Applied Soft Computing Journal 101 (2021) 107045

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

A linguistic multi-criteria decision making methodology for the evaluation of tourist services considering customer opinion value

Itzcóatl Bueno [a], Ramón A. Carrasco [b], Carlos Porcel [c], Gang Kou [d,*], Enrique Herrera-Viedma [c,*]

[a] Statistics Faculty, Complutense University of Madrid, Avenida Puerta de Hierro, s/n, 28040 Madrid, Spain
[b] Department of Management and Marketing, Complutense University of Madrid, 28223, Madrid, Spain
[c] Andalusian Research Institute of Data Science and Computational Intelligence (DaSCI), University of Granada, Granada, 18071, Spain
[d] School of Business Administration, Southwestern University of Finance and Economics, Chengdu 610074, China

ARTICLE INFO

Article history:
Received 1 August 2020
Received in revised form 15 November 2020
Accepted 14 December 2020
Available online 28 December 2020

MSC:
00-01
99-00

Keywords:
Fuzzy linguistic modeling
Customer opinion value
Multi-Criteria Decision-Making
Evaluation of tourist services

ABSTRACT

As a consequence of the exponential growth in online data, tourism sector has experimented a radical transformation. From this large amount of information, opinion makers can be benefited for decision making in their purchase process. However, it can also harm them according to the information they consult. In fact, being benefited or harmed by the information translates into greater or lesser satisfaction after the purchase. This will largely depend on the published opinions that they take into account, which in turn depend on the value of the opinioner who publishes said information. In this paper, the authors propose a methodology that integrates multiple decision-making techniques and with which it is intended to obtain a ranking of hotels through the opinions of their past clients. To do this, the customer value is obtained using the Recency, Frequency, Helpfulness model. The information about the users found in the social networks is managed and aggregated using the fuzzy linguistic approach 2-tuples multi-granular. In addition, we have verified the functionality of this methodology by presenting a business case by applying it on TripAdvisor data.

1. Introduction

Due to the expansion of new technologies and the intensive use of social networks, clients have increased the publication of their opinions in forums and social networks to value products and services, this has strengthened the relationship with the client through the Internet [1]. The electronic word of mouth (eWOM) in this type of customer reviews has a great impact on the process of purchase or selection of services by the consumer [2–5]. The customer has access to an infinity of other users opinions who have enjoyed of the products and services in which they are interested. For that reason, it is worth considering the customer's value based on their online opinions and how the rest of the users notice them as useful or reliable.

These opinions are published on the web mainly through two formats. Firstly, satisfaction surveys [6–8] are a common tool used in different comparison shopping websites, sellers websites, etc. to obtain an overall assessment of the product and/or its characteristics, which normally represents the degree of agreement or disagreement with each characteristic through Likert scales [9, 10]. This type of scale was defined by [11]. Users respond to each question with their degree of agreement or disagreement. In [12], the Likert scale is described as a set of items with approximately the same number of positive and negative possibilities and a central point. For example, on a 5-point scale, responses are usually "strongly disagree", "disagree", "neutral", "agree" and "strongly agree". However, these responses are characterized by the uncertainty and blurriness of the perception they represent [13], since the same concept can indicate very different perceptions [14]. Some authors [15] consider that an approach based on the use of linguistic evaluations would be better to model these human perceptions than that of conventional numbers (crisp).

However, not all characteristics that influence customer satisfaction can be identified using questionnaires or surveys [16]. This limitation could be overcome by analyzing the opinions provided by the users in natural language. In [17], it is stated that there are 3 main online review formats. Firstly, in the 'Pros and Cons' format, the reviewer describes pros and cons of the product or service separately. Secondly, in the 'Pros, Cons and detailed review' format, the reviewer describes pros and cons of the product or service separately and include a detailed review. Finally, in the 'Free format' there is no separation between pros and cons, so the reviewer can freely write their opinion.

* Corresponding authors.
E-mail addresses: kougang@swufe.edu.cn (G. Kou), viedma@decsai.ugr.es (E. Herrera-Viedma).

https://doi.org/10.1016/j.asoc.2020.107045

Although the reviews provide a more complete information about a product or service and their sub-categories, surveys are the way traditionally used to measure customer satisfaction. These post-purchase surveys ask to rate the customer satisfaction with the product or areas of the product. The answers do not provide as rich data as reviews, but they will give the elements to build quality ratings. However, not all of these ratings come from questionnaire responses. An example of these second type of ratings is the TripAdvisor's Popularity Index, which includes the frequency of opinions among its elements. Other scales such as the star system for hotels or the credit rating (AAA, AA +, …) were used before these quality index were developed, however in some domains they are being replace by these new ratings that incorporate elements such as the customer satisfaction which is measured through their questionnaire answers or the sentiment analysis of their reviews [15]. Between both alternatives, users prefer more to express their experience posting comments than filling out satisfaction questionnaires. It could be said that this new way of measuring the quality of a product or service is of great economic importance since that as it is suggested in [18] there is a strong relationship, in the European countries studied in their work, between consumer satisfaction and the expenditure it makes. Furthermore, their results show the importance of satisfied customers in the economy as a whole.

These ratings are a fundamental guide in the purchasing process for customers. During the process, it is usual not only to look at the ratings (hotel stars, global scores, etc.) but also to obtain information through the responses that other users give to the satisfaction questionnaires and by reading their opinions expressed in natural language. In the literature [19–21] has been widely discussed the different importance of each user. Therefore, there are users more influential in the purchase process of other customers, that is, the value perceived as an opinioner is important for their opinion be more or less valued. For tourism sector, [19] showed that if a user is recognized as experienced and trustworthy, their opinions can be very influential in the purchase decisions made by other users. Many studies have been done on the influence that user-generated content has on the tourism sector [22–24].

The different platforms obtain these ratings with the aggregations of these questionnaires. However, they do not take into account the influence that customers may have as opinion-makers or the content of the opinions they express in natural language. As we have mentioned, natural language posts provide richer information and more extensive than the questionnaires. Therefore, taking into account both is a considerable improvement from our point of view. For this reason, in this work we set ourselves the goal of building a hotels evaluation model that integrate both factors, providing a more reliable measurement of hotels.

Both types of information, opinions expressed in natural language and satisfaction questionnaires, are not lacking uncertainty and exist diverse approaches in the literature in which they have been modeled by fuzzy linguistic variables [25–27]. Specifically, the 2-tuple and multi granular model meets a double objective: it allows to model the associated fuzziness to the items in the questionnaires and to the output of the sentiment analysis processes applied to the opinions posted in natural language, and also the aggregations of this information can be done without loss of information. Therefore, it is considered an adequate framework to propose our model.

On the other hand, there are various models to obtain the perceived value of the opinioner, many of them inherited from more general models of customer value. One of the most relevant is the RFM model [28–30] that values a customer based on the recency, frequency and monetary value of the purchase, which adapted to the perceived value of the opinioner is RFH model: recency, frequency and helpfulness of their opinions [31,32]. The main advantage of the RFM model is its easy linguistic interpretation. However, it lacks great precision. Therefore, using the RFM 2-tuples model it is possible to maintain the linguistic interpretation and improves the precision of the base model.

These three variables are relatively simple to obtain, so we are going to base on RFM model. However, we will make a more precise version based on the idea of expanding the RFM in [5] to obtain the perceived value of the opinioner. A frequent problem in this type of models is to aggregate the three variables in an unique perceived value. In order to make this integration we will consider a valid Multi-criteria decision making (MCDM) [33–37] for these purposes with the Analytical Hierarchy Process (AHP) [38–41].

Once the perceived value of customer is obtained, we face a second problem: integrating the diverse questionnaires' items online answered by the customers and the results of the sentiment analysis models obtained from the natural language opinion posted by these clients. One more time, we can consider this problem as a MCDM for which we will consider again the AHP and, of course, the perceived value of the opinioner previously obtained.

With the methodology that we propose in this work, we address both problems. On the one hand, we integrate both types of information through multi-granular language modeling, and on the other, we obtain the value of customer opinion using the RFH model. The result is a ranking that is not based on the characteristics of the hotel (price, stars, location, etc.), but would be generated from the opinions of its past clients. Being the opinion makers themselves who influence the value of a hotel for its positioning in the ranking, we consider that it will produce changes in the online purchase dynamics. In such a way that users will trust more in online recommendations, facilitating their purchase process, predisposing them to share their experiences, and also, the best-rated hotels could increase their sales. In the literature there are proposals that have used the RFM [28] for different purposes in the tourism sector and even its adaptation, the RFH [32]. On the other hand, there are also contributions that incorporate fuzzy linguistic modeling in their proposals [42] and others that use AHP or extensions of it [43]. However, we consider that our methodology is novel since, to the best of our knowledge, all these techniques are combined in order to obtain a more accurate hotel ranking. Moreover, its applicable in any web application given it just need the input data and the set up of an expert or user. It makes us consider that our methodology in addition to being innovative would be highly applicable in different business cases.

We consider that the previous work has not been able to achieve the idea that we proposed, or at least completely. For example, in [44] the AHP method is used to calculate the weight of the criteria that the hotel guests influence in the decision-making process. His approach is to assess the TripAdvisor criteria from the consumer's point of view. However, it does not fully capture the value of the client, by using limited information. It only takes into account the answer to questionnaires. In [28], the RFM model is used for the clustering of hotel clients. This represents an advance, by incorporating the RFM model. However, both the information from reviews and questionnaires are omitted, only the characteristics of the reservations are taken into account. Also, this is done from the hotels point of view to improve their strategies. In the works of Carrasco et al. [5,45], fuzzy methods are incorporated into the RFM model, but continues to use the model to improve business decisions. Regarding helpfulness, we have been able to find little that can be related to our work. On the one hand, [30] is about predicting the helpfulness of reviews. On the other hand, [32] is the work that is closest to our proposal. The authors segment the reviewers using k-means based on the RFH model. After this, the reviewers' value is calculated using fuzzy AHP and the segments are ranked. However, it continues to omit all information from reviews or questionnaires and its objective is to rank users. Instead, our goal is to provide a more reliable ranking of hotels in order to facilitate customer decision-making.

The remainder of this paper is organized as follows. In Section 2, a state of the art presenting how customer value is defined in the literature, how utility opinions influence consumer value and some ways to measure it. In Section 3 is presented the elements necessary to carry out our proposal. In Section 4, the proposed multi-granular RFH 2-tuples model is presented, which provide a linguistic recommendation ranking after applying an AHP model. In Section 5 we present the software implementation of our methodology. To test our proposal, we present a use case in Section 6 using TripAdvisor data. A discussion of the obtained results is made in Section 7. Finally, conclusions and future work are outlined in Section 8.

2. State of the art

In this section, we present the previous work in customer value, electronic word of mouth (eWOM) and RFM.

2.1. Customer value

In recent years, the relationship with the customer via online has been intensifying through their opinions about the business, the products and services they acquire or enjoy, etc. on company websites, forums or social networks. As Graf and Maas [46] exposed, in the literature various definitions for customer value have been given and they summarize some of them in Table 1.

However, in this new context in which online opinions of customers are growing exponentially, and with them the amount of information without enough utility to ease the final decision to other users. It is a fact that the helpfulness of these reviews is becoming a key factor in order to determine the customer value. Therefore, [51] proposed an improvement in the model developed by Mudambi and Schuf [52] to identify which reviews are useful and which are not. In addition, in their proposal they made a review of what other authors have considered that makes an opinion useful, as can be: the quality of the posted opinion (lack of spelling mistakes, readability, etc.) [53] or if it is posted by an identified user [54,55].

Instead, [30] stated that an opinion is considered useful when it satisfies two conditions. Firstly, the customer has really read the review and, after evaluating it, he/she considers it is useful. Secondly, the review can provide valuable information and impact even more in the final decision of the customer. And, the authors affirm, based on these two conditions, that the reviews with a higher number of feedback votes will be more likely to be read by other users.

