Predicting the Performance of Online Consumer Reviews: A Sentiment Mining Approach

Completed Research Paper

Mohammad Salehan
University of North Texas
1155 Union Circle #311160
Denton, TX 76203-5017
mohammad.salehan@unt.edu

Dan Kim
University of North Texas
1155 Union Circle #311160
Denton, TX 76203-5017
dan.kim@unt.edu

Abstract

Online consumer reviews (OCR) have helped consumers to know about the strengths and weaknesses of different products and find the ones that best suit their needs. This research investigates the predictors of readership and helpfulness of OCR using a sentiment mining approach. Our findings show that reviews with higher levels of positive sentiment in the title receive more readerships. Sentimental reviews with neutral polarity in the text are also perceived to be more helpful. The length and longevity of a review positively influence both its readership and helpfulness. Our findings suggest that the current methods used for sorting OCR may bias both their readership and helpfulness. This study can be used by online vendors to develop scalable automated systems for sorting and classification of OCR which will benefit both vendors and consumers.

Keywords: Online Consumer Reviews, Sentiment Mining, Helpfulness, Readership

Introduction

Businesses are now using social media to promote their products and services. Many companies maintain Facebook and Twitter accounts to keep in touch with their customers. Customers also use social media to receive information about products/services. In many ways the Internet in general and social media in particular, have changed the way customers shop for goods and services. It's now quite normal for people to enter brick-and-mortar stores, find the product they want, and then order it online. Moreover, online consumer reviews (OCR) have helped customers to learn about the strengths and weaknesses of different products and to find the ones that best suit their needs. Some studies suggest that customers show more interest toward user-generated product information on the Internet than the vendor information (Bickart and Schindler 2001). A recent study shows that OCR are the second most-trusted source of product information after recommendations from family and friends (Nielsen 2012). Compared to vendor-generated product descriptions, OCR are more user-oriented and describe the product in terms of different usage scenarios and assess it from the user's perspective (Chen and Xie 2008). Thus, it has even been suggested that consumers who write OCR serve as "sales assistants" for online retailers (Chen and Xie 2008).

The process of analyzing OCR can be broken into two steps: the decision to read the review, and the actual processing of the information in the review that deems the decision to use it based on perceived helpfulness of the review (Ahluwalia 2000). Focusing on the two steps of processing OCR, this study proposes two fundamental research questions in the context of OCR:

RQ1: Which factors determine the likelihood of a user paying attention to a review?

RQ2: Which factors determine the perceived helpfulness of a review?

The first research question looks into the characteristics of OCR that absorb the attention of online consumers. Many products receive too many reviews that make it difficult for consumers to read all of them. Thus, most consumers have to read them selectively. In summary, the first research question explores the determinants of the readership of OCR. Although reading a review is the first step in determining its helpfulness, previous research overlooks the readership of OCR.

While the first research question remains largely unexplored, the second one has received some attention. Previous research has found strong evidence that OCR influence product sales (Chevalier and Mayzlin 2006; Dellarocas, Zhang and Awad 2007; Duan, Gu and Whinston 2008; Forman, Ghose and Wiesenfeld 2008; Ghose and Ipeirotis 2011; Godes and Mayzlin 2004; Liu 2006; Mudambi and Schuff 2010). However, three important aspects of the second research question require further investigation. First, most studies only use the numerical review ratings (e.g., the number of stars) and the length of the reviews in their empirical analysis, without formally incorporating the information contained in the text of the reviews. Therefore, a deeper analysis of the textual information contained in the OCR can provide greater insight into what constitutes a useful online review (Mudambi and Schuff 2010).

Second, previous research shows that positive statements are considered to be more helpful by consumers (Schindler and Bickart 2012). However, research has produced conflicting results regarding the negative comments. For example, Sen and Lerman (2007) found that in the case of utilitarian products, negative comments are more useful than positive ones, whereas Schindler and Bickart (2012) failed to find any significant relationship between negative statements and helpfulness of a review. By investigating the effect of sentiment polarity (i.e., negative, positive, or neutral) on helpfulness, this study provides insights into the performance of OCR.

Finally, previous research tends to identify the predictors of the performance of OCR without providing many practical solutions for online vendors. Human subjects are largely employed for categorization of OCR based on the textual information contained in them. Although this method provides a descriptive view of the performance of OCR, it does not facilitate the development of scalable automated systems for classification of OCR.

This study investigates the performance of OCR through analysis of textual information contained in the reviews. Previous research has indicated that computer-mediated communications (CMC) can effectively transfer emotions. Moreover, the emotions contained in a message transferred through CMC significantly influence how the message is processed and interpreted by the receiver (Berger and Milkman 2012; Riordan and Kreuz 2010; Walther and D'Addario 2001). Sentiment mining can be used for emotional analysis of textual information. Sentiment mining refers to the use of natural language processing and computational linguistics to find and extract subjective information from text data. Sentiment mining is usually done using automated tools that provides benefits such as scalability, effective information retrieval, automated cyber risk management, and increased business profits (Bai 2011). It also facilitates processing of large amounts of data. Social media is an important source of big data and is quite suitable for text mining purposes (Russell 2013).

This study contributes to the existing body of knowledge in three unique ways. First, it provides a research model that predicts the performance of OCR in terms of readership and helpfulness of reviews. Second, it uses automated tools to analyze a set of secondary OCR data composed of reviews collected from Amazon.com website. Finally, it provides insights to online business managers regarding the design and implementation of scalable automated systems to improve classification and sorting of OCR, which will eventually help them achieve increased sales.

Theoretical Background

Online review performance measures

Different measures have been used by previous studies to evaluate the performance of OCR. Most studies use helpfulness as the single performance measure of OCR (Aral 2013; Mudambi and Schuff 2010; Sen and Lerman 2007). Helpfulness has also been referred to as the value of the review (Schindler and Bickart 2012). For studies that use secondary data, helpfulness is measured by dividing the number of people who find a review helpful by the total number of people who voted for that review (Mudambi and Schuff 2010; Sen and Lerman 2007).

Purchase intention is another measure of performance. Customer purchase intention is influenced by both quantity and quality of the reviews (Park, Lee and Han 2007). Some studies use sales revenue as a performance measure for online reviews. One study used online reviews from Yahoo Movies website to predict box office revenues and found strong evidence that online reviews influence movie sales (Liu 2006). Online reviews have also been used to predict the online sales of books (Chevalier and Mayzlin 2006). Product ratings also indirectly influence sales through sentiment (Hu, Koh and Reddy 2013).

