Spring 2014

On the Helpfulness of Analyzing Online Sentiments

Eric Alexander Domenic Morgera
Florida State University, em10g@my.fsu.edu

Abstract:

(reviews, helpfulness, sentiment)

The helpfulness dimension of online user reviews has a substantial mediating effect on the consumer buying decision process and holds potential to influence sales. Looking past the disputed numeric star rating element, this study empirically examines the variance of sentiment levels and the embedded emotional load contained within a review's textual body to uncover why certain reviews receive more helpfulness votes than others. Using a dataset comprised of nearly 5,000 hotel reviews analyzed by an emergent sentiment analysis tool, the findings represent that sentiment levels are indeed an extremely influential component in determining the helpfulness vote count. Likewise, reviews with increased emotional intensity receive more helpfulness votes than others, indicating that people tend to pay greater attention to opinions that are passionate in nature. The broader theoretical, methodological and practical implications of these findings are discussed.

THE FLORIDA STATE UNIVERSITY

COLLEGE OF BUSINESS

ON THE HELPFULNESS OF ANALYZING ONLINE SENTIMENTS

By
ERIC ALEXANDER DOMENIC MORGERA

A Thesis submitted to the
Department of Entrepreneurship, Strategy, and Information Systems
in partial fulfillment of the requirements for graduation with
Honors in the Major

Degree Awarded:
Spring, 2014

The members of the Defense Committee approve the thesis of Eric Alexander Domenic Morgera defended on April 17, 2014.

Dr. Iris Junglas
Thesis Director

Dr. David Paradice
Committee Member

Dr. Charles Hofacker
Outside Committee Member

To my parents, Salvatore and Mischa, who always believe and inspire.

To my father, whose parents arrived to Ellis Island less than 95 years ago, owning nothing more than the belongings in their pockets and a dream to build. Your insistence on the necessity of education and optimism have allowed me to become someone of substance.

To my mother, a first-generation American who suffered the loss of her family, friends, and culture during the Iranian Revolution to immediately thereafter pick the pieces up and start anew. Your perseverance, compassion, and love have allowed me to become someone of worth.

I hope I have made you proud.

Acknowledgements

My sincerest appreciation goes to my Thesis Director, Dr. Iris Junglas, for her encouragement, extraordinary direction, and attention to detail throughout the research timeline. Through her tremendous commitment and effectiveness as a mentor, she has quietly and graciously opened doors for my professional growth that would have otherwise remained latched shut. She truly posses the patience, authenticity, and, above all, the passion that leaves a positive impact on the lives of all students with whom she comes in contact; I know she has mine.

Heartfelt thanks goes to Dr. Charles Hofacker and Dr. David Paradice for their willingness to serve on this committee, their sagacity, and support. An appreciation to Florida State University's College of Business and the Honors Program for enabling an environment that inspires undergraduates to think, achieve, and become.

Table of Contents

1. Introduction

In May 2010, JordynC disclosed to the TripAdvisor.com community a most nauseatingly frightful experience, “If you are looking for a hotel with chewing tobacco spit oozing down the halls and corridors; spiders actively making webs in every corner of your room; carpeting so greasy and dirty you wouldn't want to sit your luggage down—let alone walk around barefoot; and a room so putrid and smelly it causes a gag-reflex when you walk in…by all means, stay at The Grand Resort” (Farnham 2013).

Her review, marked helpful by dozens of readers, and the profusion of repeatedly negative narratives soon landed the hotel atop TripAdvisor’s 2011 Dirtiest Hotels in America listing. Dubbed the dirtiest, the hotel’s dismal reputation permeated across the Internet’s blogosphere and quickly ushered the property into sudden and certain foreclosure by year’s end (CBSNews 2011; Bly 2012; Farnham 2013).

The Grand Resort’s downhill spiral first precipitated by JordynC’s review illustrates the astronomical impact that a single online review and a review’s helpfulness dimension can have on a business—now imagine the number of these reviews multiplied by thousands, posted each minute on the Internet. With the proliferation of online review systems, consumers are actively and rapidly taking to websites such as TripAdvisor.com, Amazon.com, and Yelp.com to share and exchange their opinions on businesses, services, and products in a real time environment. In turn, as relatable consumers seek assistance during the buying decision process, these existing reviews have been shown to hold potential for significantly impacting future sales and profit (Chen, Fay and Wang 2003; Chevalier and Mayzlin 2006; Duan, Gu and Whinston 2008; Chen, Dhanasobhon and Smith 2008).

The current research landscape has investigated the numeric star rating that accompanies reviews within most systems, but studies widely differ in their acceptance of the star rating's impact on helpfulness and sales. Some have indicated that even a one-star improvement in ratings can increase sales by a minimum of 5 percent (Chevalier and Mayzlin 2006; Luca 2011); others, peculiarly even those analyzing similar data from identical review systems and product lines, reject any impact of the star rating on sales (Duan et al. 2008; Hu, Koh and Reddy 2014). Therefore, relying on a review's star rating is not an effective panacea.

Given the disparity of previous research, new studies now look to examine and mine the hitherto untapped rich textual body of online reviews, rather than the star rating. This written component has been found to contain the stylistic elements, semantic characteristics, and argumentation that result in improved predictions when compared to those derived from numeric star ratings as indiscriminately chosen by reviewers (Ganu, Elhadad and Marian 2009; Willemsen, Neijens, Bronner and de Ridder 2011; Schindler and Bickart 2012). Particularly, the included sentiment, defined as certain subjective attitudes expressed through affective word choice based upon opinion or feeling, has become an area of increased focus and analysis (Hu et al. 2014). These sentiments, frequently categorized into three types of positive, negative and neutral emotional levels, are deemed paramount to understanding and fully decoding the perceived helpfulness of reviews (Ganu et al. 2009; Schindler and Bickart 2012; Kennedy 2012).

Empirical research has defined this helpfulness dimension as an individual's subjective endorsement in response to reading and agreeing with the textual portion of another's review, routinely captured by a question phrased "Was this review helpful?" (Li, Huang, Tan and Wei 2013). This simple question has proven to become one of the most profitable and valuable marketing strategies in existence. For Amazon.com, the introduction of this question alone is responsible for $2.7 billion of additional revenue each fiscal year (Mudambi and Schuff 2010; Spool 2009).

Yet, the profoundly crucial relationship between sentiment and helpfulness is poorly understood and underreported, though few existing studies introduce the possibility of a direct relationship between these variables (e.g., Schindler and Bickart 2012; Hu et al. 2014). Findings have also begun to postulate that reviews containing the most sentiment and most helpfulness votes command the greatest role in determining sales (Hu et al. 2014). While online reviews richly contain sentiment information, it has proven previously difficult and laborious to analyze the frequently ambiguous textual information from reviews, especially for larger datasets (Cao, Duan and Gan 2011). This time-consuming challenge may account for why extant studies have left this relationship in an inconclusive state. However, new automated text mining technologies have emerged with proprietary and open-source sentiment analysis tools, enabling speedy sentiment evaluation for thousands of reviews with ease and within seconds.

