binomial regression was used to model count variables with excessive zeros and over-dispersed count outcome variables. With this model, the excess zeros were assumed to be generated through a separate process from the count values and therefore should be modeled as independently as possible.

The results showed that positive sentiment had a negative effect on gaining useful votes for positive reviews but no significant effect on negative reviews. Poor readability had a negative effect on gaining useful votes and was not moderated by the review star ratings.

These findings yield considerable managerial implications. The results are helpful for online websites when analyzing their review guidelines and identifying useful reviews for their business. Based on this study, positive reviews are not

....

···

necessarily helpful; therefore, restaurants should consider which type of positive review is helpful for their business. Second, this study is beneficial for businesses and website designers in creating review mechanisms to know which type of reviews to highlight on their websites and which type of reviews can be beneficial to the business. Moreover, this study highlights the review systems employed by websites to allow their customers to post rating reviews.

1. Introduction

With the ubiquitous nature of e-WOM these days, more and more online retailers rely on the influencing power of customer product reviews to improve and assist the online shopping experience(Chae et al., 2015; Choi and Lee, 2011; Chun and Ahn, 2015). E-commerce companies like Amazon.com and Yelp.com use online customer reviews to provide quality information to consumers and feedback to their affiliates and partners. Review mechanisms put up by websites not only attempt to influence consumer purchasing decisions but it also increases a website's online reputation and presence. Kumar and Benbasat (2006) showed that the presence of customer reviews on a website increases the customer's perception of the usefulness and social presence of the site. Product reviews also increase "stickiness", the ability to increase time spent on the website, attract consumer visits, and help create a sense of community among customers (Mudambi and Schuff, 2010).

Nonetheless, aside from making it easier for companies to provide accurate information about their products, past research have shown that customer reviews can have positive influence on sales (Chen et al., 2008; Chevalier and Mayzlin, 2006; Clemons et al., 2006; Ghose and Ipeirotis, 2006). For this reason, online retailers have since been proactive in providing reviews about their products online. Resnick et al. (2000) used the website, eBay.com, and showed that seller reviews affect the probability of a sale, while Chevalier and Mayzlin (2006) found that book sales were influenced by reviews on Amazon.com. Chen et al. (2008) found that higher quality reviews have a stronger effect than overall reviews which means customers mitigate the high search costs of reading through a lot of reviews by focusing on few highlighted reviews.

Consequently, the prevalence of reviews on websites push the customer to sift through them and change the focus from mere search of customer reviews to identifying which of the available reviews are valuable and useful for the purchase decision at hand. To identify useful reviews, websites develop different mechanisms to give customers an option to evaluate existing reviews. Ratings makes decision-making easier

(Dabholkar, 2006) and positively affects the consumer's attitude about shopping online (Jiang and Benbasat, 2007). Amazon.com uses a ratio-type helpfulness index (= Number of useful votes / (Number of useful votes + Number of not useful votes)) while Yelp.com uses count-type usefulness index (= Number of useful votes). This practice of evaluating useful/helpful reviews increased the dimensions available for investigation regarding the effect of customer reviews on actual sales and which characteristics of a review is actually helpful for a customer.

This paper investigated the effects of sentiment and readability on count type usefulness index for customer reviews on Yelp.com－an online review website for business and establishments. To extend the investigation on customer reviews, this paper does not focus on review's effect on sales, rather it focuses on what makes a review helpful by comparing a review's sentiment, readability, and count type usefulness index. This paper uses Yelp.com's data to investigate the relationship between sentiment, readability, and usefulness votes. Control variables used were review length, store reputation, and popularity, the independent variables are sentiment and readability, while the dependent variable is the number of helpful votes.

2. Review of Related Literature and Framework

2.1. Consumer Reviews

Online consumer reviews have become important in consumer decision making because, as

one form of online word of mouth, it provides evaluations of a product or service and recommendations from a previous customer's experience (Baum and Spann, 2014). It also supplements product information provided by the website and personalized advice generated by automated recommender systems. Because of its first-hand nature, consumer-generated recommendations are viewed to be more credible and more trustworthy than marketer-generated recommendations (Bickart and Schindler, 2001; Park et al., 2007). Various studies have proven that online consumer reviews affect the entire purchase decision-making process of the customer, from the moment the need is recognized (Schindler and Bickart, 2005), to his or her information search behavior (Baek et al., 2012-13; Mudambi and Schuff, 2010), to evaluating product alternatives (Baek et al., 2012-13), to making the actual purchase decision (Chevalier and Mayzlin, 2006), until the after-purchase phase.

Aside from its influences and benefits to the consumer, customer reviews also provide considerable advantages to the website. Product reviews are said to increase a website's "stickiness"－ it's ability to increase time spent on the site, attract consumer visits, and assist in creating a sense of community among customers (Mudambi and Schuff, 2010). To add to that, the presence of customer reviews on a website increases the customer's perception of the usefulness and social presence of the site (Kumar and Benbasat, 2006).