Predictors of online review performance

Different measures have been used to predict the performance of OCR. Some studies have focused on the numeric star rating and word count of the reviews to predict their performance. For example, extreme numerical ratings are positively related to sales of books (Chevalier and Mayzlin 2006). Reviews with extreme numerical ratings are also considered more helpful (Mudambi and Schuff 2010). Length of a review may also predict its performance. Length of reviews for a book significantly predict its sales on Amazon.com (Chevalier and Mayzlin 2006). Length of a review is also positively related to its helpfulness (Mudambi and Schuff 2010; Schindler and Bickart 2012). Empirical analysis, however, fails to find any significant relationship between review length and sales of books on Barnes & Noble website (Chevalier and Mayzlin 2006).

The difference between the performance of positive and negative reviews is controversial research avenue in the context of OCR. Drawing upon the negativity bias theory, some studies propose that negative reviews are considered more helpful than positive ones. According to the negativity bias theory, people face difficulty in making inferences about the actions of an actor when the actor behaves in an expected fashion. The inference is easier when the actor departs from the norms of behavior (Jones and Davis 1965; Kelley 1973). Thus, some researchers have argued that negative comments should be considered more helpful than positive ones because they deviate from the accepted norm of staying positive (Sen and Lerman 2007). Several studies test the negativity bias theory in the context of OCR. Their findings, however, are contradictory. While some studies show that consumers find negative reviews more valuable, others find no significant difference between positive and negative reviews. Some others even find that positive reviews are more helpful than negative ones. Sen and Lerman (2007) find that positive reviews are generally more helpful than negative ones. However, negative reviews are more useful for utilitarian products (products that focus on task performance) and positive comments are more useful for hedonic products (products that deal with pleasure). Schindler and Bickart (2012) find that the number of positive statements is a significant predictor of the value of a review while the effect of the number of negative statements on value is not significant. A recent study suggests that consumer reviews may be subject to positive social influence bias (Aral 2013). Through an experiment, the author manipulated the helpfulness of OCR and found out that positive manipulations create a positive social influence that last several months. The negative manipulations, however, are offset by a correction effect that neutralizes the manipulation. The findings suggest that corporations can easily manipulate OCR and the reviews should be considered with some level of skepticism.

Recent studies use text of a review to predict its performance. Schindler and Bickart (2012) look at the wording of online reviews rather than their source. They divide wording factor into two categories: content and style. They define the content as the information the review provides. Style, by contrast, is defined as the choice of words used to convey the information. They find that proportion of product-descriptive statements and proportion of reviewer-descriptive statements are significant predictors of the value of the review. They also find that while use of negative style characteristics decrease the value of the review, use of positive style doesn't improve its performance. Finally, they didn't find any significant relationship between negative evaluative statements and value of the review.

Review sentiment is another measure used by previous studies to predict the performance of computer-mediated communications (CMC) in general, and OCR in particular. Previous research has indicated that CMC can effectively transfer emotions. The receiver of a message can detect the sender's emotions through verbal cues such as emotion words as well as nonverbal cues such as emoticons (Harris and Paradice 2007). Moreover, the emotions contained in a message transferred through CMC significantly influence how the message is processed and interpreted by the receiver (Riordan and Kreuz 2010; Walther and D'Addario 2001). Different methods can be used to extract sentiment out of text. While some studies use human subjects to extract the sentiment of OCR, others use automated sentiment mining to extract sentiment from the text of reviews (Bai 2011; Schindler and Bickart 2012; Sen and Lerman 2007).

Moderators of Performance

Product type is an important factor in the context of online reviews which has mostly been viewed as the variable moderating the relationship between independent variables and helpfulness. For example, reviews on hedonic products are less likely to be perceived as helpful (Sen and Lerman 2007). There are also differences between helpfulness of search and experience goods. For search goods, consumers can obtain information about product quality prior to purchase while evaluation of experience goods requires sampling or purchase. While reviews with extremely high or low star rating are considered more helpful for search goods, the effect is different for experience goods. For experience goods, extreme reviews are considered less helpful than the ones with moderate rating (Mudambi and Schuff 2010). Product type also moderates the relationship between the number of positive and negative statements with the helpfulness of a review (Sen and Lerman 2007). The length of a review has greater positive impact on its helpfulness for search goods compared to experience goods (Mudambi and Schuff 2010).

Theories related to message selection

The theory of selective attention posits that people respond to messages selectively because of limited information processing capacity (Treisman 1969). When people receive multiple stimuli at once, they need to filter some of the messages because they have limited processing capacity. Same thing happens in the context of online reviews. Many products have large number of reviews which makes people to selectively read online reviews. The theory of selective perception helps explain the mechanism underlying selective attention. According to selective perception, people develop belief structures which are a simplified representation of the world and use these structures to filter and interpret information (Walsh 1988).

Attribution theory explains how people analyze different behaviors. Based on the theory, people identify two types of explanation for different behaviors: actions related to internal (personal) factors and the ones that result from external environment (situational) (Folkes 1988; Heider 1958). The theory largely explains the predictors of source credibility and other areas related to consumer perceptions and inference formation (Dholakia and Sternthal 1977). It has also been used to explain the perceptions of the consumers regarding the helpfulness of OCR. Readers consider the motivation of the author of a review when deciding about whether or not to use the information contained in the review. Readers may attribute a review to either external (product-related) or internal (reviewer-related) reasons. If the reader feels that the review is based on external reasons, they are more likely to accept it. If reader believes that the review is based on internal reasons, they are more likely to disregard it (Sen and Lerman 2007).

Research Model and Hypothesis

Following the directions provided by Mudambi and Schuff (2010) regarding the need for the analysis of the textual information contained in OCR, we utilize sentiment mining to analyze the text of the reviews and to investigate our two research questions. Each Amazon review has a title and a body text. We extract the sentiment of both the title and the text of reviews and use them to predict performance. We have two measures for OCR performance, readership and helpfulness, each of which corresponds to one of our research questions. Figure 1 shows the proposed research model. In our model, sentiment refers to the total amount of sentiment that exists in a text, both positive and negative. Polarity refers to the direction of the sentiment which can be positive, negative, or neutral. More information on our measures can be found in the methodology section.

