Citation for published item:
Moro, S. & Esmerado, J. (2021). An integrated model to explain online review helpfulness in hospitality. Journal of Hospitality and Tourism Technology. 12 (2), 239-253

Further information on publisher's website:
10.1108/JHTT-01-2020-0026

An integrated model to explain online review helpfulness in hospitality

Abstract

Purpose

This study proposes a model to explain online review helpfulness grounded on both previously identified constructs (e.g., review length) and new ones which have been analyzed in other online reviews' contexts, but not to explain helpfulness.

Design/methodology/approach

A total of 112,856 reviews published in TripAdvisor about 21 Las Vegas hotels were collected, and a random forest model was trained to assess if a review has received a helpful vote or not.

Findings

After confirming the validity of the proposed model, each of the constructs was evaluated to assess its contribution to explain helpfulness. Specifically, a newly proposed construct, the response lag of manager's replies to reviews, was among the most relevant constructs.

Originality/value

The achieved results suggest that hoteliers should invest not only in responding to the most interesting reviews from the hotel's perspective, but also that they should do it quickly to increase the likeliness of the review being considered helpful to others.

Keywords

Online reviews; helpfulness; hospitality; social media; TripAdvisor.

Introduction

To increase the reputation of online review systems, many online review platforms adopted a feature named helpful votes which enables other consumers to vote a review as being helpful. When a reader signals a review as being helpful, he/she is also signaling other readers that the review provides useful information, motivating future readers to pay more attention to that review in detriment of others. Therefore, unlocking the drivers that lead a reader to consider a review as helpful is key. Within hospitality, TripAdvisor was among the first platforms to adopt helpful votes for reviews. It is considered one of the richest information content platforms (Moro et al., 2017), making of it ideal to evaluate online review models through its features. Accordingly, Kwok and Xie (2016) proposed a model to explain review helpfulness based on the following constructs: (i) review length, (ii) score, (iii) reviewer characteristics, and (iv) manager's response. They adopted regression to validate their model using secondary data. However, a large majority of studies evaluating review helpfulness in several contexts has adopted a purely data-driven approach without proposing a conceptual model (e.g., Singh *et al.*, 2017). Such profusion of studies suggests the lack of a comprehensive model encompassing the distinct features deemed as relevant across the literature.

This study proposes an integrated conceptual model including constructs grounded on previous online reviews' research to evaluate the characteristics that distinguish a helpful from a non-helpful review in hospitality. The proposed model adopts a comprehensive set of features from which hypotheses are raised and tested. Additionally, by adopting TripAdvisor, this study takes advantage of a large volume data source containing several important features that characterize the online review, the travel context, and the reviewer.

One of those features is the speed to which unit managers respond to online reviews. This unchartered feature in the current body of knowledge is controllable by managers, unlike others related to the review (e.g., review length) or reviewer characteristics (e.g., reviewer experience). The model is evaluated through a data-driven approach based on data mining techniques to assure a robustness fit of the gathered data. Thus, the main contributions of the proposed conceptual model are:

- Including relevant constructs to online review helpfulness which are currently dispersed across the literature;
- Discovering new features that influence review helpfulness (e.g., the time that a manager takes to respond to a review).

Theory and conceptual model

Online review helpfulness

Electronic word-of-mouth has become an everyday reality since the Web 2.0 rose to empower consumers (Wang and Kubickova, 2017). As a result, social media emerged to provide information exchange platforms such as online review websites where users can freely write their opinions, as well as read what others have published (Mirzaalian and Halpenny, 2019). The hospitality industry has been among the first to adopt online reviews, paving the way through innovative platforms such as TripAdvisor and Yelp. These platforms have developed new features besides the usual review text and score and, currently, users can rate different items on TripAdvisor and are encouraged to write reviews through gamification features such as the user contributor level (Yang *et al.*, 2017). Additionally, readers can mark a review from another user as being helpful.

Helpfulness is recognized in existing literature as a key feature that brings value to users' online reviews about a product or service. Such perceived value by other users increases their trust in the opinions expressed in reviews deemed helpful, which can trigger decisions in the decision-making process (Mudambi and Schuff, 2010). Its importance has been assessed by Filieri (2015) and Singh *et al.* (2017), with both studies concurring to its contribute to influence users during the decision-making process. Therefore, a few studies have proposed conceptual models to explain the main constructs influencing the helpfulness of a review. Kuan *et al.* (2015) tested their model on a dataset of reviews collected from Amazon. They included constructs such as the review length, the sentiment valence, and the reviewer credibility. Siering *et al.* (2018) also evaluated their model using Amazon's reviews and included additional constructs such as the reviewer's experience and the month of review. Both studies applied regression to a large set of reviews. Yet, using secondary data restricts model's evaluation to the features offered by the source platform, and therefore the tested models cannot go beyond such scope.

Hypotheses and proposed model

Current state-of-the-art research to understand the drivers of a helpful review is based on primary data collected by inquiring users (Filieri, 2015) or on secondary data directly harvested from the studied platforms (Zhang and Tran, 2010). Especially due to the latter where data is restricted to what is available, the existing theoretical models propose several constructs which are not necessarily convergent, depending on the available features that enable to test such constructs. While review length is a widely used construct, some studies include the quantitative review score or total number of helpful votes the reviewer has received (Yang *et al.*, 2017), while others focus on the expressed sentiments (Singh *et al.*, 2017). Regarding the sentiment influence, Zhang and Lin (2018) identified "that consumers may not favor a very positive or slightly negative review but a very negative or slightly positive review" (p. 8). On the opposite, Chatterjee (2020) found that a higher polarity makes the review to be considered less helpful. There is not a consensus regarding the impact of the sentiment score computed from the textual contents of online reviews when compared to other features (e.g., those related to reviewer characteristics) (Lee *et al.*, 2018). Such inconsistent results provide evidence of a lack of a convergent path toward understanding review helpfulness. Nevertheless, the following individual hypotheses are raised based on the findings of the abovementioned studies:

H1: A lengthier review is more likely to be considered helpful when compared to shorter reviews.

H2: Reviews that have received lower scores are the most helpful ones.

