The Dimensions of Review Comprehensiveness and Its Effect on Review Usefulness: A Latent Dirichlet Allocation Approach

Emergent Research Forum Paper

Navid Aghakhani
University of Tennessee at Chattanooga
navid-aghakhani@utc.edu

Mohammad Salehan
California State Polytechnic University, Pomona
msalehan@cpp.edu

Sonya Zhang
California State Polytechnic University, Pomona
xszhang@cpp.edu

Abstract

Online review sites like Yelp.com, TripAdvisor.com, and AngiesList.com provide values to both business and consumers. A large body of literature investigates drivers of online review usefulness. Review comprehensiveness has been identified as one the most important dimension of review quality and an important predictor of review usefulness. This study contributes to the literature by crafting and operationalizing review comprehensiveness using a text mining approach. We also empirically test the effect of the operationalized review comprehensiveness construct on review usefulness. In practice, online review providers, such as Yelp.com, can benefit from this study by integrating review comprehensiveness in their sorting algorithms.

Keywords

Restaurant reviews, Yelp.com, review comprehensiveness, review usefulness.

Introduction

Online review sites like Yelp.com, TripAdvisor.com and AngiesList.com provide values to both business and consumers. As customers increasingly rely on other consumers' online reviews to make purchase decisions, businesses also actively monitor and manage their online reviews to build brand trust and increase the likelihood of purchases. As of December 31, 2016, Yelp had 121 million reviews on over 2.8 million businesses (of which 18% were restaurants), attracting over 160 million monthly users via mobile and desktop (e.g., Yelp.com 2017). A recent study found a one-star increase in Yelp rating leads to a 5-9% increase in revenue for independent restaurants (Luca 2011). Another study found an extra half star rating on Yelp causes restaurants to sell out 19% more frequently (Anderson and Magruder 2012).

Most products / services have a large number of reviews (Salehan and Kim 2016). However, consumers find some reviews more useful than the others. Hence, a large body of literature investigates drivers of review usefulness (e.g., Connors, Mudambi and Schuff 2011; Mousavizadeh, Koohikamali and Salehan 2015; Mudambi and Schuff 2010; Salehan and Kim 2016). While early studies in this stream mainly focused on the effect of review rating or reviewers' characteristics on review usefulness, the new stream of research tend to investigate how the content of online reviews influence review usefulness (e.g., Baek, Ahn and Choi 2012; Cao, Duan and Gan 2011; Salehan and Kim 2016). In particular, the use of text mining and sentiment mining techniques is deemed important to automatically examine content of online reviews and study how the review text affects review usefulness (Cao et al. 2011). This study contributes to this new line of literature by crafting and operationalizing review comprehensiveness using a text mining approach. Review comprehensiveness has been identified as one the most important dimension of review

quality and an important predictor of review usefulness (Cheung, Lee and Rabjohn 2008). To this end, the theoretical and practical contributions of this study are twofold. In theory, this study identifies various dimensions of review comprehensiveness and examines their effect on review usefulness. To the best of our knowledge, this is the first attempt to quantify the review comprehensiveness based on actual reviews and by drawing on text mining techniques. Thus, this study will contribute to extant body of research on how review quality affects review usefulness. In practice, online review providers, especially Yelp.com, can benefit from this study by integrating review comprehensiveness in their sorting algorithms.

Literature Review

Factors of Restaurant Reviews

Previous studies have identified numerous factors used to characterize restaurant reviews (Bakhshi, Kanuparthy and Gilbert 2014; López and Farzan 2015; Pettijohn, Pettijohn and Luke 1997; Yüksel and Yüksel 2002). While restaurant attributes such as product quality, service quality, and price affect recommendations, external factors such as demographics and weather can also influence reviews significantly (Bakhshi et al. 2014). In addition, consumers do not use all available information and are more responsive when a rating contains more information, or when a restaurant has higher number of reviews, or when the reviewers are certified as "elite" by Yelp (Lucas, 2015). Users are able to detect reviews written by knowledgeable locals, and they perceive reviews provided by locals more useful because they are more trustworthy (Lopez, 2015). Table 1 shows a list of selected factors that can be measured by automated software applications.