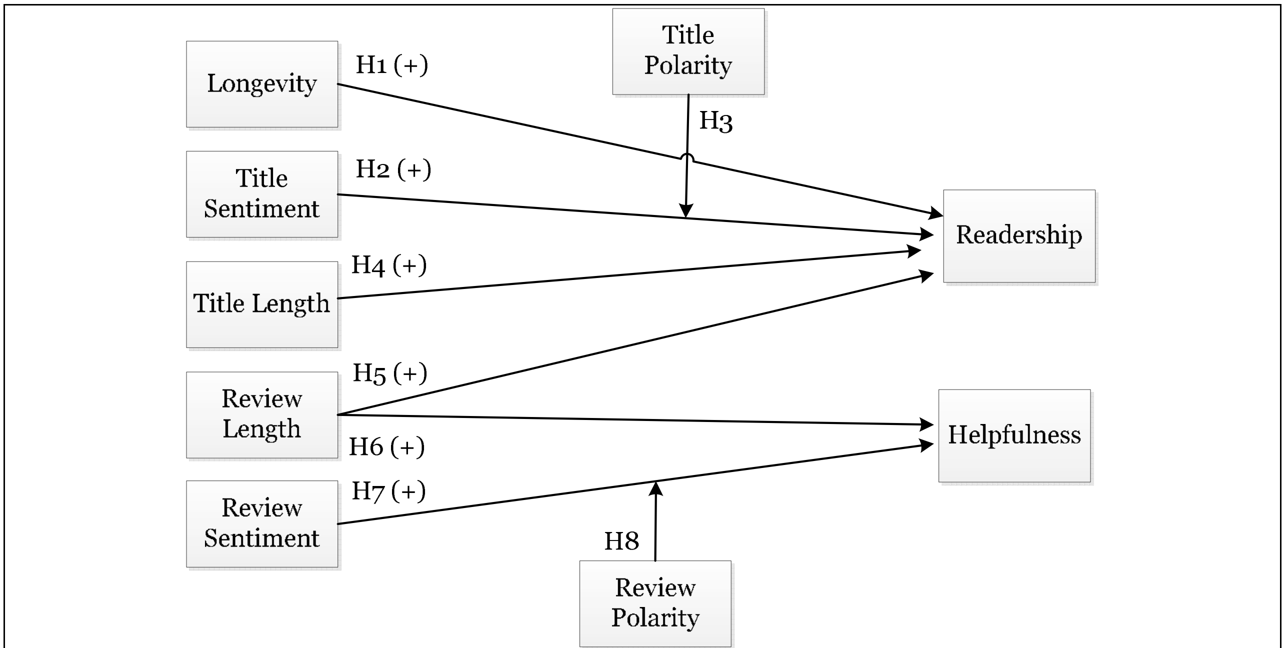

Figure 1-Research Model

We expect longevity of a review to be positively related to its readership. Older reviews have been on the website for a longer time and hence have higher chance of being viewed and read by consumers. Thus, we can expect older reviews to receive more readership than newer ones. Consequently, we hypothesize the following:

H1: Longevity of a review has a positive effect on the readership of the review.

According to the theory of selective attention, people respond to messages selectively because individuals possess limited information processing capacity (Treisman 1969). We believe that OCR are no exception to this theory. Many products / services have large number of reviews and reading all of them is very difficult and time consuming for consumers. Hence, people have to pay selective attention toward OCR. We expect people to look for quick signals that enable them to decide about reading a review. Title is a small but important part of a review and provides users with instant information about the general theme of the review. Thus, we expect the title of a review to be an important predictor of its readership. Moreover, sentiment of a message can be effectively communicated through the text and significantly influences the perceptions of the reader (Harris and Paradice 2007; Riordan and Kreuz 2010; Walther and D'Addario 2001). Finally, affective language in the online environment receives more attention and feedback compared to neutral language (Huffaker 2010). Hence, we expect the total sentiment contained in the title of a review to be positively related to its readership. Consequently, we hypothesize that:

H2: The larger the total amount of sentiment (positive or negative) the title of a review exhibits, the more readership it receives.

According to the theory of selective perception, people use their mental structure to filter information. People tend to absorb information that matches their mental model and filter out the information that don't fit their mental structure (Walsh 1988). Similarly, we expect people to pay selective attention toward OCR based on their initial attitude about the product. Those with a positive attitude toward the product are expected to look for positive information and those who are more suspicious are expected to look for negative information. Many consumers read OCR to receive reassurance that they have made a good choice (Bailey 2005; Goldsmith and Horowitz 2006; Hennig-Thurau, Gwinner, Walsh and Gremler 2004). Reassurance will be gained through positive information rather than negative information. Consequently, we expect a larger number of consumers look for positive reviews rather than negative ones. Thus, we expect reviews with positive titles attract more attention and readership. Considering the above arguments, we expect positive sentiment to have a larger effect on the readership of online reviews than negative sentiment. This leads to the following hypothesis:

H3: Title polarity moderates the effect of title sentiment on the readership of the review. The effect will be larger for positive titles than negative and neutral ones.

While the total amount of sentiment in the title of a review is an important source of information for consumers, length of the title can also be an important factor related to its readership. Short titles are usually not very informative and do not contain much information. A short title usually communicates the general idea of the author about the product such as “I love it” or “not very high quality.” In contrast, longer titles will give the reader more information about the content of the review which subsequently increases the likelihood of it being read by the consumers. For example, a review entitled “the best LED-lit television at a budget price” signals the reader that it contains information about both quality and pricing of the product which can motivate more people to read it. Moreover, length of the title may decrease consumer’s search costs through increased information diagnosticity (Johnson and Payne 1985). Hence, we expect length of the title of a review to be related to its readership performance. Consequently, we hypothesize that:

H4: The length of the title of a review has a positive effect on the readership of the review.

Length of a review is an important predictor of its performance (Mudambi and Schuff 2010; Schindler and Bickart 2012). Short reviews are more likely to be shallow and lack the comprehensive evaluation of product features. Longer reviews, in contrast, contain more information and are more likely to contain deep analysis of the product, its features, and the context in which it was used. Longer reviews are more likely to receive attention from users. Reading longer reviews may decrease consumer’s search costs through increased information diagnosticity (Johnson and Payne 1985). Moreover, longer reviews are more likely to be perceived helpful. An individual’s argument is more persuasive when it provides larger amount of information (Schwenk 1986). Increased number of reasons for a choice escalates the decision maker’s confidence (Tversky and Kahneman 1974). Hence, we expect longer reviews to receive more readership as well as more perceptions of helpfulness compared to the shorter ones. As a result, we suggest the following hypotheses:

H5: Length of a review has a positive effect on the readership of the review.

H6: Length of a review has a positive effect on the helpfulness of the review.

People tend to find reviews with extreme numerical ratings more helpful (Mudambi and Schuff 2010). We can expect that reviews with extreme ratings also contain more sentiment because the author is either very satisfied or very unsatisfied. The extreme levels of satisfaction or dissatisfaction are very likely to turn into strong emotions and consequently strong sentiment. The sentiment of a message can be effectively communicated through the text and significantly influences the perceptions of the reader (Harris and Paradice 2007; Riordan and Kreuz 2010; Walther and D’Addario 2001). One can argue that the sentiment contained in the review is the driver of the perceptions regarding its helpfulness rather than just the numerical rating. Different people have different experiences with the same product. Sentiment is the vehicle for people to convey their emotions to others through text. Thus, sentimental reviews are better conveyer of the experience with the product. Consequently, we can expect high-sentiment reviews to be perceived more helpful by the consumers because they are more likely to convey the experience with the product. Thus, we hypothesize that:

H7: The larger the total amount of sentiment (positive or negative) the text of a review exhibits, the more helpful it is perceived to be.