Existing studies have also shown the positive influence of customer reviews on sales (Chen et

al., 2008; Chevalier and Mayzlin, 2006; Clemons et al., 2006; Ghose and Ipeirotis, 2006). To investigate the effects of reviews on sales, Resnick et al. (2000) used eBay.com to show how seller reviews affect the probability of a sale, while Chevalier and Mayzlin (2006) investigated Amazon.com and found that reviews affect book sales. For this reason, companies have placed primary focus on putting up a review mechanism on their website in an attempt to influence consumers' purchasing decisions.

2.2. Review Helpfulness

To identify useful reviews, websites develop different mechanisms to give customers an option to evaluate existing reviews. Ratings makes decision-making easier (Dabholkar, 2006) and positively affects the consumer's attitude about shopping online (Jiang and Benbasat, 2007). Amazon.com uses a ratio-type helpfulness index. In this index, Amazon has implemented a system where customers are allowed to vote whether a particular review has been helpful or not. The fraction of customers who thought that the review was helpful is displayed with the total number of votes and the actual text of the review. Several studies have used Amazon's ratio-type helpfulness index to investigate the helpfulness of a review (Korfiatis et al., 2012; Chen et.al., 2008; Chevalier and Mayzlin, 2006).

On the other hand, Yelp.com uses count-type usefulness index. To our knowledge, no other paper has investigated whether count-type usefulness index where customers can only rate a review as useful or completely disregard giving a

rating if it's not. Yelp's count-type usefulness index differs from Amazon's ratio-type helpfulness index in that Yelp only shows the number of people who rated the review as useful without any information on how many people actually tried to rate the review. Yelp.com also includes other choices for rating a review like "Funny" and "Cool" but for the purpose of this paper, only the "Useful" rating is used.

This practice of evaluating useful/helpful reviews increased the dimensions available for investigation regarding the effect of customer reviews on actual sales and which characteristics of a review is actually helpful for a customer. In particular, Clemons et al. (2006) found that strongly positive ratings positively impact the growth of sales. However, Mudambi and Schuff (2010) investigated whether those findings have limitations based on the type of product being reviewed. While Chen et al. (2008) showed in their study that the quality of a review as measured by helpfulness votes also positively impacts sales. To investigate reviewer behavior, Shen et al. (2015) compared how reviewers on Amazon.com and Barnes and Noble.com strategize their reviews to gain attention and reputation.

2.3. Hypothesis development

This paper investigates the effects of sentiment and readability on count-type useful votes for customer reviews on Yelp.com. To extend the investigation on customer reviews that uses count-type usefulness index, this paper does not focus on a review's effect on sales, but rather it

examines what makes a review useful and how a review gets more useful votes by comparing a review's sentiment, readability, and count-type useful votes. For the purpose of this paper, helpfulness and usefulness votes are used in a similar context. The difference lies in that Amazon.com uses the term "helpful" while Yelp.com uses "useful" but they are both used to describe a review that has been identified as beneficial for the customer. This paper uses Yelp.com's data to examine the relationship between (positive) sentiment, readability, and usefulness votes.

Based on previous studies, the sentiment of a message can be communicated effectively through the text and can extensively influence the interpretations of the reader (Harris and Paradice, 2007; Riordan and Kreuz, 2010). Jiang and Srinivasan (2015) expressed that the use of e-WOM has now turned into one of the major sources of product information to reduce uncertainties about price, quality, shipment, etc. Therefore, consumer-generated WOM has now become as one of the most credible forms of advertising for most consumers. Because of this developed dependence of consumers to user reviews, online retailers have become active solicitors of user-generated content in addition to passively hosting it. More and more retailers have started to give benefits in different forms to people who would provide user reviews to their purchases. Aside from that, users are encouraged to provide feedback, as well as peer endorsement, to other user reviews and reviewers through comments and votings on reviews. To add to that,

in the field of customer reviews, Mudambi and Schuff (2010) and Clemons et. al. (2006) found that people find reviews with extreme numerical ratings and strongly positive ratings to be more helpful.

Due to aforementioned reasons, it is valid to argue that people look for positive sentiment in reviews regarding their choices to convince them to make a purchase or visit a store. Therefore;

H1: Positive sentiment of the review text has a positive effect on the usefulness of the review.

The theory of selective attention posits that people respond to messages selectively because of limited information processing capacity (Treisman, 1964). People are usually limited in time and resources to sift through dozens of available reviews. Therefore, they opt for reviews that are easier to read. This can be theorized in the cognitive effort level in terms of the text's cognitive fit to a typical consumer with an average amount of knowledge about the restaurant/or product being evaluated in the review (Park and Kim, 2009).

Readability of textual data indicates the amount of effort that is needed by a person of a certain age and education level to understand a piece of text. This paper computes the readability of a review as calculated by Coleman-Liau index using korPus package in R (Michalke, 2012). The results of the Coleman-Liau index approximates the U.S. grade level thought necessary to comprehend the text.

