An Examination of the Joint Impacts of Review Content and Reviewer Characteristics on Review Usefulness—the Case of Yelp.com

Emergent Research Forum papers

Li Jing
The Hong Kong Polytechnic University
aaron.li@connect.polyu.hk

Xu Xin
The Hong Kong Polytechnic University
xin.xu@polyu.edu.hk

Eric Ngai
The Hong Kong Polytechnic University
mswtngai@polyu.edu.hk

Abstract

We examine the interaction effects of review content (certainty) and reviewer characteristics (popularity and expertise) on consumer judgment of review usefulness. Utilizing an extended Poisson regression model, we empirically tested the joint impacts based on 5426 reviews from Yelp.com about 968 restaurants. Our results indicate that (1) reviewer popularity negatively interacts with certainty to affect review usefulness and (2) in contrast, reviewer expertise positively interacts with certainty to affect review usefulness. These findings add new insights into online review research and offer practical implications for online review platforms.

Keywords

Review usefulness, certainty tone, reviewer popularity, reviewer expertise, social media.

Introduction

With the proliferation of online communications facilitated by social media, nowadays a vast number of customer reviews for a wide range of products and services are generated online. Given that the increasing availability of a large scale of reviews could create information overload for consumers (Jones et al. 2004), social media platforms such as Yelp.com provide peer voting system to allow reviewers to grant 'useful' votes to a review. Although this system can identify useful reviews in an ex post way, the accumulation of votes needs time and may delay the right information—here the useful reviews— from reaching the right audience. An ex ante approach to the prediction of the usefulness of a review will help social media platforms to better screen and select e-word-of-mouth (e-WOM) to feed online visitors with limited time and cognitive resources.

Existing research on review usefulness / helpfulness mostly focus on its determinants, including review characteristics—e.g., emotions (Yin et al. 2014) and reviewer characteristics—e.g., self-identity disclosure (Forman et al. 2008). While these studies have been instrumental in enhancing our understanding of review usefulness, several interesting issues are still not well answered. First, these studies only verified the solitary impacts of review and reviewer features, little is known about their interaction effects. Second, the majority of existing studies measure the review helpfulness as the percentage of helpful votes (e.g., Mudambi & Schuff 2010), few directly models the count of helpfulness votes (e.g., Wei et al. 2014). Second, to the best of my knowledge, few studies directly investigating the role of certain tone as a predictor of review usefulness, except for Yin et al. (2014) who indirectly tested the mediation mechanism of certainty in order to differentiate impacts of anger (high certainty) and anxiety (low certainty) on the review helpfulness. To fulfill the identified literature gaps, our paper intends to see whether the number of usefulness votes of a review in certainty tone is moderated by the popularity and expertise of a reviewer. Moreover, the method of predicting the count of usefulness votes in existing studies is not well implemented. Wei et al. (2014) did not match the timeline between the content characteristics and the source characteristics. Given the methodological limitation, we intend to make improvements in estimation by using a more appropriate method.

Given the research gaps and methodological limitations identified above, our research aims to examine the joint effects of both content features and source characteristics on the count of review usefulness votes with a refined empirical method. In particular, in addition to the existing sentiment features of the review content examined (e.g., anger and anxiety), we add one more content feature—certainty—and examine the interaction effects between certainty and source features (e.g., reviewer expertise, popularity and status) on review usefulness. To examine review usefulness, we integrate dual-process theory (Angst and Agarwal 2009) with social influence theory (Kelman 1961) to explain the underlying rationales behind our proposed relationships.

Our empirical model is validated based on 5426 reviews in the last quarterly of 2013 from 968 restaurants in city Phoenix from Yelp.com using stepwise zero-inflated negative binomial (ZINB) Poisson regression with variable length of observation periods. We find that reviewer expertise (no. of reviews written) enhances the persuasion power of the certainty tone in the reviews while interestingly, reviewer popularity (no. of fans) weakens the impact of certainty. Our study and findings contribute to the existing literature on the review usefulness (e.g., Mudambi and Schuff 2010).