The decision regarding the helpfulness of a review is made after the person reads the review (Ahluwalia 2000). Reviews with larger number of descriptive statements are considered more helpful (Schindler and Bickart 2012). One can argue that a good description of an object contains different aspect of it including both positive and negative qualities. This implies that a good review should also contain both positive and negative statements and consequently both positive and negative sentiment. Moreover, 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 (Schindler and Bickart 2012). In terms of review polarity, this will translate into leaning toward neutral polarity. Neutral polarity does not imply that sentiment is non-existent in the text. It implies that there are balanced levels of positive and negative sentiment. Hence, we expect sentimental reviews with neutral polarity to be perceived more helpful than positive and negative ones. This leads to the following hypothesis:

H8: Review polarity moderates the effect of review sentiment on the helpfulness of the review. The effect will be larger for neutral reviews than for positive and negative ones.

Methodology

Measurement

Sentiment mining was done using SentiStrength software (Thelwall, Buckley, Paltoglou, Cai and Kappas 2010)e. The software is free for academic research and has been tested and validated in previous research (Garcia and Schweitzer 2011; Gruzd, Doiron and Mai 2011; Stieglitz and Dang-Xuan 2013; Thelwall and Buckley 2013; Thelwall, Buckley and Paltoglou 2012; Thelwall et al. 2010). SentiStrength is capable of processing different types of information contained in the text including analysis of emoticons and booster words, correction of spelling due to repeated letters, and use of negative words (e.g., not) to flip emotions.

SentiStrength reports two separate numbers for positive and negative emotions. The positive number ranges from 1 (not positive) to 5 (extremely positive). The negative number ranges from -1 (not negative) to -5 (extremely negative). Because both numbers should be considered when evaluating the sentiment of a statement, we use the approach used by Stieglitz and Dang-Xuan (2013) to combine the two numbers. We calculate the sentiment polarity of each statement by creating the following measure, which determines the direction of the sentiment as well as its strength:

$$\text{polarity} = \text{positive sentiment} + \text{negative sentiment},$$

Because positive sentiments range from 1 to 5 and negative ranges from-1 to -5, polarity will have a range of -4 to 4. The other approach to combine the positive and negative numbers is to calculate the total amount of sentiment in a statement regardless of the polarity of positive and negative statements. To attain this, the absolute value of positive and negative sentiments should be added up using the following formula:

$$\text{sentiment} = (\text{positive sentiment} - \text{negative sentiment}) - 2$$

Positive sentiment ranges from 1 to 5 and negative sentiment ranges from-1 to -5. Thus, total sentiment has a range of 2 to 10. Therefore we subtracted 2 from (positive – negative) to normalize the range from [2, 10] to [0, 8].

Longevity was measured by counting the number of the days since the review was created. Length of title and review were measured by counting the number of words in the text. Helpfulness of a review was measured as the ratio of “helpful votes” to “total votes” cast by readers in each review (Forman et al. 2008; Ghose and Ipeirotis 2011; Mudambi and Schuff 2010). Because we cannot directly measure the readership of OCR, we use the total number of votes of a review as a measure for its readership. People usually vote for a review after they read it. Hence, total votes can be a good estimate of readership.

Data Collection

A group of 35000 online reviews of 20 different products were collected from Amazon.com website using the crawler software developed by the authors. We selected products that had at least 100 reviews. The examined product types were mobile phones, TVs, laptops, tablets, and TV mounts. We eliminated the reviews that had less than 4 votes to ensure that there is a minimum number of votes accumulated for the review (Ghose and Ipeirotis 2011). The final sample consisted of 2616 reviews. **Figure 2** shows the system design of sentiment extraction and mining process.

Data Analysis and Results

We first checked the descriptive statistics of our data to determine the proper data analysis approach. Table 1 shows the descriptive statistics of our sample. Variables in our model including review length and total votes represent nonnegative and integer data and their standard deviation is larger than their mean. Hence, the analysis needs to be adjusted for overdispersion using log-transformation (Cameron and Trivedi 2013). We also checked the distribution of the data. Review polarity, review sentiment, and title polarity follow a normal distribution. However, title sentiment follows a negative binomial distribution with r=1 (Hilbe 2011). Negative binomial distribution measures the number of failures before a certain number of successes (i.e., r) are achieved. Figure 3 shows the distribution of our different measures. To analyze hypotheses 1 through 5, we tested the following regression model:

$$\log\ (Total\ Votes)\ \% = \beta_0 + \beta_1 Title\ Sentiment\ +\ \beta_2\ TITLE\_POSITIVE +\ \beta_3\ Title\ Length\ +\ \beta_4\ Title\ Sentiment \times TITLE\_POSITIVE +\ \beta_5\ log(Review\ Length) +\ \beta_6\ log(longevity)$$

Equation 1-Predictors of readership

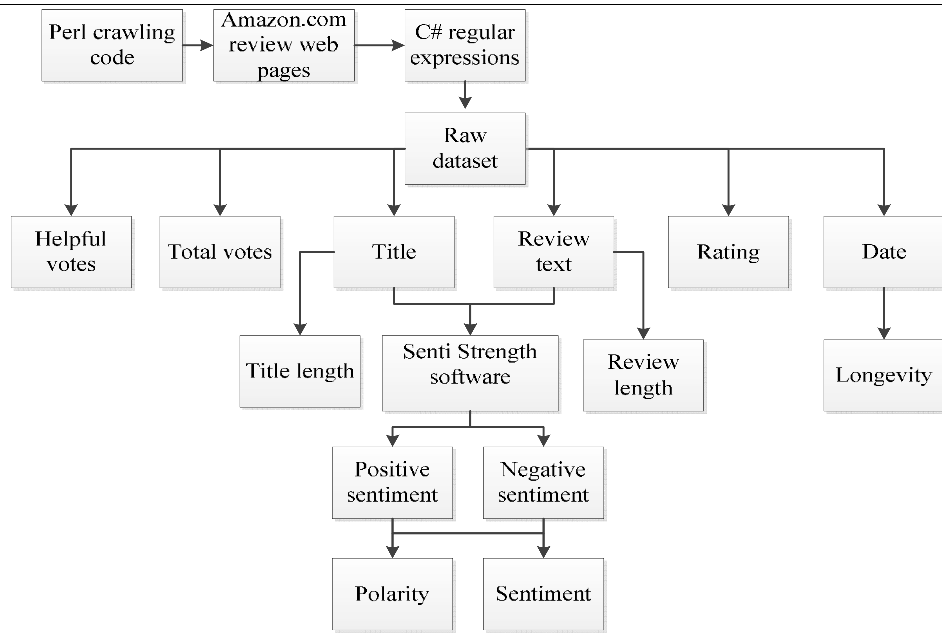

Figure 2-System design of sentiment extraction and mining process

We used negative binomial regression to test the first model in order to control for overdispersion assuming that the data follows a negative binomial distribution (Hilbe 2011). Title sentiment refers to the total sentiment (positive and negative) available in the title of the review. We created one dummy variable, TITLE_POSITIVE, to test the moderation effect of polarity on the effect of title sentiment on readership of reviews. The variable has a value of 1 for titles with positive polarity and a value of zero for the others.