Segment-specific Satisfaction

While a majority of previous studies have explored customer satisfaction at aggregate market level, (Yüksel and Yüksel 2002) suggest that there are different market segments seeking different sets of benefits, and each segment base their satisfaction judgments on different factors. Using factor analysis and cluster analysis of survey data from 449 tourists, they identified five customer clusters based on similarities and differences in the benefits they seek from restaurants: *value seekers, service seekers, adventurous food seekers, atmosphere seekers,* and *adventurous food seekers*. When making their restaurant selection decision, *adventurous food seekers* sought tasting new, interesting and local dishes, and are less concerned with the nutritiousness and healthiness of food as the *healthy food seekers*. *Atmosphere seekers* desired a pleasant dining ambience and a good time, and less concerned with the service quality as the *service seekers*, or menu diversity as the *value seekers*.

Research Model

In our research model, we test the effect of review comprehensiveness on review usefulness while controlling for other variables. Our control variables encompass factors related to review (i.e., length, longevity, and readability), reviewer characteristics (i.e., number of reviews, number of tips, number of friends, and number of followers), and factors related to restaurant (i.e., average rating and number of reviews). The effect of these factors on review usefulness have been investigated by prior studies (Baek et al. 2012).

A comprehensive review is the one that discusses all the factors that matter to consumers. Hence, in this study, we define review comprehensiveness as the degree to which a review discusses all the relevant factors about a restaurant. This study investigates the effect of review comprehensiveness on review usefulness. Different people look for different information in reviews (Bailey 2005). Furthermore, restaurant customers can be categorized into different segments and people in each segment put emphasis on different characteristics of a restaurant (Yüksel and Yüksel 2002). Hence, reviews that cover a wider range of characteristics of a restaurant will be able to satisfy a wider range of consumers and consequently be perceived as more useful than other reviews. Therefore, we propose the following hypothesis:

> H1: A review's comprehensiveness is positively related to its usefulness.

Methodology

Data

We plan to use data from Yelp Dataset Challenge (Huang, Rogers and Joo 2014). The dataset encompasses businesses, reviews, users, and check-in data. We will use the subset of the dataset that contains the restaurant category.

Factor	Definition	Study
Service quality and staff attitude	If the review talks about the standards, consistent quality, courtesy of the service; friendliness, knowledge, willingness to help, communication, competency, and attentiveness of the restaurant staff.	Yüksel and Yüksel (2002); Pettijohn et al. (1997); Bakhshi et al. (2014)
Product/food quality	If the review talks about the quality, portions, tastiness, temperature, presentation, preparation consistency and non-greasiness of food.	Yüksel and Yüksel (2002); Pettijohn et al. (1997)
Menu diversity	If the review talks about the menu variety, or availability of menu items, dishes or beverages liked, local dishes, and health food choice.	Yüksel and Yüksel (2002); Pettijohn et al. (1997)
Hygiene/ cleanliness	If the review talks about the cleanliness of the restaurant, utensils, or restaurant staff.	Yüksel and Yüksel (2002); Pettijohn et al. (1997)
Convenience and location	If the review talks about the location, crowd level, and operating hours of the restaurant.	Yüksel and Yüksel (2002); Pettijohn et al. (1997)
Noise	If the review talks about the quietness of the restaurant and surroundings.	Yüksel and Yüksel (2002)
Service speed	If the review talks about the waiting time for dishes or efficiency of service.	Yüksel and Yüksel (2002); Pettijohn et al. (1997)
Price and value	If the review talks about the prices and value of food.	Yüksel and Yüksel (2002); Pettijohn et al. (1997); Bakhshi et al. (2014)
Facilities/ Special features	If the review talks about the children facilities.	Yüksel and Yüksel (2002); Bakhshi et al. (2014)
Atmosphere/ ambience	If the review talks about the atmosphere or ambience in the restaurant.	Yüksel and Yüksel (2002); Pettijohn et al. (1997)
Advertisement	If the review talks about the online advertising such as being featured, or promotions such as coupons of the restaurant.	Bakhshi et al. (2014)

