Modelling Online Complaining Behaviour in The Hospitality Industry: an Application of Data Mining Algorithms

Raksmey Sann, Pei-Chun Lai, Shu-Yi Liaw and Chi-Ting Chen

October 13, 2021

Modelling online complaining behaviour in the hospitality industry: An application of data mining algorithms

Sann, Raksmey
Tourism Management Division
International College
Khon Kaen University

Lai, Pei-Chun (Corresponding Author)
Department of Hotel and Restaurant Management
National Pingtung University of Science and Technology

Liaw, Shu-Yi
College of Management, Director of Computer Center
National Pingtung University of Science and Technology

Chen, Chi-Ting
Department of Hospitality Management
Ming Chuan University

Abstract:

This study aims to predict the complaint attributions significantly differing from various hotel classes (i.e. *higher star-rating* and *lower star-rating*) of travelers related to their online complaining behavior. For this, Decision Tree Algorithm was conducted. Findings reveal that guests from higher star-rating hotels are most likely to give online complaints on *Service Encounter* and tend to stayed at large size hotel. Additionally, guests of lower star-rating hotel are most likely to give online complaints on *Cleanliness*, and are inclined to stay at small size hotel.

Keywords: Online Complaining Behavior (OCB), Decision Tree Algorithms (DTAs), Hotel Class, Online Complaining Attributes (OCAs), Data Mining Algorithms (DMAs)

1. Introduction

Today choosing a hotel or restaurant via seeking online recommendations is one of the most challenging tasks for customers when it goes with a big quantity of online reviews on the web that can be acquired easily. However, customers tend to select some of them to reduce the consideration set of the possible alternatives. When reading the online reviews, customers evaluate the overall rating 66%, review valence (positive and negative) 63%, review detail 62%, and reviewer's status 40% as the top four factors for consideration (Guerreiro & Moro, 2017). In terms of review valence (positive and negative), negative information is easier for consumers to perceive than positive information according to the theory of negative effects; thus, negative information can have a stronger negative effect on purchase decisions (Tsao, Hsieh, Shih, & Lin, 2015). And also, online reviews have the power to procure 30 times more consumers (Abubakar & Ilkan, 2016), thus this study follows the emerging style of research using user generated data by looking at complaint reviews and attempt to understand their online complaining behavior (**OCB**) differing between hotel classes. For this contribution, the main purpose of this study is to predict the complaint attributions significantly differing from various hotel classes (i.e. *higher star-rating* and *lower star-rating*) of travelers related to their OCB. The main contributions of this study lie in the fact that this is one of the innovative papers to predict OCB from different classes of hotel guests by utilizing Data Mining Algorithms.

2. Literature Review

Recently, many research scholars have been utilizing data mining (***DM***) procedures in conducting their studies on the tourism and hospitality industry. For instance, Golmohammadi, Jahandideh, & O'Gorman (2012) studied the application of DM, specifically using Decision Tree (***DT***) modeling the tourists' behavior in the online environment. DM has also been studied in terms of its importance and influence in the hotel's marketing field, and how this approach can help the company to reach their potential customers, know them and their behavior (Moro, Rita, & Coelho, 2017). Thus, DM techniques focusing on the analysis of the textual contents from travellers' reviews/feedbacks have been used in the publication of many papers (Moro et al., 2017). With the unique abilities of DM approach, hoteliers can receive invaluable information which enables them to have a better insight about customer behavior and to develop effective customer retention strategies (Golmohammadi et al., 2012).

3. Methodology

3.1. Data Collection and Sample

353 hotels, ranked from *2*- to *5*-star based on British's hotel rating system, were randomly selected from a population of 1,086 listed on TripAdvisor's site (TripAdvisor, 2018). In total, 1992 valid complaint reviews were collected for the analysis. These complaints were classified into two groups: *higher star-rating* hotels and *lower star-rating* hotels.

3.2. Coding and Reliability of Online Complaining Attributes

By developing the coding categories, content analysis of texts was manually applied. Then the coding subjects were independently categorized into various complaint attributes and items. The test of reliability adopted from Cenni & Goethals (2017) two-step inter-code reliability test, which the both coding grids were > 90%, was judged acceptable.