H3: The sentiment score of a written review influences less the helpfulness of the review when compared to other features.

H4: A reviewer with more overall helpful votes as a TripAdvisor user is more likely to have its next written review being considered as helpful.

Table I exhibits eleven constructs based on existing literature, from which the first eight are known to influence review helpfulness. Those constructs are drawn from four studies based on Amazon (Singh *et al.*, 2017; Kuan *et al.*, 2015; Siering *et al.*, 2018; Ngo-Ye and Sinha, 2014, the latter also analyzed Yelp data) and the remaining three on TripAdvisor (Yang *et al.*, 2017; Filieri, 2015; Kwok and Xie, 2016). Yet, some of them have been evaluated through primary data collected using questionnaires, such as the case of the scores granted to specific categories, which was analyzed by Filieri (2015). This study discovered their relevance to review helpfulness, but did not scrutinize each individual category, which can be directly extracted from TripAdvisor (i.e., value, location, sleep, rooms, cleanliness, service). Other studies adopting the six TripAdvisor categories and corroborating the influence in the overall guest satisfaction include the one by Rhee and Yang (2015). Therefore, an extended hypothesis to the one confirmed by Filieri (2015) would validate if each of the individual categories rated in TripAdvisor influence helpfulness:

H5: Each of the six rating categories in TripAdvisor (e.g., value, location, sleep, rooms, cleanliness, service) positively influences the review helpfulness.

Table I

However, existing online review literature has found evidence of other constructs which can be extracted from TripAdvisor that influence traveler's satisfaction, such as the user experience as a TripAdvisor member, and the travel type (the last three from Table I). Thus, the proposed model extends previous literature by including such constructs, which can also play a role in explaining review helpfulness, i.e., users reading reviews are also influenced by them. By linking the review score, which is considered a proxy for guest satisfaction (Moro *et al.*, 2017), with the review helpfulness (Yang *et al.*, 2017), one would expect that the reviewer's experience would influence a written review helpfulness. This is in line with the findings by Ngo-Ye and Sinha, (2014), Moro *et al.* (2017), and Yang *et al.* (2017), who identified an influence on TripAdvisor score of the number of reviews written by a reviewer, membership years as TripAdvisor member, and contributor level, respectively. Additionally, specifically for the case of contributor level, Yang *et al.* (2017) found it to be less relevant than both review length and review score by analyzing 1,158 reviews from a single hotel located in New York City. Therefore, the following hypotheses may be raised:

H6: The reviewer contributor level is less relevant to review helpfulness when compared to review length and score.

H7: The number of reviews contributed by a user influences a written review helpfulness.

H8: The membership years as a TripAdvisor member influences a written review helpfulness.

Seasonality is a widely studied construct in tourism literature. Regarding review helpfulness, Siering *et al.* (2018) found that reviews written during August are less helpful than those written in other months. They suggest that such result is explained by reviewers spending less time in writing during August, which is a vacation month in most Western economies. Thus, the following hypothesis can be raised:

H9: The month influences the helpfulness of a review, with reviews written during August being less helpful.

Different types of travelers were found to grant different scores and reviews according to their type. Chang *et al.* (2019) identified in TripAdvisor reviews that business travelers use more negative words, while couples tend to grant higher scores, corroborating similar findings by Radojevic *et al.* (2018). Nevertheless, as the latter authors mention, business travel usually also involves some leisure activities, limiting the influence of travel type in data-driven models. Thus, it can be argued that:

H10: The influence of travel type to understanding review helpfulness is limited.

Additionally, Kwok and Xie (2016) found that reviews with a manager's response are more likely to receive helpful votes. The model proposed in this study extends such construct by also considering the response lag, i.e., how much time elapsed between the review's data of publication and the manager's response, for the cases where there is a response. Thus, the feature is extended from binary ("yes", if the manager responded, "no" otherwise) to a categorical one with four categories: (i) the manager responded within a day of the review's publishing date; (ii) the manager responded within a week;

(iii) the manager responded but only more than a week of publishing date; or (iv) he/she never responded. The rationale for testing such construct is that reviews on TripAdvisor are shown from the most recent to the older ones and, although it offers several filtering options, there are no other ordering possibilities. Thus, readers are more likely to read the most recent reviews, and managers delay in responding to reviews may affect the visibility of their response. This is a previously unchartered feature in current state-of-the-art literature. Based on the above premises, the following hypothesis can be proposed:

H11: Reviews to which managers take longer to respond tend to be considered less helpful.

The theoretical model proposed by Chatterjee (2020), which already includes the review (e.g., sentiment score extracted from the text; quantitative rating) and reviewer (e.g., user's helpful votes) dimensions, was adopted. Then, based on the highlighted constructs from Table I, such model is extended to include also the manager's response, and travel context, among others. The proposed extended model is shown in Figure 1, organized under three dimensions: (i) the review itself (including both user inputs and the manager's response), (ii) user information, and (iii) the travel context.

Figure 1

Also, the response by managers is further scrutinized by considering the response lag instead of just if a review has a response by a manager. Additionally, the proposed model includes constructs known from previous studies to influence review helpfulness, such as the expressed sentiments and the user contributor level, a gamification feature of TripAdvisor to promote user participation in reviewing hotels.

Data and approach

The case of Las Vegas Strip

Considering the hospitality industry is highly affected by both the hotel unit and its location, the Las Vegas Strip was chosen, which is the main avenue in Las Vegas, US, where the largest resorts holding casinos are located. Hence, the relative homogeneity (large units with casinos) in the offer can help in framing the results and limit the effect of each unit's service which influences guest satisfaction. The 21 chosen hotels were the same as in the study by Moro *et al*. (2017) and are characterized in Table II. Yet, instead of the 504 reviews collected manually by the aforementioned authors, a large dataset of 112,856 online reviews was collected through web scraping, characterized by the 16 variables that translate each of the constructs identified on Table I.

Table II

Furthermore, as also highlighted by Moro *et al*. (2017), the destination image associated with the Strip, as well as the types of amenities offered by its hotels, enables to restrict constructs related to both location and hotel unit which could affect the model. Hence, the model can be framed under the identified constructs (Figure 1) without the risk of losing accuracy derived from such uncontrolled variables.