Empirical Model and Method

Empirical Model

We use the ZINB model to estimate the impacts of content features, reviewer characteristics, and the interaction effects between the two on the count of usefulness votes. This approach is appropriate to model the over-dispersed count data with excess zeros. We also control for the variable length of the observation periods of the reviews. The NB model is to model the number of usefulness votes as specified in equation (1) while the logit model is to model the likelihood of a review receiving zero votes as specified in equation (2).

$$Log(Y_{i,j,k}) = \beta_0 + \beta_1 Cer_{i,j,k} + \beta_2 Rev_k + \beta_3 Fans_k + \beta_3 Stat_k + \beta_4 Inter_{i,j,k} + \beta_5 Contr_{i,j,k} + \varepsilon_{i,j,k} \quad (1)$$

$$Logit(Y_{i,j,k}{}^{*}) = \alpha_0 + \alpha_1 Cer_{i,j,k} + \alpha_2 Controls_{i,j,k}{}' + \varepsilon_{i,j,k}{}' \quad (2)$$

where

$Y_{i,j,k}$ is the expected number of usefulness votes on review i of business j by reviewer k;

$Y_{i,j,k}{}^{*}$ is the probability of zero usefulness votes on review i of business j by reviewer k;

$Cer_{i,j,k}$ is the certainty of review i of business j by reviewer k;

Rev_k is the number of reviews written by reviewer k;

$Fans_k$ is the number of fans of reviewer k;

$Stat_k$ is the status of reviewer k;

$Interaction_{i,j,k}$ represents a matrix of interaction terms for review i of business j by reviewer k, including $Cer_{i,j,k}*Fans_k$, $Cer_{i,j,k}*Rev_k$ and $Cer_{i,j,k}*\ Stat_k$;

$Controls_{i,j,k}$ represents a matrix of control variables for review i of business j by reviewer k , including $rating_{i,j,k}$, squared term of $rating_{i,j,k}$, $length_{i,j,k}$, $readability_{i,j,k}$, $anger_{i,j,k}$, $anxiety_{i,j,k}$, $timespan_{i,j,k}$, $Yelping\_time_k$, and $restaurant\_reputaion_j$, $restaurant\_popularity_j$, $price_j$;

$Controls_{i,j,k}{}'$ represents a matrix of control variables for review i of business j by reviewer k , including $rating_{i,j,k}$, squared term of $rating_{i,j,k}$, $length_{i,j,k}$, $anger_{i,j,k}$, $timespan_{i,j,k}$, $Yelping\_time_k$;

$\varepsilon_{i,j,k}$ is the residual error term of equation (1);

$\varepsilon_{i,j,k}{}'$ is the residual error term of equation (2).

Note that Rev_k $Fans_k$ $Stat_k$ are treated as static variables—i.e., we assume no significant changes in these reviewer characteristics in our three-month short observation time window.

Variables and Measurement

The dependent variable is operationalized as the number of usefulness votes of a review. The independent variables of interests are review certainty, reviewer expertise, reviewer popularity, reviewer status and their interactions. The control variables are other characteristics of review content, reviewer, and business. For review content, we control for readability (Yin et al. 2014), review length, review rating and squared terms of rating (Mudambi and Schuff 2010), anger, anxiety (Yin et al. 2014), review timespan (see Table 1 below; Racherla and Friske 2012). For reviewer characteristics, we controlled for yelping time. Finally, for restaurant characteristics, we control for price range of a restaurant (see Table 1). In addition, we also control for the average rating and the total amount of reviews a restaurant received. The former captures the reputation while the latter represents the popularity of a restaurant. The operationalization and descriptive statistics of variables are displayed in Table 1.

Variable Type	Variable Name	Operationalization	Mean	Std. Dev
DV	Review Usefulness	Count of usefulness votes	0.73	1.54
IV	Certainty	(Certainty-related words/total words in a review)*100	1.68	1.91
	Expertise	Number of previous reviews written by a reviewer	74.31	174.75
	Popularity	Number of fans of a reviewer	3.83	18.94
	Status	A dummy variable, titled as "elite" or not	0.17	0.37
Control	Rating	Star rating (1-5) given by a reviewer	3.84	1.28
	Length	Number of words in a review	124.33	106.10
	Readability	Gunning Fog Index=0.4*(average words per sentence+ count of the word with more than six characters).	13.58	8.78
	Anger	(Anger-related words/total words in a review)*100	0.24	0.95
	Anxiety	(Anxiety-related words/total words in a review)*100	0.13	0.58
	Timespan	Number of weeks lapsed since a review posted	10.82	3.79
	Yelping time	Number of weeks lapsed since a reviewer registered	124.51	85.90
	Reputation	Average star rating of a restaurant	3.86	0.52
	Popularity	Total number of reviews obtained by a restaurant	171.16	207.56
	Price	Price level ranging from $, $$, $$$ to $$$$.	1.77	0.58

Table 1. Operationalization and Descriptive Statistics of Variables

Data Collection

Our research context is a popular online review website Yelp.com founded in October 2004, which covers a broad range of 22 product and service categories such as restaurants, shopping, beauty & spas, public services, etc. On this website, once a review is written, anyone (with or without an account) can read the review and give a vote, including 'useful', 'funny' or 'cool'.