2.2. eWOM

Due to the growth that communication on the Internet has experienced, the number of user publications has also increased rapidly in volume. This has brought word of mouth (WOM) to social networks, generating eWOM [56]. In [6], eWOM is defined as "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet". eWOM is a main key in the customers online purchasing process decision. In fact, according to Search Engine Lab,[1] 92% of users take into account the online reviews in order to determine if the business are good or not. Moreover, 87% of consumers will not consider businesses with low ratings. Finally, Search Engine Lab reported that 80% of users considered online reviews as important as personal recommendations.

This stats are close to those reported by TripBarometer[2] 2017/18, a global report published by TripAdvisor. It reports that 86% of travelers will not book accommodation without reading reviews first and 89% usually do some research on a destination's activities and restaurants before travel. Additionally, the report states that *'Word of mouth recommendations are one of the more influential sources of information despite being used by a relatively small proportion of travelers'*. Moreover, TripAdvisor can be considered, in tourism sector, as one of the most visited online review sites [57]. In summary, many studies have been done on the impact of eWOM in tourism and hospitality sectors [58–60]. In tourism sector, how the eWOM influence hotel booking intentions has been widely studied [61–63].

Helpfulness of the reviews is an eWOM attribute that can contribute to influencing in customers booking intention [2]. In [64] helpfulness was proposed for the first time as a component of reviews. The evaluation system of reviews in TripAdvisor allows the users to vote if they consider useful a post. This allows users to filter the large number of reviews based on their usefulness [30,65]. Moreover, travelers consider that the most valuable reviews should be the most reliable and therefore useful [66–68]. Therefore, the utility is linked to the reservation intention [69–71].

Finally, it is important to note that all this information can also be modeled with fuzzy methods. For example, in [72] a fuzzy analytic hierarchy process approach is proposed. It is done in order to determine customer's most preferred movie reviews platform. Also in [73] the Fuzzy AHP (FAHP) method was adopted. In this case, determine the importance of criteria in evaluating customer's trust towards online SNSs sellers. In addition, [74] presents a new collaborative recommendation approach. It uses the fuzzy linguistic approach to represent multicriteria user–item preference ratings, and computes recommendations using fuzzy aggregation-based approach. In Table 2 we show some of the work developed on applications of the fuzzy approach to eWOM in the tourism sector.

2.3. RFM

A way to determine the customer value is through the RFM model, which tries to predict consumer purchases based on their past transactions. These transactions are characterized by three indicators that summarize the purchase behavior (recency, frequency and monetary (economical value of the transaction)), combining them a global score is obtained that allows to segment the customers and take different marketing strategies based on their customer values. Ngo-Ye and Sinha [31] have applied this model to evaluate the engagement characteristics on online review helpfulness. In fact, Aakash y Jaiswal [32] replace one of the RFM model indicators introducing the helpfulness indicator. Thus, they proposed to segment the customers based on a RFH model, whose indicators are: recency, frequency and helpfulness.

Customer value is usually measured on numerical scales, being the result of combining some factors such as the total number of transactions, average value of purchase, the frequency of purchase, total number of acquired products, etc. An example of this type of numerical scales is the value which is given to

[1] https://searchengineland.com/87-percent-customers-wont-consider-low-ratings-228607 (accessed October 2020).

[2] https://mk0tainsightsjao4bom.kinstacdn.com/wp-content/uploads/2018/10/TripBarometer-2017-2018.pdf (accessed October 2020).

Table 1
Definitions of customer value (Graf & Maas, 2008) [46].

Zeithaml (1988) [47]	"Perceived value is a customer's overall assessment of the utility of a product based on perceptions of what is received and what is given."
Gale (1994) [48]	"Customer value is market perceived quality adjusted for the relative price of your product. [It is] your customer's opinion of your products (or services) as compared to that of your competitors."
Holbrook (1994) [49]	Customer value is "a relativistic (comparative, personal, situational) preference characterizing a subject's [consumer's] experience of interacting with some object ...i.e., any good, service, person, place, thing, event, or idea."
Woodruff (1997) [50]	Customer value is a "customer's perceived preference for and evaluation of those product attributes, attribute performance, and consequences arising from use that facilitate (or block) achieving the customer's goals and purposes in use situations."

Table 2
Previous fuzzy methods for eWOM analysis in tourism sector.

Author	Description
Chou, T.Y et al. (2008) [25]	A fuzzy multi-criteria decision making (FMCDM) model for international tourist hotel location selection is presented.
Nilashi, M et al. (2019) [75]	They propose a new recommender system for hotel recommendations in e-tourism platforms. According to their experiments, it improves the quality of recommendations in tourism domain.
Doğan, S et al. (2020) [76]	Authors use fuzzy rule-based system (FRBS) which hotel attributes used in Travel 2.0. data impact on price–performance (PP).

customers through the application of the RFM methodology [77–79]. An advantage of this type of scales is its easy interpretation from the point of view of the user, being on an n-point scale, normally 1 means the worst client for a given variable, while n corresponds to the best client. However, they are not very accurate. In [5] is proposed the use of 2-tuple model in order to obtain more accuracy without loss of information. For this reason, we consider it is appropriate to use this philosophy in our methodology to obtain the customer opinion value. As we have mentioned before, the RFM model is characterized by its easy linguistic interpretation, although it does not have good accuracy. For this reason, we consider it appropriate to use the 2-tuples approach in our methodology to obtain the customer opinion value. Then, using the RFM 2-tuples model, we will achieve that easy linguistic interpretation while improving the precision of the RFM model.

3. Preliminaries

In this section we present the main components in which is based our proposal: Fuzzy Linguistic Modeling, RFM model and its 2-tuple extension, and AHP.

3.1. Fuzzy linguistic modeling

Fuzzy logic is presented as an alternative to traditional logic, with the aim to introduce grades of uncertainty to the statements that it interprets [80]. The information we handle in the real world may have different ranges of valuation, and the values may have a different nature. For this reason, at times, it may not be easy to assess accurately it using a quantitative value, however, it may be feasible to do so qualitatively. In this case, adopting a linguistic approach usually offers better results than applying a numerical one. There are situations in which information, by its own nature, cannot be quantified and, therefore it is needed to be valued through the use of linguistic terms. For example, when we make an assessment of a book we have read, we usually use terms such as good, regular or bad. In other cases, working with precise information in a quantitative way is not possible. Either because the necessary elements are not available to carry out an exact measurement of that information, or because the computational cost is too high and with the application of an approximate value is enough for us. In this sense, the use of fuzzy set theory has given very good results for the qualitative treatment of information [80,81]. Fuzzy Linguistic Modeling is a tool that allows representing qualitative aspects. It is based on the concept of linguistic variables, that is, variables whose values are not numbers, but words or statements expressed in natural or artificial language [80]. Each linguistic value its characterized by a syntactic value or label, and a semantic value or meaning. The label is a word or statement that belongs to a set of linguistic terms and the meaning is a fuzzy subset in a discourse universe.

3.1.1. 2-tuple fuzzy linguistic modeling

Herrera and Martínez developed in [82] a fuzzy linguistic representation model in which a pair of values (s, α) defined as 2-tuple represent the linguistic information. This pair is formed by s that is a linguistic label, and by α that represents the value of the symbolic translation.

In their paper, they consider that let a linguistic term set $S = \{s_0, \dots, s_\kappa\}$ and a value that represent the result of a symbolic aggregation operation ($\beta \in [0, \kappa]$), then the 2-tuple expressing the equivalent information to β is obtained by:

$$\Delta[0,\kappa] \longrightarrow S \times [-0.5, 0.5)$$
$$\Delta(\beta) = (s_i, \alpha), \text{ with } \begin{cases} s_i & i = round(\beta) \\ \alpha = \beta - i & \alpha \in [-0.5, 0.5) \end{cases} \quad (1)$$

where $round(\cdot)$ is the round operation, s_i has the closest index label to β and α is the value of the symbolic translation. Moreover, from a 2-tuple is always possible to return its equivalent numerical value $\beta \in [0, \kappa]$ by a Δ^{-1} function. Other useful operator is the known as negation operator, which is described as:

$$neg((s_i, \alpha)) = \Delta(\kappa - \Delta^{-1}(s_i, \alpha)) \quad (2)$$

Finally, in our model we just use one aggregation operator which is the weighted average and it is defined as:

Definition 1. Let S = $\{(s_1, \alpha_1), \dots, (s_n, \alpha_n)\}$ be a set of linguistic 2-tuple and $\Omega = \{\omega_1, \dots, \omega_n\}$ be their associated weights. The 2-tuple weighted average $\bar{S}^\omega$ is:

$$\bar{S}^\omega[(s_1, \alpha_1), \dots, (s_n, \alpha_n)] = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot \omega_i}{\sum_{i=1}^{n} \omega_i}\right) \quad (3)$$

For further information on the 2-tuple linguistic representation model, see [82] and [83].

3.1.2. Multi-granular fuzzy linguistic modeling

The cardinality of the linguistic term set S, also known as the 'granularity of the uncertainty', is an important parameter to be set up in any approach. A linguistic term set has more or less terms relying on the uncertainty grade expressed by an expert on

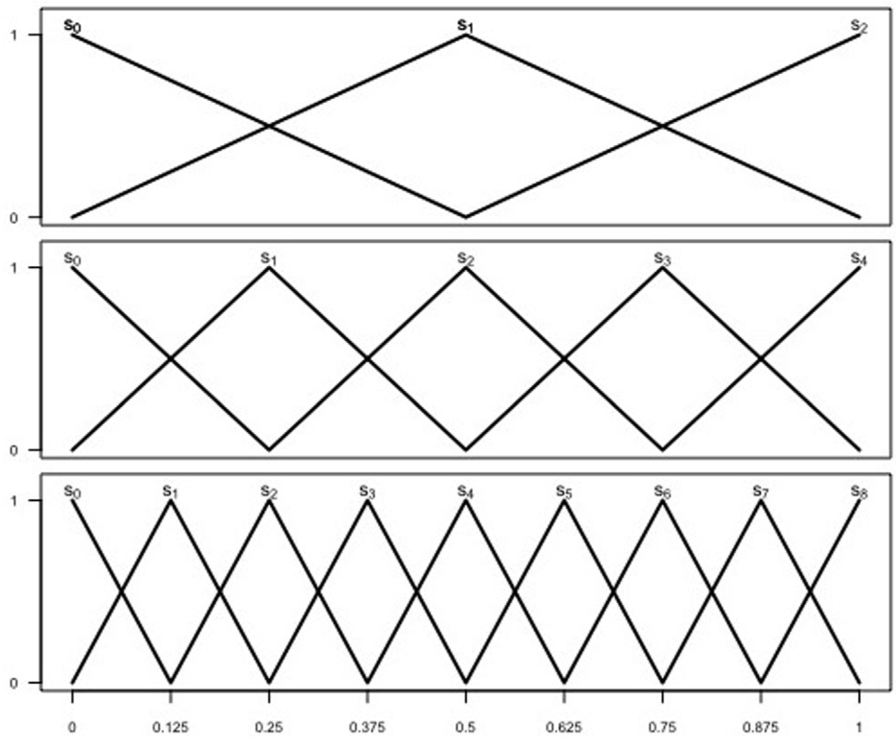


Fig. 1. Linguistic Hierarchy of 3, 5 and 9 terms.

the study case. Occasionally, experts have different uncertainty grades for a given phenomenon or they have to value distinct ideas. When one of these situations occurs, it is necessary to use various linguistic term sets with different granularities. A tool to manage the multi-granular information in those situations is proposed in [83], they present the 2-tuple multi-granular fuzzy linguistic modeling based in the concept of linguistic hierarchy.