Table 1 - The factors used to characterize restaurant reviews

In order to automatically classify the reviews based on topics discussed in the table 1, we will use Latent Dirichlet Allocation (LDA) (Blei 2012). LDA is a probabilistic topic modeling algorithm that assumes writers (i.e., reviewers in our study) write documents (i.e., reviews in our study) by selecting a mixture of topics and drawing a word from vocabulary of each topic (Debortoli, Junglas, Müller and vom Brocke 2016; Vakulenko, Müller and Brocke 2014). In this regard, LDA is an unsupervised classification technique as there are no predefined labels for the documents. However, the model that is initially built on a set of documents (i.e., the training set) can be applied to uncover the distribution of topics in a test set. Since the main purpose of this study is to operationalize review comprehensiveness, we propose a three-step procedure to incorporate LDA in measuring review comprehensiveness. For this purpose, we

randomly split our empirical dataset into three subsets (subset 1, subset 2, and subset 3) and each single subset will be used in different steps.[1]

Measurement of Review Comprehensiveness

Step 1 (exploratory congruency analysis): In this step, we parse the reviews in subset 1 to LDA and we set the number of topics to 10. The 10 topics are identified by Yüksel and Yüksel (2002) are: (1) service quality and staff attitude, (2) product / food quality, (3) menu diversity, (4) hygiene / cleanliness, (5) convenience and location, (6) noise, (7) service speed, (8) price and value, (9) facilities / special features, and (10) atmosphere / ambience. These 10 topics serve as our gold standard scheme to examine the degree of congruency between the LDA output and our categories. Therefore, this step requires us to examine the semantic similarity between the top words that appear in the LDA topics and our predefined categories. Thus, the purpose of this stage is to identify the commonalities between LDA suggested topics and our gold standard topics. The common topics later will be used in step2 for topic selection process.

Step 2 (topic selection process): We use the model built in step1 to predict the distribution of the topics in the reviews in subset 2. By doing so, LDA will identify the distribution of each topic for every single review. Suppose that from step 1, we identify 7 overlap topics. Then, instead of reinitiating step 1 with 7 topics, we simply ignore the remaining 3 topics in this step. The main purpose of this step is to examine the direct effect of each single topic on review usefulness. For instance, suppose that in subset 2, on average, 60% percent of reviews are related to price and value. However, at review level, price and value topic may not influence review usefulness. In order to address this issue, we estimate the equation (1). On Yelp.com, usefulness is reported as the number of people who find a review useful. Because there is huge variation in the number of people who vote for a review (ranging from zero to thousands), we expect usefulness to suffer from over-dispersion. The best method to analyze an over-dispersed count variable is to utilize a negative binomial model with natural logarithm as the link function (Fox 2015; Hilbe 2011). Furthermore, we expect many reviews to have received fewer than 5 votes, which makes readership unobserved for those reviews. Truncated regression is the appropriate method of analysis when values below or above a threshold are considered as unobserved (Birkinshaw 2000; Fox 2015). Hence, truncated negative binomial regression is the proper method for examining the direct effect of the topics on review usefulness.

$Usefulness = \beta_0 + \beta_1\ Topic\ (1) + \beta_2\ Topic\ (2) +.... + B_i\ Topic\ (i) + Controls+ \varepsilon$ (Equation 1)

Where:

    a) $0 \le i \le 10$

    b) *Topic (i) is the percentage score of each topic in every single review*

Step 3 (review comprehensiveness operationalization): Using equation 1, we identify the topics that have a direct effect on review usefulness. Thus, the insignificant topics will be excluded from review comprehensiveness operationalization. In this step, again, we apply the model built in step 1 to reviews in subset 3 to identify the percentage score of each topic. However, the disqualified topics in step 1 and step 2 will be ignored in operationalization of review comprehensiveness. We define review comprehensiveness as the number of topics discussed in a review. Thus, for a particular review, each dimension's score (e.g., menu diversity) will be set to one if the review discusses that topic, and set to zero otherwise. By doing so, review comprehensiveness score would be calculated as the sum of number of topics (i.e., dimensions) discussed in a review. We estimate equation (2) using the same regression method used to estimate Equation 1 as follows:

$Usefulness = \beta_0 + \beta_1\ Comprehensiveness + Controls+ \varepsilon$ (Equation 2)

---

[1] Please note that due to space limitations we didn't include the text preparation steps and we just focused on the main algorithm.

