Abstract

With the prevalence of online reviews, an unwieldy glut of information may be in some cases presented to the consumers. To address the problem of information overload, the voting mechanism for online review helpfulness has been used by many websites. This study focuses on the effects of emotional valence and arousal in review content on consumer perceptions of online review helpfulness, which has not received enough attentions before. Using real-world data and econometric model, we found emotions within review content significantly influence review helpfulness. Furthermore, the interaction effect of emotional valence and arousal is investigated in details. This study makes a theoretical contribution to online review helpfulness research and also provides guidance to e-commerce companies in managing online reviews.

Keywords

Online review, helpfulness, emotion, valence, arousal.

Introduction

Nowadays, many websites including e-commerce, social networking services, and third-party online review sites, allow consumers to give star rating and write comments on products or services. A Nielsen (2012) report surveying more than 28,000 Internet users in 56 countries found that online consumer reviews are the second most-trusted source of brand information (after recommendations from friends and family). With the prevalence of online reviews, an unwieldy glut of information may be in some cases presented to the consumers (Godes & Silva 2012; Mudambi & Schuff 2010). Therefore, the social voting mechanism has been introduced to many online reviews platform allowing users to give “helpful/useful” votes to the reviews they read in order to signal the credibility of the reviews (Ghose & Ipeirotis 2011; Otterbacher 2009), which help consumers to search and view the reviews to overcome the information overload (Willemsen, Bronne, & Ridder 2011).

In contrast to the research on the outcomes of online reviews (e.g., Chevalier & Mayzlin 2006; Duan, Gu, & Whinston 2008), recent studies have shifted attention to the antecedents of online reviews (Cheung & Thadani 2012; Stephen & Lehmann 2009), particularly investigating what characteristics lead to a review that is perceived more helpful by online consumers (Baek, Ahn, & Choi 2013; Ghose & Ipeirotis 2011; Mudambi & Schuff 2010; Yin, Bond, & Zhang 2013). However, there are a few of shortcomings in prior research. First, these studies focused on quantitative characteristics of individual review, such as the numerical rating and the length (Korfiatis, García-Bariocanal, & Sánchez-Alonso 2012; Mudambi & Schuff 2010), but ignored the qualitative or textual characteristics of review content, especially emotions within it. Second, most of existing studies focus on the context of physical goods instead of services. Due to
higher uncertainty and risk associated with services, consumers rely heavily on consumer reviews to assess services quality prior to purchase (Racherla & Friske 2012).

To address the above limitations, drawing from the circumflex structure of affect (Russell 1980), this paper focused on the effects of emotional valence, arousal and the interaction between them on review helpfulness in the context of hotel service. Then, the 16929 reviews of 307 hotels from the Yelp.com were used to test the hypotheses empirically. The rest of the paper is organized as follows. Section 2 gives the summary of the existing literatures. The research model and hypotheses development are presented in Section 3. The research method, including data collection, and econometric model are described in Section 4. Section 5 presents the preliminary results of econometric analysis. Finally, the discussion and conclusion are given in Section 6.

Literature Review

Studies on Online Review Helpfulness

The review helpfulness is defined as a measure of perceived value for consumers in the decision-making process (Mudambi & Schuff 2010), and can also be viewed as a reflection of the diagnosticity of the information (Jiang & Benbasat 2004). A rich stream of studies has been made about ‘what makes a helpful online review’. While most of them focus on two dimensions, say review characteristics and reviewer characteristics.

For review characteristics, it is confirmed positive relationship between review rating/length and perceived review helpfulness (Pan & Zhang 2011). Schlosser (2011) researched on rating valence/extremity, review length, and the two-sidedness of review. Recently, Text mining has also been used for the research of review helpfulness (Beak 2013; Cao, Duan, & Gan 2011).

For reviewer characteristics, a stream of studies focuses on ‘who said it’ (e.g., Cheung, Sia, & Kuan 2012; Racherla & Friske 2012). Ghose & Ipeirotis (2011) found that the disclosure of reviewer's identity would significantly impact the helpful votes received. Connors and Mudambi (2011), and Schlosser (2011) also found significant influence of expertise of the reviewer on review helpfulness. Based on dual process theories, a few studies investigate the effects of reviewer credibility including the expertise and trustworthiness on review helpfulness (Baek et al. 2013; Cheung et al. 2012; Racherla & Friske 2012).