Previous limitations and challenges considered, this study seeks to uncover why certain online reviews receive more helpfulness votes than others by deciphering the relationship of sentiment levels as expressed in a review's textual body. More particularly, an empirical examination will investigate how the variance of sentiment levels—positive, negative or neutral emotion—along with their embedded emotional load, capturing the intensity of emotions irrespective of their direction, affects the quantity of helpfulness votes a review receives. Using an extensive dataset containing nearly 5,000 reviews collected from a prominent crowdsourcing website, an emerging sentiment analysis tool will be applied within this study to automate the systematic extraction of sentiment contained within each review.

2. Literature Review and Model Development

2.1 The Effect of Online Reviews on Sales

Online reviews have become an increasingly commonplace element on e-commerce platforms and crowdsourcing websites, their content positioned at the intersection of sales and publicity. According to a recent survey conducted by Dimensional Research, an overwhelming 90 percent of Americans that recollected browsing online reviews claimed the presence of positive reviews influenced their buying decision, whilst 86 percent declared their decision was influenced by negative reviews. Of these same respondents, 95 percent would later share their own negative experiences in an online review, with 87 percent sharing good experiences (Gesenhues 2013; Zendesk 2013). This growing online review audience and the authoritative influence it exerts on the buying decision process is reflected by a thriving body of research that examines the correlation between online reviews and sales.

Over the past decade, studies bridged between the information systems and marketing disciplines have devoted considerable attention to document whether and to what degree the existence of online reviews and their associated numeric star ratings affect product sales. The hypothesis that online reviews in general influence future sales is largely endorsed (Senecal and Nantel 2004; Chevalier and Mayzlin 2006; Zhang and Dellarocas 2006; Li and Hitt 2008; Zhu and Zhan 2010; Archak et al. 2011; Chintagunta et al. 2010; Duan, Gu and Whinston 2008; Chen, Fay and Wang 2003; Chen, Dhanasobhon and Smith 2008; Berger 2012), supporting a positive relationship. For example, Senecal and Nantel (2004) determined that individuals consulting product recommendations would purchase these products twice as often as those who did not consult this information. Chevalier and Mayzlin (2006) discovered online reviews and review volume significantly increased book sales, while Zhu and Zhan (2010) obtained similar results in the video game industry and Zhang and Dellarocas (2006) in the movie industry. Additional independent studies also confirm that online reviews increase sales across the board, supplementing an additional 32 percent to 52 percent in sales (Berger 2012).

In contrast, studies dissecting the relationship between numeric star ratings, gauged frequently on a scale of one to five stars, and product sales have mixed results. Whereas Luca (2011) finds a one-star improvement in a restaurant's Yelp rating can increase sales by at least 5 percent, Hu et al. (2014) argues a review's star rating does not impact sales, providing instead that a review's helpfulness vote count and sentiment are the greatest factors affecting future sales. Duan et al. (2008) also observed the lack of any relationship between a review's star rating and sales.

Thus, the existence of online reviews clearly impacts sales, but the relationship between the numeric star rating dimension and sales is challenged and divergent. Literature hints towards the fact that other dimensions, particularly the helpfulness of online reviews, can hold a mediating effect on the buying decision process and sales forecasting—in addition to just volume and star rating.

2.2 About the Helpfulness of Online Reviews

Emerging research has begun to recognize that the value of an online review is often judged by its votes of helpfulness (Mudambi et al. 2010; Cao et al. 2011; Ghose and Ipeirotis 2010; Ghose, Ipeirotis and Li 2012).

Frequently provided in an effort to help digest the sheer volume of feedback available, websites provide consumers the possibility of marking a review as helpful, capturing their response through a question routinely phrased as "Was this review helpful?" (Li, Huang, Tan and Wei 2013). Online review systems, such as Amazon.com or Yelp.com, then quantify and present this data in summarized form for quick consumer digestion and relay the information either above or below the review's textual body as, for example, "30 out of 50 people found this review helpful" (Cao et al. 2011). These systems subsequently rank and feature the most helpful of reviews, or consumers may be offered the choice to sort reviews by their associated level of helpfulness (Siering and Muntermann 2013).

Helpfulness, captured as the number of votes or as a percentage thereof, signifies a subjective endorsement and is produced in response to reading the textual body of a review (Mudambi et al. 2010; Cao et al. 2011; Ghose and Ipeirotis 2010; Siering and Muntermann 2013; Kim, Pantel, Chklovski and Pennacchiotti 2006). A single vote endorses the semantic information expressed in the review and indicates the weight a consumer places on this information as part of the buying decision process (Cao et a. 2011).

Voting a review as helpful carries multiple assumptions. First, it implies that a consumer has actually read the review; he or she has to have processed the review and examined, at least in parts, its content for positive and negative evaluations. Second, it indicates that the review provides valuable information that is deemed influential for others and their decision-making process (Weiss et al. 2008). Third, it suggests that the information provided could potentially be above and beyond of what is already known to influence the buying decision process. Finally, it implies that this particular review provides more or better information relative to others (Weiss et al. 2008). As a result, a helpfulness vote increases the likelihood of a review to be read by others.

Bearing in mind the significant impact review helpfulness has on perceived weight and value during the information search phase of the buying decision process, this study sets out to better understand its predictors. A clear definition of the relationship stemming from these predictors will provide needed insight into what consumers are looking for in reviews and elucidate the reasons behind the phenomenon of why certain online reviews receive more helpfulness votes than others.

2.3 The Relationship between Helpfulness and its Precursors

Online reviews contain a multiplicity of characteristics that can contribute to and heavily influence the helpfulness dimension (Ganu, Elhadad and Marian 2009; Schindler and Bickart 2012; Willemsen, Neijens, Bronner and de Ridder 2011). Past studies have generally opted to characterize these distinct characteristics into the following common groups: structural elements (review lengthiness, word count, and readability), sidedness (bias and message subjectivity), reviewer identity (expertise and personality) and sentiment (positive, negative, or neutral words) (Cao et al. 2011; Kim et al. 2006).

Although this paper focuses primarily on dissecting the relationship between a review's helpfulness dimension and the sentiment contained within the rich textual body, it is important to survey and detail all review characteristics that have been dissected in previous studies prior to engaging in further discussion. In this section, the paper explores each of the aforementioned characteristic groups and then round out our discussion by focusing on sentiment—one characteristic in dire need of expanded investigation. Additionally, a new conceptualization for sentiment, through the use of an emotional load variable, will also be discussed.

2.3.1 Structural Elements Influencing Helpfulness

Several studies have looked to determine if the structural characteristics of a review effect helpfulness. Kim et al. (2006), Liu et al. (2011), Forman et al. (2008), Mudambi et al. (2010), Pan and Zhang (2011) all conclude that helpfulness is positively affected by the length of a review, each agreeing that the effect of length is more substantial for search goods (i.e., products or services that can be easily evaluated) versus that of experiential goods (i.e., products and services that have to be experienced before an individual can properly judge them) (Nelson 1970; 1974).