Definition 2. A linguistic hierarchy (LH) [83] is a set of levels, which are denoted as $l(t, n(t))$, where each level t is a linguistic term set with a different granularity $n(t)$. The levels are ordered by their granularity, such that for two consecutive levels t and $t+1$, we have that $n(t+1) > n(t)$, and we can define one based on the previous one such that $l(t, n(t)) \rightarrow l(t+1, 2 \cdot n(t) - 1)$. Thus, a linguistic hierarchy is defined as the union of all levels t

$$LH = \bigcup_t l(t, n(t)) \tag{4}$$

An example of a linguistic hierarchy with 3 levels and 3, 5 and 9 terms in each one is show in Fig. 1 and it would be defined as $l(t, n(t)) = \{l(1,3), l(2,5), l(3,9)\}$. In each level the linguistic terms would be the following: $S^3 = \{s_0^3, s_1^3, s_2^3\}$, $S^5 = \{s_0^5, s_1^5, s_2^5, s_3^5, s_4^5\}$ and $S^9 = \{s_0^9, s_1^9, s_2^9, s_3^9, s_4^9, s_5^9, s_6^9, s_7^9, s_8^9\}$

In [83] a group of transformation functions between labels of different levels, in both directions, is defined to integrate multi-granular linguistic information without loss of information.

Definition 3. Let $LH = \bigcup_t l(t, n(t))$ be a linguistic hierarchy whose linguistic term sets are denoted as $S^{n(t)} = \{s_0^{n(t)}, \ldots, s_{n(t)-1}^{n(t)}\}$. The transformation function between a 2-tuple that belongs to level t and another 2-tuple in level $t' \neq t$ is defined as:

$$TF_{t'}^{t} : l(t, n(t)) \longrightarrow l(t', n(t'))$$
$$TF_{t'}^{t}(s_i^{n(t)}, \alpha^{n(t)}) = \Delta\left(\frac{\Delta^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \cdot (n(t') - 1)}{n(t) - 1}\right) \tag{5}$$

Finally, to set up the computational model, it is necessary to choose a level in order to uniform the information (e.g., the highest granularity level), and then we are able to use the operators defined in Section 3.1.1

3.2. RFM 2-tuples

In 1994, Hughes [84] proposed a model used to analyze consumer behavior known as RFM. This segmentation technique consists of three dimensions that gives the initials that make up its name: Recency (R) represents the length of a time period since the most recent acquisition or visit to the establishment. It is measure in time units such as days, months, years, etc. Frequency (F) represents the total number of acquisitions or establishment visits during the studied period. Finally, Monetary (M) represents the total economic value of the acquisitions made during the studied period.

Once the analysis period is chosen, the aforementioned dimensions are gathered at a user level, so the RFM information is summarized in a table which will contain: user identifier, recency, frequency and monetary values for each user. Additionally, the users are arranged according to each RFM measure and are grouped in equal size classes, usually in quintiles. Thus, the measures (recency, frequency and monetary) are mutated into ordinal scores. Finally, in order to obtain a unique judge that describes jointly the RFM scores, a RFM Score is calculated. It is a weighted average of the R, F and M scores by the user-defined weights, and it is determined as follows:

$$RFM_i = \omega_R \times R_i + \omega_F \times F_i + \omega_M \times M_i \tag{6}$$

Incorporating the 2-tuple model it solves the main limitation of RFM model, which is the lack of accuracy in the scores calculation. First, the symmetric and uniformly distributed domain S is defined using 5 linguistic labels. The labels have a semantic meaning, depending on the use case, for the variables of the RFM model. Thus, let $S = \{s_0, \ldots, s_\kappa\}$ with $\kappa = 4$ its definition is showed in Fig. 2.

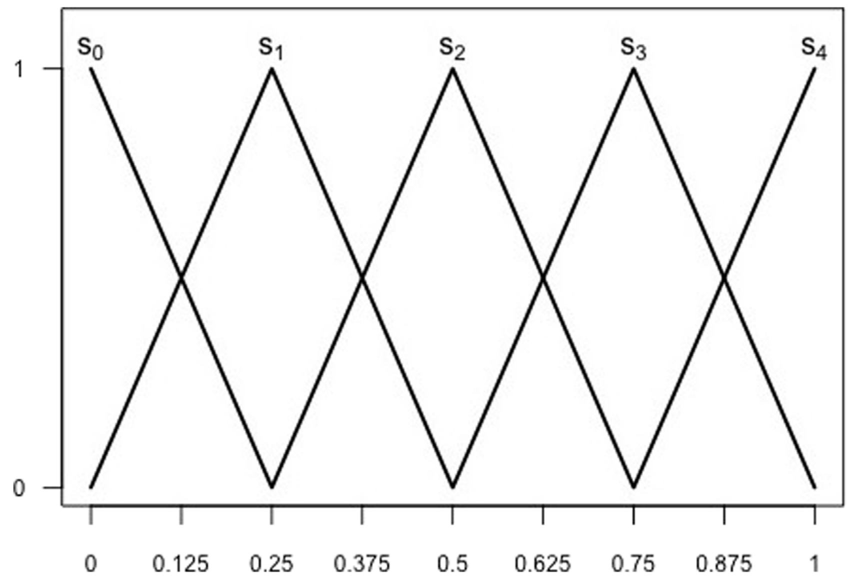


Fig. 2. Definition of the set S.

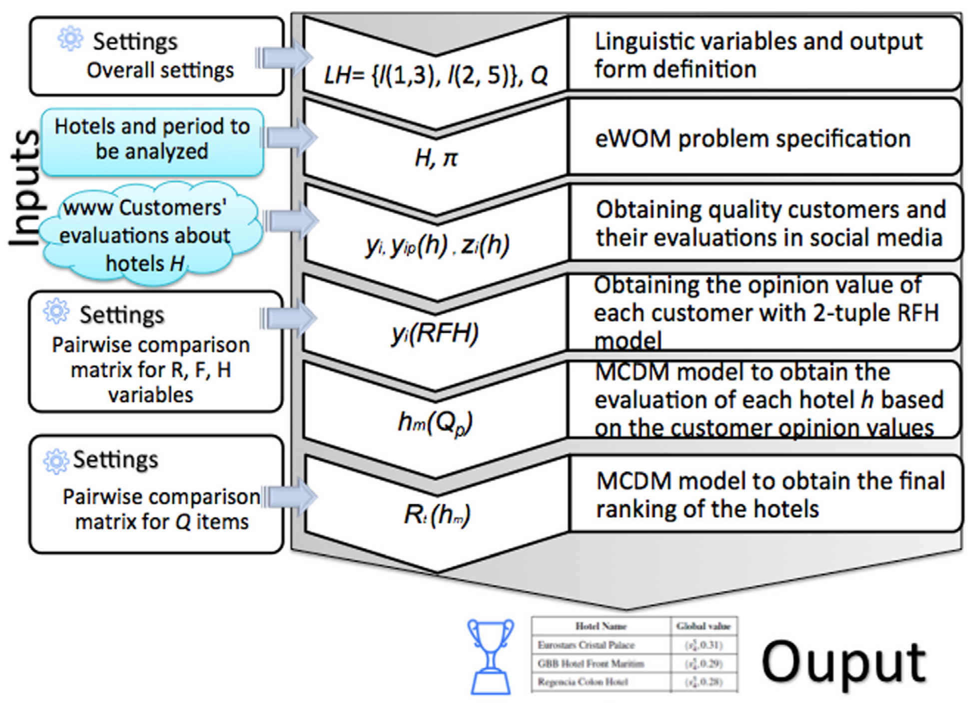


Fig. 3. Steps of our methodology.

Consequently, for each user we have $U_i = (U_{1i}, U_{2i}, U_{3i})$ $i = 1, \dots, n$ where U_{1i} is the recency score, U_{2i} is the frequency score and U_{3i} is the monetary score for user i. In a first step, users are arranged in ascending order according to each RFM measure. In a next step, the ranking of each user regard to each of the three measures is defined in the following way:

$$PercRank_{ij} = \frac{rank_{ij} - 1}{n - 1} \tag{7}$$

with $rank_{ij} \in, \{1, \dots, n\}$, $n > 1$, $PercRank_{ij} \in [0, 1]$, $i = 1, \dots, n$ and $j = \{1, 2, 3\}$. Therefore, the 2-tuple score U_{ij} is determined as:

$$U_{ij} = \begin{cases} \Delta(PercRank_{ij}) & \text{if } j \neq 1 \\ neg(\Delta(PercRank_{ij})) & \text{if } j = 1 \end{cases} \tag{8}$$

where $neg(\cdot)$ and $\Delta(\cdot)$ have been defined in Section 3.1.1. The negation operator is used on Recency because the highest scores mean the most recent users. Finally, the 2-tuple RFM Score, which describes jointly the RFM scores, is obtained using Eq. (3) for each user i as:

$$RFM_i^{Score} = \bar{S}^{\omega}[U_{ij}] \tag{9}$$

with the weights $\Omega = \{\omega_R, \omega_F, \omega_M\}$ defined by the user.

3.3. Analytical hierarchy process (AHP)

AHP was presented by Saaty [85–87]. Through a pairwise comparison matrix [88,89], this technique tries to provide a priority scale to a set of alternatives based on expert judgment on different criteria. A scale of absolute judgments (Table 3) is used to make the comparison of criteria, it is interpreted as how much more one criterion dominates another.

However, these judgments could be inconsistent [90,91] and therefore it is necessary to check their logic through the Consistency Ratio that is determined by $CR = \frac{CI}{RI}$ where CI is the Consistency Index obtained as in Eq. (10) and RI is the Random Index, which represents consistency of a randomly generated pairwise comparison matrix.

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{10}$$

Being λ_{max} the highest eigenvalue of pairwise comparison matrix A. Particularly, if $CR \leq 0.1$ the inconsistency is tolerable, and so a trustworthy result will be awaited from AHP model. Once the consistency of each pairwise comparison matrix is checked, the priority scale in each level of the hierarchy structure is obtained. Finally, these priority scales are combined by multiplying them by the priority of their parent nodes and adding for all such nodes.

4. Methodology

As we have mentioned, there are many ways to measure customer value. We want to take that value into account in order to rank stores, hotels, restaurants, products, etc. Therefore, in this section, we present the methodology of our proposal. Our model combines various decision-making techniques, achieving a novel way of measuring user satisfaction with a product or service.

In Fig. 3 we can see the inputs that each of the phases of the model receives either via parameterization or from the previous phase output. Before starting the process, hotel reviews and questionnaire answers are obtained. This can be done by crawling, scraping, web APIs, buying them, etc. It should be done daily or, at least, every so often according to the range defined by the user. This data collection phase includes the filter of not valid registers according to the criteria defined by the user. If we analyze the scheme step by step we will see that in the first phase the input are general settings such as the type of scales or data that the model will receive, and that the output returned will be the linguistic hierarchy (LH) to use and the questions (Q) of the questionnaires incorporated as inputs. The second step introduces a set of hotels (H) on which the questions (Q) obtained in the previous step must be measured. In addition a period of analysis (π) is defined, it will reduce the set H to the hotels that have received opinions in that period, this group of hotels is the exit of this phase. In the third step, the evaluations given by the opinion makers on the hotels contained in the set H during the π period are introduced for the different Q metrics that have been defined in the previous steps. This produces the following outputs: the information related to an opinioner (y_i), the rating that the opinioner gives to each of the questions in the questionnaire for a given hotel ($y_{ip}(h)$) and its opinion expressed in natural language about a hotel ($z_i(h)$). In step 4, three metrics of each opinioner are used as input, which can be obtained from the dataset already generated in previous steps. These inputs are: how many days ago they posted their last opinion, the frequency with which they have given their opinion during the analysis period and how useful they are perceived by the other opinion makers. Additionally, a matrix (A) of preferences in relation to these three dimensions must be parameterized to weight what is more relevant when evaluating an opinioner. After applying the RFH model, the output generated by this phase is a global customer value ($y_i(RFH)$). In step 5, we calculate the average value of the customer ratings $y_{ip}(h)$ obtained in survey questions for each hotel. Providing as output the values $h_m(Q_p)$, that represents the average value of the assessment for question Q_p of the hotel h_m. Finally, we use as input the average rating for each question on the questionnaire and for the opinions in natural language that each hotel receives. After that, we parameterize a matrix of preferences on the importance of the aspects to which it refers each question with respect to the others. Once we have the inputs and the setup, in this last step AHP is applied to obtain the final ranking of hotels ($R_t(h_m)$).