The Role of Emotion in Online Content

Emotion refers to a mental state of readiness that arises from cognitive appraisals of events or thoughts' (Bagozzi, Gopinath, & Nyer 1999), which are often associated with specific resulting action tendencies and behaviors (Lerner & Keltner 2000). There is rich study on how should emotions be conceptualized. Modern research has reconceptualized emotion as having two dimensions, valence (positive or negative) and arousal (high or low) (Bradley & Lang 1994, Cacioppo, Petty, Losch, & Kim 1986, Russell 1980). For example, a consumer presented with product review showing anger, may report the combination of negative valence and high arousal.

Emotion is a major factor in providing valuable implicit or explicit information for making fast and optimal decision (Bechara & Damasio 2005). And as one type of effect, it has the characteristics of triggering a short but intense effect on the individual (Frijda 1994). It has been found that emotionally charged messages tended to be shared more often (Stieglitz & Xuan 2013). In addition, people may share emotionally charged content to make sense of their experiences, reduce dissonance, or deepen social connections (Festinger, Riecken, & Schachter 1956; Peters & Kashima 2007).

Recently, Yin (2013) confirmed the effect of discrete emotion of anxiety and anger on perceived helpfulness above and beyond ratings or information content alone. Cao (2011) used text mining to explore determinants of helpfulness, which showed emotion played a role.

However, most of them focused on the effect of single dimension of emotions. Some of them even arrive at mixed conclusion. Yin (2013) found anxiety within a review represent more involved efforts of reviewer, thus the review could be perceived more helpful than angry review. While Ahmad (2013) demonstrated reviews containing more anxiety would be perceive less helpful because it is deem as less certain. Discrete
emotion fail to capture the continuous change of emotional valence and arousal, which may explain why they come to mixed results.

**Theory and Hypotheses**

The past works have shown that, emotion can substantially influence the way that reviews are processed (e.g., Kuan, Hui, Prasarnphanich, & Lai 2011), and rich information embedded in review text can be useful in explaining what constitutes a ‘helpful’ review (Cao et al. 2011; Pavlou & Dimoka 2006). Numerous researchers have suggested fundamental dimensions for the classification of emotions. Among these dimensions, valence and arousal have been consistently identified as most accepted (Niedenthal 2008; Russell 1980). Valence describes the extent to which an experience is pleasant (positive valence) or unpleasant (negative valence), while arousal describes the extent to which an actor is activated or deactivated (Niedenthal 2008).

It has been demonstrated that content containing more positive emotion is more viral and more likely to be shared (Berger and Milkman 2011). Thus, this kind of reviews could be perceived more helpful. Yin(2013) confirmed that negative emotion (anger & anxiety) have impact on online review helpfulness, and these two enhance review helpfulness overall. Heath, Bell and Sternberg (2001) found information spreading widely usually contain negative emotion. This leads us to hypothesize,

H1: Review containing more positive or negative emotion would be perceived more helpful.

High-arousal state has been shown to increase action related behaviors such as getting up to help others (Gaertner and Dovidio 1977) and responding faster to offers in negotiations (Brooks and Schweitzer 2011). Berger and Milkman (2011) found that arousal can affect social transmission, which is that content evoking more of low-arousal emotion is less likely to be shared, and content evoking high-arousal emotion is more viral. Berger (2011) found arousal increases social transmission of information. This leads us to hypothesize,

H2: Review evoking high arousal would be perceived more helpful.

There is long debate about interaction between valence and arousal (Eder & Rothermund 2010). However, for researches on emotion for review helpfulness, the majority of them only focused on valence, ignoring the effect of arousal and the interaction between them. Berger and Milkman (2011) found that the boost in arousal mediated the effect of the amusement condition on sharing. There is interaction between arousal and other kind of emotion for sharing. In addition, people passed along information that was exaggeratedly negative (Heath 1996). In our context, reviews containing both overmuch negative emotion as well as high arousal show irrationality thus could be perceived less helpful. This leads us to hypothesize,

H3: Reviews containing increasingly negative emotion would be perceived less helpful in the presence of higher arousal.

Prior research in psychology and organizational studies reveal that people respond differently to positive and negative stimuli, and negative events tend to elicit cognitive responses than neutral or positive events (Taylor 1991). Furthermore, researchers have argued that positive and negative emotion cannot be considered as a single continuum, but rather must be thought of as qualitatively distinct phenomena (Berscheid 1983). In research for helpfulness, a common finding is negativity bias, whereby negative reviews tend to be more influential. Negative information is also weighted more heavily in the attribution of evaluations to others (Abelson & Kanouse 1966). Given that, the study of the interaction between valence and arousal focuses on reviews appearing negative emotion.