Methodology

A web scraping script was developed using the R statistical language (specifically, the "rvest" package was adopted) to extract a total of 112,856 online reviews written between January/2015 and May/2018 from TripAdvisor related to the 21 hotels located on the Strip, Las Vegas (Table II). Several of the constructs included are translated by variables that needed to be computed based on the information extracted, while the remaining could be directly used to validate the model. Table III highlights each of the variables used by showing the same names as those previously described in Table I (the column "computed indicates which were computed). With the exception of the review's sentiment score (variable "sent"), which was obtained using the "sentimentr" package, all the remaining were computed using the base package from the R statistical tool. The six specific category scores were transformed from numeric (the original format on TripAdvisor, $\in\{1,2,3,4,5\}$) into categorical, considering that: (i) this is not a mandatory feature for publishing a review, thus there are many missing values (those were marked as "Undefined") and (ii) most of the granted scores were above 3. A review was considered helpful if it received at least one helpful vote (23,008 from the total).

Table III

Previous studies have chosen to evaluate their conceptual models using regression (Kwok and Xie, 2016). To evaluate the proposed model, two data mining techniques, random forest (RF), and neural networks (NN), were chosen. The RF is an ensemble of decision trees that make individual contributions to an overall model, thus improving the overall performance by benefiting from the heterogeneity of several decision trees. The NN attempts to mimic the human brain by connecting neurons (or nodes) fed by input variables in a network to reach a decision about an output variable (Denton *et al.*, 1990). The multilayer perceptron was adopted, which is a popular architecture with one hidden layer composed by several hidden nodes, with each node triggered through an activation function. Each node's output is computed by weighting previous nodes' outputs.

To validate the models, the 10-fold cross-validation scheme was adopted, where a subset 1/10 of reviews is iteratively selected by first being removed from the training set of the model, and then to test the obtained model. There are several metrics to measure the performance of classification models. Two of the most robust, according to Moro *et al.*

(2014), were adopted: (i) the area under the Receiver Operating Characteristic curve (AUC), and (ii) the area under the cumulative Lift curve (ALIFT).

By assessing the data mining model's accuracy in modeling review helpfulness, it can be confirmed that the proposed constructs can help in explaining helpfulness. Thereafter, the need of understanding the contribution of each individual feature was mandatory. To do so, the data-based sensitivity analysis (DSA) was adopted, which provides the percentage of relevance that each feature has in explaining helpfulness (Cortez and Embrechts, 2013). DSA assesses how sensitive the model is to varying each input variable by changing simultaneously several features randomly selected from the dataset. All modeling experiments were accomplished using the R statistical tool and the "rminer" package.

Results and analysis

Table IV exhibits standard statistics for understanding the distribution of values per variable. Table V shows the performance results for the models built using RF and NN. By comparing both, it is clear that RF achieved the best performance by far.

Table IV

Table V

The obtained results from Table V enable to validate the RF as capable of modeling the helpfulness based on the proposed constructs (Moro *et al.*, 2014, achieved an AUC below 80%). Thus, the proposed model (Figure 1) is considered a valid representation of reviews' helpfulness. Figure 2 exhibits the whole model signaling the relevance computed through DSA in percentage of individual constructs using a gray scale where the most relevant construct has a black background (i.e., "nword"), and the least relevant has a white background (i.e., "score.loc"). Since each construct is translated into a variable (Table III), with the exception of specific score categories, which are six and surrounded by a dark gray square, Figure 2 shows just the variable names, for display purposes only.

Figure 2

Three out of the four most relevant constructs were also found the most relevant for review helpfulness by Yang *et al.* (2017), namely, the review length, the overall score, and the total number of helpful votes the user has ever received. However, the order is different, since the aforementioned study ranked first the overall score, then the total number of reviews of the user, and finally, the total number of words, which was considered the most relevant in the present study. The specific data used plays a key role in the obtained results and can help to justify the differences. Yang *et al.* (2017) collected only 1,158 reviews of a single hotel located in New York, not benefiting from a large dataset.

To further scrutinize how each of the four most relevant analyzed constructs affect helpfulness, the DSA can be used to plot graphics (i.e., variable effect characteristic curves) that relate each variable with the goal (the probability of a review being considered helpful) (Cortez and Embrechts, 2013). Thus, Figures 3 to 6 enable to explain how the review helpfulness is influenced by each of the three constructs.

Figure 3

Figure 4

Figure 5

Figure 6

Discussion and conclusions

Conclusions

In this section, the raised hypotheses from the conceptual model (Figure 1) are assessed, enabling to draw conclusions from the findings. The result shown on Figure 3 partially confirm H1. For reviews up to 300 words, a lengthier review increases its helpfulness. Yet, helpfulness gets reduced for reviews above 500 words, contracting the finding by Yang *et al.* (2017). It should be stated, however, that those authors considered only three levels of review length, with the lengthier being reviews with more than 61 words. The 112k reviews used in this study have in average 125 words, with a standard deviation of 142, suggesting that the three levels used by Yang *et al.* (2017) are insufficient to capture review length's influence. Furthermore, the presented results on Figure 3 are corroborated by Kuan *et al.* (2015, p. 63): "a longer review attracts more attention and motivates reading up to a certain point, and discourages processing when the additional cognitive effort anticipated exceeds the incremental value expected from extra length". Also, Kwok and Xie (2016) did not corroborate such finding. They hypothesize that the fact a reader needs to click the "more" link to see more than five lines of a review may have shadowed the relevance of this construct. Yet, they then debate that their result contradicts the uncertainty reduction theory, which states that review readers value information rich reviews (Daft and Lengel, 1986). While they suggest future research in other e-commerce settings to further validate such theory, both ours and the abovementioned studies' findings contradict the results by Kwok and Xie (2016), showing there is still road to cover in evaluating the relevance of online review length.