Kim et al. (2006), for example, studied online reviews for consumer electronic search goods, determining that sentence length and other structural components were imperative to review helpfulness. Forman et al. (2008) ascertained that review readability positively impacted perceived helpfulness, while spelling mistakes negatively impacted helpfulness. Dissimilarly, yet important to highlight for this paper's purposes, Korfiatis et al. (2012) found that semantic elements, such as sentiment and emotional word choice, were more impactful on helpfulness than the length of the textual body.

2.3.2 Sidedness Influencing Helpfulness

Concerning the impact of review sidedness and bias on helpfulness, studies are contradictory. Forman et al. (2008) empirically investigated reviews for three product types on Amazon.com, finding that reviews containing a blend of subjective and objective statements were more helpful to consumers than reviews not open to two-sidedness. Willemsen et al. (2011) and Schlosser (2011) rebut, determining that one-sided reviews, primarily those containing negative information, are more powerful in afflicting review helpfulness than two-sided reviews.

2.3.3 Reviewer Identity Influencing Helpfulness

Reviewer identity and expertise characteristics influence the helpfulness dimension as well (Pan and Zhang, 2011; Forman et al. 2008; Chen and Xie 2008; Willemsen et al. 2011). When a reviewer's personality characteristics are revealed or their identity is described in the textual body, consumers find these reviews to be more helpful than those that do not reveal any details or few details (Forman et al. 2008; Willemsen et al. 2011).

Surprisingly, Chen and Xie (2008) determined that using endorsement-format advertising, such as displaying third-party product award logos, can hurt the helpfulness dimension. Their findings indicate that this is caused by an interaction between the pricing and advertising reactions by product competitors. While review-endorsed advertising will improve the winning product's advertising effectiveness, it simultaneously motivates price-cutting by the losing products. As a result, consumers turn their interest to these reduced-priced products, translating the winning products' helpfulness dimension into an opportunity loss.

2.3.4 Sentiment Strength Influencing Helpfulness

The sentiment strength included in a review's textual body has been defined as certain subjective attitudes expressed through affective word choice based upon a consumer's opinion or feeling as a result of using a product or service. These sentiments are frequently categorized into three emotional levels: positive (e.g., success), negative (e.g., failure), and neutral (e.g., history) (Hu, Koh and Reddy 2014; Schindler and Bickart 2012; Kennedy 2012; Stravroula, Vinson, and Vigliocco 2009).

Being that sentiment has become a recent area of interest within the larger research landscape, examining the existing, albeit limited literature is important. Studies performed by Mudambi et al. (2010), Cao et al. (2011), Pan and Zhang (2011), Siering and Muntermann (2013), and Hu et al. (2014) have explored the relationship between sentiment strength and review helpfulness, each using datasets that contain reviews for a variety of goods in order to attempt to define this relation.

For example, Hu et al. (2014) examined a dataset containing reviews from Amazon.com for 4,405 books and developed a model to examine the interrelationship between ratings, sentiments, and sales. The study concluded that a review's sentiment most significantly impacted overall sales for this product type. Moreover, reviews voted most helpful by consumers were crucial in determining sales.

Mudambi et al. (2010) analyzed 1,587 reviews from Amazon.com for six goods (i.e., cell phones, digital cameras, laser printers, MP3 players, music CDs, and video games). The study concluded that the strength of sentiment word choice affected the perceived helpfulness of a review, primarily for search goods.

Conversely, Willemsen et al. (2011) suggest the exact opposite of the all the above, indicating that an inverse relationship exists between sentiment and helpfulness. Li and Zhan (2011) also evidenced that extreme sentiment strength adversely affects helpfulness, and further hurts the perceived competence and credibility of the reviewer's identity. Schindler and Bickart (2012) found that negative sentiment is detrimental, but positive sentiment is beneficial to review helpfulness.

Based on the preceding review of literature, it can be summarized that the effects of sentiment strength on review helpfulness have not been entirely agreed upon or understood thus far. Furthermore, none have explored the relationship using reviews from complex experiential goods, such as hotel stays or hospital visits. These complex goods, however, may contain the key to defining the relationship between sentiment strength and helpfulness.

As Mudambi et al. (2010) explains, obtaining information on product quality prior to interaction for a search good is relatively easy; product attributes are readily comparable, and a strong need to use one's senses to evaluate quality is not needed. For complex experiential goods, however, it becomes difficult and costly to obtain product quality information prior to interaction; key product attributes are not easily comparable, and a need to use one's senses to evaluate product quality exists.

Therefore, in response to a need to better clarify the effect of sentiment strength on review helpfulness, this paper explores complex experiential goods and postulates that the presence of sentiments within a review's textual body positively impacts the helpfulness dimension. Thus, the following hypothesis is formulated:

H1: Positive and negative sentiment strengths contained within an online review's textual body will positively influence a review's helpfulness dimension.

2.3.5 Emotional Load Influencing Helpfulness

The concept of sentiment strength, as hypothesized above, provides an indication of a review's overall emotional direction, resulting either in a positive, negative, or neutral outcome. However, within one review—and even within one sentence—a spectrum of sentiments might be present. For example, one sentence may exude a positive sentiment about a hotel room and its amenities, whereas another sentence within the same review laments about the lousy service. The notion of sentiment strength, as illustrated above, would capture this aspect by summing the various sentiment values and aggregating them into one overall assessment value. For example, if a review contains equally positive (e.g., +2) and negative (e.g., -2) sentiment, the sentiment strength variable would summarize these values into a neutral sentiment output (e.g., 0).

However, it is important to have more information than the simple summation provides, especially in determining if the full extent of a review's emotional wordage and passion, regardless if positive, negative or neutral, can impact perceived helpfulness. This effect is particularly evident in reviews that contain high levels of sentiments to either polarity. For instance, a review that has an extremely positive (e.g., +5) as well as an extremely negative (e.g., -5) sentiment likely contains more passionate—and thus more useful—information on a consumer's experience than a review with only slight levels of both, but would still result in a neutral sentiment strength output overall (e.g., 0). Therefore, analyzing the sentiment strength variable alone can lead to a misleading analysis and provides little information on how emotionally loaded a review truly is. While the overall strength of a sentiment embedded in a review is important, capturing the intensity of its emotionalism might be equally vital in its effect on helpfulness. This postulation is supported by a study conducted by Stieglitz and Dang-Xuan (2013) who analyzed 165,000 tweets on Twitter and concluded that it is critical to examine overall emotion expressed within tweets versus the sentiment strength variable alone.

Thus, in order to capture the intensity to which sentiments are expressed and their subsequent impact on helpfulness, this study introduces the notion of emotional load. Emotional load is defined as the aggregate of both positive and negative sentiment strength valences, resulting in a variable that captures the overall intensity of emotions expressed within a review, irrespective of polarity. The purpose of the emotional load variable is to represent the full intensity of a review's emotional wordage, summing the absolute values of positive and negative sentiment strengths into a combined output. Citing the above example, a review now containing extremely positive (e.g., +5) and extremely negative (e.g., |-5|) sentiment would output as an emotionally loaded review (e.g., +10) and be deemed to have higher emotional load than a review with only slight levels of both.