Helpfulness is measured as the proportion of helpful votes out of total votes. Hence, we used binomial regression with logit transformation to examine hypotheses 6 through 8 using the following regression equation (Baum 2008):

$$\frac{Votes\ Helpful}{Votes\ Total}\ \% = \beta_0 + \beta_1 Review\ Sentiment + \beta_2\ REVIEW\_NEUTRAL +\ \beta_3\ log\ \ (Review\ Length) +\ \beta_4\ Review\ Sentiment \times REVIEW\_NEUTRAL + \beta_5\ \ log(Longevity)$$

Equation 2 - Predictors of helpfulness

We created a dummy variable, REVIEW_NEUTRAL, to test the moderation effect of polarity on the relationship between review sentiment and helpfulness. The variable has a value of 1 for reviews with neutral polarity (i.e., polarity=0) and a value of zero for the others. We also controlled for the effect of longevity of the review on its helpfulness by including the log transformation of longevity in the model.

Variable	Range	Median	Mean (SD)
Rating	1-5	2	2.71 (1.70)
Longevity	15-2196	336	433 (362)
Total Votes	4-912	7	22.18 (66.89)
Helpful Votes	0-873	4	16 (60)
Title Sentiment	0-8	1	1.02 (1.02)
Title Polarity	-4-4	0	0.18 (1.26)
Review Sentiment	0-8	3	2.78 (1.48)
Review Polarity	-4-4	0	0.19 (1.32)
Title Length	1-24	4	5.06 (3.42)
Review length	16-2369	84	149.9 (202.8)

Table 1-Descriptive statistics

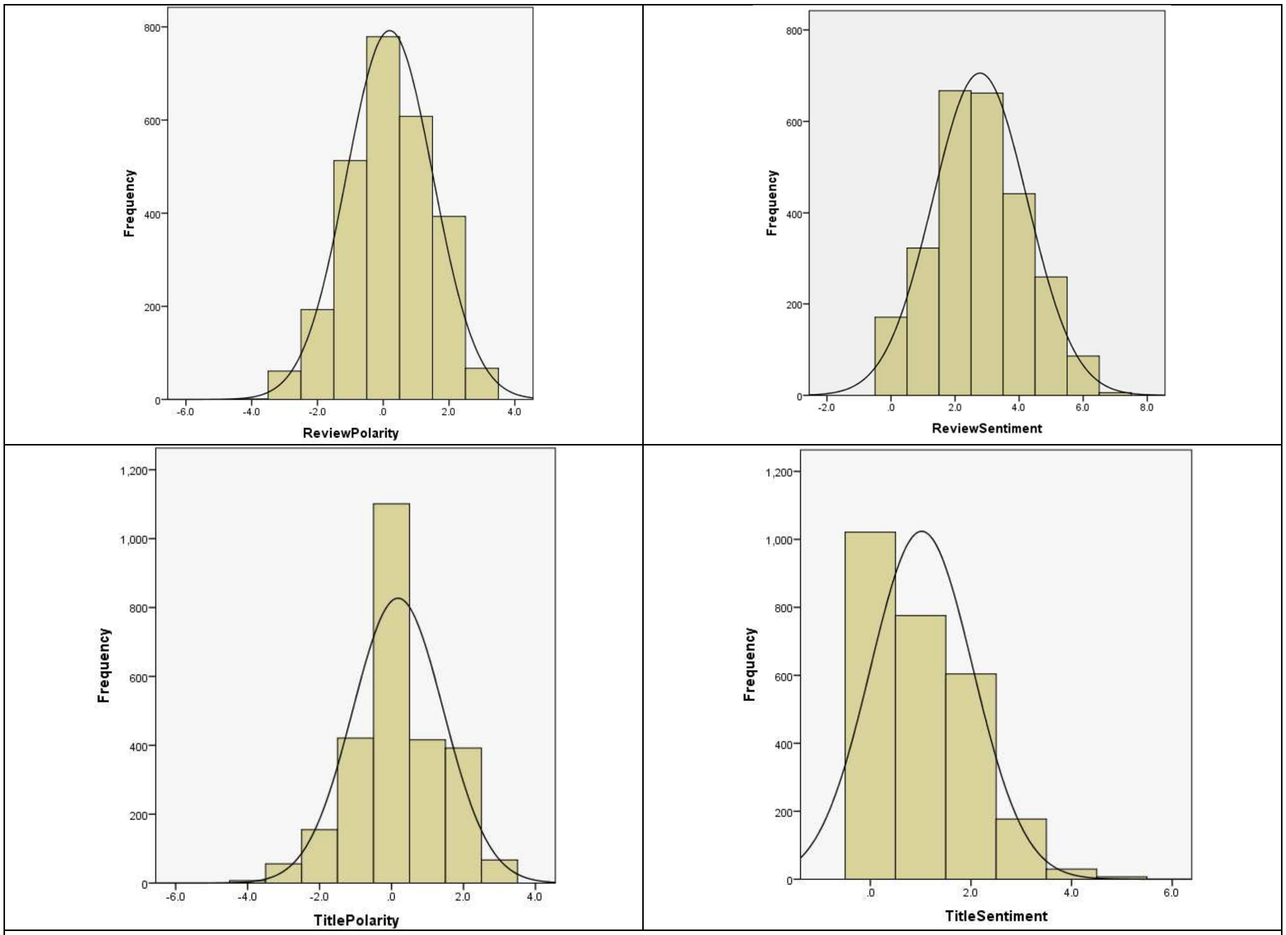

Figure 3 -Distribution of the measures

We first checked the correlation matrix for independent variables to test for multicollinearity. Table 2 shows the correlation matrix for each equation. Because we observed relatively high correlations among some variables, we check the VIF of independent variables. The result of the analysis showed that multicollinearity is not an issue in this study.

Then we proceeded with model analysis. Both models were significant at 0.000 level. The first model shows a goodness of fit of 1.160 indicating that negative binomial regression was a good choice for analysis of the proposed model. The proposed relationship between longevity and total votes was significant ($b = 0.393$, $p < 0.001$), thus supporting H1. The relationship between title sentiment and total votes was significant but the coefficient was negative ($b = -0.087$, $p < 0.001$). Therefore we don't find support for H2. However, the coefficient for the interaction term Title Sentiment × TITLE_POSITIVE was positive and significant (b = 0.439, $p < 0.001$). Thus we find support for H3. The relationship between title length and total votes was significant but negative (b = -0.017, $p < 0.01$). Thus, H4 is not supported. Review length was found to be a significant predictor of total votes (b = 0.355, $p < 0.001$), providing support for H5.

