Topic Modelling Application for Luxury Hotel Reviews

Lüks Otel Yorumları için Konu Modellemesi Uygulaması

Feray ADIGÜZEL [1]
Moamen ELSHERBINY [2]
Juana Teresa Aguiar QUINTANA [3]
Christian GONZÁLEZ-MARTEL [4]

1. Introduction

Electronic word-of-mouth has been quite popular in recent years, and travellers are increasingly depending on digital feedback by observing reviews posted on travel portals. In the lodging sector, online customer reviews (OCRs) are considered as a critical source of information that has a direct impact on travellers' decision-making (Baka, 2016; Hu, Zhang, Gao, & Bose, 2019; Mariani, Borghi & Gretzel, 2019; Mellinas, Nicolau and Park, 2019). Online reviews summarize guests' hotel stays and demonstrate their attitudes toward the hotel and their stay experiences (Xu & Li, 2016). Due to high competition in the lodging industry, hotels must focus on customer satisfaction as a crucial asset to create loyalty and to promote a positive attitude (Cetin and Dincer, 2014). The first step for hoteliers is to identify the important topics in OCRs that lead to customer satisfaction and dissatisfaction (Xu & Li, 2016, Sparks and Browning, 2011). Due to the nature of the hospitality industry, sentiment analysis and topic modeling approaches are growing rapidly as a novel research technique to analyze consumers' in-depth opinions, as online feedback and comments from customers have a significant effect on their purchasing behaviour (Ban and Kim, 2019). Despite a review text can contain much more information than a review rating, topics in review text have not attracted as much research attention as volume, and valence.

In this research, topic modelling was employed to extract hidden topics in consumer reviews. Topic modelling is defined as being a machine learning tool for discovering hidden topics appearing in a large collection of documents (Blei, Ng, & Jordan, 2003). Since the determinants of consumer satisfaction ratings differ with the hotel type (Xu & Li, 2016), luxury hotel reviews from the same chain will be investigated in this research. This research addresses two main questions: (1) What are the important topics that luxury hotel guests articulate through their online reviews? (2) How do these topics, together with other review characteristics such as emotions embedded in a review text, influence satisfaction ratings?

2. Literature Review

2.1. Determinants of customer satisfaction

Understanding the determinants that influence hotel guest satisfaction has long been a popular area of research. Cleanliness, location (in terms of close to attraction and accessibility with public transportation), room (such as size), service and value (such as room price) are listed as essential

1 Luiss University, ferayadiguzel14@hotmail.com
2 Luiss University, melsherbiny@luiss.it
3 Universidad de Las Palmas de Gran Canaria, teresa.aguuiar@ulpg.es
4 Universidad de Las Palmas de Gran Canaria, christian.gonzalez@ulpgc.es

attributes to reduce dissatisfaction for luxury hotels (e.g. 4-5 stars hotels) (Li et al. 2020; Zhou et al. 2014). Content analysis or surveys were preferred ways to listen OCRs in literature.

Li et al. (2013), for example, used text mining and content analysis to analyse hotel reviews. Accommodation, transportation accessibility, proximity to tourist attractions, and value for money were listed as customer satisfaction factors (Alrawadieh et al. 2018). The determinants of customer satisfaction and dissatisfaction changed depending on the type of hotel (Kim et al. 2015). Gu and Ryan (2008) determined seven factors that positively affect consumers' overall satisfaction: Bed comfort, bathroom cleanliness, room size and condition, location and accessibility, food and beverage quality, auxiliary service, and staff performance.

The content analysis of dissatisfied online reviews for luxury hotels in Jordan categorized these themes: service quality, the efficiency of hotel facilities, cleanliness and hygiene, quality of food and beverage, noisiness and crowdedness, high pricing, hotel's physical characters (décor, location, etc.), policies applied by the hotel, overbilling and refunding related issues, lack of sufficient safety measures, and misleading information (Dincer & Alrawadieh, 2017).