Regarding H2 related to the influence of review score, the result from Figure 4 is a confirmation of the corresponding hypothesis raised and validated by Kwok and Xie (2016), in which lower rated reviews tend to be considered more helpful. Nevertheless, although reviewer experience has shown to play a significant role to helpfulness in prior studies (Kwok and Xie, 2016; Ngo-Ye and Sinha, 2014), none has shown how the total number of reviews influences the probability of a review being considered helpful. In what concerns to H3, the sentiment score appears with a reduced influence of just 3.89%, when compared to other features. This is especially true when considering the review related features included in this study, review length and review score. Thus, H3 is also confirmed.

Regarding H4, Figure 6 shows that new users with few helpful votes are the ones whose reviews are especially considered helpful, with the probability decreasing around 12% for a reviewer who have reviewed 100 helpful votes. After this threshold, review helpfulness gradually increases. Therefore, H4 is only partially confirmed. One hypothesis for this result for novice users can be related to a propagating effect of increased trustworthiness by these users who usually also have written fewer reviews and thus who are less likely to write incentivized reviews by hotel managers (Filieri, 2016). Yet, this discovery calls for further research in other contexts to assess if the same result is found.

Another interesting finding relates to the six specific score categories. Filieri (2015) suggested their relevance through an assessment based on questionnaires. Nevertheless, this is the first study proposing a model that measures the relevance of this construct. H5 proposes that each of the six TripAdvisor categories positively influences review helpfulness. The presented results in Figure 2 only partially corroborate H5, suggesting context information may play a key role in the relevance that each category has to the review perceived helpfulness. For example, the significantly low relevance of location and value can be most likely explained as a side-effect of choosing the Strip in Las Vegas. Besides the location being the same, the 21 chosen resorts are very large and hold a casino, constituting a similar perceived offer by guests, which is even overshadowed by the strong destination image associated with Las Vegas (Moro *et al.*, 2017).

Previously, the influence of the number of helpful votes a user has already received was scrutinized under H4 assessment. Regarding the remaining three reviewer variables (i.e., usr.level, usr.reviews, and usr.member.months), related to hypotheses H6, H7, and H8, H6 is confirmed by comparing usr.level relevance of 2.79% to both review length and score, the two most relevant features to review helpfulness, with relevance above 15% each. However, both H7 and H8 are only partially confirmed, with relevance around 5% each.

The remaining two new constructs introduced by the proposed model, the travel type, and the month of review (see Table I), play a smaller role in explaining helpfulness, both with a relevance of around 5%. Thus, H9 is only partially confirmed. However, H10, which posits that the influence of travel type is limited, is confirmed, based on the argument that the influence of travel type is overshadowed by the difficult in distinguishing leisure from business, limiting its contribution to understanding review helpfulness (Radojevic *et al.*, 2018).

One of the most interesting contributions relates to the discovered relevance of the newly proposed construct related to manager’s response lag to online reviews. In fact, this variable rises in third place, contributing to explain more than 13% of helpfulness. Figure 5 shows how such influence is exerted. This result not only confirms that a manager’s response positively affects review helpfulness (Kwok and Xie, 2016), but also quantifies such influence. Therefore, H11 is confirmed. As reviews on TripAdvisor are ordered by recency (the most recent ones appear first), its visibility likely decreases once new reviews are written and take the place of older ones. Accordingly, it is imperative that not only managers respond to the most interesting reviews from the hotel's perspective, but also that they respond quickly. This is an important managerial contribution, suggesting that managers can still play an influential role in an online world of empowered consumers.

Theoretical implications

This study proposes a conceptual model including both quantitative (e.g., review length) and qualitative features (e.g., sentiment score). According to the recently published study by Chatterjee (2020), there are "very few papers, which combine information generated from both qualitative data of the textual review" (p. 2). Further, the model has proven to be accurate in explaining review's helpfulness by using a comprehensive set of 16 features related to the review (contributed by both the reviewer and the unit manager), to the reviewer as a TripAdvisor member, and to the travel context.

The major contribution of this study to the existing body of knowledge concerning review helpfulness is the discovery that manager's response lag to a review influences its perceived helpfulness by other users. This is a previously unchartered finding that requires further discussion to deepen its understanding. Kwok and Xie (2016) have previously identified that managers' responses play a role in review helpfulness but did not assess the impact of the response speed. Additionally, Min *et al.* (2015) found that the speed of the response does not turn the review more useful. However, the same authors developed a questionnaire where they have asked the participants about the relevance of the response speed instead of directly quantifying the real response speed. Thus, there must be another reason that justifies that the response speed was found to be highly relevant on the present study. The most plausible justification lies in the TripAdvisor website paging mechanism. When viewing the reviews of a hotel, TripAdvisor shows by default only the five most recent reviews. There are some options to help in filtering the reviews, such as by quarter of stay, by travel type, by language, and by a reduced set of keywords they named "popular mentions". Yet, there is not a free text input filter to select only reviews containing user written keywords. Therefore, the most recent reviews are the most read, with older reviews appearing in higher page numbers. Thus, the longer a manager takes to respond to a review, the less visible is his/her response, turning it into less likely that the review with a response is valued by readers, since only those that iterate through the pages will read it.

Practical implications

Hospitality unit managers are currently aware of the importance that online reviews have for their potential future guests. When a review is considered helpful by a reader, its perceived usefulness for other readers is also amplified due to the electronic word-of-mouth effect (Racherla and Friske, 2012). Thus, by understanding what it takes for a review to be considered helpful, managers can intervene in the only variable they control that has previously been proven to influence the review helpfulness, which is the response to the guest's review (Filieri, 2016). From a practical standpoint, this study proves that not only is it important that managers answer reviews, but also that they do it quickly. Since in TripAdvisor the reviews are sorted by date (with the most recent first), in a pagination system with five reviews per page, with no other ordering option and with limited filters, a review being answered quickly implies that the manager's response is more likely to be seen by a wider audience. Hence, if managers take too long to answer, the visibility of the response becomes limited, justifying the finding. Another interesting discovery is that reviews published by novice users (in writing reviews) are more easily considered as being helpful. This finding calls for more research, showing there is still many to uncover when it comes to user behavior in online reviews, despite the large number of related studies.