3.3. Knowledge Modellings of Decision Tree Algorithms

In this step, **CHAID** DT algorithm were employed. This algorithm was tested on the output variable (*Hotel Class* as dependent variables) and a total of 11 inputs (*Hotel Size*, *Room Issue*, *Hotel Facility*, *Cleanliness*, *Service Encounter*, *Location Accessibility*, *Value for Money*, *Safety*, *Miscellaneous Issue*, *Room Space* and *F & B Issue* as independent variables) by using holdout samples. The dependent variable as target was put into the models as binary variables. To test classification models, SPSS Modeler 18 was utilized.

4. Results

From *Figure 1*, five descriptors splitting nodes were "*Hotel Size*", "*Service Encounter*", "*Cleanliness*", "*Value for Money*", and "*Room Space*". Among the hotel guests (N = 1,992), 57.63% indicated guests made online complaints coming from *higher star-rating* hotels, whereas 42.37% of them give online complaints coming from *lower star-rating* hotels.

The first splitting complaining attribute was *"Hotel Size"* (x^2 = 279.20, *d.f.* = 2, p = .000). In Node 1, 81.73% of *higher star-rating* hotel guests made online complaints are staying at large size hotel whereas only 18.27% from *lower star-rating* hotel guests. Similarity in Node 2, 73.70% of the *higher star-rating* hotel customers give online complaints are staying at medium size hotel but around 26% from *lower star-rating* hotel guests. On contrary, 61.42% of *lower star-rating* hotel guests made online complaints are coming from small size hotel while around 38.58% from *higher star-rating* hotel customers.

The second pruning tree of Node 1 was based on the complaining attribute of "*Service Encounter*" (x^2 = 10.97, *d.f.* = 1, *p* = .001). Node 1 was diverged into Node 4 and Node 5. In Node 5, 88.89% of those who give online complaining on *Service Encounter* were from *higher star-rating* hotel guests; while only 11.11% are from *lower star-rating* hotel guests.

The second split of Node 2 was "*Service Encounter*" (x^2 = 19.32, *d.f.* = 1, *p* = .000). Node 2 (*N* = 608) was pruned into Node 6 (*N* = 385) and Node 7 (*N* = 440). In Node 7, about 80% of *higher star-rating* hotel guests give online complaints on *Service Encounter* but around 20% of them are complained by *lower star-rating* hotel guests. Node 7 was diverged into Node 12 and Node 13, which is "*Value for Money*" (x^2 = 3.97, *d.f.* = 1, *p* = .046). In Node 13, approximately 90% of *higher star-rating* hotel guests give online complaints on *Value for Money* but around 10% are coming from *lower star-rating hotel*.

The last pruning tree of Node 3 was "*Service Encounter*" (x^2 = 31.95, *d.f.* = 1, *p* = .000). Node 3 was split into Node 8 and Node 9. In Node 9, 50.43% of *higher star-rating* hotel guests were complained on *Service Encounter* during their stay experience; on the other hand, about 49.57% of them are complained by *lower star-rating* hotel guests. Node 9 was also split into Node 16 and Node 17, which is "*Room Space*" (x^2 = 10.85, *d.f.* = 1, *p* = .001). In Node 17, approximately 73.33% of *higher star-rating* hotel guests give online complaints on *Room Space* but about 26.67% are coming from *lower star-rating* hotel.