This new, unique conceptualization of emotional load is introduced to provide clarity on whether the presence of emotional passion within a review is important to consumers and to the perceived helpfulness of a review. Thus, the following hypothesis is posited:

> **H2:** Emotional load, containing the absolute values of both positive *and* negative sentiment strengths, will have an overall positive impact on the review's helpfulness dimension.

2.4 Measuring and Analyzing Sentiment

With the proliferation of online review systems, such as TripAdvisor.com and Yelp.com, social media and other crowdsourcing websites, consumers are actively taking to these channels and harnessing public opinion and sentiment to facilitate the buying decision process. However, mining the reviews found on these websites and distilling the rich sentiment contained within proves itself a daunting task due to the dissimilarity of data formatting across these different websites (Tang, Tan and Cheng 2009; Duan, Cao, Yu and Levy 2013; Mishra and Jha 2012).

The manual human evaluator will face complications to properly identify relevant websites and accurately summarize the sentiment found within the review's content (Pang and Lee 2008; Mukherjee 2012). Furthermore, extant studies have established that human evaluation is prone to considerable partiality and biases. Particularly, individuals will give greater consideration to sentiment that coincides within their own personal preferences and beliefs and will also face increased difficulty in maintaining consistency with results as the dataset size grows, due to general physical and mental limitations (Pang and Lee 2008; Tang et al. 2009).

Thus, the need for using automated sentiment mining and summarization systems is paramount, as complications faced by human evaluators can be defeated through the use of an objective sentiment analysis system. These sentiment analysis systems utilize NLP (Natural Language Processing), statistics, and machine learning methods to extract, identify, and characterize people's opinions, attitudes, appraisals, and emotional feelings toward products, services, businesses, events, topics and their attributes (Nanli, Ping, Weiguo and Meng 2012).

Over recent years and months, such systems have emerged in the forms of both proprietary and open-source software, transforming the previously arduous, problematic process of using human evaluators to perform sentiment analysis into a technically unchallenging and practical process (Mullen and Collier 2004). Businesses are now able to easily dissect public and consumer opinion regarding their products or services. This study will employ the use of these new, automated sentiment analysis technologies to accurately and effectively determine the findings to the hypotheses established.

2.5 Research Model

Figure 1 depicts the research model proposed for this study. First, it will measure the effect of sentiment strengths on a review's helpfulness dimension. And second, it will scrutinize the effect of emotional load, measuring the intensity of emotions, on the helpfulness dimension.

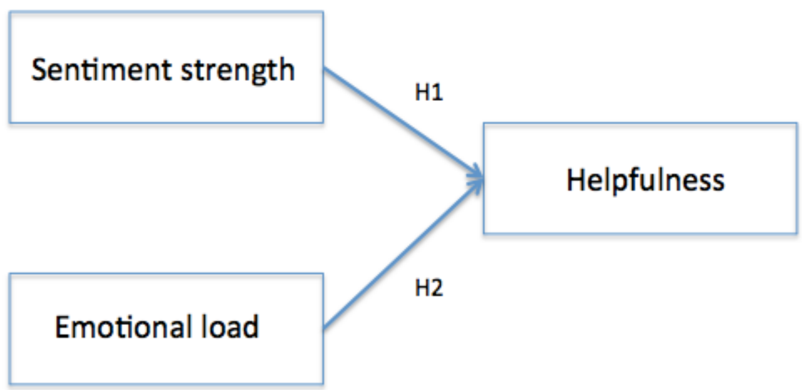

Figure 1: Research model

3. Research Method

This study utilizes panel data, drawn from a large aggregator platform that provides consumer reviews for the travel industry. The dataset is comprised of 4,419 reviews posted over a two-month period and pertains to a total of 100 hotels located within the United States. For each review, the review's title, textual body, numeric star rating and associated helpfulness vote count was reported.

Table 1 and Figure 2 provide the frequencies pertaining to the numeric star rating, and Table 2 and Figure 3 displays the frequency distribution for helpfulness votes.

Table 1: Frequency of reviews, classified by their numeric star rating

Numeric Star Rating	1	2	3	4	5	Total
Number of reviews	190	209	556	1,451	2,013	4,419

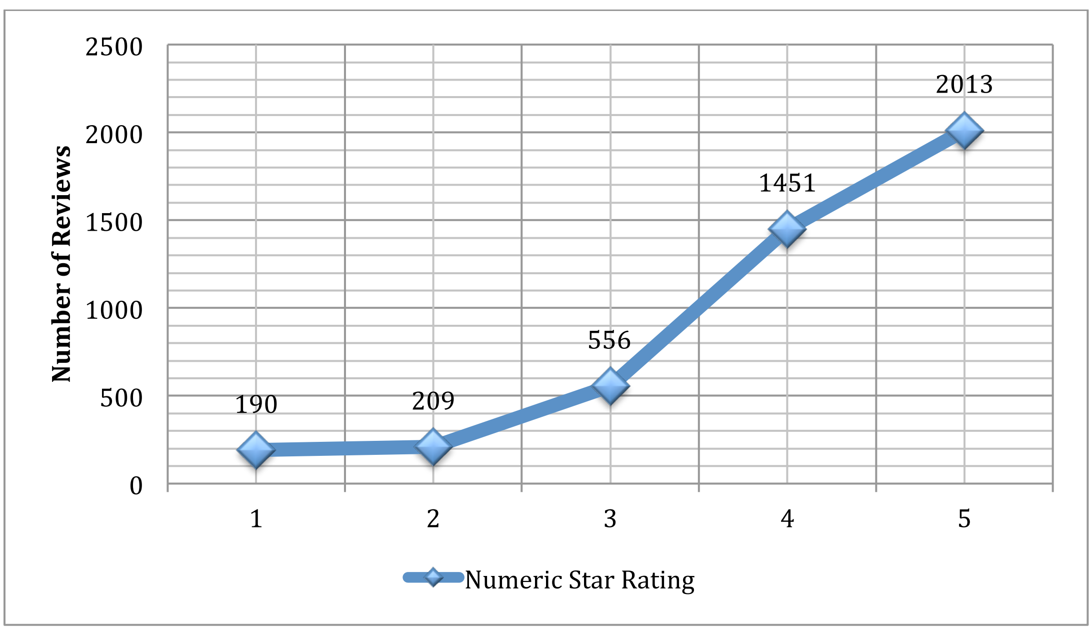

Figure 2: Numeric star rating frequency

Table 2: Frequency of helpfulness votes

Helpfulness votes	0	1	2	3	4	5	6	7	8	9	11	12	14	15	16	22	Total
Number of reviews	2,289	1,172	515	224	109	52	18	15	13	2	2	1	1	3	2	1	4,419