Readability, therefore, is how easy it is to read and comprehend a piece of text containing a review related to the restaurant being evaluated (Korfiatis et al., 2012). Related to H1, we posit that readability of a review will be considered a deciding factor for rating a review is useful or not. Reviews that are easier to read and understand will lead to more useful votes for that particular review. Therefore, we hypothesize that;

H2: The level of readability of the review text has a positive effect on the usefulness of the review (the more understandable the text, the more useful the review).

Star ratings are a reflection of the deviation from the midpoint of an attitude scale or the "attitude extremity" (Krosnick et al., 1993). Clemons et al. (2006), in their study, found that strongly positive ratings can affect the growth of product sales positively. Meanwhile, there has been conflicting results on the helpfulness of extreme ratings compared to moderate ratings. Previous research investigates the effectivity of extreme ratings and moderate ratings and some of them found that two-sided or moderate arguments were more credible (Schlosser, 2005; Crowley and Hoyer, 1994) while some other studies found that one-sided extreme ratings were more useful (Pavlou and Dimoka, 2006; Forman et al., 2008). Mudambi and Schuff (2010) investigated the role of product type in review extremity on the helpfulness of reviews and found that for experience goods, review with extreme ratings are

less helpful than reviews with moderate ratings. Nonetheless, the moderating effect of star ratings on review sentiment have not been investigated as of date. This study aims to find out whether positive (4, 5 stars) or negative (1,2 stars) ratings will have a moderating effect on review sentiment whether a review will be deemed helpful or not. Therefore we hypothesize that;

H3: The effect of review sentiment on usefulness of a review is moderated by the ratings given to that particular review by the reviewer

The theory of selective attention expresses the customers' limited resources in processing reviews (Treisman, 1964). In line with this, customers will still choose which reviews they choose to read and focus on. Moreover, Schindler and Bickart (2012) found that the use of positive or negative style characteristics does not increase the value of a review and may even decrease the perceptions regarding its helpfulness. This study also attempts to investigate whether ratings affect the usefulness of readable reviews. We assume that readers' choices on which readable reviews to rate useful can be affected by whether they prefer to read positive or negatively star-rated reviews. Therefore we hypothesize that;

H4: The effect of readability on usefulness of a review is moderated by the ratings given to that particular review by the reviewer

2.4. Theoretical Framework
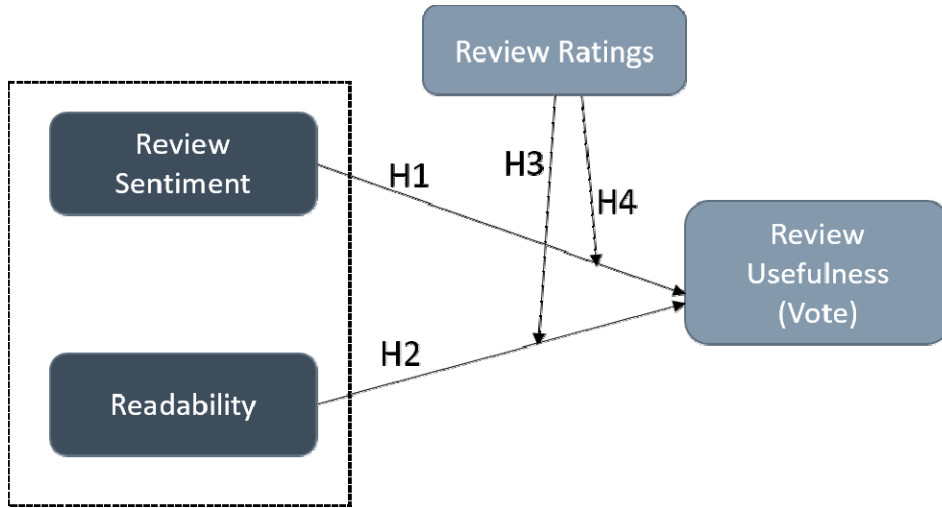
〈Figure 1〉 Theoretical Framework: the Influences of review sentiments and readability on count-type usefulness votes

3. Data and Methodology

We collected data for this study using Yelp.com's online customer reviews[1]. We selected customer reviews for stores in the beverage and food categories. In total, we had 170,294 reviews which contains information on a store's reputation and popularity.

3.1. Variables

This paper used the following variables: control variables used were review length, store reputation, and popularity, the independent variables are sentiment and readability, while the dependent variable is the number of helpful votes. Review rating is the moderating variable for review sentiment and readability.

Length of a review is the number of characters

in the review. Popularity is the number of reviews for a store and Reputation is the general average of all ratings of reviews for a store. Readability of a review was calculated by Coleman-Liau Index from korPus package in R (Michalke, 2012). Sentiment of a review is a positivity score for the review calculated from SentiWordNet (Baccianella, Esulti, and Sebastiani, 2010). Review rating is a preference score selected from 1 to 5 (stars) by the review author.