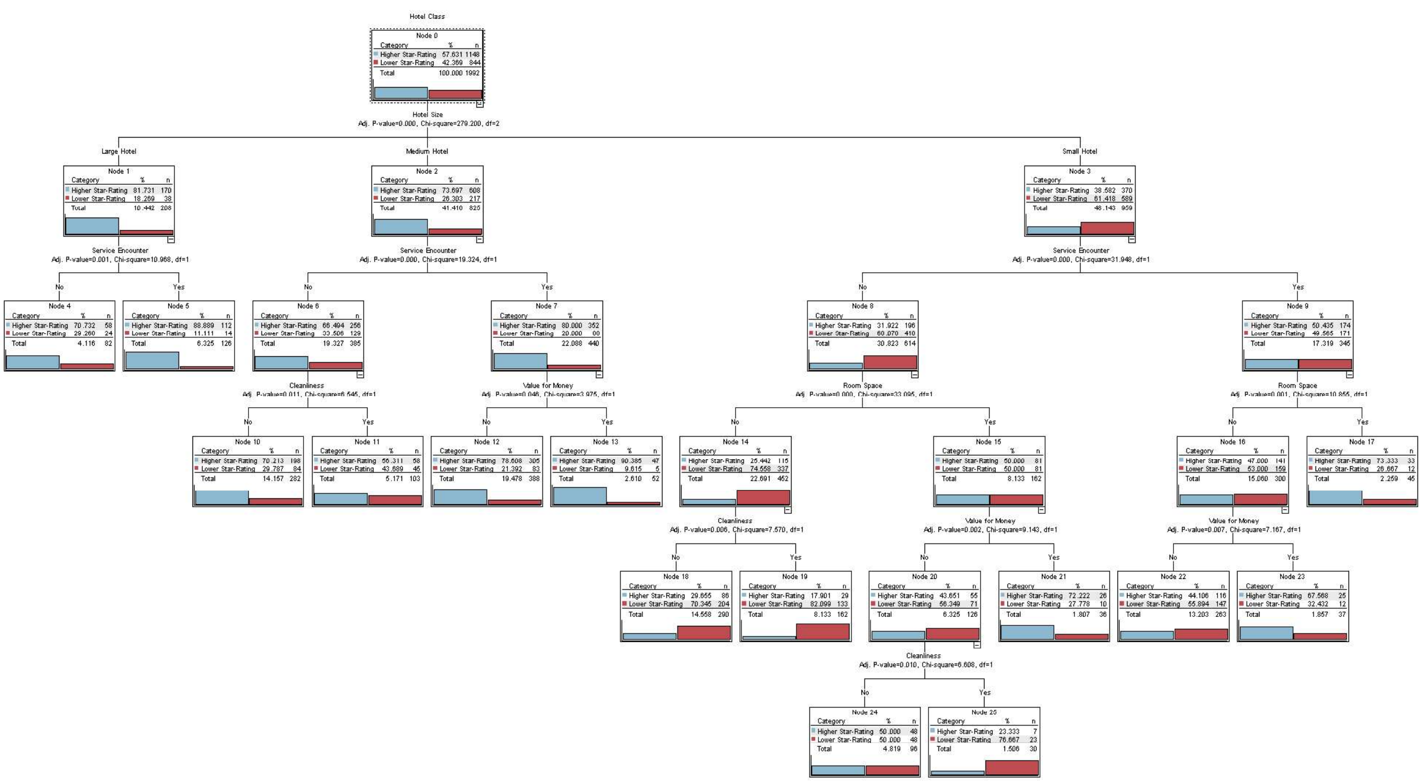

Figure 1. Online complaining behavior for different hotel classes using whole dataset (100%).

5. Discussion and Conclusion

This study aims to enrich literature on Big Data Analytics and Data Mining to the field of hospitality and tourism industry by predicting the complaint attributions significantly differing from different hotel classes of travelers related to their OCB. The study achieved this goal by applying the classification models to analyze TripAdvisor complaint reviews in the United Kingdom. Due to the methodology advantages of manual content analysis and Data Mining algorithms, this research not only corroborate, but also go beyond the conclusions reached by previous studies by revealing the significant differences in OCB from various hotel classes.

The main contribution of this study lie in the fact that this is one of the innovative papers to predict the OCB in the tourism and hospitality industry by utilizing machine learning algorithms, while previous studies most often relied on traditional methods (e.g. surveys or questionnaires). By analyzing the real world data (i.e. complaint reviews) allows researchers to discover additional empirical and quantitative study; specifically, DT algorithms.

REFERENCES

Abubakar, A. M., & Ilkan, M. (2016). Impact of online WOM on destination trust and intention to travel: A medical tourism perspective. *Journal of Destination Marketing & Management, 5*(3), 192-201.

Cenni, I., & Goethals, P. (2017). Negative hotel reviews on TripAdvisor: A cross-linguistic analysis. *Discourse, Context & Media, 16*, 22-30.

Golmohammadi, A. R., Jahandideh, B., & O'Gorman, K. D. (2012). Booking on-line or not: A decision rule approach. *Tourism Management Perspectives, 2-3*, 85-88.

Guerreiro, J., & Moro, S. (2017). Are Yelp's tips helpful in building influential consumers? *Tourism Management Perspectives, 24*, 151-154.

Moro, S., Rita, P., & Coelho, J. (2017). Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip. *Tourism Management Perspectives, 23*, 41-52.

TripAdvisor (2018). *London 2018: Best of London, England Tourism—TripAdvisor Hotel Statistics*. Available at: https://www.tripadvisor.com/Tourism-g186338-London_England-Vacations.html, Accessed on 9-Nov-18.

Tsao, W. C., Hsieh, M. T., Shih, L. W., & Lin, T. M. Y. (2015). Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management, 46*, 99-111.