

Understanding Cruise Tourists' Satisfaction by Analysing Their Online Ratings and Reviews

Susan (Sixue) Jia¹

¹ School of Business and Management, Shanghai International Studies University, Shanghai, China

Correspondence: Susan (Sixue) Jia, School of Business and Management, Shanghai International Studies University, Shanghai, 200083, China. Tel: 86-21-3537-2000.

Received: May 25, 2019

Accepted: June 14, 2019

Online Published: June 23, 2019

doi:10.5430/jms.v10n4p1

URL: <https://doi.org/10.5430/jms.v10n4p1>

Abstract

Associating online ratings with reviews provides a potentially cost-effective way for cruise tourism managers to capture satisfaction and quality improvement information from customers so as to facilitate the sustainable development of the tourism industry. For this purpose, this study proposes an integrated approach that leverages text mining and empirical modeling to quantitatively correlate online ratings with reviews. From TripAdvisor.com, 4248 pairs of rating and review regarding Chicago's First Lady Cruises were examined, with major topics identified. Subsequently, multilinear regression was employed to screen out the most impactful factors that influence overall ratings. Managerially, the idea of triggering the synergistic benefit from customer ratings and reviews is referential for market practitioners both within and beyond the tourism industry.

Keywords: tourism, river cruise, online rating, online review, text mining, big data

1. Introduction and Related Works

Because of its high commercial value and rich business context, tourism study has aroused great interest among economics, management, and information researchers (Litvin et al., 2008; Chang et al., 2014; Shapoval et al., 2018). One key issue of tourism research is detecting, understanding, and enhancing tourist satisfaction (Tan, 2017; Tsang et al., 2017; Yoo et al., 2017), because satisfaction undoubtedly leads to positive word-of-mouth and increased sales (Han & Ryu, 2009; Liu & Jang, 2009).

Tourists have been posting online ratings and online reviews to narrate, praise, or criticise their travel experiences. Such user generated content (UGC) is a valuable information source from which tourism managers can capture quality improvement hints. According to Ye et al. (2011), online rating and review are two of the most typical forms of UGC. Rating, usually based on a five-point scale, quantitatively tells whether and to what extent a customer is satisfied with the provider (Anderson & Magruder, 2012). Unlike rating, review is a piece of textual comment, which describes the experience of the customer with the provider in a qualitative manner (Lee & Bradlow, 2011). Studied together, rating and review can construct a complete depiction of customer response and can synergistically maximize research effectiveness (Büschken & Allenby, 2016; Hao et al., 2017).

Based on the above background, this study aims at detecting tourists' satisfaction by mining UGC on the internet. More specifically, this paper unfolds as follows. The 'data and method' section describes the source of tourist UGC, as well as the acquisition, processing, and analysis of these data using statistical and text mining tools. Subsequently, the 'result and discussion' section reports the findings from the data, and further interprets the findings with methodological and managerial implications. Finally, the 'conclusion' section summarizes the whole paper by revisiting the key issues of this study and providing future study suggestions.

The methodological contribution of this paper is that it establishes a standardized protocol of mining and correlating the massive quantitative and qualitative information available online, bridging the gap between the two sets of entirely different types of data. Meanwhile, this paper also has managerial contribution in that it explicitly addresses the most impactful factors that influence tourism customer satisfaction, offering valuable improvement suggestions to tourism operators.

2. Data and Method

2.1 Data Source

The focus of this tourism study is Chicago's First Lady Cruises (CFLC), which was initiated by Chicago Architecture Foundation (CAF). According to TripAdvisor (2018), world's largest travel site and crowd-sourced online review community, CFLC is the No.1 tour in the city based on user reviews. CFLC provides the best architecture tour experience, with highly trained and passionate volunteer docents detailedly narrating Chicago's various architectural styles and the stories of the people who designed and built the city.

The UGC regarding CFLC all comes from TripAdvisor.com. Typically, a CFLC customer provides the following information regarding the tour: (1) overall rating, a five-point scale ('5ps' for short, with 1–5 corresponding to poor, average, good, very good, and excellent), and (2) review, a piece of textual comment (mostly in English). Whereas (1) is structured data that can be measured directly, (2) is unstructured data that has to be analysed using text mining.