As a review-specific factor, the role of emotions embedded in reviews on consumers' satisfaction ratings has not been investigated sufficiently. In leisure services, such as amusement parks, theatres, resort hotels, consumers consume service to stimulate emotions (Otto & Ritchie,1996). The degree of arousal or excitement experienced by customers when consuming a service may be a major determinant of pleasure and satisfaction yielded through the experience (Russell & Pratt, 1980; Mano & Oliver, 1993). Thus, joy in a review text, as a positive and high arousal emotion, might have a positive impact on satisfaction ratings, while anger, as a negative and high arousal one, might have a negative impact.

2.2. Topic modeling and sentiment analysis

Text data analysis can be classified into two categories based on what people read in the text: topic modeling and sentiment analysis. The term "topic modeling analysis" refers to a set of techniques for determining what the text is about, whereas "sentiment analysis" refers to a set of techniques for determining the emotions or sentiments that present in the text (Blei al. 2003). Sentiment analysis has been quite important in recent years in the service environment. Customer sentiment in most situations relates to the feelings conveyed in text reviews by customer (Geetha et al., 2017). While emotions are related to distinct human experiences such as trust, sadness, anger, fear and surprise, the sentiments are related to the polarity (positive, neutral, negative) and intensity (Kirilenko et al., 2018).

The application of topic modelling on hotel reviews has been investigated scarcely. Latent Dirichlet allocation (LDA) was initially created by Blei et al. (2003) with the purpose of discovering latent semantic structures in a textual document collection. The fundamental notion is that each paper has a mix of latent subjects, in which each subject differs from document to document, defined by distribution across the word (single words in document collection).

Guo, Barnes, and Jia (2017) identified 19 topics that consumers frequently mentioned in hotel reviews using LDA and examined the effects on satisfaction ratings. While homeliness, strong event management capability, and pet-friendliness were topics that appear in 5-star hotels, the quality from perspectives of hotel and resort facilities, food quality, room size, and decoration were the topics in 4 and 4.5-star hotels. In sharing economy context, Zhang (2019) conducted topic modelling to extract 16 topics in Airbnb reviews and investigated the influence on property's listing performance. Although online review ratings have been already proven to influence sales significantly (e.g., Godes & Mayzlin, 2004), the predictive power of topics extracted from luxury hotel reviews on customer satisfaction have not been investigated in depth. This study contributes to literature applying LDA to extract topics in reviews and examines the influence on satisfaction ratings for luxury hotels.

Method:

This empirical research is descriptive and relies on 8376 TripAdvisor hotel reviews of three Spanish luxury hotels from the same chain between 2002-2019. The star ratings of the selected hotels range from four to five stars, indicating that the sample represents a homogeneous group of hotels in terms of service ratings. These hotels are similar in terms of being in the same area, having comparable room rates, to have at least 1000 reviews. Three hotel reviews were extracted which are situated in Canary Islands, one of the most visited regions of Spain: Lopesan Costa Meloneras Resort, Lopesan Baobab Resort, and Lopesan Villa del Conde Hotel Resort. Lopesan is the leading Canarian tourist company in the Islands and one of the top ten in Spain.

The data collected has been analysed using the open-source software Python. We have utilized statistical text processing techniques to assess the chance that topics are latent throughout the whole document using structurally modified materials for a topic modeling. In this paper, we utilized the latent Dirichlet allocation (LDA) approach, the most recognized modeling method. Emotional analysis was utilized in the text to generate good and negative terms in consumers' emotional words. The techniques used in topic modeling are divided into three stages. First, we collect the data that will be used for topic modeling. During the second stage, pre-processing analysis were conducted to convert unstructured data into data suitable for topic modeling. Data analytics were the last stage.