Limitations and future research

This study has some limitations. While choosing the Strip in Las Vegas enabled to frame the undertaken experiments, it also raises the limitation of being restricted to a single location. Such limitation has overshadowed the possible influence of the specific location score category. Furthermore, while several of the hotels from the dataset collected are among the largest in the world, the number of rooms per unit ranges from 188 to 4027, which influences each one's service due to the available resources to meet guests' expectations. Also, an extended set of features could include reviewers' cultural background based on the country of origin, which is a field available in TripAdvisor users (albeit not being mandatory). Additionally, while secondary data available from online reviews has several advantages (e.g., it is already available, it can be retrieved in large volumes), it also has some drawbacks. Namely, it is limited to the features available on the data source (in this case, TripAdvisor). Therefore, future research may test and enrich the proposed conceptual model based on primary data collected through instruments such as questionnaires.

References

Chatterjee, S. (2020), "Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach", *International Journal of Hospitality Management*, Vol.85, 102356.

Chang, Y.C., Ku, C.H. and Chen, C.H. (2019), "Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor", *International Journal of Information Management*, Vol.48, pp.263-279.

Cortez, P. and Embrechts, M.J. (2013), "Using sensitivity analysis and visualization techniques to open black box data mining models", *Information Sciences*, Vol.225, pp.1-17.

Daft, R.L. and Lengel, R.H. (1986), "Organizational information requirements, media richness and structural design", *Management Science*, Vol.32 No.5, pp.554-571.

Denton, J.W., Hung, M.S. and Osyk, B.A. (1990), "A neural network approach to the classification problem", *Expert Systems with Applications*, Vol.1 No.4, pp.417-424.

Filieri, R. (2015), "What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM", *Journal of Business Research*, Vol.68 No.6, pp.1261-1270.

Filieri, R. (2016), "What makes an online consumer review trustworthy?", *Annals of Tourism Research*, Vol.58, pp.46-64.

Kuan, K.K., Hui, K.L., Prasarnphanich, P. and Lai, H.Y. (2015), "What makes a review voted? An empirical investigation of review voting in online review systems", *Journal of the Association for Information Systems*, Vol.16 No.1, pp.48-71.

Kwok, L. and Xie, K.L. (2016), "Factors contributing to the helpfulness of online hotel reviews: Does manager response play a role?", *International Journal of Contemporary Hospitality Management*, Vol.28 No.10, pp.2156-2177.

Lee, P.J., Hu, Y.H. and Lu, K.T. (2018), “Assessing the helpfulness of online hotel reviews: A classification-based approach”, *Telematics and Informatics,* Vol.35 No.2, pp.436-445.

Liu, S., Law, R., Rong, J., Li, G. and Hall, J. (2013), “Analyzing changes in hotel customers’ expectations by trip mode”, *International Journal of Hospitality Management*, Vol.34, pp.359-371.

Min, H., Lim, Y. and Magnini, V.P. (2015), “Factors affecting customer satisfaction in responses to negative online hotel reviews: The impact of empathy, paraphrasing, and speed”, *Cornell Hospitality Quarterly*, Vol.56 No.2, pp.223-231.

Mirzaalian, F. and Halpenny, E. (2019), “Social media analytics in hospitality and tourism”, *Journal of Hospitality and Tourism Technology,* Vol.10 No.4, pp.764-790.

Moro, S., Cortez, P. and Rita, P. (2014), “A data-driven approach to predict the success of bank telemarketing”, *Decision Support Systems,* Vol.62, pp.22-31.

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

Mudambi, S.M. and Schuff, D. (2010), “Research note: What makes a helpful online review? A study of customer reviews on Amazon.com”, *MIS Quarterly,* Vol.34 No.1, pp.185-200.

Ngo-Ye, T.L. and Sinha, A.P. (2014), “The influence of reviewer engagement characteristics on online review helpfulness: A text regression model”, *Decision Support Systems*, Vol.61, pp.47-58.

Racherla, P. and Friske, W. (2012), “Perceived ‘usefulness’ of online consumer reviews: An exploratory investigation across three services categories”, *Electronic Commerce Research and Applications*, Vol.11 No.6, pp.548-559.

Radojevic, T., Stanisic, N., Stanic, N. and Davidson, R. (2018), “The effects of traveling for business on customer satisfaction with hotel services”, *Tourism Management*, Vol.67, pp.326-341.

Rhee, H.T. and Yang, S.B. (2015), “Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers’ overall ratings”, *Computers in Human Behavior*, Vol.50, pp.576-587.

Siering, M., Muntermann, J. and Rajagopalan, B. (2018), “Explaining and predicting online review helpfulness: The role of content and reviewer-related signals”, *Decision Support Systems*, Vol.108, pp.1-12.

Singh, J.P., Irani, S., Rana, N.P., Dwivedi, Y.K., Saumya, S. and Roy, P.K. (2017), “Predicting the "helpfulness" of online consumer reviews”, *Journal of Business Research*, Vol.70, pp.346-355.

Wang, C.R. and Kubickova, M. (2017), “The impact of engaged users on eWOM of hotel Facebook page”, *Journal of Hospitality and Tourism Technology*, Vol.8 No.2, pp.190-204.

Yang, S.B., Shin, S.H., Joun, Y. and Koo, C. (2017), “Exploring the comparative importance of online hotel reviews’ heuristic attributes in review helpfulness: a conjoint analysis approach”, *Journal of Travel & Tourism Marketing*, Vol.34 No.7, pp.963-985.

Zhang, R. and Tran, T. (2010), “Helpful or Unhelpful: A Linear Approach for Ranking Product Reviews”, *Journal of Electronic Commerce Research*, Vol.11 No.3, pp.220-230.

Zhang, Y. and Lin, Z. (2018), “Predicting the helpfulness of online product reviews: A multilingual approach”, *Electronic Commerce Research and Applications*, Vol.27, pp.1-10.

Table 1 - Constructs proposed based on existing literature.