<Table 1> shows the description of the data set used in this study. Establishments included in the

〈Table 1〉 Data description

	Average	Standard deviation
Length	814.7	614.47
Popularity	129.10	140.70
Reputation	3.76	0.49
Rating	3.75	1.18
Readability	6.32	1.96
Sentiment	2.87	3.17

---

1) https://www.kaggle.com/c/yelp-recsys-2013

〈Table 2〉 Correlations of the variables

	Length	Popularity	Reputation	Rating	Readability	Sentiment
Length	-					
Popularity	0.063	-				
Reputation	-0.004	0.311	-			
Rating	-0.120	0.130	0.418	-		
Readability	-0.004	0.013	0.061	0.083	-	
Sentiment	0.508	0.095	0.105	0.179	-0.006	-

〈Table 3〉 Values of variance inflation factor(VIF)

	Poisson regression	Negative binomial regression
Length	1.443	1.471
Popularity	1.321	1.116
Reputation	1.113	1.320
Rating	1.324	1.330
Readability	1.008	1.009
Sentiment	1.467	1.492

study had an average of 129.10 reviews with an average reputation of 3.76. Reviews had an average of 814.7 characters, with a readability rating of 6.32, and a sentiment graded as 2.87. Reviews have an average rating of 3.75. <Table 2> shows the correlation coefficients among variables and <Table 3> shows values of variance inflation factor(VIF). There are no correlation coefficients greater than 0.7 and values of VIF are less than 2. We concluded that there are no multicollinearity issues and performed further analyses.

3.2. Approach & Methods

The dependent variable, usefulness votes, used in this study is a count variable and therefore Poisson regression model, a common model used to account for the discrete and nonnegative nature of the count data, is applied for the analyses. In this study, we assume that the increase in helpful votes follows a Poisson distribution. Since the Poisson model assumes equal mean and variance and our data has over-dispersion, we estimate a negative binomial distribution model that allows for over-dispersion of the count variable (Hausman et al., 1984). We also used the Zero-inflated Negative Binomial Regression to model count variables with excessive zeros and over-dispersed count outcome variables. Using this model, we assume that the excess zeros are generated through a separate process from the count values and therefore model the excess zeros as independently is possible.

4. Results

<Table 4> shows the results of the Zero-inflated Negative Binomial Regression. The Count Model results show which variables affect a review's tendency to get more usefulness votes. Model 1 results show that the control variables, review length, reputation, and popularity, are significant

〈Table 4〉 Results of Zero-inflated Negative Binomial Regression

	Model I (All)	Model II (All)Count Model	Model III (All)
Length	0.0005***	0.0006***	0.0006***
Popularity	0.0001***	0.0001***	0.0001***
Reputation	0.0617***	0.0660***	0.0675***
Rating	-0.0051	0.0063*	-0.0246**
Readability		-0.0289***	-0.0386***
Sentiment		-0.0112***	-0.0332***
Rating*Readability			0.0025
Rating*Sentiment			0.0061***
Log(theta)	0.2039***	0.2106***	0.2105***
Zero-Inflation Model			
Length	-0.0097***	-0.0103***	-0.0104***
Popularity	-0.0011***	-0.0011***	-0.0011***
Reputation	0.0152	0.0288	0.0375
Rating	0.2447***	0.2288***	0.1978***
Readability		-0.0473***	-0.0456***
Sentiment		0.1613***	0.1740***
Theta	1.2261	1.2344	1.2343
Log-likelihood	-264,900	-264,700	-264,600

predictors for getting more useful votes while rating, the moderating variable, is insignificant. While model 2 shows all variables show to be significant, including the independent variables. On Model 3, we include moderating effects of rating to the model. Rating*Readability is not significant, which means that the readability is not moderated by review rating. Focusing on our dependent variable, usefulness votes and sentiment have a negative relationship in the count model results. This means that positive sentiments, since our measure for sentiment is based on positivity, have a negative effect on gaining usefulness votes on the website. For Rating*Sentiment on the count model, the result show that rating has a positive moderating effect on sentiment. The Zero-inflation

Model results show which variables affect a review's tendency to get a usefulness vote from zero. As shown on Table 4, Reputation is not significant in all zero-inflation models. While Readability has negative significant effect on usefulness votes on all count and zero-inflation models. This means that higher readability of a review (higher level of thought needed to understand the review) lessens the chances of that review to get a useful vote.