Equation 1					Equation 2			
	Title Length	Title Sentiment	Review Length	Longevity		Review Sentiment	Longevity	Review Length
Title Length	1				Review Sentiment	1		
Title Sentiment	0.086**	1			Longevity	-0.009	1	
Review Length	.152**	.093**	1		Review Length	.115**	-0.012	1
Longevity	-0.057**	-0.001	-0.012	1				

* significant at the 0.05 level, ** significant at the 0.01 level, and *** significant at the 0.001 level

Table 2-Correlation Matrix of Independent Variables

In the second model, review length is a significant predictor of helpfulness (b = 0.407, $p < 0.001$) providing support for H6. The proposed relationship between review sentiment and helpfulness is significant but negative (b = -0.068, $p > 0.001$). Thus, H7 is not supported. However, the coefficient for the interaction term Review Sentiment × REVIEW_NEUTRAL is significant and positive (b = 0.162, $p < 0.001$). Thus, we find support for H8. Surprisingly, our control variable, longevity, had significant relationship with helpfulness (b = 0.416, $p < 0.001$). Figure 4 shows the results of the research model analysis. Table 3 summarizes the hypothesis testing.

Because the effect of sentiment on performance did not completely match our expectations, we ran post-hoc analyses to explore the difference between positive, negative, and neutral sentiment. We tested the first model for reviews with positive, neutral, and negative title polarity separately. Then we ran the second model for reviews with positive, neutral, and negative sentiment. The results are summarized in Table 4. Positive sentiment is an important predictor of readership but not helpfulness. Negative sentiment, in contrast, has negative non-significant effect on both measures of performance. The results regarding neutral sentiment are mixed. Neutral sentiment in the text of the review improves perceptions of helpfulness. However, reviews with neutral titles attract fewer readerships.

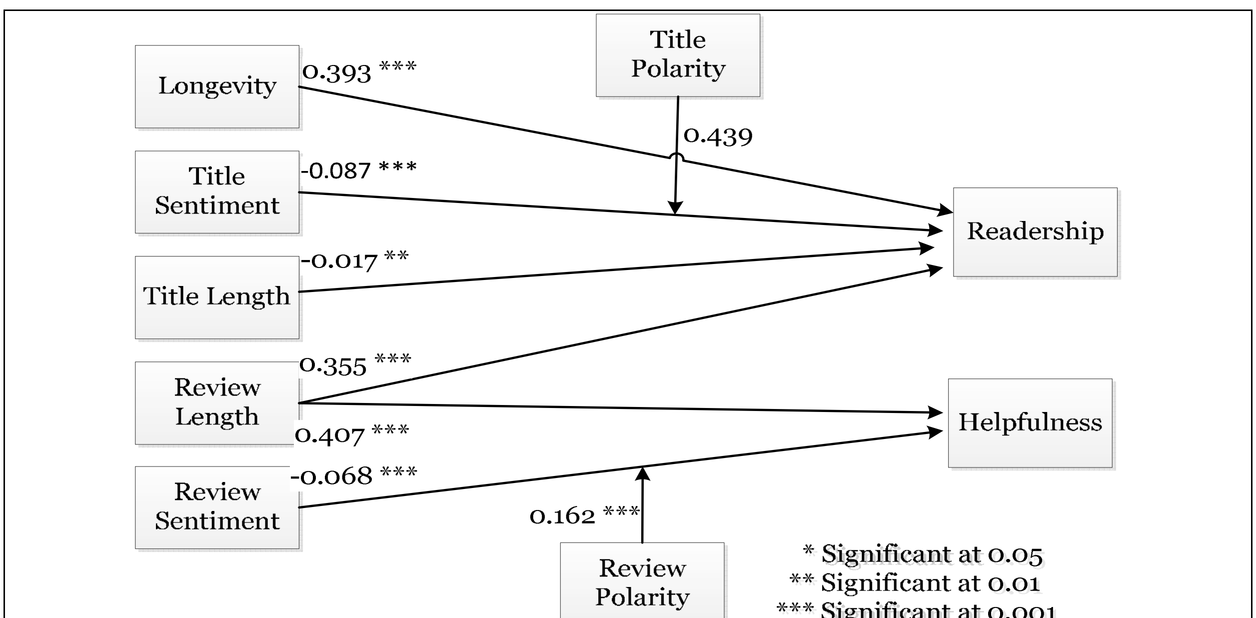

Figure 4- Research model analysis

Hypothesis	Hypothesized Relationship	Estimates (Wald Chi-square)	Results
H1	Longevity → Readership	0.393 (227.71) ***	Supported
H2	Title sentiment → Readership	-0.087 (11.67) ***	Not Supported
H3	Title sentiment X Title positive → Readership	0.439 (53.27) ***	Supported
H4	Title length → Readership	-0.017 (7.05) **	Not Supported
H5	Review length → Readership	0.355 (311.53) ***	Supported
H6	Review length → Helpfulness	0.407 (1031.03) ***	Supported
H7	Review sentiment → Helpfulness	-0.068 (37.21) ***	Not Supported
H8	Review sentiment X Review neutral → Helpfulness	0.162 (161.38) ***	Supported
* Significant at 0.05, ** Significant at 0.01, and *** Significant at 0.001			

Table 3-Hypothesis testing results

Performance Measure	Readership			Helpfulness		
Sentiment Polarity	Positive title	Neutral title	Negative title	Positive review	Neutral review	Negative review
Title sentiment	0.302***	-0.209***	-0.070			
Title length	-0.048***	-0.003	0.014			
Review length	0.424***	0.379***	0.233***	0.382**	0.475***	0.450***
Review sentiment				-0.027	0.066***	-0.005
Longevity				0.461***	0.361***	0.312***
* Significant at 0.05, ** Significant at 0.01, and *** Significant at 0.001						

Table 4-Group analysis

We also tested the second model for different categories of total votes to control for the effect of readership on helpfulness. We categorized the reviews in three categories based on total number of votes: 1-9 votes, 9-99 votes, and 100 votes or higher. We ran the model for the three categories separately and did not find significant differences between them. The only exception is that sentiment does not have a significant effect on reviews with medium number of votes (10-99). Table 5 shows the results of the analysis.