2.2 Identifying Review Topics

To quantify the qualitative review data, one needs to first identify topics from reviews and then characterize those reviews with the identified topics. Latent Dirichlet allocation (LDA) approach is used for topic identification. The approach is based on the assumption of probabilistic topic model (Blei et al., 2003; Blei, 2012), which assumes the word generation in a document as a two-stage process:

- (1) Randomly choose a distribution of topics.
- (2) For each word in the document,
 - (a) Randomly choose a topic from the distribution of topics in (1).
 - (b) Randomly choose a word from the corresponding distribution of the vocabulary.

In real situations, neither the distribution of topics over documents nor the distribution of words over topics is known a priori; only the documents are observed. The connection between hidden and observed variables is the joint distribution expressed using Equation (1).

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^K p(\beta_i) \cdot \prod_{d=1}^D p(\theta_d) \cdot \sum_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \quad (1)$$

β_i : distribution of word in topic i , altogether K topics

θ_d : proportions of topics in document d , altogether D documents

z_d : topic assignment in document d

$z_{d,n}$: topic assignment for the n^{th} word in document d , altogether N words

w_d : observed words for document d

$w_{d,n}$: the n^{th} word for document d

The identification of topics and words is thus a posteriori estimation (Equation 2) using Gibbs sampling (Griffiths & Steyvers, 2004), without the trouble of creating a user dictionary in advance (Haselmayer & Jenny, 2017). In this study, the estimation was realised using Python LDA 1.0.5 (The Python Community, 2018).

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D} | w_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})} \quad (2)$$

2.3 Rating and Review Correlation

The above LDA-based text mining helps identify the topics that CFLC customers most care about regarding their tour experience. However, text mining alone does not tell how these topics are associated with attitude or satisfaction. Therefore, it is necessary to establish the relation between review and rating, creating a bridge that links the factual data (i.e., reviews) with attitudinal data (i.e., ratings).

For this purpose, multilinear regression model was applied to estimate the rating-review relation similar to B üschken and Allenby's (2016) study. Moreover, because treating the words in one topic together as a single factor results in coarse grain clusters and thus limited managerial insights (also see B üschken and Allenby, 2016), the topics were manually broken and reorganized into smaller word groups, with each group representing one fine grain factor, as is in Jia's (2018) research, see Equation (3).

$$RATING = \beta_0 + \sum_{i=1}^N \beta_i \times WOF_i + \varepsilon \quad (3)$$

RATING: Rating for a review

β_0 : Intercept of the model

N : Total number of factors

β_i : Coefficient for the i^{th} factor

WOF_i : Weight of the i^{th} factor in the review, calculated by Equation (4)

$$WOF_i = \frac{f_i}{\sum_j f_j} \tag{4}$$

f_i : Number of times a word in the review is allocated to factor i

$\sum_j f_j$: Number of words in the review that are allocated to a factor

3. Result and Discussion

3.1 Basic Statistics

A total of 4248 pairs of rating and review were obtained from TripAdvisor.com, commenting on CFLC. The distributions of UGC over ratings have been summarized in Table 1, with the majority of the users assigning 5 star, another one eighth of them assigning 4 star, and the rest giving 3, 2 or 1 star. The existence of a small fraction of lower ratings (i.e., 3, 2 or 1 star) provides the opportunity to explore the reasons why some tourists are unsatisfied, from which subsequent managerial implications can be drawn.

Table 1. Distributions of UGC over ratings

	5 star	4 star	3 star	2 star	1 star	Total
Number	3635	501	79	21	12	4248
%	86%	12%	2%	<1%	<1%	100%

3.2 Topics in Review

From the reviews written by the tourists, 10 topics were successfully extracted using LDA, as are shown in Table 2. The first topic is the general information of CFLC. From the words in the topic, it is clear that CFLC is a 90 minute boat tour along a river with docent providing guidance. The second topic deals with Chicago Architecture Foundation (CAF), which runs CFLC. The third topic includes some general comments regarding CFLC, e.g., 'great', 'learn', 'wonderful', and 'relaxing'. The fourth topic suggests that a large number of tourists would 'definitely' recommend CFLC to others as a 'must-visit' tour.