The first step is to remove stopwords in the preparation of the data. The algorithm extracts the words of the document (this step is called tokenization), cleans the data to remove the irrelevant words (i.e. “a”, “is”, etc...) and reduces inflectional form and derivationally related forms of a word to a common base form (stemming and lemmatization). As a last descriptive analysis, we have employed a Bag of Word (BoW) model. In this approach, we look at the histogram of the words within the text, considering each word count as a feature [see Zhang, Jin & Zhou (2010) and Wu, Hoi, Yu (2010) for further information].

Review polarity was measured as sentiment intensity and its value ranges from –1 (extremely negative) to 1 (extremely positive). This is done using VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon (Gilbert & Hutto, 2014). The NRC Word-Emotion Association Lexicon (Mohammad & Turney, 2010) was used to obtain review length, and classify words into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Prior average ratings and variance were measures of social influence from prior reviews. Latent Dirichlet allocation (LDA) was applied to extract and label the dimensions of customer satisfaction across reviews.

Ordinal logistic regression was used to understand the effects of predictors on customer satisfaction ratings. The predictors are review polarity and length, prior average rating and variance in ratings, joy and anger emotions, hotel dummies (Costa Moleneras was the reference one), and topic dummies. Models were run with and without the topics and compared.

Results:

Twenty topics were decided as the best representing review texts of three luxury hotels (Table 1). The decision of the optimal number of topics relies on coherence and perplexity statistics. Each topic was labelled with the extracted the top 10-words and their relative weight. For instance, topic 13 was labelled as *pay problems* with top-ten words, namely *room, pay, hotel, free, charge, safe, extra, drink, price, day,* which are ordered according to their weight. The model with topics was significantly better than the model without topics (R2=0.16 vs 0.11).

The results of the ordinal regression model indicated that polarity, joy emotion, prior average rating and variance were significantly positively associated with satisfaction ratings; however, review length and anger emotion were significantly negative associated. Hotel dummies were also significant: Baobab has higher satisfaction than Villa del Conde and Costa Meloneras. The significant negative topics were 13(pay problems) and 18(room services). The most important topics were 11(hotel lovers), 9(staff quality/attitude), 6(returning customers), 5 (convenience of location), and 2(service quality). The least important one was topic 15 (nightlife and hotel entertainment). Among twenty topics, nonsignificant topics were topics 10 (restaurant quality), 12 (hotel complainers), 19 (problem issues) compared to the reference topic 20 (crowdedness/hotel guest quality). Remaining topics and their labels were 1(view and romantic surroundings), 3(hotel size and design), 4(checking service), 5(convenience of location), 7(family with kids), 8(hotel facilities, especially pool), 14(satisfied hotel guests), 16(hotel amenities), and 17(breakfast and food).

Table 1: Topics with LDA

Topic # and label	Words
1: View and romantic surroundings	Room, pool, view, area, restaurant, hotel, sea, main, large, quiet
2: Service quality	Good, excellent, service, quality, spa, high, facility, standard, restaurant, location
3: Hotel size and design	Area, large, resort, pool, garden, style, feel, huge, design, comfortable
4: Checking service	Room, reception, check, day, arrive, leave, book, give, service, <u>wait</u>
5: Convenience of location	Walk, beach, restaurant, shop, bar, hotel, minute, close, front, plenty
6: Returning customers	Hotel, visit, stay, year, time, return, holiday, staff, fantastic, back
7: Family with kids	Child, family, pool, kid, great, couple, adult, love, young, time
8: Hotel facilities (especially pool)	Good, nice, pool, great, hotel, lot, big, clean, place, food
9: Staff quality/attitude	Staff, friendly, clean, food, helpful, excellent, choice, ground, plenty, return
10: Restaurant quality	Restaurant, food, hotel, dinner, evening, good, pool, eat, week, cold
11: Hotel lovers	Great, amazing, fantastic, staff, lovely, food, beautiful, back, perfect, love
12: Hotel complainers	People, time, review, <u>bad</u>, issue, point, <u>problem</u>, thing, read, <u>complain</u>
13: Pay problems	Room, pay, hotel, free, <u>charge</u>, safe, <u>extra</u>, drink, price, day
14: Satisfied hotel guests	Hotel, stay, lovely, week, night, find, back, time, feel, eat
15: Nightlife and Hotel entertainment	Entertainment, evening, night, good, bar, show, hotel, German, watch, bit
16: Hotel amenities	Hotel, pool, bed, find, towel, sunbed, area, day, people, plenty
17: Breakfast and food facilities	Breakfast, restaurant, night, bar, food, dinner, drink, meat, eat, selection
18: Room services	Room, day, bed, water, change, double, end, towel, bottle, clean
19: Problem issues	Room, shower, open, reception, floor, bathroom, bath, light, door, <u>noise</u>
20: Crowdedness or hotel guest quality	Hotel, guest, make, star, feel, experience, place, staff, stay, number