Construct	**Variable**	**References**
Review length	nword	(Singh *et al.*, 2017) (Kuan *et al.*, 2015) (Yang *et al.*, 2017)
Expressed sentiments (computed sentiment score through sentiment analysis)	sent	(Singh *et al.*, 2017) (Kuan *et al.*, 2015)
Overall review score granted	overall.score	(Kwok and Xie, 2016) (Yang *et al.*, 2017)
Reviewer contributor level (from TripAdvisor)	usr.level	(Yang *et al.*, 2017)
Total nr. helpful votes the user has received	usr.helpful	
Seasonality	month	(Siering *et al.*, 2018)
Reviewer's experience in writing reviews	usr.reviews	(Ngo-Ye and Sinha, 2014) (Kwok and Xie, 2016)
Scores granted to specific categories help consumers to learn about the quality of a product as they summarize reviewers' evaluations of the main features of a product (i.e., cleanliness; location; rooms; service; sleep; value)	score.clean score.loc score.rooms score.service score.sleep score.value	(Filieri, 2015)
Travel type: {as a couple, on business, solo, with family, with friends}	travel.type	(Liu *et al.*, 2013)
If and when a manager responds to a user (within a: {day, week, more, never})	response.lag	(Kwok and Xie, 2016)
For how long the user has been a TripAdvisor member	usr.member.months	(Moro *et al.*, 2017)

Table 2 - Constructs proposed based on existing literature.

Hotel	Stars	Nr. Rooms
The Venetian Las Vegas Hotel	5	4027
Excalibur Hotel & Casino	3	3981
Bellagio Las Vegas	5	3933
Circus Circus Hotel & Casino Las Vegas	3	3773
Caesars Palace	5	3348
The Palazzo Resort Hotel Casino	5	3025
Monte Carlo Resort &Casino	4	3003
The Cosmopolitan Las Vegas	5	2959
Paris Las Vegas	4	2916
Treasure Island- TI Hotel & Casino	4	2884
Wynn Las Vegas	5	2700
Encore at Wynn Las Vegas	5	2034
Tropicana Las Vegas - A Double Tree by Hilton Hotel	4	1467
Trump International Hotel Las Vegas	5	1282
Hilton Grand Vacations on the Boulevard	4	1228
The Westin Las Vegas Hotel Casino & Spa	4	826
Wyndham Grand Desert	3	787
Marriott's Grand Chateau	4	732
Tuscany Las Vegas Suites & Casino	3	716
Hilton Grand Vacations at the Flamingo	3	315
The Cromwell	4	188

Table 3 - Collected variables.

Dimension	Variable	Computed?
Review	nword	Y
	sent	Y
	overall.score	N
	score.clean	Y
	score.loc	Y
	score.rooms	Y
	score.service	Y
	score.sleep	Y
	score.value	Y
	response.lag	Y
User	usr.level	N
	usr.helpful	N
	usr.reviews	N
	usr.member.months	Y
Travel context	month	N
	travel.type	N

Table 4 - Variables analysis.

Numeric Variables	Data description		
	Type	**Mean**	**SD**
nword	Integer (number of words of the review)	125.1	142.3
sent	Numeric (sentiment classification, with 0 being neutral sentiment)	0.238	0.203
overall.score	Integer, {1,2,3,4,5}, (score granted)	4.15	1.12
usr.level	Integer, {0,1,2,3,4,5,6}	2.61	2.13
usr.helpful	Integer	19.6	73.6
usr.reviews	Integer	40.1	96.5
usr.member.months	Integer	39.1	39.8
Categorical Variables	**Distribution of occurrences per category**		
score.clean	"3 or less"=6,748; "4"=10,506; "5"=35,673; "Undefined"=59,929		
score.loc	"3 or less"=5,112; "4"=10,079; "5"=37,583; "Undefined"=60,082		
score.rooms	"3 or less"=9,109; "4"=10,652; "5"=31,217; "Undefined"=61,878		
score.service	"3 or less"=15,739; "4"=17,032; "5"=47,574; "Undefined"=32,511		
score.sleep	"3 or less"=7,799; "4"=10,365; "5"=33,029; "Undefined"=61,663		
score.value	"3 or less"=12,740; "4"=13,216; "5"=26,627; "Undefined"=60,273		
response.lag	"day"=30,633; "week"=31,464; "more"=5,776; "never"=44,983		
month	Jan=7,746; Feb=8,701; Mar=10,499; Apr=11,175; May=10,820; Jun=9,391; Jul=9,825; Aug=9,666; Sep=9,589; Oct=9,514; Nov=7,367; Dec=8,563		
travel.type	"as a couple"=45,808; "on business"=16,502; "solo"=5,373; "with family"=25,088; "with friends"=20,085		

Table 5 - Models' performance metrics.