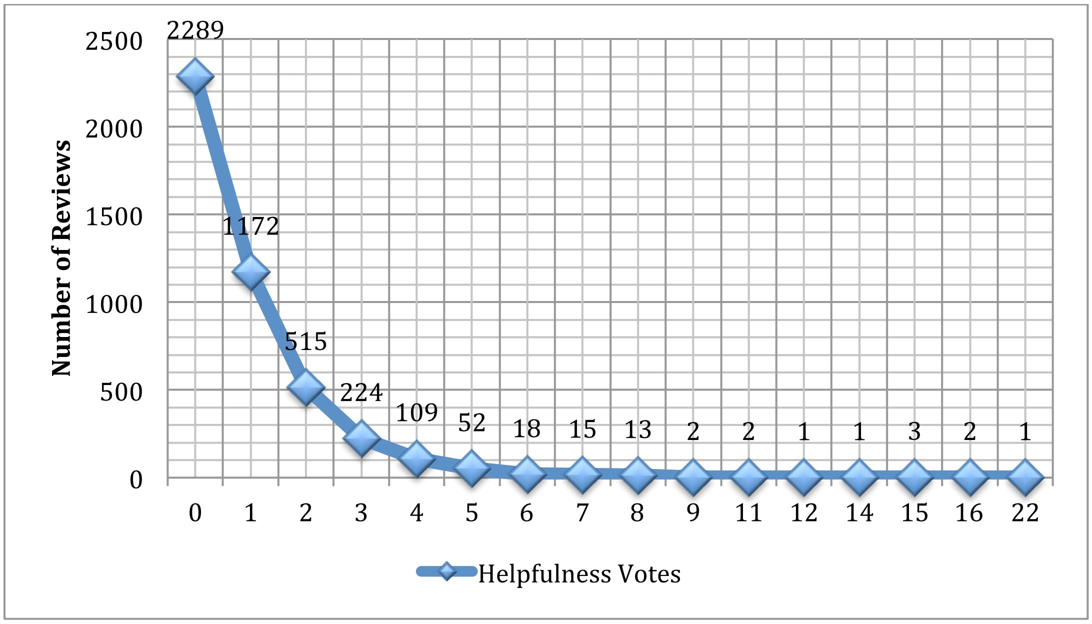

Figure 3: Helpfulness votes frequency

3.1 Variables and Measures

The dependent variable, *helpfulness*, was measured by the total number of helpfulness votes a review received within the specified sampling frame.

Two independent variables were calculated: *sentiment strength* and *emotional load*. *Sentiment strength* represents the summary emotion expressed in a review. More specifically, this variable represents the calculated difference between the positive and the negative sentiment of a review, as determined by the sentiment analysis tool SentiStrength and explained later.

Emotional load represents the intensity of emotion displayed in a review, which combines both polarities (positive or negative) of a review's sentiment; in other words, it focuses on the valence of the emotion expressed, irrespective of its direction. The variable is calculated as the sum of the absolute values for both positive and negative sentiments, as will be detailed later.

In order to account for extraneous effects other than those caused by sentiments, the numeric star rating was designated as the control variable. As mentioned in the literature review, prior research has indicated that numeric star rating may impact on helpfulness (Baek, Ahn and Choi 2012; Korfiatis, Garcia-Bariocanal and Sanchez-Alonso 2012; Forman et al. 2008; Li and Zhan 2011; Pan and Zhang 2011; Mudambi and Schuff 2010; Pavlou and Dimoka 2006). Table 3 represents the descriptive statistics for all variables used in this study.

Table 3: Descriptives

	N	Minimum	Maximum	Mean	Standard Deviation
Helpfulness	4419	0	22	.92	1.444
Sentiment strength	4419	-4.00	4.00	1.1503	1.29037
Emotional load	4419	.00	7.00	2.6384	1.28949
Star rating	4419	1	5	4.11	1.071

3.2 Sentiment Tool Procedure

In order to properly evaluate the sentiment embedded within a review's textual body, this study makes use of sentiment analysis tools, a new breed of software that is able to produce an estimate of the emotional content of textual data. Specifically, this study employs SentiStrength 2.0, an open-source sentiment analysis tool[1].

SentiStrength utilizes a lexical approach in combination with its contained knowledge of grammatical structures. The core of the algorithm comprises a lexicon (i.e., a set of pre-defined words) along with sentiment polarities for each. The lexicon is a combination of various lists,

[1] SentiStrength is available freely for academic research at http://sentistrength.wlv.ac.uk.

including, for example, the Linguistic Inquiry and Word Count (LIWC)[2], developed by Pennebaker, Booth, and Francis (Pennebaker, Mehl and Niederhoffer 2003), and the General Inquirer[3] list of sentiment terms (Stone, Dunphy, Smith and Oglivie 1966). SentiStrength also considers additional rules, such as rules regulating negations (e.g., "not cool"), booster words (e.g., "very cool"), amplifications (e.g., "coooool"), emoticons (e.g., ;-)), and spelling errors (Stieglitz and Dang-Xuan 2013). The latter are automatically corrected.

Since research has shown that individuals are able to simultaneously experience positive and negative attitudes (Norman et al. 2011), SentiStrength is one of the few software tools that generates two outputs: a value for the positive sentiment detected in the text (captured in the variable *sentpos*) and a value for the negative sentiment detected in the text (captured in the variable *sentneg*). The variables *sentpos* and *sentneg* range from +1 to +5 and -1 to -5, respectively. Per definition of the tool, the value +1 indicates that the review does not contain any positive sentiment, whereas a review of +2 contains some positive sentiment, and a value of +5 indicates an extreme positive sentiment. The same applies for negative values, -1 indicating no negative sentiment whereas -5 represents extreme negative sentiment.

The SentiStrength tool has gained some popularity among practice and research. Yahoo!, for example, is one of its users (Thelwall, in press), and a recent study in the IS field analyzing the news clippings during a German election has also utilized the SentiStrength tool (Stieglitz and Dang-Xuan 2013). Its performance is comparable to that of human raters, as demonstrated by Thelwall (in press): "SentiStrength has near-human accuracy on general short social web texts but is less accurate when the texts often contain sarcasm, as in the case of political discussions." (p. 12, Thelwall, in press).

---

[2] See also: http://www.liwc.net

[3] See also: http://www.wjh.harvard.edu/~inquirer/

In order to better illustrate the interworking of SentiStrength, below is an example of a review's textual body marked up with SentiStrength's algorithm for analytical purposes:

> *Spent 3 nights here (+1;-1). Liked that the bedroom was apart from the rest of the room (+3;-1). Great setup with kitchenette (+3;-1). They kept breakfast well stocked (+1;-1). Staff was helpful and friendly (+2;-1). Nice, safe area (+3;-1). Away from main area of town but quick to get anywhere you need (+1;-1).*

SentiStrength Output: *sentpos = +3; sentneg = -1*

The markings, in the form of underlines, denote the words SentiStrength would identify as emotional words, based on its dictionaries. Moreover, SentiStrength treats every text as a collection of sentences, and analyzes each sentence separately. *Sentpos*, or the overall positive estimate for a review, is calculated as the highest positive sentiment expressed across all sentences. Analogously, *sentneg*, or the overall negative estimate for a review, is calculated as the highest (in absolute terms) negative sentiment expressed across all sentences.