The direction of the effects of sentiment on useful votes in <Table 4> is opposite as we hypothesized in H1. To investigate the effects in detail, we separated reviews with positive and negative ratings and analyzed again. <Table 5> shows the results of the Zero-inflated Negative

⟨Table 5⟩ Results of Zero-inflated Negative Binomial Regression

	Model II (All)	Model II (Rating 4,5)Count Model	Model II (Rating 1,2)
Length	0.0006***	0.0007***	0.0004***
Popularity	0.0001***	-0.00002	0.0012***
Reputation	0.0660***	0.1795***	-0.1114***
Rating	0.0063*	-0.0075	-0.0327*
Readability	-0.0289***	-0.0282***	-0.0291***
Sentiment	-0.0112***	-0.1761***	-0.0033
Log(theta)	0.2106***	0.1773***	0.3895***
Zero-Inflation Model			
Length	-0.0103***	-0.0110***	-0.0066***
Popularity	-0.0011***	-0.0015***	-0.0905*
Reputation	0.0288	0.4145***	0.1731
Rating	0.2288***	-0.3849***	-0.0376
Readability	-0.0473***	-0.0374*	-0.0169
Sentiment	0.1613***	0.1078**	0.1476
Theta	1.2344	1.194	1.4762
Log-likelihood	-264,900	-175,400	-45820

Binomial Regression on positive and negative reviews. To obtain the results, reviews with positive ratings (4, 5 stars) were separated from reviews with negative ratings (1, 2 stars). The Count Model for all reviews shows that all variables are significant. For positive reviews (Rating 4, 5), positive sentiment has a highly negative effect on getting more useful votes, while the harder it is to understand a review (readability), the less likely for people to vote for it to be useful. For negative reviews (Rating 1,2), sentiment is not significant, while readability has the same effect as above. Results show that for positive reviews (reviews with 4, 5 ratings), the more positive sentiments it contains, the less likely it is for readers to give it a useful vote. It is interesting to note that despite having a high rating for the review, readers might still look for more

objective opinion than a very positive judgment. For the Zero-inflation model, reputation is only significant for positive reviews. For positive reviews (Rating 4, 5), positive sentiment has a significant effect in gaining a useful vote, while reviews which are difficult to read have a negative effect in gaining a useful vote. For negative reviews, sentiment and readability are both insignificant.

For the results of the negative binomial regression for reviews which have at least one useful votes on <Table 6>, we assume that the dependent variable, usefulness votes, is over-dispersed and does not have an excessive number of zeros. This test only included those reviews with usefulness votes therefore reviews with zero votes were eliminated from the sample. The negative relationship between sentiment and

⟨Table 6⟩ Results of Negative Binomial Regression with reviews (useful votes 〉 0)

	Model I (All)	Model II (All)	Model III (All)
Length	0.0003***	0.0003***	0.0003***
Popularity	0.00004*	0.00005**	0.00005**
Reputation	0.0495***	0.05198***	0.0530***
Rating	0.0043	0.0121***	-0.0138
Readability		-0.0189***	-0.0288***
Sentiment		-0.0072***	-0.0213***
Rating* Readability			0.0002*
Rating* Sentiment			0.0003***
AIC	371,117	370,876	370,829
Log-likelihood	-185,052	-185,429	-185,404

⟨Table 7⟩ Results of Negative Binomial Regression with reviews (useful votes 〉 0)

	Model II (All)	Model II (Rating 4,5)	Model II (Rating 1,2)
Length	0.0003***	0.0004***	0.0002***
Popularity	0.00005**	-0.00005*	0.0007***
Reputation	0.05198***	0.1213***	-0.0518***
Rating	0.0121***	-0.0087	-0.0265*
Readability	-0.0189***	-0.0179***	-0.0204***
Sentiment	-0.0072***	-0.0108***	-0.0027
AIC	370,876	244,641	66,328
Log-likelihood	-185,429	-122,312	-33,156

usefulness votes show that positive sentiments have negative effect in gaining usefulness votes. The negative relationship between readability and usefulness votes show that reviews that are difficult to read have a negative effect on useful votes. Rating*Readability and Rating*Sentiment results show that rating has a moderating effect on readability and sentiment for reviews with at least 1 useful vote.

For <Table 7>, results of the negative binomial regression of reviews which have at least one useful votes show the effect of rating (positive and negative reviews). To obtain the results, reviews with positive ratings (4, 5 stars) were separated from reviews with negative ratings (1, 2 stars). For positive reviews (Rating 4, 5), readability and sentiment are both highly negatively significant. For positive reviews, positive sentiment has a negative effect in gaining usefulness votes. So positive feedback on restaurants are not necessarily useful. For negative reviews, sentiment is not significant while readability has the same effect as in positive reviews.

5. Discussions & Conclusions

In this paper, we empirically examine the effects of sentiment and readability on count type usefulness index for customer reviews. This subject matter is important for researchers and businesses alike because the quality of a review as measured by helpfulness votes was also found to influence sales (Chen et al. 2008). To the best of our knowledge, this is the first attempt to investigate customer reviews' usefulness using a count-type index as affected by the reviews' sentiment and readability moderated by review ratings. Our study adds to the literature by providing empirical evidence on how customers' rating on reviews as useful that utilizes count-type usefulness index depending on the positivity of the sentiment and readability.

Prior literature focuses mainly on using "physical" attributes of a review like review length to measure its usefulness and the impact of online reviews on product sales. Moreover, the website used in this study is a restaurant review website, a different category compared to existing studies that use product reviews and online shopping websites such as Amazon. Going a step further from previous studies, this paper investigates the nature of sentiment and readability that influence customers to rate an online review as useful. More importantly, our analysis also controlled review length, reputation, and popularity. Our results suggest that positive sentiment on a review about a specific restaurant or store does not necessarily result to a useful review based on customers' usefulness votes.