Following prior work on review helpfulness, our analysis controlled for a series of relevant variables.

For review characteristics, review length is measured by word counts (Pan & Zhang 2011; Chevalier & Mayzlin 2006; Gupta and Harris 2010). Rating representing particular personal experience has influence on helpfulness (Forman, Ghose & Wiesenfeld 2008; Mudambi and Schuff 2010). Readability, measured by Coleman–Liau Index (CLI), is related to representational quality of the user review (Otterbacher 2009).
For reviewer characteristics, the number of social ties should be considered as a pivotal factor for consumers’ perception of review helpfulness (Racherla & Friske 2012; Wang 2010). Yelp can designate prolific reviewers with elite badge, which enhance the social presence of the reviewer (Forman et al. 2008; Racherla & Friske 2012). These two factors helps them gain creditability thus promotes their review helpfulness. (Shen 2010; Tang, Gu, & Whinston 2012).

Some control variables, the elapsed days of a review since posted (Pan & Zhang 2011), the popularity measured by the review counts of the hotel (Ghose & Ipeirotis 2011), and the average rating (Yin, Bond, & Zhang 2013) are offered.

### Research Method

#### Data and Variables

Data for this research were collected from Yelp.com in September 2012, which is one of the top ratings and review sites with more than 50 million unique visitors a month. Yelp.com was chosen because the website has extensive social networking features and provides the entire history of all reviews posted for restaurants, hotels, and a variety of other local business services. We collected the entire history of review data up to September 2012 for all the hotels in San Francisco. In the end, we collected a total of 16269 reviews across the 307 hotels.

The valence and arousal of emotion within the content of review are calculated by Revised Dictionary of Affect in Language (Whissell 2009). Words in this dictionary are given value of valence and arousal, which based on abundant empirical studies. We averaged the value of words’ valence for any words within review fall in the scope of the dictionary. Then the averaged value is regard as the valence of the review. The process of calculating review's arousal is same as valence’s.

Whissell (2009) provided normative values for natural English texts and had a valence mean of 1.85. So we extract data in condition that reviews’ valence is significantly less than 1.85 in whole data, which results in negative emotion dataset including 4128 reviews. For better studying the effect of negative emotion, we calculated the absolute value of valence minus 1.85, which is measured as the review valence in this dataset. Then in this dataset, the larger the valence is, the more negative the emotion within the review is. The largest valence of 0.85 means most unpleasant and the valence of 0 means neutral emotion.

We summarize all variables and their operationalization in Table 1, and also illustrate how the data were used to operationalize the variables.

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Variables</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable</strong></td>
<td>Review helpfulness</td>
<td>The number of ‘useful’ votes on each review;</td>
</tr>
<tr>
<td><strong>Independent Variables</strong></td>
<td>Valence</td>
<td>Calculated by Revised Dictionary of Affect in Language; (3= most pleasant, 1= most unpleasant)</td>
</tr>
<tr>
<td></td>
<td>Arousal</td>
<td>Calculated by Revised Dictionary of Affect in Language; (3= highly activated, 1= no activation)</td>
</tr>
<tr>
<td><strong>Control Variables</strong></td>
<td>Review length</td>
<td>The number of words in a review;</td>
</tr>
<tr>
<td>– Review Characteristics</td>
<td>Readability</td>
<td>Coleman-Liau Index of Readability;</td>
</tr>
<tr>
<td></td>
<td>Rating</td>
<td>The rating given by the reviewer;</td>
</tr>
<tr>
<td><strong>Control Variables</strong></td>
<td>Reviewer elite badge</td>
<td>Whether or not a review posted by the “elite” reviewer; (1= yes, 0= no; )</td>
</tr>
<tr>
<td>– Reviewer Characteristics</td>
<td>Number of social ties</td>
<td>The number of reviewer’s friends;</td>
</tr>
<tr>
<td><strong>Control variables</strong></td>
<td>Elapsed days</td>
<td>The difference between the date</td>
</tr>
</tbody>
</table>
Table 1. Variables and Operationalization