Based on the outputs for *sentpos* and *sentneg*, *sentiment strength* can be calculated as *sentpos+sentneg,* and *emotional load* as |*sentpos*|+|*sentneg*|-2, as emotional load uses a transformed scale[4]. Applied to the textual body examined above, *sentiment strength* as well as *emotional load* would be calculated as +2.

Another illustrative example is listed below for further evidence:

> *The hotel was so bad that I cannot even describe it (+1; -2). Everything went wrong (+1;-2). Absolutely everything (+1;-1). I simply hate this hotel (+1, -4).*

SentiStrength Output: *sentpos = +1; sentneg = -4*

Analogously, the calculations for *sentiment strength* and *emotional load* are -3 and +3, respectively.

---

[4] Note: In order to improve the interpretability of the scale, the emotional load variable's response scale was transformed to conform to a range from 0 to +8. This was accomplished by subtracting a value of two from the original range from +2 to +10.

3.3 Analytical Procedure

The study applies a generalized linear model and uses SPSS for analytical purposes. Due to the high count of reviews with zero votes, a negative binominal distribution was used to represent the number of helpfulness votes. Sentiment strength, emotional load and numeric rating were included in the model as predictor variables, conforming to the following regression equation:

$$Log(helpfulness) = \beta_0 + \beta_1 * sentiment\ strength + \beta_2 * emotional\ load + \beta_3 * numeric\ rating + e$$

Accordingly, the following code was generated by SPSS:

```
GENLIN num_helpful_votes WITH sentitext ratings.overall
 /MODEL sentitext ratings.overall INTERCEPT=YES
  DISTRIBUTION=NEGBIN(1) LINK=LOG
  /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL
MAXITERATIONS=100 MAXSTEPHALVING=5
    PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012
ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD
    LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION
(EXPONENTIATED).
```

3.4 Findings

The following findings were produced by SPSS. Table 4 summarizes the model specification, Table 5 alludes to the number of cases processed in the data set, and Table 6 provides descriptive information about the variables used. Table 7 determines if the negative binominal distribution was an appropriate choice to transform the number of helpfulness votes. The values 0.945 and 1.096 are sufficiently close to 1, indicating that the model was not over-specified and that the binomial distribution was the correct choice. A Poisson distribution was fitted into the model, yielding values of 1.688 and 2.101, respectively, indicating over-specification.

Table 4: Model Information

Dependent Variable	Helpfulness
Probability Distribution	Negative binomial (1)
Link Function	Log

Table 5: Case Processing Summary

	N	Percent
Included	4419	100.0%
Excluded	0	0.0%
Total	4419	100.0%

Table 6: Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	**Helpfulness**	4419	0	22	.92	1.444
Covariate	**Sentiment Strength**	4419	-4.00	4.00	1.1503	1.29037
	Emotional Load	4419	.00	7.00	2.6384	1.28949
	Star Rating	4419	1	5	4.11	1.071

Table 7: Goodness of Fit

	Value	df	Value/df
Deviance	4172.714	4415	.945
Scaled Deviance	4172.714	4415	
Pearson Chi-Square	4836.765	4415	1.096
Scaled Pearson Chi-Square	4836.765	4415	
Log Likelihood[b]	-5792.883		
Akaike's Information Criterion (AIC)	11593.765		
Finite Sample Corrected AIC (AICC)	11593.774		
Bayesian Information Criterion (BIC)	11619.340		
Consistent AIC (CAIC)	11623.340		

Dependent Variable: Helpfulness
Model: (Intercept), sentiment strength, emotional load, star ratings
a. Information criteria are in small-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Apart from assessing the appropriateness of the statistical model, SPSS calculated the results for the logistic regression model. Table 8 shows the significance of the overall model, indicating that the overall model was significant. Likewise, Table 9 displays the significance levels for each independent variable, indicating that sentiment strength, emotional load and the numeric rating were influential factors in predicting the number of helpfulness votes (at a significance level of 5 percent).

Since a logarithmic transformation was applied, beta coefficients, as represented in Table 10, column 2, cannot be interpreted as specified. They must be exponentiated, as computed in column 6, “Exp(B)”. For example, the beta coefficient of -.088 for sentiment strength is computed to $e^{-.088}$= .915. Likewise, coefficients for .39 and -.139 are calculated as $e^{.039}$ = 1.039 and $e^{-.139}$ = .870, representing the exponentiated beta coefficients for emotional load and star rating, respectively.

Table 8: Omnibus Test

Likelihood Ratio Chi-Square	df	Sig.
141.629	3	.000

Dependent Variable: helpfulness
Model: (Intercept), sentiment strength, emotional load, star rating
a. Compares the fitted model against the intercept-only model.

Table 9: Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	22.606	1	.000
Sentiment strength	21.693	1	.000
Emotional load	5.539	1	.019
Star rating	40.817	1	.000

Dependent Variable: helpfulness
Model: (Intercept), sentiment strength, emotional load, star rating

Table 10: Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	.454	.0954	.267	.641	22.606	1	.000	1.574	1.306	1.897
Sentiment strength	-.088	.0190	-.126	-.051	21.693	1	.000	.915	.882	.950
Emotional load	.039	.0164	.006	.071	5.539	1	.019	1.039	1.006	1.073
Star rating	-.139	.0218	-.182	-.097	40.817	1	.000	.870	.833	.908
(Scale)	1[a]									
(Negative binomial)	1[a]									

Dependent Variable: helpfulness
Model: (Intercept), sentiment strength, emotional load, star rating
a. Fixed at the displayed value.

As Table 10 indicates, both sentiment strength and emotional load do affect the number of helpfulness votes. More specifically, and all else being equal, if sentiment strength increases by one unit, the number of helpfulness votes increases by 0.915. For emotional load the coefficient is slightly higher. One unit increase in emotional load causes the number of helpfulness votes to increase by 1.039. The numeric rating is also significant in its effect. One unit increase in numeric rating leads to an increase of 0.870 helpfulness votes. When compared in size, emotional load as well as sentiment strength has a stronger influence on the number of helpfulness votes than numeric rating.

4. Discussion

Online reviews are a critical component to the consumer buying decision process and their included characteristics, primarily sentiment, have recently gained great attention in the academic community. Existing studies have established the need to better understand how consumers digest the sentiment and full intensity of emotions embedded within the textual body of online reviews and under what circumstances these are likely to impact the perceived helpfulness of a review. The findings furnished by this paper contribute to an emerging set of sentiment analysis research by addressing and answering these fundamental, yet poorly understood relationships using a dataset containing nearly 5,000 reviews for hotels, an experiential good.

The first portion of the study involved analyzing the sentiment strength included within the textual body and found that, in general, sentiments expressed in reviews are positively associated with the number of helpfulness votes a review receives, thus supporting H1.