This paper utilized multiple tests in identifying the effects of review sentiment and readability to usefulness votes. As presented in the results, the review sentiment and readability have very consistent effects on gaining usefulness votes for reviews in both zero-inflated negative binomial regression and negative binomial regression tests.

Similar to other studies on readability and helpfulness using Amazon.com (Ghose and Ipeirotis, 2011; Korfiatis et al., 2012; Chen et al., 2008; Chevalier and Mayzlin, 2006), results have been consistent regarding predicting power of readability towards the usefulness of a review. To explain it simply, a review that is easier to comprehend and read has more value to a potential consumer in evaluating a product/ restaurant. Aside from the measure used in this paper, other measures have been used to assess readability, including the Automated Readability Index (ARI), Flesch Reading Ease (FRE), FleschKincaid Grade Level, and the Gunning fog index. All these measures, including the Coleman-Liau Index, have their limitations that have been addressed by developments in the computational linguistics domain. This paper showed that similar to results of papers that investigate the ratio-type usefulness index of Amazon.com, the count-type usefulness index of Yelp.com, despite the differences in obtaining the usefulness score, still yield the same results for readability and usefulness relationship.

Studies regarding sentiment, on the other hand, have mixed results. While existing studies have shown the positive influence of reviews to sales (Chen et al., 2008; Chevalier and Mayzlin, 2006; Clemons et al., 2006; Ghose and Ipeirotis, 2006)

and despite the results of Clemons et. al (2006) that strongly positive ratings positively impact the growth of sales, Baek et al. (2012-13) found that helpfulness ratings increased when product reviews contained a higher percentage of negative words. Specific, as opposed to general reviews, have also been shown to be positively related to online product review usefulness. Results of this study found that positive sentiment on positive reviews have negative effect on gaining usefulness votes. This may be due to very positive reviews being doubtful and suspicious for customers. Moreover, positive reviews (those with 4, 5 ratings) might be better for customers if reviewers provide more detailed justifications for their numeric rating.

Possible extensions for this study is to delve deeper into which type of words are contained on useful reviews. Moreover, positivity, negativity, and other aspects of sentiment can also be quantified and investigated as to how they affect the usefulness of a customer review. Because this paper only uses positivity as a measure of sentiment, this aspect can be expanded and investigated further. To add to that, topics of useful reviews can also be studied to explain why certain reviews get the useful vote. Moreover, the relationship of variables to each other can also be studied to see which combination of significant variables result to a useful review. Reviewer rating and reward systems can also affect the quality of a review and therefore behavior of online reviewers with considerable online reputation can be investigated as to whether their reviews result to a useful vote.

This study yields considerable managerial

implications. Results of this study is helpful for online websites in analyzing their review guidelines and to identify useful reviews for their business. Based on this study, positive reviews are not necessarily helpful and therefore restaurants should take into account which type of positive review is helpful for their business. Second, this study is beneficial for businesses and website designers in creating their review mechanisms to know which type of reviews to highlight on their websites and which type of reviews can be beneficial to the business. Moreover, this review puts the spotlight on the review systems employed by websites to allow their customers in rating reviews. Yelp.com only consider usefulness votes as compared to other websites who employ usefulness votes and not-useful votes to rate reviews. Third, results of this study will also help people who maintain an online reputation on websites. A lot of people currently benefit from writing reviews about products and restaurants. Because this paper shows that positive sentiment is not exactly recognized as useful by customers, online reviewers can identify their strategies in making a review and striking a balance between positive traits and negative traits to write about. More than that, positive reviews that give more details might be useful then generic ones. Also, striking a good balance between length and readability will also be helpful in writing reviews. Giving a good rating also showed to be important but have a negative effect for lower ratings.

Our use of reviews for beverage and food stores represents one limitation of this study. As stated in Data & Methodology Section, we only used the

reviews in those categories. Yelp.com is popular for discussing businesses in those categories. The results in this study may be different with reviews for other businesses or products. It should be careful to apply the findings in this study for other business categories or product reviews. Future research could therefore extend data sets into other categories and compare the effects among the categories. Another limitation is that this study investigated the effects of sentiment and readability on count type usefulness index for customer reviews. Analysis results with data of ratio type usefulness index may be different from the findings in this study.

References

Baccianella, S., A. Esuli, and F. Sehastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, (2010), 2200~2204.

Baek, H., J. Ahn, and Y. Choi, "Helpfulness of online consumer reviews: Readers' objectives and review cues," *International Journal of Electronic Commerce*, Vol.17, No.2(2012–13), 99~126.

Baum, S. and M. Spann, "The Interplay Between Online Consumer Reviews and Recommender Systems: An Experimental Analysis," *International Journal of Electronic Commerce*, Vol.19, No.1(2014), 129~162.