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hotel level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Popularity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 presents a summary of descriptive statistics for the variables in the sample.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.Valence</td>
<td>1.89</td>
<td>0.06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.Arousal</td>
<td>1.63</td>
<td>0.05</td>
<td>0.30</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.Social ties</td>
<td>119.77</td>
<td>404.29</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.Elite badge</td>
<td>0.24</td>
<td>0.43</td>
<td>-0.02</td>
<td>-0.05</td>
<td>0.34</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.Review length</td>
<td>165.50</td>
<td>130.43</td>
<td>-0.22</td>
<td>-0.03</td>
<td>0.06</td>
<td>0.09</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.Rating</td>
<td>3.61</td>
<td>1.22</td>
<td>0.41</td>
<td>-0.01</td>
<td>0.06</td>
<td>0.09</td>
<td>-0.14</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>7.Readability</td>
<td>5.44</td>
<td>2.61</td>
<td>0.04</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.03</td>
<td>0.02</td>
<td>0.08</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Econometric Model

The dependent variable is helpful votes received. As Figure 2 shows, a large proportion of reviews (47.34%) did not receive a single helpful vote; the data demonstrates ‘over-dispersion’ with the variance of the ‘helpful’ votes much larger than its mean. To handle this, the negative binomial model, one of Poisson model variations, will be fitted to model this dataset to address above problem (Greene 1994; Stieglitz & Xuan 2013).

Figure 1. Distribution of ‘helpful’ Votes for the Review

(Mean = 1.37; Variance = 6.15)
For the whole dataset, to test our hypotheses, the squared valence is added to the model. The negative binomial model is as follows,

\[
\mu_{jk} = \beta_0 + \beta_1 \text{Valence}_{jk} + \beta_2 \text{Squared Valence}_{jk} + \beta_3 \text{Arousal}_{jk} \\
+ \beta_4 \ln(\text{Network Centrality})_{jk} + \beta_5 \text{Elite badge}_{jk} + \beta_6 \ln(\text{Depth})_{jk} + \beta_7 \text{Rating}_{jk} \\
+ \beta_8 \text{Easy of Understanding}_{jk} + \beta_9 \ln(\text{Elapsed Days})_{jk} + \beta_{10} \text{Popularity}_{jk} \\
+ \beta_{11} \text{Average Rating}_{jk} + \epsilon_{jk}
\]

Where \( X_{jk} \) is a set of explanatory variables (including control variables) of review k of hotel j, that determine the conditional probabilities (\( \mu_{jk} \)) in the negative binomial process, \( \epsilon_{jk} \) is the error term and \( \beta \) is the vectors of coefficients need to be estimated by econometric regression.

For negative emotion dataset, to test our hypotheses, the interaction between negative valence and arousal is added to the model. The negative binomial model is as follows,

\[
\mu_{jk} = \gamma_0 + \gamma_1 \text{Valence}_{jk} + \gamma_2 \text{Arousal}_{jk} + \gamma_3 \text{Valence} \ast \text{Arousal}_{jk} \\
+ \gamma_4 \ln(\text{Network Centrality})_{jk} + \gamma_5 \text{Elite badge}_{jk} + \gamma_6 \ln(\text{Depth})_{jk} + \gamma_7 \text{Rating}_{jk} \\
+ \gamma_8 \text{Easy of Understanding}_{jk} + \gamma_9 \ln(\text{Elapsed Days})_{jk} + \gamma_{10} \text{Popularity}_{jk} \\
+ \gamma_{11} \text{Average Rating}_{jk} + \epsilon_{jk}
\]

Where \( X_{jk} \) is a set of explanatory variables (including control variables) of review k of hotel j, that determine the conditional probabilities (\( \mu_{jk} \)) in the negative binomial process, \( \epsilon_{jk} \) is the error term and \( \beta \) is the vectors of coefficients need to be estimated by econometric regression.

**Preliminary Results**

To test the hypotheses, the econometric model is analyzed using STATA program. In the whole dataset analysis, as Figure 2 shows, negative binomial model fits much better than Poisson model.
Figure 2. Fitting Curve of Negative Binomial Regression and Poisson for Whole Dataset

Table 3 shows that H1 hypothesis is supported ($\beta_1 = -11.070$, $p<0.05$ & $\beta_2 = 2.831$, $p<0.05$). This indicates that reviews containing more positive or negative emotion are perceived more helpful. H2 hypothesis also is supported ($\beta_3 = 0.776$, $p<0.01$), which suggests that high arousal increases perceived review helpfulness.