Theoretical and Practical Contributions

From a theoretical perspective, this study will contribute to extant body of review usefulness studies. We, in particular, aim to advance the new stream of research studying how the content of online reviews affect review usefulness. By using a text mining approach and actual online reviews, we propose a new method of crafting and operationalizing review comprehensiveness as one of the important dimensions of review quality. From a practical perspective, the review comprehensiveness construct developed by this study can be utilized in recommendation systems by online review websites such as Yelp.com.

References

Anderson, M., and Magruder, J. 2012. "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database," *The Economic Journal* (122:563), pp 957-989.

Baek, H., Ahn, J., and Choi, Y. 2012. "Helpfulness of online consumer reviews: Readers' objectives and review cues," *International Journal of Electronic Commerce* (17:2), pp 99-126.

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

Bakhshi, S., Kanuparthy, P., and Gilbert, E., 2014. "Demographics, weather and online reviews: A study of restaurant recommendations," Proceedings of the 23rd international conference on World wide web, ACM, pp. 443-454.

Birkinshaw, J., 2000. *Entrepreneurship in the global firm: enterprise and renewal* Sage.

Blei, D.M. 2012. "Probabilistic topic models," *Communications of the ACM* (55:4), pp 77-84.

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

Cheung, C.M., Lee, M.K., and Rabjohn, N. 2008. "The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities," *Internet research* (18:3), pp 229-247.

Connors, L., Mudambi, S.M., and Schuff, D., 2011. "Is it the review or the reviewer? A multi-method approach to determine the antecedents of online review helpfulness," System Sciences (HICSS), 2011 44th Hawaii International Conference on, IEEE, pp. 1-10.

Debortoli, S., Junglas, I., Müller, O., and vom Brocke, J. 2016. "Text Mining For Information Systems Researchers: An Annotated Topic Modeling Tutorial," *Communications of the Association for Information Systems* (39:1), p 7.

Fox, J., 2015. *Applied regression analysis and generalized linear models* Sage Publications.

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

Huang, J., Rogers, S., and Joo, E., 2014. "Improving restaurants by extracting subtopics from yelp reviews," iConference 2014 (Social Media Expo).

López, C., and Farzan, R., 2015. "Exploring the mechanisms behind the assessment of usefulness of restaurant reviews," Proceedings of the 7th International Conference on Communities and Technologies, ACM, pp. 19-27.

Luca, M. 2011. "Reviews, reputation, and revenue: The case of Yelp. com," *Harvard Business School*).

Mousavizadeh, M., Koohikamali, M., and Salehan, M., 2015. "The Effect of Central and Peripheral Cues on Online Review Helpfulness: A Comparison between Functional and Expressive Products," ICIS 2015.

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

Pettijohn, L.S., Pettijohn, C.E., and Luke, R.H. 1997. "An evaluation of fast food restaurant satisfaction: determinants, competitive comparisons and impact on future patronage," *Journal of Restaurant & Foodservice Marketing* (2:3), pp 3-20.

Salehan, M., and Kim, D.J. 2016. "Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics," *Decision Support Systems* (81), pp 30-40.

Vakulenko, S., Müller, O., and Brocke, J.v., 2014. "Enriching iTunes App Store categories via topic modeling," ICIS 2014.

Yelp.com. (2017). An introduction to Yelp metrics as of December 31, 2016. Retrieved from https://www.yelp.com/factsheet

Yüksel, A., and Yüksel, F. 2002. "Measurement of tourist satisfaction with restaurant services: A segment-based approach," *Journal of vacation marketing* (9:1), pp 52-68.