On the one hand, it should be noted that any parameterization would be done only once in the application of the model. On the other hand, it is also important to mention that our methodology is flexible and that it can be applied in a general way. It would simply be necessary to change the inputs (information obtained from social networks, the analysis period and the set of hotels) to run the model, in order to obtain the ranking of hotels.

The authors present a novel model formed by various components. Thanks to them, our model is characterized by obtaining the most reliable evaluation of hotels. Firstly, the combined use of natural language opinions that are treated with sentiment analysis and the opinions of the opinion-makers through forms as inputs to the model. All this information, which is imprecise, is modeled through the fuzzy linguistic representation, highlighting the use of the 2-tuple model that allows working without loss of information. Furthermore, as their linguistic representations belong to different groups, it has been possible to combine both types of information through a linguistic hierarchy. Once both types of information are conjugated through linguistic modeling without loss of information, obtaining the value of the client as an opinioner is done in a more accurate way thanks to the inclusion of the novel RFH 2-tuple model. Finally, by means of the AHP, the importance of each dimension of the RFH model can be parameterized to obtain a customer value according to the business specifications, as well as the importance of the criteria valued for each hotel in order to obtain a global rating, also attending to said business specifications. Using AHP ensures that all these settings are consistent by being in accordance with the commercial criteria specified by the experts.

4.1. Step 1: Linguistic variables and output form definition

In our methodology we are going to work with two types of information that we will model linguistically. On one hand, information obtained through forms. On the other hand, information in natural language obtained through the opinions of users. As part of setting up the methodology, experts have to decide the linguistic hierarchy. In the literature, it has been found that linguistic scales are usually always with an odd number of elements. On the one hand, for satisfaction surveys, the responses usually use Likert scales with 5 or 7 labels [92–94]. On the other hand, the direction of sentiment is usually classified in 2-points (positive and negative) or in 3-points (positive, neutral and negative) [95–97]. In addition, it is not recommended that the number of categories be very high according to [98]. Thus, for each problem based on the inputs, the linguistic hierarchy is designed.

An example of form that provide the first type information are those used by TripAdvisor (Fig. 4). In this type of forms, a group of aspects (Q) of the product, hotel, store, etc. is requested to be evaluated on a given scale. In the case of TripAdvisor, it is a value between 1 and 5 with the following linguistic translation: Terrible, Poor, Average, Very Good, Excellent as can be seen in the bubbles of Fig. 4 and they have been used in other works [99–101]. The information obtained from the forms will be modeled using a scale with an odd number of categories in order to have an equal number of positive and negative assertions on the attitude towards the object to be valued.

For example, the questionnaire presented in Fig. 4 taken from TripAdvisor uses for each aspect of a hotel a 5-point scale in

Table 3
The fundamental scale of absolute numbers.
Source: Saaty (2008) [87]

Intensity of importance	Definition	Explanation
1	Equal Importance	Two activities contribute equally to the objective
2	Weak or slight	
3	Moderate importance	Experience and judgment slightly favors one activity over another
4	Moderate plus	
5	Strong importance	Experience and judgment strongly favors one activity over another
6	Strong plus	
7	Very strong or demonstrated importance	An activity is favored very strongly over another; its dominance demonstrated in practice
8	Very, very strong	
9	Extreme importance	The evidence favoring one activity over,another is of the highest possible order of affirmation
Reciprocals of above		If activity i has one of the above non-zero numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i.
1.1–1.9	If the activities are very close	May be difficult to assign the best value but when compared with other contrasting activities the size of the small numbers would not be too noticeable, yet they can still indicate the relative importance of the activities.

Fig. 4. TripAdvisor form.

which each point represents one of the linguistic labels aforementioned and that has an equal number of positive and negative labels and other for a mean value. Therefore, we model these questions through sets, as the showed in Fig. 2, of five linguistic terms such that $S^5 = \{s_0^5, \ldots, s_\kappa^5\}$, $\kappa = 4$ and $s_0^5 = Terrible = TE$, $s_1^5 = Poor = PO$, $s_2^5 = Average = AV$, $s_3^5 = Very\ Good = VG$ and $s_4^5 = Excellent = EX$.

Furthermore, the opinions in natural language posted by users will be modeled through another set with three terms that has the following linguistic labels: negative, neutral and positive. We choose this linguistic representation for the sentiment in the opinions given that it is usually used in the literature of sentiment analysis classification algorithms [97,102,103].

In order to model it, we will use a set of three linguistic terms such as $S^3 = \{s_0^3, s_1^3, s_\kappa^3\}$, $\kappa = 2$ and $s_0^3 = Negative = NEG$, $s_1^3 = Neutral = N$ and $s_2^3 = Positive = POS$. Given the need to integrate it with the 5-level sets used to model the questions in the questionnaires, we will use a linguistic hierarchy, such as those presented in Section 3.1.2, in the form $LH = \{l(1,3), l(2,5)\}$.

4.2. Step 2: eWOM problem specification

In this step we define the elements or inputs of our model and some of the outputs we pretend to obtain during the process. Firstly, let $H = \{h_1, \ldots, h_M\}$ a set of hotels, the goal to obtain the valuations of the users over these hotels during a period of time (π) from TripAdvisor website. The period to be studied should be selected by years given the seasonal behavior of the tourist sector. We could consider that the valuations that users give at TripAdvisor are a questionnaire as the mentioned in step 1, thus we can define a hotel evaluation questionnaire as one of the inputs of this step as can be seen in Fig. 3, which in the following will be denoted as $Q = \{Q_1, \ldots, Q_p\}$. For the business case presented in this paper, $p = 8$ and Q would be composed by Q_1 = Value, Q_2 = Location, Q_3 = Rooms, Q_4 = Cleanliness, Q_5 = Check in Front Desk, Q_6 = Service, Q_7 = Business Service and Q_8 = Overall Sentiment.

4.3. Step 3: Obtaining quality customers and their evaluations in social media

During the studied period we obtain all the users that have posted over one of the hotels in H. In the application on a real case the data could be obtained through web API like the one offered by TripAdvisor. However, we apply some quality criteria to ensure that we only take users meaningful for the model. These criteria, for example, could be: the posting of a minimum number of opinions during the period to be considered a member of the sample, the elimination of opinion-makers who post an unusual number of opinions, the inclusion only of opinion-users with an identified user, etc.

The information obtained after cleaning and filtering the dataset is those related with the evaluations done by the opinioners. We will get a new dataset in which its elements will be denoted as $y_{ip}(h) = \{y_{1p}(h), \ldots, y_{np}(h)\}$, $i = 1, \ldots, n$, $\forall p \in \{1, \ldots, P\}$, with $P = 7$. Each element is formed by the valuations of the opinioner y_i during the analyzed period over a hotel h and for each aspect Q_p with $p = 1, \ldots, P$ in the questionnaire defined in step 1.

Moreover, we have the natural language opinion $z_i(h)$ that a user gives to the hotel h. In order to convert this opinions into another metric of the questionnaire, an overall valuation of the user will be extracted through sentiment analysis techniques. This kind of analysis can be done by dictionary-based or machine learning based approaches [104–107]. We get the direction of the sentiment (negative, neutral or positive) implicit in the user's opinion as the aspect Q_8 and will be recorded for each user i that post about the hotel h as $y_{i8}(h)$.

4.4. Step 4: Obtaining the opinion value of each customer with 2-tuple RFH model

Once we filter the opinioners set, we get information about their posting frequency, the date of their last publication, and the usefulness that other users believe their publications have. With these information we get the relative importance of each opinioner expressed in 2-tuples through the model RFH 2-tuples: $y_i(R)$, $y_i(F)$, $y_i(H)$ and $y_i(RFH)$.

Based on RFM 2-tuple could be defined the RFH 2-tuple model. The change that differentiates them is the third dimension used, instead of taking into account the Monetary factor (M), it is replaced by the helpfulness (H) of the opinions posted by the users. The RFH model is used when recency (R) and frequency (F) are considered useful to adapt them to the value of the customer as an opinioner. However the monetary component (M) is not useful since there is no implicit monetary value in the opinion, and therefore we replace it with utility (H). From now on we will always refer to the RFH 2-tuple model, in which the utility can include the style in which the revision is written, its extension or, as mentioned in Section 2, the lack of misspellings, readability, etc. Obtaining the customer opinion value in this way has the advantage of giving a low value to fake users of recent creation.

The easiest way to weight the three dimensions of the model to obtain a unique opinion value is to apply the same weight to each one. However, it is more realistic that the weight of each factor involved in transactions is adapted according to the business case in which the model is applied. Therefore, to obtain the weights of the RFH 2-tuple model that we use in this work, we use the AHP presented in Section 3.3. We define the pairwise comparison matrix (A) that contains how the experts value some criteria with respect to others according to the Saaty scale presented in Table 3, doing it in this way we ensure that the matrix used in the business case is consistent according to the consistency ratio specified in Section 3.3. Finally, from the matrix A the vector of weights Ω will be obtained to weight each criterion to take into account.

$$\mathrm{A} = \begin{array}{c} \\ R \\ F \\ H \end{array} \begin{array}{c} \begin{array}{ccc} \;R\; & \;F\; & \;H\; \end{array} \\ \begin{pmatrix} 1 & a_{12} & a_{13} \\ a_{21} & 1 & a_{23} \\ a_{31} & a_{32} & 1 \end{pmatrix} \end{array}$$

$$\Omega = (\omega_R, \omega_F, \omega_H) \tag{11}$$

4.5. Step 5: MCDM model to obtain the evaluation of each hotel h based on the customers opinion values

At this point, we have a dataset in which each hotel h has a number of users evaluations of it, that is, not a unique questionnaire for each hotel but some of different users. Therefore, in this step we want to obtain a unique mean questionnaire value for each hotel. We do that through the well known Simple Additive Weighting (SAW) method [108,109].

In order to obtain our goal, the elements we use in the SAW method are: the RFH 2-tuple score obtained in the step 4 for each user and the evaluation for each aspect of a hotel h which is rated by an user. Thus, we use the RFH 2-tuple score as the weight or importance (ω_i) of each user in the global evaluation of each hotel, and finally we obtain for each aspect a global value such that $h_m(Q_p)$ is the result of applying Eq. (3) over each Q_p for the users who has evaluated the hotel h_m.

4.6. Step 6: MCDM model to obtain the final ranking of the hotels

Finally, a rank of the hotels is obtained. An AHP is used in order to obtain the position of each hotel h according to some criteria. In our model a set of hotels, $h = \{1, \ldots, H\}$, are the alternatives to be chosen and the criteria are the eight aspects asked on the questionnaires. So, a preferences matrix of these criteria is built based on the literature, experts opinion, etc. and the result will be a ranking of the hotels $R_t(h_m)$ in which the higher the position in the rank, the more recommended the hotel is. In this case, a pairwise comparison matrix as the shown in Section 4.4 would be defined but it would be of dimension 8×8 since it represented the relative importance of each question Q_p of the questionnaire with respect to others.