Table 2. Topics identified using LDA

Topic	Word								
1. General information	90	minutes	tour	cruise	hour	half	long	river	
	entire	time	stop	boat	guide	trip	docent		
2. Chicago Architecture Foundation	Chicago	Architect	Foundation	River	First	tours	Lady	cruise	
	Architectu	tour	one	boat	best	many	CAF		
3. General comment	way	city	Chicago	great	see	history	learn	architectu	
	wonderful	get	buildings	beautiful	river	relaxing	spend		
4. Recommendation	recommen	tour	would	highly	Chicag	must	anyone	visiting	
	definitely	everyone	go	cruise	visitors	take	enough		
5. Ticket	tickets	get	early	time	line	day	online	sure	

	boat	want	advance	minutes	book	buy	tours	
6. Price	worth	well	\$	tour	money	price	little	time
	would	bit	think	kids	person	children	cost	
7. Weather	day	weather	sunny	beautiful	perfect	cold	went	tour
	cruise	hot	bring	hat	sun	warm	sunscre	
8. Boat	boat	bar	board	nice	ride	drinks	well	comforta
	also	available	clean	full	snacks	water	staff	
9. Docent	guide	tour	knowledgea	informati	docent	interesti	extreme	entertaini
	well	fun	made	enjoyable	volunte	funny	excellen	
10. Sites	river	Michigan	tour	boat	Chicag	around	Lake	Navy
	Pier	us	along	cruise	back	walk	go	

The rest of the topics provide more detailed information about CFLC. The fifth topic lists several ways to obtain tickets, including 'online booking'. The sixth topic talks about the price of the tour, with most tourists thinking it 'worth'. The seventh topic demonstrates the various weathers tourists experienced on board, e.g., hot, warm, cold, et al. The eighth topic describes the boat as 'comfortable', 'clean', 'nice', et al. The ninth topic depicts the docents as 'knowledgeable', 'informative', and 'funny'. Finally, the tenth topic mentions several of the sites along the tour such as 'Lake Michigan' and 'Navy Pier'.

As is previously discussed, because treating the words in one topic together as a single factor results in coarse grain clusters and thus limited managerial insights, the topics were manually broken and reorganized into smaller word groups, with each group representing one fine grain factor, see Table 3.

Table 3. Potentially influential factors

bar	boat	book	CAF	children	cold	docent
drinks	history	hot	kids	line	online	price
snacks	staff	sunny	tickets	volunteer	warm	weather

3.3 Rating-Review Relation

Multilinear regression was applied to estimate the rating-review relation, with the result summarized in Table 4. Notice that one of the factors, 'staff', is removed from the model to avoid multicollinearity (Büschken & Allenby, 2016). In Table 4, negative coefficients indicate that if the corresponding factor is mentioned in the review (compared with not-mentioned), the corresponding rating is likely to be lower. In other words, the factor (e.g., 'online') is to be improved. On the contrary, positive coefficients indicate CFLC have good performance in the corresponding aspects.

Table 4. Regression result for rating-review relations

Factor	snacks	CAF	volunteer	warm	docent	bar	history
Parameter estimation	0.245	0.184	0.118	0.093	0.092	0.090	0.080
Factor	drinks	weather	staff	sunny	boat	book	cold
Parameter estimation	0.043	0.039	—	-0.001	-0.007	-0.039	-0.048

Factor	line	tickets	kids	price	hot	children	online
Parameter estimation	-0.053	-0.093	-0.156	-0.164	-0.288*	-0.433*	-0.454**

* $p < 0.05$, ** $p < 0.01$

According to Table 4, no factor has a significantly positive coefficient (i.e., $p < 0.05$), meaning that it is not likely that mentioning any factor in a review can lead to an increase in rating. Considering the fact that up to 86% of tourists gave 5 stars, such results are reasonable. On the other hand, Table 4 reveals the factors that tourists have been unsatisfied with, including hot weather, children related issue, or online booking. CFLC managers are strongly encouraged to focus on the improvement of these factors to lessen customer complaints.

4. Conclusion

Exploiting the business value of tourism customers' online UGC requires studying quantitative rating and qualitative review together to trigger their synergistic effect. This study successfully bridged the typological and dimensional discrepancies of rating and review data and established their relations. The unstructured review data was analyzed with LDA for topic identification. Quantified with weight of factor, the review data was then associated with rating values by leveraging multilinear regression. Finally, the relative impact of each factor on tourism customers' satisfaction was compared by investigating the estimated coefficients.

Methodologically, this paper contributes by establishing a standardized protocol of mining and correlating the massive quantitative and qualitative information available online, bridging the gap between the two sets of entirely different types of data. Meanwhile, managerially this paper explicitly addresses the most impactful factors that influence tourism customer satisfaction, offering valuable improvement suggestions to tourism operators.