We used Yelp Academic Data Set (https://www.yelp.com/academic_dataset) released in January 2014. Based on the dataset, we built up our own research sample. First, we selected restaurants as the research object because restaurant is typical experience goods whose quality can't be thoroughly inspected before purchasing. Second, we selected Phoenix as the target city which is most popular with the largest number of reviews in the dataset. Third, we only focused on reviews within the three months before the released time, that is, from Oct 2013 to Dec 2013. As explained earlier, reviewer characteristics change over time. In order to match the timeline between the content characteristics and the source characteristics, we focus on this short observation time window and assume no significant changes in reviewer characteristics. Finally, we only examined restaurants still open till the date of data collection. Thus, our sample has 5426 reviews from 968 restaurants by 3537 reviewers from Oct 2013 to Dec 2013.

Data Preparation

To capture the textual characteristics of reviews, we conducted content analysis using the linguistic inquiry word count (LIWC) program (Pennebaker et al. 2007), which determines the frequency of words related to different categories, including affective processes (e.g., positive emotions, anxiety, anger), cognitive processes (e.g., certain), linguistic processes (e.g., word count, adverbs) and so forth. LIWC calculates the total number of times the dictionary words appear in a category, divided by the total number of words in the review, to determine the percentage of the text that falls into a specific process. And the reliability and validity of LIWC program has been extensively investigated (Pennebaker et al. 2007). And Correlations of variables are in Table 2.

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Usefulness	1													
Certainty	-0.01	1												
Expertise	0.32***	-0.03*	1											
Popularity	0.37***	-0.02	0.65***	1										
Rating	-0.01	0.04**	-0.00	0.01	1									
Length	0.03*	-0.08***	0.02	-0.00	-0.01	1								
Anger	-0.01	-0.01	-0.00	0.00	-0.06***	0.01	1							
Anxiety	0.01	0.00	0.01	0.02	-0.01	0.01	0.01	1						
Readability	0.03*	-0.06***	0.02	-0.00	0.00	0.82***	0.01	0.01	1					
Timespan	0.01	0.00	0.01	0.00	0.02	-0.02	0.02	-0.03	-0.02	1				
Yelping time	0.17***	-0.01	0.38***	0.26***	-0.01	0.03	-0.02*	0.01	0.03*	-0.03	1			
Status	0.33***	-0.01	0.52***	0.34***	0.01	0.03*	0.02	0.01	0.02	-0.00	0.38***	1		
Restaurant reputation	0.05***	0.07***	0.01	0.02	0.41***	-0.01	-0.14***	-0.03*	-0.01	0.03*	0.05***	0.02	1	
Restaurant popularity	-0.02	0.04**	-0.03*	-0.03*	0.10***	0.06***	-0.03*	0.00	0.05***	0.03	0.012	-0.01	0.25***	1

****p*<0.001; ***p*<0.01; **p*<0.05

Table 2. Correlations of Variables

Analysis and Estimation

SAS Proc Genmod was used for empirical analysis. To clarify predictive effects, we used stepwise regression composed of three blocks of variables—controls, linear effects, and interaction effects. Estimation results are displayed in Table 3. Because of space limitation, we only present the results of interaction block including all the variables. First, logit results reveal that natural log of elapsed weeks of a review (α=3.4360, p<.001) increases the probability of a review getting zero usefulness votes while reviewer yelping time (α=-0.0933, p<.01) decreases the probability. Second, NB results show reviewer popularity (β=0.0138, p<.001) and status (β=0.9166, p<.001) significantly predicted the number of usefulness votes. Besides, reviewer expertise strengthened the power of content certainty (β=0.0003, p<.05) reviews while reviewer popularity weakened the persuasion of certain reviews (β=-0.0041, p<.01). And results involving control variables were somewhat consistent with existing literature (Mudambi & Schuff 2010). Lower rating signaled a more useful review (β=-0.2687, p<.05) and reviews by restaurants with better reputation (β=0.2489, p<.001) were more useful.