Conclusions/Future Research/Limitations:

This study contributes to literature about consumer satisfaction in the hospitality sector by using LDA to explore topics embedded in online customer reviews. The extracted topics were related both with the determinants of customer (dis-)satisfaction and different customer segments jointly. The extracted topics explained satisfaction ratings. Topic extraction is an alternative method to surveys or content analysis to listen customers' voice. It can help managers to identify service issues that customers talk about and make informed decisions regarding service improvement.

Most of extracted topics were similar to determinants of consumer (dis-)satisfaction identified in literature (Guo, Barnes, & Jia, 2017; Zhang 2019; Dincer & Alrawadieh, 2017). Pay problems (topic 13) and insufficient hotel amenities such as water, towel etc. (topic 18) seem to create dissatisfaction for luxury hotel guests. Compared to previous studies, we did not obtain any cleanness issue or value for money in our extracted topics. This might be caused by using only luxury hotel reviews. When joy emotion increases, satisfaction ratings increase and anger emotions is the opposite. Consumers are influenced by other ratings (i.e. prior average rating) and variability in their ratings (i.e. distribution of previous ratings) when they give a satisfaction rating, similar to Lee et al. (2015). One of limitation of this study was using three luxury hotels from the same chain. More luxury hotels can be further added to examine the stability of the findings. As a future research, reviewer characteristics can be added to explain satisfaction ratings and to profile the extracted topics in more detail.

References

Alrawadieh, Z., & Law, R. (2019). Determinants of hotel guests' satisfaction from the perspective of online hotel reviewers. International Journal of Culture, Tourism and Hospitality Research.

Alrawadieh,Z., Dincer,M.Z., Istanbullu Dincer F., Mammadova,P. (2018) "Understanding destination image from the perspective of Western travel bloggers: the case of Istanbul", International Journal of Culture, Tourism and Hospitality Research, 12(1)

Baka, V. (2016). The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. *Tourism Management*, *53*, 148–162.

Ban, H. J., & Kim, H. S. (2019). Understanding customer experience and satisfaction through airline passengers' online review. Sustainability, 11(15), 4066.

Bastani, K., Namavari, H., & Shaffer, J. (2019). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. Expert Systems with Applications, 127, 256-271.

Bilgihan, A., Seo, S., & Choi, J. (2017). Identifying restaurant satisfiers and dissatisfiers: Suggestions from online reviews. Journal of Hospitality Marketing & Management, 27(5), 601–625.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. the Journal of machine Learning research, 3, 993-1022.

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. Journal of Hospitality Marketing & Management, 26(7), 675-693.

Cetin, G. and Dinçer, F.I. (2014), "Influence of customer experience on loyalty and word-of-mouth in hospitality operations", Anatolia, Vol. 25 No. 2, pp. 181-194.

Dinçer, M. Z., & Alrawadieh, Z. (2017). Negative word of mouse in the hotel industry: A content analysis of online reviews on luxury hotels in Jordan. Journal of Hospitality Marketing & Management, 26(8), 785–804.