The second portion of the study introduced a new conceptualization of an emotional load variable to probe whether the overall intensity of emotions, both positive and negative, within a review affects helpfulness. Results revealed that emotional load (or review passion) is a strong predictor of review helpfulness, thus supporting H2. Specifically, review readers largely prefer reviews with a greater emotional load, to either the positive or negative sentiment strength polarities.

The conclusions in the latter aspect of this study suggest that reviews with a stronger overall emotional intensity would gain more helpfulness votes than those that contain only a strong sentiment. In reality, review readers may appreciate the emotional intensity of a product or service review, thus making it more helpful than one that details either the positive or negative aspects alone.

Both analyses further indicate that sentiment strength and emotional load were better predictors of review helpfulness than the numeric star rating, supporting the already challenged body of research for this review element.

4.1 Managerial Implications

The findings from this study provide several important managerial implications. The growing popularity of online reviews and the extreme impact that sentiment expressed within the textual body of reviews has on helpfulness suggests that businesses should provide an easy avenue and mechanism for consumers to provide their textual feedback for reviews. Yelp.com, TripAdvisor.com, and Amazon.com are several examples of businesses that provide an intuitive interface that supports this expression of consumer sentiment.

Furthermore, implications for website developers are established as adjustments can be made to improve review entry and helpfulness voting mechanisms to encourage more emotionalism in reviews. By adjusting the wording that solicits reviewers to share their experiences from, for example, “How would you describe the hotel’s service?” to a more emotionally engendering question, for example, “How did your experience with the hotel’s service staff make you feel?” reviewers will likely share more passionate answers and provide more information, thus improving the helpfulness of reviews.

Businesses should also support the information search phase of a consumer’s buying decision process by taking note of how consumers evaluate products or services on their website. They may then wish to design an easily accessible method to assist consumers in searching, evaluating, and making a product or service choice; this would assist in better gaining and retaining future customers.

4.2 Limitations and Future Research

This study presents several limitations and creates opportunity for future research. The panel of data analyzed within this paper encompasses reviews from hotel visits, an experiential good, and, as a result, may not be generalizable to reviews for products or services that do not require such emotional investment or evaluation (Mudambi and Schuff 2010). However, the review characteristics used in this study are universal for online user reviews and, as such, the analytical methodology can be easily reproduced and applied to various other products or services. Future research is encouraged to explore reviews across other industries and of other goods to more closely examine and explore the impact of sentiment on review helpfulness.

Additionally, alternative operationalizations are conceivable than the one presented in this study. Instead of applying measures for sentiment strength and emotional load, which aggregate the sentpos and sentneg variables, future research might wish to consider looking at sentpos and sentneg individually as well as explore their interaction effect. Differentiating between positive and negative sentiments might allow researchers to better dissect their respective impact on helpfulness, as it seems conceivable that negative sentiments could behave differently in their effect on helpfulness than positive sentiments.

Furthermore, future studies might wish to look at star rating not only as a linear effect, but also as a quadratic effect; inn other words, researchers may want to hypothesize that low ratings or high ratings hold strong effects on the helpfulness dimension whereas middle ratings have a comparatively lower effect. This alternative model may encompass an equation such as:

$$Log(helpfulness) = \beta_0 + \beta_1 * sentpos + \beta_2 * sentneg + \beta_3 * sentpos * sentneg + \beta_4 * star + \beta_5 * star^2 + e$$

Finally, new sentiment analysis technologies were employed in this paper and proved to be an extremely effective, tractable, and accurate method to extract and analyze a review's sentiment characteristics. While the majority of past research have developed hand-coded methods or used human evaluators to analyze reviews, future studies are encouraged to continue exploring the use of sentiment analysis tools. However, despite advances in sentiment analysis to capture and summarize sentiments efficiently, its limitations on accuracy are unknown. While the tool used within this paper, SentiStrength 2.0, has gained popularity and acceptance within the existing research landscape, future work may wish to incorporate a cross-validation method to enhance confidence in the employment of sentiment analysis tools to extract sentiments.

5. Conclusion

As seen with Amazon.com's influx of $2.7 billion in additional revenue by asking consumers the "Was this review helpful?" question, understanding the helpfulness dimension of online user reviews and harnessing the power of review sentiment is essential. This study compliments previous work on the helpfulness of online reviews and sentiment, answering several extenuating questions posed by researchers and providing explanation to the phenomena of why certain reviews receive more helpfulness votes than others.

The findings represent that sentiment strength and emotional load found within a review's textual body are the root explainers. Specifically, it was revealed that sentiment levels are indeed an extremely influential component in determining a review's helpfulness vote count. The stronger the sentiment strength within a review's textual body, the more helpfulness votes that review tends to receive. Likewise, the results also found that reviews featuring greater overall emotional intensity have a higher probability of receiving more helpfulness votes than others, indicating that people pay greater attention to opinions that are passionate in nature.

By presenting a novel approach to analyzing online review data through the use of SentiStrength, an emergent sentiment analysis tool, and introducing a new and unique conceptualization of sentiment, by way of an emotional load variable, additional areas for future research development were established, thus enabling researchers to continue dissecting and elucidating the relationship between review helpfulness and sentiment.

6. References

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews, *Management Science*, 57(8), 1485-1509.

Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues, *International Journal of Electronic Commerce*, 17(2), 99-126.

Berger, J. (2012). Bad Reviews can Boost Sales. Here's Why, *Harvard Business Review*, The Magazine, March, Retrieved March, 2014, from http://hbr.org/2012/03/bad-reviews-can-boost-sales-heres-why/ar/1

Bly, L. (2012). Does TripAdvisor play dirty with hotel reviews? *USAToday*, October 24, Retrieved March, 2014, from http://travel.usatoday.com/destinations/dispatches/post/2011/10/does-tripadvisor-play-dirty-with-hotel-reviews/557148/1

Cao, Q., Duan, W. J., & Gan, Q. W. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach, *Decision Support Systems*, 50(2), 511-521.

CBSNews (2011). "Dirtiest Hotel in America" Looks to Clean Up, *CBSNews*, January 26, Retrieved March, 2014, from http://www.cbsnews.com/news/dirtiest-hotel-in-america-looks-to-clean-up/

Chen, P.-Y., Dhanasobhon, S. & Smith, M. D. (2008). All Reviews Are Not Created Equal: The Disaggregate Impact of Reviews on Sales on Amazon.com, *Working Paper*, Retrieved March, 2014, from http://ssrn.com/abstract=918083

Chen, Y., Fay, S. & Wang, Q. (2003). Marketing Implications of Online Consumer Product Reviews, *Working Paper*, University of Florida, Retrieved March, 2014, from http://plaza.ufl.edu/faysa/review.pdf

Chen, Y., & Xie, J. (2008). Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix, *Management Science*, 54(3), 477–491.

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews, *Journal of Marketing Research*, 43(3), 345-354.

Chintagunta, P. K., Gopinath S., & Venkataraman S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets, *Marketing Science*, 29(5), 944-957.

Duan, W. J., Gu, G., & Whinston, A. B. (2008). Do online reviews matter? - An empirical investigation of panel data, *Decision Support Systems*, 45(4), 1007-1016.

Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality, *System Sciences*, 3119-3128

Farnham, A. (2013). Filthiest Hotel Not Defamed, Judge Rules, *ABC News*, September 4, Retrieved March, 2014, from http://abcnews.go.com/Business/tripadvisor-calling-hotel-dirtiest-defamation-judge-rules/story?id=20155418

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets, *Information Systems Research*, 19(3), 291–313.

Ganu, G., Elhadad, N. & Marian, A. (2009). Beyond the Stars: Improving Rating Predictions using Review Text Content*, Twelfth International Workshop on the Web and Databases* (WebDB 2009), June 28, Providence, Rhode Island, USA, Retrieved March, 2014, from http://people.dbmi.columbia.edu/noemie/papers/webdb09.pdf

Gesenhues, A. (2013). Survey: 90% of Customers say Buying Decisions are Influenced by Online Reviews, *Marketing Land*, April 9th, Retrieved March, 2014, from http://marketingland.com/survey-customers-more-frustrated-by-how-long-it-takes-to-resolve-a-customer-service-issue-than-the-resolution-38756

Ghose A., & Ipeirotis, P.G. (2010). Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics, *Knowledge and Data Engineering, IEEE Transactions,* 23(10), 1498-1512.

Ghose A., Ipeirotis P.G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowd-Sourced Content, *Marketing Science*, 31(3), 493-520.

Hu, N., Koh, N. S., & Reddy, S. K. (2014). 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, 42-53.

Hu, N., Bose, I., Gao, Y., & Liu, L. (2011). Manipulation in digital word-of-mouth: A reality check for book reviews, *Decision Support Systems*, 50(3), 627–635.

Kennedy, H. (2012). Perspectives on Sentiment Analysis, *Journal of Broadcasting & Electronic Media*, 56(4), 435-450.

Kim, S. M., Pantel P., Chklovski T., & Pennacchiotti M. (2006). Automatically Assessing Review Helpfulness, *Conference on Empirical Methods in Natural Language*, 423-430.

Korfiatis, N., Garcia-Bariocanal, E., & Sanchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content, *Electronic Commerce Research and Applications*, (11:3), 205-217.

Li, X. & Hitt, L. (2008). Self-Selection and Information Role of Online Product Reviews. Information Systems Research, *Information Systems Research*, 19(4), 456-474.

Li, J., & Zhan, L. J. (2011). Online Persuasion: How the Written Word Drives WOM Evidence from Consumer-Generated Product Reviews, *Journal of Advertising Research*, 51(1), 239-257.

Li, M., Huang, L., Tan, C.-H., & Wei, K. K. (2013). Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features, *International Journal of Electronic Commerce*, 17(4), 101-136.

Luca, M. (2011). Reviews, Reputation, and Revenue: The Case of Yelp.com, *Working Paper*, 12-016, Harvard Business School, Retrieved March, 2014, from http://www.hbs.edu/faculty/Publication%20Files/12-016.pdf

Mishra, N., & Jha, C.K. (2012). Classification of Opinion Mining Techniques, *International Journal of Computer Applications*, 53(13), pp. 1-6.

Mudambi, S. M., & Schuff, D. (2010). What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com, *MIS Quarterly*, 34(1), 185-200.

Mukherjee, S. (2012). Sentiment Analysis: A Literature Survey, *Working Paper*, Indian Institute of Technology, Bombay, Retrieved March, 2014, from http://www.cfilt.iitb.ac.in/resources/surveys/SA-Literature%20Survey-2012-Subhabratam.pdf

Mullen, T., & Collier, N. (2004). Sentiment Analysis Using Support Vector Machines With Diverse Information Sources, *Proceedings of EMNLP 2004*, 412-418.

Nanli Z., Ping Z., Weiguo L., & Meng C. (2012). Sentiment Analysis: A Literature Review, *Management of Technology* (ISMOT), 572-576.

Nelson, P. (1970). Information and Consumer Behavior, *Journal of Political Economy*, 78(2), 311-329.

Nelson. P. (1974). Advertising as Information, *Journal of Political Economy*, 82(4), 729- 754.

Norman, G. J., Norris, C., Gollan, J., Ito, T. A., Hawkley, L. C., Larsen, J. T., Cacioppo, J. T., & Berntson, G. G. (2011). Current emotion research in psychophysiology: The neurobiology of evaluative bivalence, *Emotion Review*, 3, 3349-359.

Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews, *Journal of Retailing*, 87(4), 598-612.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis, *Foundations and Trends in Information Retrieval*, 2(1), 1-135.

Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation, *Information Systems Research,* 17, 392-414.

Pennebaker, J., Mehl, M., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547-577.

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), 234-243.

Schlosser, A. E. (2011). Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments, *Journal of Consumer Psychology*, 21(3), 226–239.

Senecal, S. & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices, *Journal of Retailing*, 80(2), 159-169.

Siering, M. & Muntermann, J. (2013). What Drives the Helpfulness of Online Product Reviews? From Stars to Facts and Emotions, *Wirtschaftsinformatik*, Retrieved March, 2014, from http://aisel.aisnet.org/wi2013/7

Spool, J. M. (2009). The Magic Behind Amazon's 2.7 Billion Dollar Question, U*ser Interface Engineering*, Retrieved March, 2014, from March 17, http://www.uie.com/articles/magicbehindamazon/

Stieglitz, S. & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior, *Journal of Management Information Systems*, 29(4), 217-247.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). T*he general inquirer: A computer approach to content analysis*. Cambridge, MA: The MIT Press.

Stravroula K., Vinson, D. P., & Vigliocco G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words, *Cognition*, 112(3), 473-481.

Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews, *Expert Systems with Applications*, 36(7), 10760-10773.

Thelwall, M. (in press). Heart and Soul: Sentiment Strength Detection in the Social Web with SentiStrength, book chapter, Retrieved March, 2014, from http://sentistrength.wlv.ac.uk/documentation/SentiStrengthChapter.pdf

Weiss, A. M., Lurie, N. H., & Macinnis, D. J. (2008). Listening to strangers: Whose responses are valuable, how valuable are they, and why?, *Journal of Marketing Research*, 45(4), 425-436.

Willemsen, L. M, Neijens, P. C., Bronner, F., & de Ridder, J. A. (2011). "Highly Recommended!" The Content Characteristics and Perceived Usefulness of Online Consumer Reviews, *Journal of Computer-Mediated Communication*, 17, 19-38.

Zendesk (2013). The Impact of Customer Service on Customer Lifetime Value, *Zendesk*, Retrieved March, 2014, from http://www.zendesk.com/resources/customer-service-and-lifetime-customer-value

Zhang, X. M. & Dellarocas, C. (2006). The Lord of the Ratings: How a Movie's Fate Is Influenced by Reviews, *Proceedings of the 27th International Conference on Information Systems (ICIS)*, Milwaukee, Association for Information Systems.

Zhu, F. & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics, *Journal of Marketing*, 74, 133-148.