<table>
<thead>
<tr>
<th>Whole dataset</th>
<th>Negative Binomial Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Variables</td>
<td>Coefficient</td>
</tr>
<tr>
<td>Valence</td>
<td>-11.070*</td>
</tr>
<tr>
<td>Squared valence</td>
<td>2.831*</td>
</tr>
<tr>
<td>Arousal</td>
<td>0.776**</td>
</tr>
<tr>
<td>Social ties</td>
<td>0.251**</td>
</tr>
<tr>
<td>Reviewer Elite badge</td>
<td>0.117**</td>
</tr>
<tr>
<td>Review length</td>
<td>0.533**</td>
</tr>
<tr>
<td>Rating</td>
<td>-0.113**</td>
</tr>
<tr>
<td>Readability</td>
<td>-0.010*</td>
</tr>
<tr>
<td>Review Elapsed Days</td>
<td>0.125*</td>
</tr>
<tr>
<td>Popularity</td>
<td>-0.001**</td>
</tr>
<tr>
<td>Average Rating</td>
<td>0.123</td>
</tr>
</tbody>
</table>

Table 3. Results of Negative Binomial Regression in Whole Dataset

In the negative emotion dataset, as Table 4 shows, H3 hypothesis is supported ($\gamma_3 = -31.671$, $p<0.05$). This means reviews in possession of negative emotion would be perceived less helpful in the presence of higher arousal significantly. The results of both valence ($\gamma_1 = 52.867$, $p<0.05$) and arousal($\gamma_2 = 2.376$, $p<0.01$) are
consistent with whole dataset analysis. In addition, all control variables, except readability and average rating, stay the same with whole dataset analysis both in significance and plus-minus sign.

<table>
<thead>
<tr>
<th>Negative emotion dataset</th>
<th>Negative Binomial Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Variables</td>
<td>Coefficient</td>
</tr>
<tr>
<td>Valence</td>
<td>52.867*</td>
</tr>
<tr>
<td>Arousal</td>
<td>2.376**</td>
</tr>
<tr>
<td>Valence*arousal</td>
<td>-31.671*</td>
</tr>
<tr>
<td>Social ties</td>
<td>0.196**</td>
</tr>
<tr>
<td>Reviewer Elite badge</td>
<td>0.130*</td>
</tr>
<tr>
<td>Review length</td>
<td>0.489**</td>
</tr>
<tr>
<td>Rating</td>
<td>-0.168**</td>
</tr>
<tr>
<td>Readability</td>
<td>-0.004</td>
</tr>
<tr>
<td>Review Elapsed Days</td>
<td>0.188**</td>
</tr>
<tr>
<td>Popularity</td>
<td>-0.001**</td>
</tr>
<tr>
<td>Average Rating</td>
<td>0.102**</td>
</tr>
<tr>
<td>Likelihood-ratio test of alpha=0: chibar2(01) = 2203.38 p &lt;0.01; *, p&lt;0.05; **, p&lt;0.01;</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Results of Negative Binomial Regression in Negative Emotion Dataset

Discussion and Conclusion

Utilizing real-world reviews data from Yelp.com, the empirical results provided supports for all three hypotheses. In the whole dataset analysis, the results are consistent with past researches. Both more positive and negative emotions enhance perceived review helpfulness. The negative emotion dataset analysis supports our hypotheses and it does make sense in our daily experience. It is hard to approve the information appearing overmuch unpleasant as well as evoking.

Focusing on emotional valence and arousal other than discrete emotions, we suggest that both of them and their interaction have effect on perceived review helpfulness. We provided theoretical contribution to online review helpfulness research, which, to our knowledge, is the first in this area. In addition, researches in this area usually focus on discrete emotions, which cannot to capture the continuous changes in emotion. So, they may come to mixed results.

For other variables, readability become insignificant but its plus-minus sign is still identical with whole dataset analysis. The average rate become significant in negative emotion dataset, this may call for further research. The value and plus-minus sign of other variables all make sense and are consistent with past findings in both two analyses, which makes our results robust.

This study makes contribution to practical uses. The results suggest that do not express desperately (overmuch unpleasant and high-arousal) if you want your information perceived more helpful or to enhance its transmission. It is particularly useful to those who want to make their voice spread more widely and approved by more in Internet.

Acknowledgements

The first author acknowledges the Program for Young Excellent Talents of UIBE for financial supports.
REFERENCES