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels – An empirical analysis. Tourism Management, 61, 43–54.

Godes, D. and D. Mayzlin (2004). "Using Online Conversations to Study Word-of-Mouth Communication," Marketing Science, 23 (4), 545–60.

Guo, Y., Barnes, J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. Tourism Management, 59, 467–483.

Gu, H.& Ryan, C. (2008). "Place attachment, identity and community impacts of tourism - the case of Beijing hutong". Tourism Management, 29, 637-647

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. Tourism Management, 72, 417-426.

Gilbert, C., & Hutto, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth international conference on weblogs and social media (icwsm-14). available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf (Vol. 81, p. 82).

Kim, W. G., Lim, H., & Brymer, R. A. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management*, *44*, 165–171.

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. Journal of Travel Research, 57(8), 1012–1025.

Kwon, H. J., Ban, H. J., Jun, J. K., & Kim, H. S. (2021). Topic modeling and sentiment analysis of online review for airlines. Information, 12(2), 78.

Lee, Y., Hosanagar,K., Tan, y. (2015). Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings. Management Science 61(9):2241-2258.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Li, H., Ye, Q.& Law, R. (2013). Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis, Asia Pacific Journal of Tourism Research, 18:7, 784-802

Li, H., Liu, Y., Tan, C.W. and Hu, F. (2020), "Comprehending customer satisfaction with hotels: data analysis of customer-generated reviews", International Journal of Contemporary Hospitality Management, Vol. 32 No. 5, pp. 1713-1735.

Mano, H. and Oliver, R.L. (1993) Assessing the Dimensionality and Structure of the Consumption Experience: Evaluation, Feeling, and Satisfaction. Journal of Consumer Research, 20, 451-466.

Mariani, M. M., Borghi, M., & Gretzel, U. (2019). Online reviews: Differences by submission device. *Tourism Management*, *70*, 295–298.

Mellinas, J. P., Nicolau, J. L., & Park, S. (2019). Inconsistent behavior in online consumer reviews: The effects of hotel attribute ratings on location. *Tourism Management*, *71*, 421–427.

Mohammad, Saif M. and Peter D. Turney (2013), "Crowdsourcing a Word–Emotion Association Lexicon," Computational Intelligence, 29, 3, 436–65.

Otto, J.E.,& Ritchie., J.R.B. (1996) The service experience in tourism Tourism Management, 17 (3), 165-174

Xu, X., and Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: a text mining approach. *Int. J. Hosp. Manag*. 55, 57–69.

Russell, J. A., & Pratt, G. (1980). A description of the affective quality attributed to environments. *Journal of Personality and Social Psychology, 38*(2), 311–322.

Sparks, B.A. and Browning, V. (2011), "The impact of online reviews on hotel booking intentions and perceptions of trust", *Tourism Management*, Vol. 32 No. 6, pp. 1310-1323.

Sutherland, I., Sim, Y., Lee, S. K., Byun, J., & Kiatkawsin, K. (2020). Topic modeling of online accommodation reviews via latent dirichlet allocation. Sustainability, 12(5), 1821.

Wu, L., Hoi, S.C. H., & YU, N.(2010). Semantics-Preserving Bag-of-Words Models and Applications. (2010). IEEE Transactions on Image Processing. 19, (7), 1908-1920. Research Collection School Of Information Systems

Zhou, L., Ye, S., Pearce, P.L. and Wu, M.Y. (2014), "Refreshing hotel satisfaction studies by

reconfiguring customer review data", International Journal of Hospitality Management, Vol. 38, 1-10

Zhang, J. (2019) "Listening to the Consumer: Exploring Review Topics on Airbnb and Their Impact on Listing Performance, Journal of Marketing Theory and Practice, 27:4,371-389

Zhang, Y., Jin,R. & Zhou, Z. (2010) Understanding bag-of-words model: A statistical framework. International Journal of Machine Learning and Cybernetics 1(1):43-52