Despite these achievements, this study is not without limitations. The topic identification result from LDA can be further improved by supplementing human screening, or by comparing with pre-defined dictionary to achieve more precise results. Currently, LDA is still the best tool for topic identification of massive text, because of its advantages of being unsupervised which saves time and labor, and that relevant knowledge is not a prerequisite. Nevertheless, future study may further investigate the possibility of perfecting the LDA method, so as to generate more meaningful and specific topics.

Meanwhile, our sample might have introduced some bias, since all the samples we examined were tourism customers who volunteered to post ratings and reviews online, meaning we seemed to have neglected tourists who were less active online. In other words, we are not absolutely confident that the studied tourists are fully representative of all the customers. To further enhance the persuasiveness of our findings as well as other text mining based studies, we suggest future studies incorporate both text mining which allows a large sample size, and traditional offline methods which help eliminate the sample bias.

References

- Anderson, M., & Magruder, J. (2012). Learning from the crowd: regression discontinuity estimates of the effects of an online review database. *The Economic Journal*, 122, 957-989. <https://doi.org/10.1111/j.1468-0297.2012.02512.x>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55, 77-84. <https://doi.org/10.1145/2133806.2133826>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, 35, 953-975. <https://doi.org/10.1287/mksc.2016.0993>
- Chang, S., Pu, C. -Y., & Hsieh, P. -J. (2014). A regional competition analysis of medical tourism industry: An example of Taiwan. *International Journal of Information and Management Sciences*, 25, 139-156.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Science*, 101, 5228-5235. <https://doi.org/10.1073/pnas.0307752101>
- Han, H. S., & Ryu, K. S. (2009). The roles of the physical environment, price perception, and customer satisfaction

- in determining customer loyalty in the restaurant industry. *Journal of Hospitality and Tourism Research*, 33, 487-510. <https://doi.org/10.1177/1096348009344212>
- Hao, H., Zhang, K., Wang, W., & Gao, G. (2017). A tale of two countries: International comparison of online doctor reviews between China and the United States. *International Journal of Medical Informatics*, 99, 37-44. <https://doi.org/10.1016/j.ijmedinf.2016.12.007>
- Haselmayer, M., & Jenny, M. (2017). Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. *Quality and Quantity*, 51, 2623-2646. <https://doi.org/10.1007/s11135-016-0412-4>
- Jia, S. (2018). Behind the ratings: text mining of restaurant customers' online reviews. *International Journal of Market Research*. <https://doi.org/10.1177/1470785317752048>
- Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48, 881-894. <https://doi.org/10.1509/jmkr.48.5.881>
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29, 458-468. <https://doi.org/10.1016/j.tourman.2007.05.011>
- Liu, Y. H., & Jang, S. C. (2009). Perceptions of Chinese restaurants in the U.S.: What affects customer satisfaction and behavioural intentions?. *International Journal of Hospitality Management*, 28, 338-348. <https://doi.org/10.1016/j.ijhm.2008.10.008>
- Shapoval, V., Wang, M. C., Hara, T., & Shioya, H. (2018). Data mining in tourism data analysis: Inbound visitors to Japan. *Journal of Travel Research*, 57, 310-323. <https://doi.org/10.1177/0047287517696960>
- Tan, W.-K. (2017). The relationship between smartphone usage, tourist experience and trip satisfaction in the context of a nature-based destination. *Telematics and Informatics*, 34, 614-627. <https://doi.org/10.1016/j.tele.2016.10.004>
- The Python Community. (2018). *LDA 1.0.5: Topic modeling with latent Dirichlet allocation*. Retrieved 25 May 2019, from <https://pypi.python.org/pypi/lda/>
- TripAdvisor. (2018). *TripAdvisor LLC*. Retrieved 25 May 2019, from <https://www.tripadvisor.com/>
- Tsang, S.-S., Wang, W.-C., & Ku, H.-H. (2017). Study on cluster analysis characteristics and classification capabilities: A case study of satisfaction regarding hotels and bed and breakfasts of Chinese tourists in Taiwan. *Intelligent Automation and Soft Computing*, 23, 103-108. <https://doi.org/10.1080/10798587.2016.1139285>
- Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behaviour*, 27, 634-639. <https://doi.org/10.1016/j.chb.2010.04.014>
- Yoo, C. W., Goo, J., Huang, C. D., Nam, K., & Woo, M. (2017). Improving travel decision support satisfaction with smart tourism technologies: A framework of tourist elaboration likelihood and self-efficacy. *Technological Forecasting and Social Change*, 123, 330-341. <https://doi.org/10.1016/j.techfore.2016.10.071>