5. Methodology implementation

In this section, we present how our methodology would be implemented in a software. Its implementation is direct in any web application given that it is just necessary the input data. The presented business case has been developed in *R version 4.0.2*. We present a global view of the process through different pseudo codes that reflect the operations performed by our R code. The most expensive part of our model can be solved with aggregation functions (sum, count, grouping, etc.), which can be parallelized with paradigms like MapReduce. Of course, each "for" that appears in the following pseudocodes is parallelizable and distributable. Furthermore, the RFH model can also be parallelized and distributed. In fact, [110] proposes a parallel RFM customer value classification model based on the Spark framework.

In Algorithm 1 we present the order in which the algorithms are executed in our methodology.

Algorithm 1 Main Function()

```
Execute: Preprocessing algorithm [2]
Execute: RFH Data algorithm [3]
Execute: 2-tuples User Data algorithm [4]
Execute: Hotel Ranking algorithm [5]
```

In Algorithm 2, we present a pre-processing function that loads the data to filter variables and registers not necessary for the model. First, it selects the relevant variables (hotel and authors identifiers, and satisfaction variables). Then, it reduces the dataset to the specified period (π), omits missing values and discards users with less than 2 posts. Thus, after this pre-processing step we get a filtered dataset.

Algorithm 2 Preprocessing(Data, π)

```
Load Data
Select columns in: HotelID, HotelName, Author, Date, year,
  Content, Value, Location, Rooms, Cleanliness, CheckInFront-
  Desk, Service, BusinessService, NumHelpful
if year ∈ π then
    Filter data to study period
end if
for Each column p do
    if is.na(Data[,p]) then
        Omit missing registers
    end if
end for
Count registers grouped by author
if Frequency posting < 2 then
    Omit those authors
end if
return Filtered Data
```

Once the data has been filtered and the weights are obtained using the pairwise comparison matrix defined in the setting up phase, we can obtain the RFH dataset. It is important to bear in mind that all SQL operations executed in this algorithm can also be parallelized, improving scalability and model execution times [111]. In Algorithm 2 the filtered data and the weights obtained through the AHP methodology are loaded as inputs. Using the filtered data we obtain the RFH dimensions for each user. Firstly, Recency is obtained as the difference between the last posting day and the more recent day in the dataset. Given it is measure in negative. Secondly, Frequency and Helpfulness are obtained as the count of all their posts and the sum of all the votes their receives, respectively. Finally, the RFH Score is obtained through a weighted mean using the imputed weights. Thus, this script returns a dataset with the author identifier, the score for recency, frequency, helpfulness and a global score. Moreover, all this variables are obtained in its 2-tuple representation. It is done by applying Δ function defined in Eq. (1).

Algorithm 3 RFH.Data(H, Ω_{RFM})

Set reference date:
  reference.date = MAX(H[,date])
Select data from H dataset:
  RFM.User $\rightarrow$ **SELECT** R = MAX(Date), F = COUNT(Author.ID), M = SUM(Helpfulness)
    FROM H **GROUP BY** Author.ID
Mutate new variables:
  RFM.User $\rightarrow$ R = diff(reference.date, R, units='days'), ScoreR = neg(Δ(percrank(R))),
    ScoreF = Δ(percrank(F)), ScoreM = Δ(percrank(M)),
    ScoreRFM = $\Delta(\omega_R \times$ScoreR + $\omega_F \times$ScoreF + $\omega_M \times$ScoreM)
return $y_i(RFH)$ = [Author, ScoreR, ScoreF, ScoreM, ScoreRFM]

In Algorithm 4, filtered data is loaded and, for each question in the survey its 2-tuple transformation is obtained. Moreover, it is analyzed the sentiment in the content of the posts, obtaining its polarity. This polarity is transformed into linguistic label converting it into its sentiment direction, and the direction is transformed into its 2-tuples representation. Finally, the linguistic set for sentiment direction is transformed from $L(1,3)$ level to $L(2,5)$. Once we get all the variables in the same linguistic level, Algorithm 5 is executed. In this script, firstly, a global value for each of the 8 aspects of each hotel is obtained. It is done through a weighted mean. This global values are transformed into their 2-tuple representations. Secondly, we load the weights obtained through AHP for each question of the survey. Finally, the final 2-tuple ranking of hotels is obtained applying Eq. (3).

Algorithm 4 2T-User.Data(Filtered Data)

for p in 1 to 7 **do**
  Apply $\Delta(Q_p)$
end for
Analyze Content Sentiment ($z_i(h)$)
Rescale polarity from [-1,1] to [0,1]
Convert polarity to direction
Apply $\Delta(Direction)$
Transform $\Delta(Direction)$ from $L(1,3)$ to $L(2,5)$ using (5)
return y_{ip} = [Q_p] where $p = 1, \dots, 8$

Table 4
Business case dataset summary.

# items	# hotels	# opinioners	# opinions
138.973	975	115.607	138.895

Algorithm 5 Hotel Ranking(y,Ω_Q)

for id in HotelID **do**
  FILTER y by id
  SELECT: HotelID, HotelName, RFM Score, Q_1, Q_2, Q_3, Q_4, Q_5, Q_6, Q_7, Q_8
  Calculate opinion weight:
    ω_i = ScoreRFM/sum(ScoreRFM)
  for p in 1 to 8 **do**
    Calculate global value of each question for the hotel:
      $h_m(Q_p)$ = weighted.mean(Q_p,ω_i)
      Apply $\Delta(h_m(Q_p))$
  end for
end for
Calculate global 2-tuple ranking (3)
return $R_t(h_m)$

6. Use Case: TripAdvisor user's opinions utility

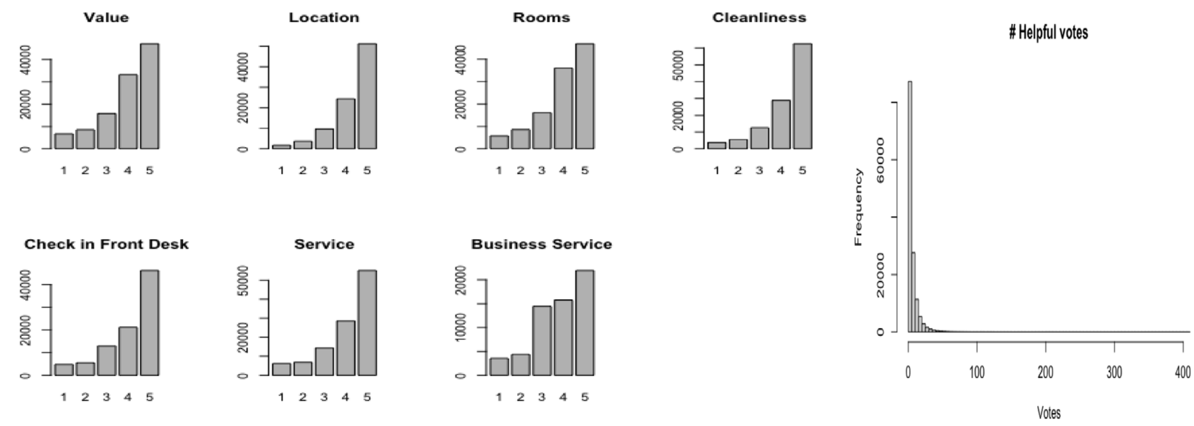
In this section we apply the methodology presented in Section 4 over the dataset used in [112,113]. The authors crawled hotel reviews from TripAdvisor for their paper, but they provide data for a longer period.[3] This dataset contains hotel information that have been collected from TripAdvisor[4] and it is formed by the following data: hotel's identifier, hotel's name, user's nickname, publication date, content of the opinion published, number of helpfulness votes given by other users to an opinion, and the punctuation given for different metrics of the hotel. Those metrics (Q_p) are represented in a Likert scale, defined in Section 2, with values between 1 and 5. In Table 4 a brief summary of the crawled data is shown. Moreover, TripAdvisor has its own API to extract information. As they describe, approved users of the Tripadvisor Content API[5] can access for accommodations, restaurants, and attractions to the following business details: name, address, coordinates, overall rating, ranking, subratings, awards, etc.

In order to delve into the data and to show the nature of the original data, an exploratory data analysis is presented in Tables 5 and 6. On the one hand, in Table 5 we present the main statistical measures for each of the quantitative variables. Moreover, we provide various plots (Fig. 5) to study the distribution of these variables. On the other hand, a summary of missing values of the categorical variables is showed in Table 6.

6.1. eWOM problem specification

The set H is formed by 975 hotels around the world (Novotel Amsterdam City, Hotel Milano, NH Mexico City...) that have received online reviews from opinioners in the aforementioned dataset. We define a period of study (π) equal to a year, and we limit the analysis to the opinions published between 2007 and 2008, with this we try to capture the seasonal component that characterizes the tourism sector.

[3] http://times.cs.uiuc.edu/~wang296/Data/.
[4] www.tripadvisor.com.
[5] http://developer-tripadvisor.com/content-api/description/.



(a) Bar plot of each questionnaire aspect

(b) Histogram of number of helpful votes

Fig. 5. Variable distribution plots.

Table 5
Main statistical measures for each of the quantitative variables.

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	# Helpful votes
Min.	1	1	1	1	1	1	1	1
1st Q	3	4	3	4	3	4	3	2
Median	4	5	4	5	5	4	4	4
Mean	3,95	4,32	3,97	4,25	4,09	4,08	3,80	6,57
3rd Q	5	5	5	5	5	5	5	8
Max.	5	5	5	5	5	5	5	407
sd	1,19	0,95	1,15	1,05	1,17	1,17	1,18	8,95
Missing values	27.796	48.430	25.727	25.794	48.337	27.839	78.889	0

Table 6
Missing values on categorical variables.

	Author	Content	Hotel name
Missing values	4.784	15	73.125

6.2. Obtaining quality customers and their evaluations in social media

In this step, we want to study the usefulness for the rest of the users of those who have published an opinion. Therefore, as we mentioned in Section 4.3, it is necessary to establish some quality filters over the opinioners. For this business case we have established the three quality criteria. Firstly, we will discard all anonymous users. Secondly, only those users who have posted at least 2 times throughout the analyze period will be taken into account. Finally, users who have not evaluated all aspects of the hotel they have reviewed will be discarded.

After this pre-processing step we obtain a dataset with the opinions chronologically ordered for each hotel such as the sample of example shown in Table 7.

From now on we will denote these hotel metrics in the way they were defined in Section 4.2. In Table 8 we show the previous data converted to its 2-tuple representation. In summary, Table 8 shows for each hotel, h, the valuations for each of those metrics, Q_p, so $y_{ip}(h)$ is the valuation of user i to the question p over the hotel h. In addition, the content of the opinion given by the user i for hotel h is presented $(z_i(h))$. The metrics are represented in 2-tuple form applying Eq. (1).

In order to obtain the sentiment that the user transmits with its post, we apply a sentiment analysis model which result will be include as a new metric that represents a sentimental global evaluation of the hotel. Therefore, it would be added as Q_8 = "Overall Sentiment" to the metrics afore defined $(Q_1, Q_2, Q_3, Q_4, Q_5, Q_6, Q_7)$. We calculate the polarity of each post using dictionary-based sentiment analysis. These dictionaries contain words with a polarity predefined by experts, they can also be combined with linguistic rules. The polarity is obtained through the presence of those words into the analyzed post, and it is measured between -1 and 1. In this business case we use the Harvard-IV dictionary[6] developed by the Harvard University, we have chosen this dictionary because it is a general-purpose dictionary, making our model applicable to problems of another domain beyond the tourist presented here. After obtaining the polarity of the post, we convert it into their corresponding sentiment direction, that is in one of the following labels: negative, neutral or positive. Thus, the dataset would be as shown in Table 9 after obtaining the polarity and the sentiment direction that the user's post provides.