	AUC	**ALIFT**
Random forest (RF)	80.75%	74.54%
Neural network (NN)	75.85%	70.58%

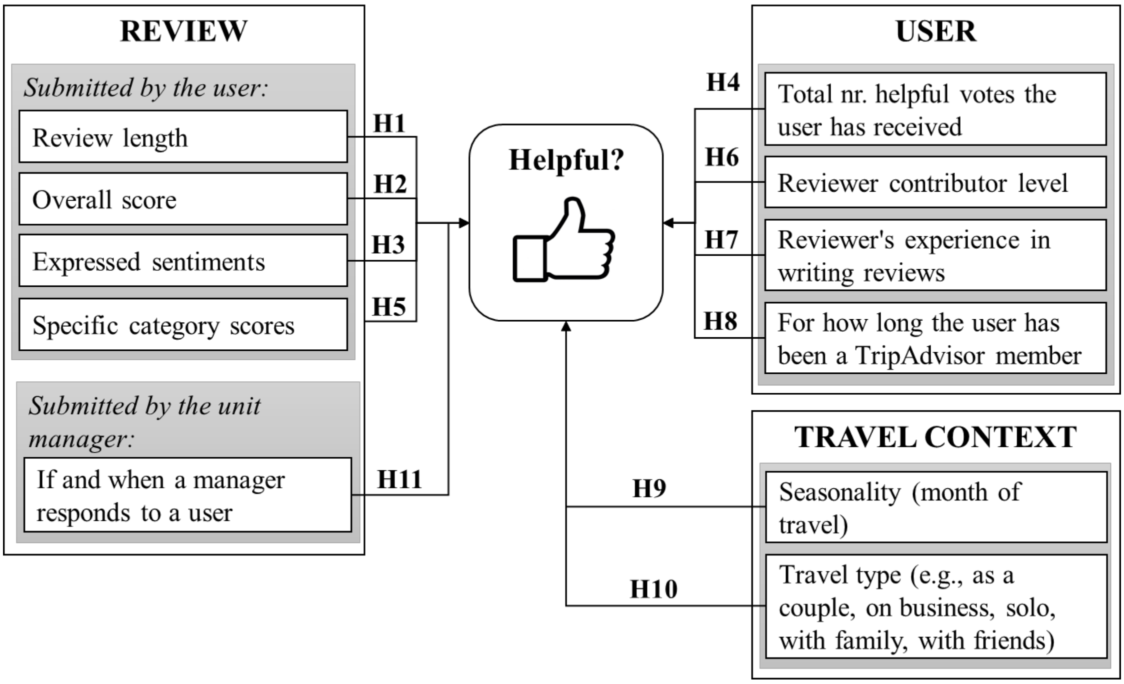

Figure 7 - Conceptual model proposed.

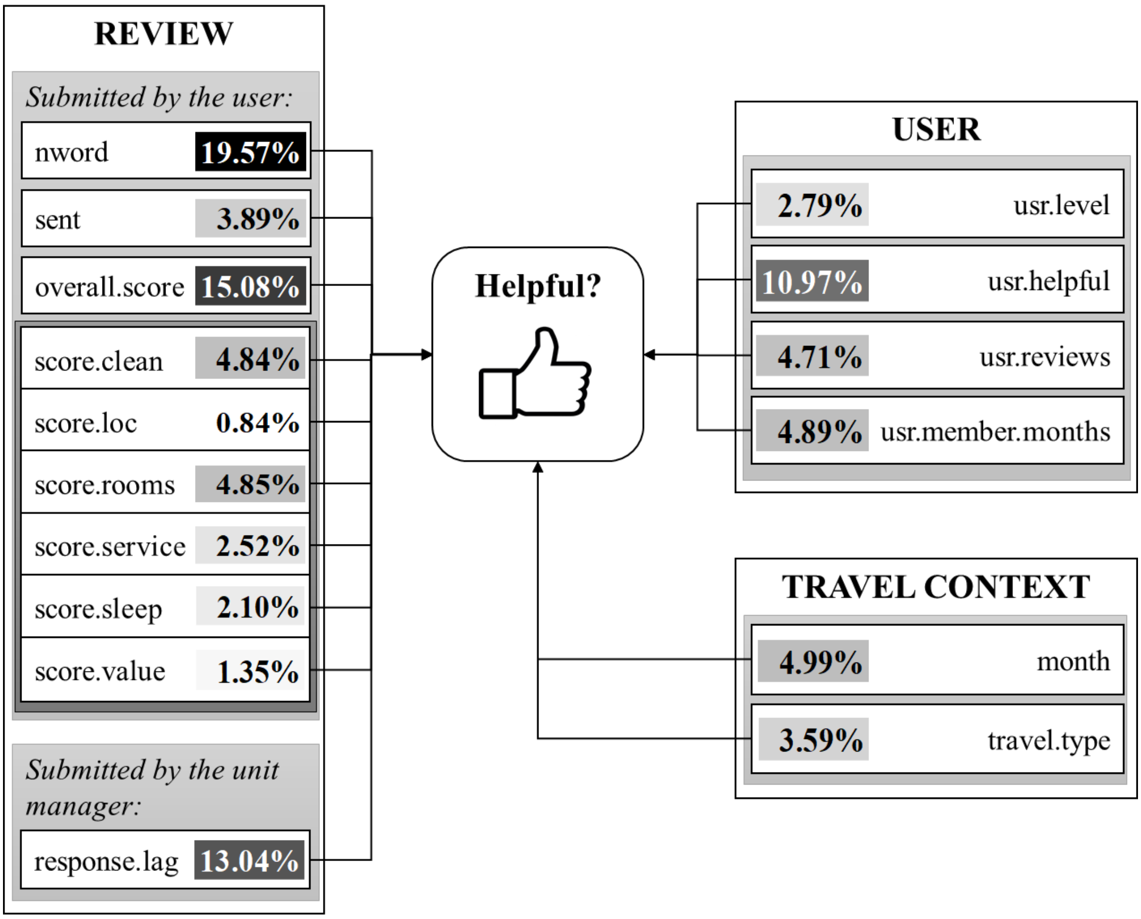

Figure 8 - Relevance of each construct to explain review helpfulness.

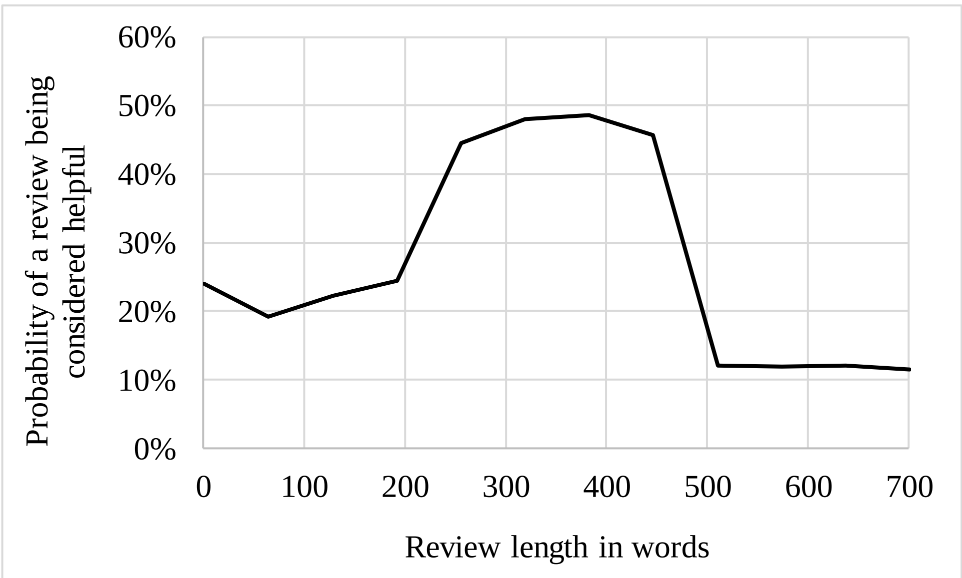

Figure 9 - Review length's influence in review helpfulness.