Bickart, B. and R. M. Schindler, "Internet forums as influential sources of consumer information," *Journal of Interactive Marketing,*

Vol.15, No.3(2001), 31~40.

Chae, S. H., J. Im, and J. Y. Kang, "A Comparative Analysis of Social Commerce and Open Market Using User Reviews in Korean Mobile Commerce," *Journal of Intelligence and Information Systems*, Vol. 21, No.4(2015), 53~77.

Chen, P., S. Dhanasobhon, and M. Smith, *All Reviews Are Not Created Equal: The Disaggregate Impact of Reviews on Sales on Amaon.com*, Working paper, Carnegie Mellon University, 2008, Available at SSRN: http://ssrn.com/abstract=918083 (Downloaded 1 November, 2015).

Chevalier, J. and D. Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research,* Vol.43, No.3(2006), 345~354.

Choi, J. W. and H. J. Lee, "The Effects of Customer Product Review on Social Presence in Personalized Recommender Systems," *Journal of Intelligence and Information Systems*, Vol.17, No.3(2011), 115~130.

Chun, B. G. and H. C. Ahn, "A Collaborative Filtering System Combined with Users' Review Mining : Application to the Recommendation of Smartphone Apps," *Journal of Intelligence and Information Systems*, Vol.21, No.2(2015), 1~18.

Clemons, E., G. Gao, and L. Hitt, "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems,* Vol.23, No.2(2006), 149~171.

Crowley, A. E. and W. D. Hoyer, "An Integrative Framework for Understanding Two-Sided Persuasion," *Journal of Consumer Research,* Vol.20, No.4(1994), 561~574.

Dabholkar, P., "Factors Influencing Consumer

Choice of a 'Rating Web Site': An Experimental Investigation of an Online Interactive Decision Aid," *Journal of Marketing Theory and Practice* Vol.14, No.4(2006), 259~273.

Forman, C., A. Ghose, and B. Wiesenfeld, "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research,* Vol.19, No.3(2008), 291~313.

Ghose, A. and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE Transactions on Knowledge and Data Engineering*, Vol.23, No.10(2011), 1498~1512.

Ghose, A. and P. G. Ipeirotis, "Designing Ranking Systems for Consumer Reviews: The Impact of Review Subjectivity on Product Sales and Review Quality," *Proceedings of the 16th Annual Workshop on Information Technology and Systems*, (2006).

Harris, R. B. and D. Paradice, "An investigation of the computer-mediated communication of emotions," *Journal of Applied Sciences Research,* Vol.3, No.12(2007), 2081~2090.

Hausman, J., B. H. Hall, and Z. Griliches, "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, Vol.52, No.4(1984), 909~937.

Jiang, Z. and I. Benbasat, "Investigating the influence of the functional mechanisms of online product presentations," *Information Systems Research*, Vol.18, No.4(2007), 454~470.

Jiang, B. J. and K. Srinivasan, "Pricing and Persuasive Advertising in a Differentiated Market," *Marketing Letters*, forthcoming

(2015), 1~10.

Korfiatis, N., E. García-Bariocanal, and S. Sánchez-Alonso, "Evaluating content quality and helpfulness of online product reviews: the interplay of review helpfulness vs. review content," *Electronic Commerce Research and Applications*, Vol.11, No.3(2012), 205~217.

Krosnick, J. A., D. S. Boninger, Y. C. Chuang, M. K. Berent, and C. G. Camot, "Attitude Strength: One Construct or Many Related Constructs?," *Journal of Personality and Social Psychology,* Vol.65, No.6(1993), 1132~1151.

Kumar, N. and I. Benbasat, "The Influence of Recommendations and Consumer Reviews on Evaluations of Websites," *Information Systems Research,* Vol.17, No.4(2006), 425~439.

Michalke, M., "koRpus – ein R-paket zur textanalyse," *Tagung experimentell arbeitender Psychologen(TeaP)*, (2012).

Mudambi, S. M. and D. Schuff, "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com," *MIS Quarterly,* Vol.34, No.1(2010), 185~200.

Park, D. H. and S. Kim, "The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews," *Electronic Commerce Research and Applications*, Vol.7, No.4(2009), 399~410.

Park, D. H., J. Lee, and I. Han, "The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement," *International Journal of Electronic Commerce,* Vol.11*,* No.4(2007), 125~148.

Pavlou, P. and A. Dimoka, "The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation,"

Information Systems Research, Vol.17, No.4(2006), 392~414.

Schlosser, A., "Source Perceptions and the Persuasiveness of Internet Word-of-Mouth Communication," in *Advances in Consumer Research* (32), G. Menon and A. Rao (eds.), Duluth, MN: Association for Consumer Research, 2005, 202~203.

Shen, W., J. H. Yu, and J. Rees, "Competing for Attention: An Empirical Study of Online Reviewers' Strategic Behaviors," *MIS Quarterly,* Vol.39, No.3(2015), 683~696.