Total Votes	1-9			10-99			>99		
	B	Wald Chi-Square	Sig	B	Wald Chi-Square	Sig	B	Wald Chi-Square	Sig
Review Length	0.30	95.36	0.00	0.43	459.10	0.00	0.28	141.97	0.00
Review Sentiment	-0.07	10.00	0.00	-0.01	0.25	0.62	-0.18	67.14	0.00
Longevity	0.29	115.08	0.00	0.30	208.64	0.00	0.16	22.09	0.00
ReviewSentiment * ReviewTextNeutralD	0.16	26.54	0.00	0.09	19.33	0.00	0.16	42.30	0.00

Table 5-Group analysis of helpfulness based on total number of votes

Discussion

This study investigates the effect of review sentiment on readership and helpfulness of online reviews. Despite the proposed positive relationship between sentiment and performance of OCR, we find that sentiment negatively influences both readership and helpfulness of OCR. Emotions have been considered to subvert the rational processes (Wang 2006). In this study we find evidence that consumers may perceive emotional content to be less rational and therefore less useful. The effect, however, is not consistent across various types of emotions. We find significant positive relationship between sentiment of title and readership of OCR. This observation can be justified using negativity bias. It is likely that people attribute negative sentiment to the author but relate positive sentiment to the product itself. Because negative emotions are attributed to the author, people are less likely to read reviews with negative titles. However, if the author expresses positive emotions in the title of the review, the likelihood of the review getting attention from consumers will increase. This observation also confirms that many consumers look for reviews that reassure them about the choice they have already made (Bailey 2005; Goldsmith and Horowitz 2006; Hennig-Thurau et al. 2004).

The effect of sentiment on helpfulness was significant and positive neutral reviews. It can be argued that a helpful review is a neutral one which communicates both positive and negative aspects of the product. Reviews that lean toward either positive or negative may be perceived by consumers as biased and consequently less helpful. Although the use of sentiment may help the authors convey their experience with the product to the audience, emotions are effective as long as they meet the expectations of the consumers.

Despite the significant effect of title sentiment on the readership of online reviews, length of the title was negatively related to it readership. Apparently people use review titles as a quick source of information about the general theme of the review. Reading and processing longer title take more time and demotivates people to read them. Unlike title length, review length had positive effect on both the readership and helpfulness of reviews. Review length can be seen as a signal about the amount of information contained in the review. Longer reviews are perceived to contain more information and thus attract more readerships. Moreover, longer reviews are more likely to analyze different aspects of the product which leads to increased perceptions regarding their helpfulness.

Surprisingly, longevity has a positive effect on the helpfulness of online reviews. In other words, older reviews are perceived to be more helpful. One reason can be the way Amazon sorts the reviews. By default, users will view the most helpful reviews although it can also be sorted to display the recent ones first. Because older reviews start receiving votes early, they have a better chance of appearing on the first page which will lead to more votes and perhaps better perceptions regarding their helpfulness. Newer reviews, however, stay at the end of the list and are less likely to receive any attention from consumers. Aral (2013) has a similar observation. He observed that positive manipulation of helpfulness of online reviews creates a positive social influence. As a result, longevity not only affects readership, it but also influences perception of helpfulness of OCR.

Implications

This study has implications for both theory and practice. From a theoretical perspective, this study introduces a new method of analyzing consumer behavior. While previous research finds that online reviews significantly influence consumer behavior, few studies have explored the issue using the textual information contained in OCR. Sentiment mining can be utilized to analyze large amount of data that is produced on the Internet every day. Sentiment mining facilitates processing of big data and is highly replicable. This study also contributes to the body of knowledge by studying both readership and helpfulness of OCR. Furthermore, although previous research has used the text contained in online reviews to investigate, to best of our knowledge this study is the first to also include the title of online reviews in the analysis. We find that title of reviews contain valuable information and should be considered in the future studies. Finally, this study can be seen as an initial step toward action research in the area of OCR. The findings of this study and future similar studies can be integrated into a single framework to provide a solution for classification and sorting of OCR.

We also have implications for practice. Most previous studies just use numerical rating and length of the review to investigate their performance. Few studies have tried to analyze the textual information contained in online reviews using human subjects. Although the use of human subjects is useful in explaining how people use online reviews, it barely provides a solution for online vendors because the process is neither scalable nor replicable through automated systems. To best of our knowledge, this study is the first to provide online vendors with a scalable replicable solution for analyzing OCR.

The findings of this study can be used by online vendors to develop systems that analyze and classify online reviews based on their potential performance. Our study finds two weaknesses of the current method used for sorting OCR. First, we find that older reviews attract more readerships. Many products have thousands of review many of which never receive any attention from consumers because they stay at the end of a long list of reviews. Out of the 35000 reviews we collected for this study, 92.5% had less than five votes and 69.7% had no votes. Second, we also observe that older reviews are more likely to be perceived helpful than the recent ones. In other words, perceptions of helpfulness are biased toward older reviews. New reviews may contain up-to-date information about the product and the recent changes made by the manufacturer such as the software updates which may be overlooked by the current sorting method. These two observations indicate that many potentially useful reviews receive little or no attention from consumers because they are barely seen by consumers. Hence, improving the sorting algorithms of online reviews seems to be necessary. An intelligent system capable of detecting potentially useful reviews can help both vendors and consumers. It helps consumers by providing them with more valuable information and by saving their time and energy. It also helps vendors to satisfy their customers' need for information which will allow customers to decide more quickly and may eventually lead to increased sales for the vendors.

Limitations and future research

Like any other studies, this study has limitations. Although automated sentiment mining is quite useful in analyzing textual information, it still has limitations. The software we used in this study can process different types of textual information. However, it lacks processing capability for alternate styles of writing such as sarcasm. While there are many areas for improvement in the field of natural language processing, future research can provide better insights regarding the information contained in online reviews using more advanced technology.

The sample used in this study was mainly limited to reviews of electronic products. While previous research shows that product type has an important moderation effect on the performance of online reviews, our conclusions may remain limited to certain product types used in this study. Future research can use a sample of different product types and test the moderation effect of product type on the performance of reviews.

Our sample also lacks language and cultural diversity. The reviews were collected from Amazon US website and all are in English. Different cultures express emotions differently. For example, it is believed that people in individualistic cultures express negative emotions more freely than those living in collectivist cultures (Takahashi, Ohara, Antonucci and Akiyama 2002). On the other hand, language as the vehicle of information may play in significant role in communicating the message to the reader. Future research may utilize reviews written in other languages to include the effect of language and culture on the performance of online reviews. Future research may also look at the differences in expression of emotions between the real life and the virtual space.

Conclusions

This research provides insights regarding the predictors of performance of online reviews using a sentiment mining approach. We find that sentiment negatively influences the performance of online reviews with two exceptions: positive sentiment in the title and neutral sentiment in the text of the review. We also find that longer reviews are more likely to attract readership and to be perceived as helpful. This research can be used by online vendors to create scalable automated systems for sorting and classification of OCR.

References

Ahluwalia, R. 2000, "Examination of psychological processes underlying resistance to persuasion," *Journal of consumer research* (27:2) pp 217-232.

Aral, S. (2013). The Problem With Online Ratings. Retrieved from http://sloanreview.mit.edu/article/the-problem-with-online-ratings-2/

Bai, X. 2011, "Predicting consumer sentiments from online text," *Decision Support Systems* (50:4) pp 732-742.