We have rescaled the polarity into the interval $[0, 1]$ and we represent its sentiment direction, their linguistic labels, converting it into its 2-tuple representation (Table 10) applying the delta function previously defined in Section 3.1.1.

However, we find in the data a linguistic hierarchy such as those presented in Section 3.1.2. While the direction of sentiment has 3 linguistic terms, the rest of the metrics have 5 linguistic terms. Thus, we have a linguistic hierarchy with two levels as shown in Fig. 6. For now on we will denote them as $L(1, 3)$ and $L(2, 5)$, and we will choose the second level, which has the highest granularity, in order to uniform the information.

Applying the transformation function defined in Eq. (5) we transfer the information from the first level to the second, such that $L(1, 3) \longrightarrow L(2, 5)$. In Table 11 we show the new tuples obtained for the Q_8 after applying the transformation function.

Transformed Q_8 together with the rest of the questions Q_p in $L(2, 5)$ evaluated on the dataset of hotels H for the study period π will form the initial dataset for the development of our model.

6 http://www.wjh.harvard.edu/~inquirer/.

Table 7
Sample of hotels with the reviews and form answers given by users and helpfulness votes received for each user.

Hotel name	Author	Date	Content	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Helpfulness votes
BEST WESTERN PLUS Executive Inn	OffTheBeatenPath	2008-02-13	I was happy with the fact that...	5	3	3	4	4	4	5	1
BEST WESTERN PLUS Executive Inn	PieroG	2008-04-07	Not bad, but not great.	4	5	5	5	5	5	3	1
Holiday Inn & Suites Phoenix Airport North	hdblue	2007-10-01	Absolutely Gross!!!	2	2	2	2	2	2	2	2
Embassy Suites Hotel Phoenix-North	Calartist	2007-10-11	I like to stay at Embassy Suites, and...	2	1	1	1	3	4	3	1
Hilton Phoenix Airport	grcas	2007-11-13	Very good value for what we paid...	4	3	4	3	4	4	4	1
Hilton Phoenix Airport	btowndude	2008-04-16	My stay was this past weekend...	4	4	4	5	5	5	4	1

Table 8
Sample of hotels with the reviews and form answers given by users and helpfulness votes received for each user represented in 2-tuples.

Hotel name	Author	Date	Content	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Helpfulness votes
BEST WESTERN PLUS Executive Inn	OffTheBeatenPath	2008-02-13	I was happy with the fact that...	$(s_4^5, 0)$	$(s_2^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_4^5, 0)$	1
BEST WESTERN PLUS Executive Inn	PieroG	2008-04-07	Not bad, but not great.	$(s_3^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_2^5, 0)$	1
Holiday Inn & Suites Phoenix Airport North	hdblue	2007-10-01	Absolutely Gross!!!	$(s_1^5, 0)$	$(s_1^5, 0)$	$(s_1^5, 0)$	$(s_1^5, 0)$	$(s_1^5, 0)$	$(s_1^5, 0)$	$(s_1^5, 0)$	2
Embassy Suites Hotel Phoenix-North	Calartist	2007-10-11	I like to stay at Embassy Suites, and...	$(s_1^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0)$	1
Hilton Phoenix Airport	grcas	2007-11-13	Very good value for what we paid...	$(s_3^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	1
Hilton Phoenix Airport	btowndude	2008-04-16	My stay was this past weekend...	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_3^5, 0)$	1

Table 9
Dataset with sentiment information.

Hotel name	Author	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Sentiment direction	Polarity
BEST WESTERN PLUS Executive Inn	OffTheBeatenPath	$(s_3^5, 0)$	$(s_1^5, 0)$	$(s_1^5, 0)$	$(s_2^5, 0)$	$(s_2^5, 0)$	$(s_2^5, 0)$	$(s_4^5, 0)$	positive	0.5079
BEST WESTERN PLUS Executive Inn	PieroG	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_1^5, 0)$	neutral	0.5000
Holiday Inn & Suites Phoenix Airport North	hdblue	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	neutral	0.5000
Embassy Suites Hotel Phoenix-North	Calartist	$(s_1^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0)$	negative	0.4744
Hilton Phoenix Airport	grcas	$(s_3^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	positive	0.5146
Hilton Phoenix Airport	btowndude	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_3^5, 0)$	positive	0.5469

Table 10
2-tuple questionnaire representation.

Hotel name	Author	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
BEST WESTERN PLUS Executive Inn	OffTheBeatenPath	$(s_3^5, 0)$	$(s_1^5, 0)$	$(s_1^5, 0)$	$(s_2^5, 0)$	$(s_2^5, 0)$	$(s_2^5, 0)$	$(s_4^5, 0)$	$(s_1^3, 0.02)$
BEST WESTERN PLUS Executive Inn	PieroG	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_1^5, 0)$	$(s_1^3, 0)$
Holiday Inn & Suites Phoenix Airport North	hdblue	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_1^3, 0)$
Embassy Suites Hotel Phoenix-North	Calartist	$(s_1^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0)$	$(s_1^3, -0.05)$
Hilton Phoenix Airport	grcas	$(s_3^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_1^3, 0.03)$
Hilton Phoenix Airport	btowndude	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_3^5, 0)$	$(s_1^3, 0.09)$

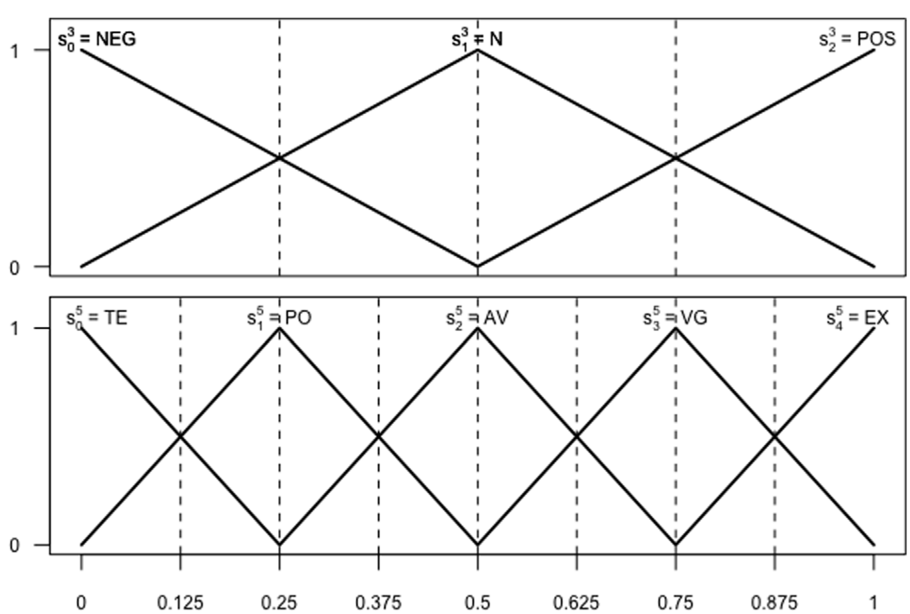


Fig. 6. Linguistic Hierarchy of our data.

Table 11
Q_8 transformation.

Hotel name	Author	Q_8	Transformed Q_8
BEST WESTERN PLUS Executive Inn	OffTheBeatenPath	$(s_1^3, 0.02)$	$(s_2^5, 0.04)$
BEST WESTERN PLUS Executive Inn	PieroG	$(s_1^3, 0)$	$(s_2^5, 0)$
Holiday Inn & Suites Phoenix Airport North	hdblue	$(s_1^3, 0)$	$(s_2^5, 0)$
Embassy Suites Hotel Phoenix-North	Calartist	$(s_1^3, -0.05)$	$(s_2^5, -0.1)$
Hilton Phoenix Airport	grcas	$(s_1^3, 0.03)$	$(s_2^5, 0.06)$
Hilton Phoenix Airport	btowndude	$(s_1^3, 0.09)$	$(s_2^5, 0.18)$

Table 12
RFH users table.

Author	Recency	Frequency	Helpfulness
OffTheBeatenPath	626	3	6
PieroG	268	2	3
hdblue	642	4	8
Calartist	447	2	4
grcas	715	2	2
btowndude	477	3	7

Table 13
RFH 2-tuple results.

Author	Recency	Frequency	Helpfulness	RFH score
OffTheBeatenPath	$(S_1^5, -0.08)$	$(S_3^5, 0.3)$	$(S_2^5, -0.07)$	$(S_2^5, -0.19)$
PieroG	$(S_3^5, 0.45)$	$(S_0^5, 0)$	$(S_0^5, 0.39)$	$(S_1^5, 0.14)$
hdblue	$(S_1^5, -0.21)$	$(S_4^5, -0.16)$	$(S_3^5, -0.39)$	$(S_2^5, 0.27)$
Calartist	$(S_3^5, -0.49)$	$(S_0^5, 0)$	$(S_1^5, -0.04)$	$(S_1^5, 0.26)$
grcas	$(S_0^5, 0.16)$	$(S_0^5, 0)$	$(S_0^5, 0)$	$(S_0^5, 0.04)$
btowndude	$(S_2^5, 0.3)$	$(S_3^5, 0.3)$	$(S_2^5, 0.31)$	$(S_2^5, 0.41)$

6.3. Obtaining the opinion value of each customer with 2-tuple RFH model

With the aim of determining the opinioner value we are going to use the RFH methodology with 2-tuples explained in Section 3.2. In this example, the three dimensions will be the number of days since the last time the user posted a review on TripAdvisor as the Recency (R). Secondly, the number of opinions posted on TripAdvisor by an user during the period analyzed will refer to the Frequency (F). Finally, the number of helpfulness votes a user receives on his posts from the rest of users during the period analyzed is the Helpfulness (H) dimension.

Once RFH variables are defined and before obtaining the user classification through the RFH 2-tuples methodology, we are going to determine the relative weights of each dimension of the client. Consequently, we use the AHP model, exposed in Section 3.3, to which we assign our preferences on what we value the most from the users in this use case. The pairwise comparison matrix is defined as:

$$A_1 = \begin{array}{c} \\ \textit{Recency} \\ \textit{Frequency} \\ \textit{Helpfulness} \end{array} \begin{array}{c} \begin{array}{ccc} \textit{Recency} & \textit{Frequency} & \textit{Helpfulness} \end{array} \\ \begin{pmatrix} 1 & 3 & 1/3 \\ 1/3 & 1 & 1/5 \\ 3 & 5 & 1 \end{pmatrix} \end{array}$$

This means that for this business case the helpfulness of user comments will have a moderate importance respect the recency they post, and a strong importance in relation with the frequency they publish an opinion. In addition, the frequency is of moderate importance with respect to the recency. In other business cases, with a total different domain than the studied in this paper, could be possible to obtain a vector of weights adapted to the problem just by setting another pairwise comparison matrix. Thus, applying the AHP with these preferences we obtain the weighting vector $\Omega = (\omega_R = 0.26, \omega_F = 0.10, \omega_H = 0.64)$.

To apply the RFH 2-tuple methodology we define a dataset that collects for each user its data about recency, frequency and helpfulness. The structure of this dataset would be as the presented below:

Next, the opinioner ranking is obtained for each variable and with them the percentage of the ranking that each user occupies is obtained using Eq. (7).