Resnick, P., R. Zeckhauser, E. Friedman, and K. Kuwabara, "Reputation Systems," *Communications of the ACM,* Vol.43, No.12(2000), 45~48.

Riordan, M. A. and R. J. Kreuz, "Emotion encoding and interpretation in computer-mediated communication: Reasons for use,"

Computers in Human Behavior, Vol.26, No.6(2010), 1667~1673.

Schindler, R. M. and B. Bickart, "Published word-of-mouth: Referable, consumer-generated information on the Internet," In C.P. Haugtvedt, K.A. Machleit, and R.F. Yalch (eds.), *Online Consumer Psychology: Understanding and Influencing Behavior in the Virtual World,* Hillsdale, NJ: Lawrence Erlbaum, (2005), 35~61.

Schindler, R. and B. Bickart, "Perceived helpfulness of online consumer reviews: The role of message content and style," *Journal of Consumer Behaviour*, Vol.11, No.3(2012), 234~243.

Treisman, A., "Selective attention in man," *British Medical Bulletin*, Vol.20, No.1(1964), 12~16.

국문요약

온라인 리뷰의 감성과 독해 용이성이 리뷰 유용성에 미치는 영향: 가산형 리뷰 유용성 정보 활용

루스 안젤리 크루즈* · 이홍주**

온라인 쇼핑몰의 상품에 대한 고객 리뷰는 구매자들의 구매 의사결정에 영향을 미치고 있으며 중요한 구전효과의 원천과 의사결정의 정보 원천의 역할을 하고 있다. 한 제품에 대한 리뷰가 무척 많기에 온라인 쇼핑몰들은 고객 리뷰 평가 방안을 도입하였고, 이를 통해 고객들에게 유용하리라고 판단되는 리뷰들을 걸러서 보여주거나 강조할 수 있게 되었다. 리뷰 평가 방안은 해당 리뷰가 도움이 되었는지 혹은 도움이 되지 않았는 지를 리뷰를 읽은 고객이 평가하게 하는 방안이다. Amazon.com은 고객 평가를 바탕으로 총 투표 수 중에서 유용하다는 투표 수의 비율을 리뷰 유용성 지표로 삼고 있으며, Yelp.com은 유용하다는 투표 수 자체를 유용성 지표로 삼고 있다.

본 연구는 고객 리뷰의 감성과 독해 용이성이 리뷰의 유용성에 미치는 영향을 파악하고자 한다. Amazon.com의 고객 리뷰 자료를 활용하여 비율형 유용성 지표를 종속변수로 하는 유사한 연구들이 수행되어 왔다. 본 연구에서는 Yelp.com의 리뷰 자료를 활용하여 가산형 리뷰 유용성 지표인 경우에도 동일한 효과가 존재하는지를 검토하고자 한다.

Yelp.com의 음료와 음식 카테고리에 해당하는 업종에 대한 리뷰를 자료로 활용하였으며, 점포의 명성과 인기도 데이터를 파악할 수 있는 170,294개의 리뷰를 분석에 활용하였다. 분석결과는 리뷰의 긍정 정도는 유용 투표수를 늘리는데 음의 영향을 미쳤다. 평가가 긍정적인 리뷰에서는 음의 영향관계가 유의 하였으나, 평가가 부정적인 리뷰에서는 리뷰의 긍정 정도가 유용 투표 수에 미치는 영향은 유의 하지 않았다. 독해 용이성은 리뷰가 읽기 어려울 수록 높은 값을 갖으며, 독해의 어려운 정도는 유용 투표수 획득에 음의 영향을 미쳤다. 독해 용이성은 긍정 리뷰, 부정 리뷰 관계없이 모두 음의 영향을 미치는 것으로 분석되었다. 이 결과는 유용 투표수가 0인 리뷰를 포함하여 영과잉 음이항 회귀분석을 수행한 경우와 유용 투표수가 0인 리뷰를 제외하고 음이항 회귀분석을 수행한 경우 모두 동일하게 파악되었다.

 * 가톨릭대학교 경영학부
** 교신저자: 이홍주
   가톨릭대학교 경영학부
   경기도 부천시 원미구 지봉로 43, 14662
   Tel: 02-2164-4009, Fax: 02-2164-4280, E-mail: hongjoo@catholic.ac.kr

저 자 소 개

Ruth Angelie B. Cruz

Ms. Cruz is currently a PhD Candidate for Business Administration at the Catholic University of Korea. She earned her MBA majoring in Management Information Systems from the same university in 2014. Her professional experience in advertising communications lead her to pursue research in social media, e-business, and applications of information systems in business.

이 홍 주

현재 가톨릭대학교 경영학전공 교수로 재직 중이다. KAIST 산업경영학과를 졸업하고 KAIST 테크노경영대학원에서 석사 및 박사학위를 취득하였다. 주요 관심분야는 데이터 분석, 지능형 정보시스템, 온라인 사용자들의 상호작용등이다.