Bailey, A.A. 2005, "Consumer awareness and use of product review websites," *Journal of interactive advertising* (6:1) pp 68-81.

Baum, C.F. 2008, "Modeling proportions," *Stata Journal* (8:2) p 299.

Berger, J., and Milkman, K.L. 2012, "What makes online content viral?," *Journal of marketing research* (49:2) pp 192-205.

Bickart, B., and Schindler, R.M. 2001, "Internet forums as influential sources of consumer information," *Journal of Interactive Marketing* (15:3) pp 31-40.

Cameron, A.C., and Trivedi, P.K., 2013 *Regression analysis of count data* Cambridge university press.

Chen, Y., and Xie, J. 2008, "Online consumer review: Word-of-mouth as a new element of marketing communication mix," *Management science* (54:3) pp 477-491.

Chevalier, J.A., and Mayzlin, D. 2006, "The effect of word of mouth on sales: Online book reviews," *Journal of marketing research* (43:3) pp 345-354.

Dellarocas, C., Zhang, X.M., and Awad, N.F. 2007, "Exploring the value of online product reviews in forecasting sales: The case of motion pictures," *Journal of Interactive Marketing* (21:4) pp 23-45.

Dholakia, R.R., and Sternthal, B. 1977, "Highly Credible Sources: Persuasive Facilitators or Persuasive Liabilities," *Journal of Consumer Research* (3) pp 223-232.

Duan, W., Gu, B., and Whinston, A.B. 2008, "Do online reviews matter?—An empirical investigation of panel data," *Decision Support Systems* (45:4) pp 1007-1016.

Folkes, V.S. 1988, "The availability heuristic and perceived risk," *Journal of consumer research* (15:1).

Forman, C., Ghose, A., and Wiesenfeld, B. 2008, "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets," *Information Systems Research* (19:3) pp 291-313.

Garcia, D., and Schweitzer, F., 2011 "Emotions in Product Reviews--Empirics and Models," Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom), IEEE, pp. 483-488.

Ghose, A., and Ipeirotis, P.G. 2011, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *Knowledge and Data Engineering, IEEE Transactions on* (23:10) pp 1498-1512.

Godes, D., and Mayzlin, D. 2004, "Using online conversations to study word-of-mouth communication," *Marketing Science* (23:4) pp 545-560.

Goldsmith, R.E., and Horowitz, D. 2006, "Measuring motivations for online opinion seeking," *Journal of interactive advertising* (6:2) pp 2-14.

Gruzd, A., Doiron, S., and Mai, P., 2011 "Is happiness contagious online? A case of Twitter and the 2010 Winter Olympics," System Sciences (HICSS), 2011 44th Hawaii International Conference on, IEEE, pp. 1-9.

Harris, R.B., and Paradice, D. 2007, "An investigation of the computer-mediated communication of emotions," *Journal of Applied Sciences Research* (3:12) pp 2081-2090.

Heider, F., 1958 *The psychology of interpersonal relations* John Wiley & Son, New York.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G., and 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* (18:1) pp 38-52.

Hilbe, J.M., 2011 *Negative binomial regression* Cambridge University Press.

Hu, N., Koh, N.S., and Reddy, S.K. 2013, "Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales," *Decision Support Systems* (57).

Huffaker, D. 2010, "Dimensions of leadership and social influence in online communities," *Human Communication Research* (36:4) pp 593-617.

Johnson, E.J., and Payne, J.W. 1985, "Effort and accuracy in choice," *Management science* (31:4) pp 395-414.

Jones, E.E., and Davis, K.E., 1965 "From acts to dispositions the attribution process in person perception," in: *Advances in experimental social psychology*, pp. 219-266.

Kelley, H.H. 1973, "The processes of causal attribution," *American Psychologist* (28:2) p 107.

Liu, Y. 2006, "Word of mouth for movies: Its dynamics and impact on box office revenue," *Journal of marketing* (70:3) pp 74-89.

Mudambi, S.M., and Schuff, D. 2010, "What makes a helpful online review? A study of customer reviews on Amazon. com," *MIS Quarterly* (34:1) pp 185-200.

Nielsen. (2012). Global consumers' trust in 'earned' advertising grows in importance. Retrieved from http://www.nielsen.com/us/en/press-room/2012/nielsen-global-consumers-trust-in-earned-advertising-grows.html

Park, D.-H., Lee, J., and Han, I. 2007, "The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement," *International journal of electronic commerce* (11:4) pp 125-148.

Riordan, M.A., and Kreuz, R.J. 2010, "Emotion encoding and interpretation in computer-mediated communication: Reasons for use," *Computers in Human Behavior* (26:6) pp 1667-1673.

Russell, M.A., 2013 *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More* O'Reilly Media, Inc.

Schindler, R.M., and Bickart, B. 2012, "Perceived helpfulness of online consumer reviews: the role of message content and style," *Journal of Consumer Behaviour* (11:3) pp 234-243.

Schwenk, C.H. 1986, "Information, cognitive biases, and commitment to a course of action," *Academy of Management Review* (11:2) pp 298-310.

Sen, S., and Lerman, D. 2007, "Why are you telling me this? An examination into negative consumer reviews on the web," *Journal of Interactive Marketing* (21:4) pp 76-94.

Stieglitz, S., and Dang-Xuan, L. 2013, "Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior," *Journal of Management Information Systems* (29:4) pp 217-248.

Takahashi, K., Ohara, N., Antonucci, T.C., and Akiyama, H. 2002, "Commonalities and differences in close relationships among the Americans and Japanese: A comparison by the individualism/collectivism concept," *International Journal of Behavioral Development* (26:5) pp 453-465.

Thelwall, M., and Buckley, K. 2013, "Topic-based sentiment analysis for the social web: The role of mood and issue-related words," *Journal of the American Society for Information Science and Technology* (64:8).

Thelwall, M., Buckley, K., and Paltoglou, G. 2012, "Sentiment strength detection for the social web," *Journal of the American Society for Information Science and Technology* (63:1) pp 163-173.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. 2010, "Sentiment strength detection in short informal text," *Journal of the American Society for Information Science and Technology* (61:12) pp 2544-2558.

Treisman, A.M. 1969, "Strategies and models of selective attention," *Psychological Review* (76:3) p 282.

Tversky, A., and Kahneman, D. 1974, "Judgment under uncertainty: Heuristics and biases," *Science* (185:4157) pp 1124-1131.

Walsh, J.P. 1988, "Selectivity and selective perception: An investigation of managers' belief structures and information processing," *Academy of Management Journal* (31:4) pp 873-896.

Walther, J.B., and D'Addario, K.P. 2001, "The impacts of emoticons on message interpretation in computer-mediated communication," *Social science computer review* (19:3) pp 324-347.

Wang, X.-T. 2006, "Emotions within reason: Resolving conflicts in risk preference," *Cognition and Emotion* (20:8) pp 1132-1152.