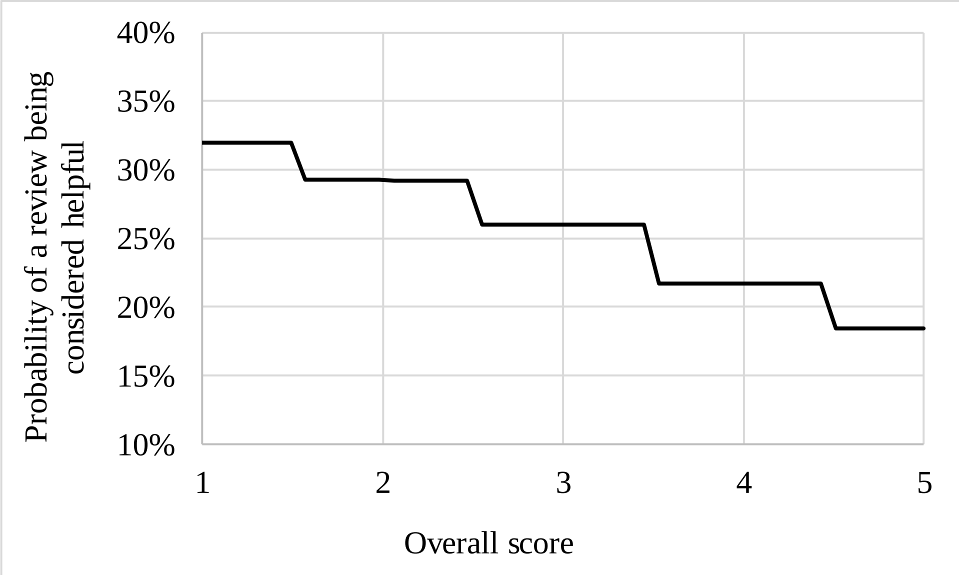

Figure 10 - Overall score's influence in review helpfulness.

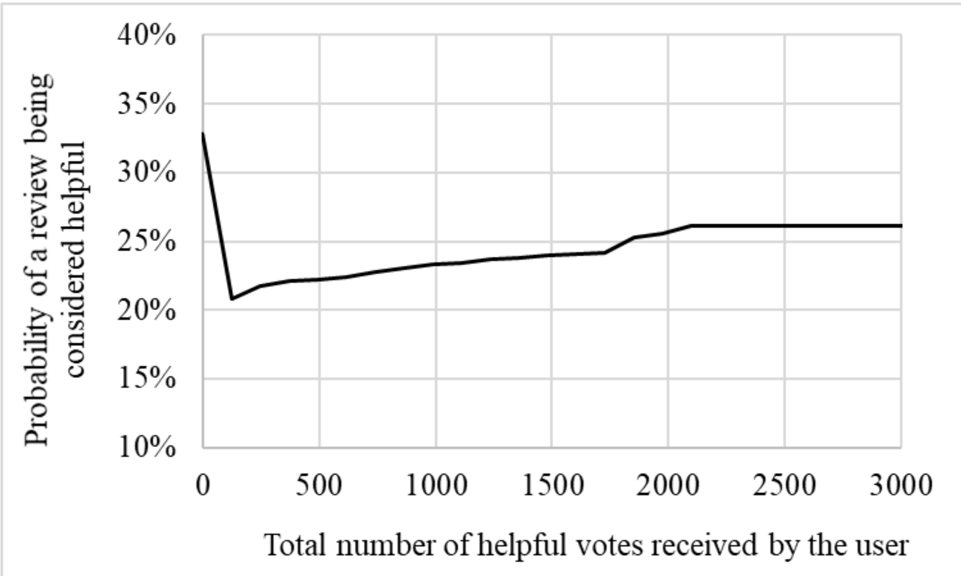

Figure 11 - User's total number of helpful votes influence in review helpfulness.

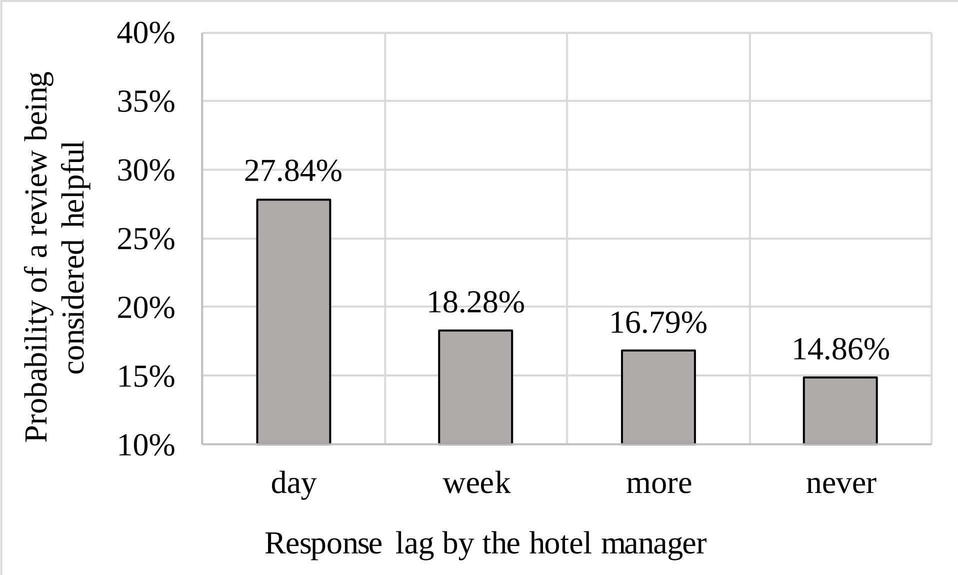

Figure 12 - Response lag influence in review helpfulness.