Finally, we apply the RFH 2-tuple methodology on the data set shown in Table 12. The idea is to convert the percentage of the ranking that each user occupies into a tuple as the defined in Section 3.1.1. It is done applying the delta function defined in Eq. (1). Thus, the obtained results are as:

In order to obtain the RFH score it is necessary to translate the tuples of each variable into their numerical representation using the Δ^{-1} function defined in Section 3.1.1. After that we calculate the weighted mean applying the weights we have determined in the previous subsection. Once we have obtained the RFH score, we just need to apply again the delta function to convert it into its linguistic representation as is shown in Table 12. According to the sample of opinioners showed in Table 13, *btowndude* is the best client. He/She is considered over the average value shifted to the right towards very good label, $(AV, +0.41)$. It is the result of being also valued over the average in how recently published and how helpfulness are his/her posts and obtaining more than a very good valoration, $(VG, +0.30)$, in how frequent he/she publishes. On the other hand, *grcas* is considered a terrible opinioner, $(TE, +0.04)$. This consideration is a consequence of being evaluated of completely terrible according to the frequency of his/her posts and the utility they provide, and its a little bit over the terrible valoration, $(TE, +0.16)$, when the users evaluate how recently post.

6.4. MCDM model to obtain the evaluations of each hotel h based on the customer opinion values

In Table 14, all the metrics evaluated for the opinioners are shown in the same linguistic level for each hotel. Additionally, the opinioner value of each user, $y_i(RFH)$, is included in order to show the relative importance of each opinioner posting about a given hotel.

In this subsection we obtain a mean value of each metric for each hotel in our dataset applying Eq. (3). Then, the evaluations that the users have given to each of the metrics are weighted using the value of the opinioner, obtained through the RFH 2-tuples model, as the weight of each user over the final value in the metrics of the hotel.

In Table 15 the average value of each metric for each hotel $(h_m(Q_p))$ taking into account just the users that had posted about them are shown.

6.5. MCDM model to obtain the final ranking of the hotels

In this final step, we rank the hotels using an AHP in which the criteria are the different metrics evaluated by the users. The

Table 14
2-tuple representation of the scores that customers have given in the hotel quality questionnaire and their customer opinion value.

Hotel name	Author	RFH score	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
BEST WESTERN PLUS Executive Inn	OffTheBeatenPath	$(S_2^5, -0.19)$	$(s_3^5, 0)$	$(s_1^5, 0)$	$(s_1^5, 0)$	$(s_2^5, 0)$	$(s_2^5, 0)$	$(s_2^5, 0)$	$(s_4^5, 0)$	$(s_2^5, 0.04)$
BEST WESTERN PLUS Executive Inn	PieroG	$(S_1^5, 0.14)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_1^5, 0)$	$(s_2^5, 0)$
Holiday Inn & Suites Phoenix Airport North	hdblue	$(S_2^5, 0.27)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_2^5, 0)$
Embassy Suites Hotel Phoenix-North	Calartist	$(S_1^5, 0.26)$	$(s_1^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_0^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0)$	$(s_2^5, -0.1)$
Hilton Phoenix Airport	grcas	$(S_0^5, 0.04)$	$(s_2^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0.06)$
Hilton Phoenix Airport	btowndude	$(S_2^5, 0.41)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_3^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_3^5, 0)$	$(s_2^5, 0.18)$

Table 15
Average values of hotel metrics.

Hotel name	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
ARCOTEL Velvet	$(s_3^5, 0.29)$	$(s_4^5, 0)$	$(s_3^5, 0.29)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_3^5, 0.29)$	$(s_4^5, -0.29)$	$(s_1^5, 0.34)$
Acevi Villarroel	$(s_3^5, 0.09)$	$(s_4^5, -0.46)$	$(s_4^5, -0.46)$	$(s_4^5, 0)$	$(s_3^5, 0.09)$	$(s_3^5, 0.09)$	$(s_4^5, -0.46)$	$(s_1^5, 0.17)$
Adolphus Hotel	$(s_3^5, -0.07)$	$(s_4^5, -0.33)$	$(s_3^5, 0.13)$	$(s_3^5, 0.14)$	$(s_3^5, 0.14)$	$(s_3^5, -0.13)$	$(s_4^5, -0.33)$	$(s_1^5, 0.34)$
Al Ponte Antico Hotel	$(s_4^5, -0.25)$	$(s_4^5, 0)$	$(s_4^5, -0.46)$	$(s_4^5, -0.25)$	$(s_4^5, 0)$	$(s_4^5, -0.21)$	$(s_3^5, 0.33)$	$(s_1^5, 0.47)$
BEST WESTERN PLUS Executive Inn	$(s_3^5, -0.39)$	$(s_2^5, -0.23)$	$(s_2^5, -0.23)$	$(s_2^5, 0.39)$	$(s_2^5, 0.39)$	$(s_2^5, 0.39)$	$(s_3^5, -0.16)$	$(s_1^5, 0.02)$
Crowne Plaza Beverly Hills	$(s_2^5, 0.49)$	$(s_3^5, 0.21)$	$(s_3^5, 0.44)$	$(s_3^5, 0.44)$	$(s_3^5, -0.01)$	$(s_3^5, -0.28)$	$(s_3^5, 0.21)$	$(s_1^5, 0.37)$

Table 16
Top 6 hotels by global valuation.

Hotel name	Global value	Global numeric value
Eurostars Cristal Palace	$(s_3^5, 0.31)$	4.31
GBB Hotel Front Maritim	$(s_3^5, 0.29)$	4.29
Regencia Colon Hotel	$(s_3^5, 0.28)$	4.28
Hotel Le Bristol	$(s_3^5, 0.26)$	4.26
Silver Cloud Hotel - Broadway	$(s_3^5, 0.25)$	4.26
Novotel Berlin Am Tiergarten	$(s_3^5, 0.24)$	4.24

Table 17
Obtained ranking vs TripAdvisor ranking.

Hotel name	Our ranking	Our numerical ranking	TripAdvisor ranking (2008)
Eurostars Cristal Palace	$(s_3^5, 0.31)$	4.31	4.0
GBB Hotel Front Maritim	$(s_3^5, 0.29)$	4.29	4.0
Regencia Colon Hotel	$(s_3^5, 0.28)$	4.28	4.0
Hotel Le Bristol	$(s_3^5, 0.26)$	4.26	4.5
Silver Cloud Hotel - Broadway	$(s_3^5, 0.25)$	4.25	4.5
Novotel Berlin Am Tiergarten	$(s_3^5, 0.24)$	4.24	4.5

value of those metrics have the helpfulness of each user implicit. This is so since we have calculated it using a weighted average in which the weights are the user value obtained through the RFH 2-tuples in Section 6.3. We have built up our pairwise comparison matrix based on the preferences that other authors have selected in their works [44,114]. Therefore, our preferences matrix is as shown below:

$$A_2 = \begin{array}{c} \\ Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \\ Q_5 \\ Q_6 \\ Q_7 \\ Q_8 \end{array} \begin{array}{c} \begin{array}{cccccccc} Q_1 & Q_2 & Q_3 & Q_4 & Q_5 & Q_6 & Q_7 & Q_8 \end{array} \\ \left(\begin{array}{cccccccc} 1 & 1 & 2 & 1/2 & 2 & 1 & 1 & 1/3 \\ & 1 & 1 & 1/3 & 1 & 1/2 & 1/2 & 1/4 \\ & & 1 & 1/3 & 1 & 1/2 & 1/2 & 1/4 \\ & & & 1 & 3 & 2 & 2 & 1/2 \\ & & & & 1 & 1/3 & 1/3 & 1/4 \\ & & & & & 1 & 1 & 1/3 \\ & & & & & & 1 & 1/3 \\ & & & & & & & 1 \end{array} \right) \end{array}$$

With those preferences the AHP provides the following weights for the criteria: ($\omega_{Q_1} = 0.1013$, $\omega_{Q_2} = 0.0657$, $\omega_{Q_3} = 0.0602$, $\omega_{Q_4} = 0.1912$, $\omega_{Q_5} = 0.0544$, $\omega_{Q_6} = 0.1162$, $\omega_{Q_7} = 0.1162$, $\omega_{Q_8} = 0.2949$). Thus, we obtain a global value for each hotel that allows us to rank them with a single criterion as is shown in Table 16. In addition to being able to do it by any of the criteria taken into account to obtain the global valuation.

7. Discussion

The application of our model has provided us with a ranking of the hotels included in the study panel. This has been possible thanks to the contributions of customers through different sources of information, either by questionnaires or by opinions published in natural language. That is why it is not a ranking based on the characteristics of the establishment (price, location, services, etc.) directly, but evaluated through the personal experience of its clients. In addition, as it has been exposed, in our methodology different data sources can be incorporated from which information is not lost when represented linguistically. Moreover, they can be integrated to form a single set of data although they are expressed in different scales or linguistic labels. In fact, in this business case, only data extracted from TripAdvisor has been used. As has been observed, it has a two-level linguistic hierarchy. However, another strength of our methodology is that information from different websites (TripAdvisor, Booking, Expedia ...) could be combined with different forms of measure customer satisfaction and would adapt easily with the inclusion of more levels in the hierarchy. Thus, for example, if we had included Booking data in our application, we should have added a third language level with 9 terms, since the Booking questionnaires measure customer satisfaction with the following labels: Poor, Disappointing, Passable, Pleasant, Good, Very Good, Fabulous, Superb, Exceptional.

Being the opinion makers the point of origin of the information used to obtain these rankings, our model facilitates the interaction of users. Additionally, it will increase their predisposition to share their experiences if they are made aware that hotel recommendations are based on their publications, and that with the comments they provide they can access personalized and interesting information.

In Table 17 we show the top 6 hotels obtained with our methodology and the value assigned in the ranking compared with the values of the hotels in TripAdvisor in 2008. As can we see, both are in the highest semantic labels of TripAdvisor's

Table 18
Questionnaire authors response for a hotel.

Hotel name	Author	RFH score	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7
Novotel Berlin Am Tiergarten	Peter123452	2.27184	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$
	prash123	2.066975	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$	$(s_4^5, 0)$

Table 19
Customer opinion value of some authors.

Author	Recency	Frequency	Helpfulness	RFH score
Peter123452	$(s_1^5, -0.17)$	$(s_0^5, 0)$	$(s_3^5, 0.23)$	$(s_2^5, 0.27)$
prash123	$(s_1^5, -0.1)$	$(s_0^5, 0)$	$(s_3^5, -0.12)$	$(s_2^5, 0.07)$

scale such that 4 and 4.5 corresponds to *Very Good* and *Excellent* valuations, respectively. However, we can see that our result is more informative than the official scale since it assigns values in intervals of 0.5. This way of measuring scale makes it difficult to compare if *Hotel Le Bristol* is better or worse than *Novotel Berlin Am Tiergarten*, for example. Thus, we see that our methodology assigns to all hotels ratings of "Very good" shifted to a greater or lesser extent to the right, obtaining a more informative ranking of hotels. It allows the user to select the best hotel given the contributions of the opinion-makers and on the other hand increases the competitiveness among the hotels by ceasing to treat all hotels that share value in the TripAdvisor ranking as equals.

This is the result of weighing users according to customer opinion value, which means that although two users give two very good ratings to a hotel, they may not have the same impact depending on the profile of the opinioner. As can be seen in Table 18, both authors response with the highest mark to all questions about *Novotel Berlin Am Tiergarten*. However, it is also shown that they do not have the same value as opinioner. Looking at Table 19 we can see that both authors have an absolutely terrible evaluation in the frequency with which they publish and that they have a quite similar evaluation in the recency of their opinions. Nevertheless one of the authors (*Peter123452*) has some opinions that other users perceive as very useful, and being valued in this dimension as (VG, +0.23). On the other hand, the other user, even being valued as a useful opinioner, (VG, −0.12), moves significantly away from the first, which makes a difference in the importance of their opinions to value a certain hotel.