Variables	Negative Binominal	Logit
Control Effects		
Rating	-0.2687*	1.9152
Squared rating	0.0299	-0.2486
Length	-0.0000	-0.2082
Anxiety	-0.0014	
Anger	0.0033	0.2242
Readability	0.0081	
Log(Review Timespan)#	1	3.4360***
Reviewer Yelping time	-0.0003	-0.0933*
Restaurant Reputation	0.2489***	
Restaurant Popularity	-0.0002	
Price: $	0.0081	
Price: $$	0.1730	
Price: $$$	0.1631	
Price: $$$$	0.0000	
Linear Effects		
Certainty	0.0040	-0.1479
Reviewer Expertise	0.0002	
Reviewer Popularity	0.0138***	
Reviewer Status	0.9166***	
Interaction Effects		

Certainty*Expertise	0.0003*	
Certainty*Popularity	-0.0041**	
Certainty*Status	-0.0127	

***$p<0.001$; **$p<0.01$; *$p<0.05$.

#As suggested by Allison (1999), we included the natural log of the review timespan as a predictor with regression coefficient equal to 1 with the purpose of incorporating variable observation periods while maintaining the Poisson error structure of the data.

Table 3. ZINB Regression Results

Discussion

We examine the joint impacts of review content and reviewer characteristics on review usefulness. We find that (1) the tone of certainty in reviews will receive less usefulness votes when written by a popular reviewer followed by many fans than when written by a less popular reviewer and (2) in contrast, reviewer expertise positively interacts with certainty to affect review usefulness. We conjecture that this is because the signal of expertise tends to enhance the source trustworthiness of the certainty tone in a review while the signal of popularity tends to raise consumers' concern about marketing manipulation thus weakening the persuasiveness of the certainty tone.

Our study has the potential to make important contributions to the e-WOM literature. First, our research adds insights to the fast-growing stream of text mining studies highlighting the role of content characteristics in influencing consumer judgement (e.g., Mudambi & Schuff 2010). Second, our findings supplement review usefulness literature (e.g., Forman et al. 2008) by not only verifying the importance of information source but also identifying their interactions with the review content. Practically, our findings can provide benefits for online third-party review websites (e.g., Yelp.com) in the screening and selection of useful information.

Limitations and Future Research

Our study has a few limitations. First, we only examined restaurant reviews in Phoenix. The generalizability to our findings to other contexts demands further empirical studies. Future research needs to consider other business categories (e.g., hotel, beauty) and extends to different cities. Second, our study examined the interaction effects between certainty and reviewer popularity and expertise mainly. Future research can examine other reviewer characteristics and textual features of reviews such as information richness, sadness and the possible interaction effects.

REFERENCES

Allison, P. D. 2012. *Logistic Regression Using SAS: Theory and Application*. SAS Institute.

Angst, C.M., and Agarwal, R. 2009. "Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion," *MIS Quarterly* (33:2), pp.339-370.

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.

Jones, Q., Ravid, G., and Rafaeli, S. 2004. "Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration," *Information Systems Research* (*15*:2), pp. 194-210.

Kelman, H.C., 1961. "Processes of Opinion Change," *Public Opinion Quarterly* (25:1), pp.57-78.

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

Pennebaker, J. W., Booth, R. J., and Francis, M. E. 2007. *Linguistic Inquiry and Word Count (LIWC2007)*, Austin, TX: LIWC (http://www.liwc.net).

Racherla, P., and Friske, W. 2012. "Perceived 'Usefulness' of Online Consumer Reviews: An Exploratory Investigation ascross Three Services Categories," *Electronic Commerce Research and Applications* (*11*:6), pp. 548-559.

Yin, D., Bond, S., and Zhang, H. 2014. "Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews," *MIS Quarterly* (*38*:2), pp. 539-560.

Wei, L., Xu, W., and Islands, C. 2014. "Exploring Heuristic Cues for Consumer Perceptions of Online Reviews Helpfulness: The Case of Yelp.com". *In Proceedings of* the 19th *Pacific Asia Conference on Information Systems*, Keng Siau, Qing Li, and Xunhua Guo (eds.),Chengdu, China,pp.52-63.