8. Conclusions and future work

Throughout this work we have raised the idea that a ranking of a set of hotels can be obtained based on customer opinions, either through questionnaires or reviews. The rankings obtained will be more reliable since they will be based on the personal experiences of the reviewers, and not so much on the characteristics of the hotels. In order to achieve it, we have developed a new methodology that integrates multiple techniques. It starts from obtaining a value of customer opinion using the recency of their opinions, their frequency of publication and the helpfulness assigned by other users, introducing the RFH 2-tuple model. Furthermore, we have incorporated multi-granular fuzzy linguistic modeling into the RFH model. It has be done in order to have the different data sources (questionnaires and opinions in natural language) without loss of information. This customer value serves as a weight when it comes to having an average rating of each hotel based on the opinion-makers who have shared their experience in it. Finally, once we have the set of hotels with the average rating that their clients provide, we can obtain a ranking by applying AHP.

We have verified the functionality of this methodology by presenting a business case. In fact, our methodology is easy to implement in any web application given it just needs the input data and the set up specifications. Thanks to our methodology we have been able to obtain a hotel ranking based on customer opinion value. This ranking has been obtained using data from TripAdvisor. This data includes opinions of the users and their answers to the hotel satisfaction questionnaires. Although there are other works in the literature that provide similar results, the novelty of our methodology is the integration of the fuzzy linguistic modeling, the weighting in different phases through the AHP and the obtaining of the opinion value to be used as a weighting through a 2-tuple RFH model. This methodology is very useful both for clients who will be better informed to make better decisions in their purchasing process, and for hotels that will increase their competitiveness, resulting in an increase in reservations.

Despite the advantages that we consider incorporating with our methodology, in the development of it we have identified certain aspects that can be improved. For example, the algorithm to be used to extract the polarity of the opinions published in natural language or the way to measure the usefulness of the publications of each opinioner [115,116]. Regarding the first, we consider that other more accurate techniques could be selected to further improve the detection of the opinion of the opinioner, while the usefulness of the posts could be measured using other scales, which we have not had into account for this job. Both questions will be addressed in future works to incorporate them into the methodology. The methodology is applicable in other domains than tourism sector. It can be used using data from any review website. For example, for valuation of video games (Steam), movies (FilmAffinity), products (Amazon), online learning (Coursera), etc.

CRediT authorship contribution statement

Itzcóatl Bueno: Writing - original draft, Investigation, Conceptualization, Formal analysis. **Ramón A. Carrasco:** Visualization, Validation. **Carlos Porcel:** Methodology, Software, Supervision. **Gang Kou:** Methodology, Writing - review & editing, Project administration, Funding acquisition. **Enrique Herrera-Viedma:** Methodology, Writing - review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper was partially supported by the Spanish State Research Agency through the projects PID2019-103880RB-I00/AEI/10.13039/501100011033 and TIN2016-75850-R, and also by grants from the National Natural Science Foundation of China (#71725001 and #71910107002), State key R & D Program of China (#2020YFC0832702) and Major project of the National Social Science Foundation of China (#19ZDA092).

References

[1] S.K. Lenka, R. Barik, Has expansion of mobile phone and internet use spurred financial inclusion in the SAARC countries? Financial Innov. 4 (1) (2018) 5.

[2] H.-C. Kuo, C. Nakhata, The impact of electronic word-of-mouth on customer satisfaction, J. Mark. Theory Pract. 27 (3) (2019) 331–348.

[3] E. Ismagilova, E.L. Slade, N.P. Rana, Y.K. Dwivedi, The effect of electronic word of mouth communications on intention to buy: a meta-analysis, Inf. Syst. Front. (2019) 1–24.

[4] E. Ismagilova, E. Slade, N. Rana, Y. Dwivedi, The effect of characteristics of source credibility on consumer behaviour: A meta-analysis, J. Retail. Consum. Serv. 53 (C) (2020) 101742.

[5] R. Alberto Carrasco, M. Francisca Blasco, J. Garcia-Madariaga, E. Herrera-Viedma, A fuzzy linguistic RFM model applied to campaign management, Int. J. Interact. Multimedia Artif. Intell. 5 (4, (Special Issue on Artificial Intelligence Applications)) (2019) 21–27.

[6] T. Hennig-Thurau, K.P. Gwinner, G. Walsh, D.D. Gremler, Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? J. Interact. Mark. 18 (1) (2004) 38–52.

[7] S.-H. Hsu, Developing an index for online customer satisfaction: Adaptation of American customer satisfaction index, Expert Syst. Appl. 34 (4) (2008) 3033–3042.

[8] R.A. Carrasco, F. Muñoz-Leiva, J. Sánchez-Fernández, F.J. Liébana-Cabanillas, A model for the integration of e-financial services questionnaires with SERVQUAL scales under fuzzy linguistic modeling, Expert Syst. Appl. 39 (14) (2012) 11535–11547.

[9] G. Albaum, The Likert scale revisited, Mark. Res. Soc. J. 39 (2) (1997) 1–21.

[10] Q. Li, A novel Likert scale based on fuzzy sets theory, Expert Syst. Appl. 40 (5) (2013) 1609–1618.

[11] R. Likert, A technique for the measurement of attitudes, Archives of psychology (1932).

[12] J. McIver, E.G. Carmines, Unidimensional Scaling, No. 24, Sage, 1981.

[13] W.-J. Deng, W. Pei, Fuzzy neural based importance-performance analysis for determining critical service attributes, Expert Syst. Appl. 36 (2) (2009) 3774–3784.

[14] H.-K. Chiou, G.-H. Tzeng, D.-C. Cheng, Evaluating sustainable fishing development strategies using fuzzy MCDM approach, Omega 33 (3) (2005) 223–234.

[15] R.A. Carrasco, J. Sánchez-Fernández, F. Muñoz-Leiva, M.F. Blasco, E. Herrera-Viedma, Evaluation of the hotels e-services quality under the user's experience, Soft Comput. 21 (4) (2017) 995–1011.

[16] X. Xu, X. Wang, Y. Li, M. Haghighi, Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors, Int. J. Inf. Manage. 37 (6) (2017) 673–683.

[17] B. Liu, M. Hu, J. Cheng, Opinion observer: analyzing and comparing opinions on the web, in: Proceedings of the 14th International Conference on World Wide Web, 2005, pp. 342–351.

[18] M.C. Yeung, B. Ramasamy, J. Chen, S. Paliwoda, Customer satisfaction and consumer expenditure in selected European countries, Int. J. Res. Market. 30 (4) (2013) 406–416.

[19] S.W. Litvin, R.E. Goldsmith, B. Pan, Electronic word-of-mouth in hospitality and tourism management, Tour. Manag. 29 (3) (2008) 458–468.

[20] A. Zareie, A. Sheikhahmadi, M. Jalili, Identification of influential users in social networks based on users' interest, Inform. Sci. 493 (2019) 217–231.

[21] A. Zareie, A. Sheikhahmadi, M. Jalili, Identification of influential users in social network using gray wolf optimization algorithm, Expert Syst. Appl. 142 (2020) 112971.

[22] J.M. Hernández, A.P. Kirilenko, S. Stepchenkova, Network approach to tourist segmentation via user generated content, Ann. Tour. Res. 73 (2018) 35–47.

[23] Y. Narangajavana Kaosiri, L.J. Callarisa Fiol, M.A. Moliner Tena, R.M. Rodriguez Artola, J. Sanchez Garcia, User-generated content sources in social media: A new approach to explore tourist satisfaction, J. Travel Res. 58 (2) (2019) 253–265.

[24] L. Mendes-Filho, A.M. Mills, F.B. Tan, S. Milne, Empowering the traveler: An examination of the impact of user-generated content on travel planning, J. Travel Tour. Mark. 35 (4) (2018) 425–436.

[25] T.-Y. Chou, C.-L. Hsu, M.-C. Chen, A fuzzy multi-criteria decision model for international tourist hotels location selection, Int. J. Hosp. Manag. 27 (2) (2008) 293–301.

[26] L.-Z. Lin, W.-C. Chen, T.-J. Chang, Using FQFD to analyze island accommodation management in fuzzy linguistic preferences, Expert Syst. Appl. 38 (6) (2011) 7738–7745.

[27] R.A. Carrasco, P. Villar, A new model for linguistic summarization of heterogeneous data: an application to tourism web data sources, Soft Comput. 16 (1) (2012) 135–151.

[28] A. Dursun, M. Caber, Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis, Tour. Manag. Perspect. 18 (2016) 153–160.

[29] A. Christy, A. Umamakeswari, L. Priyatharsini, A. Neyaa, RFM Ranking – an effective approach to customer segmentation, J. King Saud Univ. - Comput. Inf. Sci. (2018) in press, http://dx.doi.org/10.1016/j.jksuci.2018.09.004.

[30] P.-J. Lee, Y.-H. Hu, K.-T. Lu, Assessing the helpfulness of online hotel reviews: A classification-based approach, Telemat. Inform. 35 (2) (2018) 436–445.

[31] T.L. Ngo-Ye, A.P. Sinha, The influence of reviewer engagement characteristics on online review helpfulness: A text regression model, Decis. Support Syst. 61 (2014) 47–58.

[32] A. Aakash, A. Jaiswal, Segmentation and ranking of online reviewer community: The role of reviewers' frequency, helpfulness, and recency, Int. J. E-Adoption (IJEA) 12 (1) (2020) 63–83.

[33] U. Ahmad, Application of multiple criteria decision making techniques in tourism and hospitality industry: A systematic review, Transform. Bus. Econ. 15 (1) (2016) 37.

[34] A. Jusoh, A. Mardani, R. Omar, D. Štreimikienė, Z. Khalifah, A. Sharifara, Application of MCDM approach to evaluate the critical success factors of total quality management in the hospitality industry, J. Bus. Econ. Manag. 19 (2) (2018) 399–416.

[35] G. Popovic, D. Stanujkic, M. Brzakovic, D. Karabasevic, A multiple-criteria decision-making model for the selection of a hotel location, Land Use Policy 84 (2019) 49–58.

[36] S. Atmaca, H.A. Karadaş, Decision making on financial investment in turkey by using ardl long-term coefficients and ahp, Financial Innov. 6 (1) (2020) 1–22.

[37] L. Xu, L. Wu, X. Li, F. Shen, Introduction to the special issue on analytical and decision-making technique innovation in financial market, Financial Innov. (2020).

[38] W. Li, S. Yu, H. Pei, C. Zhao, B. Tian, A hybrid approach based on fuzzy AHP and 2-tuple fuzzy linguistic method for evaluation in-flight service quality, J. Air Transp. Manag. 60 (2017) 49–64.

[39] S.H. Zolfani, M. Pourhossein, M. Yazdani, E.K. Zavadskas, Evaluating construction projects of hotels based on environmental sustainability with MCDM framework, Alex. Eng. J. 57 (1) (2018) 357–365.

[40] A. Ishizaka, A. Quintano, A. Labib, A. Apostolakis, Do five-star hotel managers know their customers' priorities? An AHP-prioritised scorecard study, EuroMed J. Bus. 14 (2) (2019) 137–167.

[41] M.R. Galankashi, F.M. Rafiei, M. Ghezelbash, Portfolio selection: a fuzzy-anp approach, Financial Innov. 6 (1) (2020) 1–34.

[42] R.-x. Nie, Z.-p. Tian, J.-q. Wang, K.S. Chin, Hotel selection driven by online textual reviews: Applying a semantic partitioned sentiment dictionary and evidence theory, Int. J. Hosp. Manag. 88 (2020) 102495.

[43] C. Tian, J. Peng, W. Zhang, S. Zhang, J. Wang, Tourism environmental impact assessment based on improved AHP and picture fuzzy PROMETHEE II methods, Technol. Econ. Dev. Econ. 26 (2) (2020) 355–378.

[44] M. Zaman, L. Botti, T.V. Thanh, Weight of criteria in hotel selection: An empirical illustration based on TripAdvisor criteria, Eur. J. Tour. Res. 13 (1) (2016) 132–138.

[45] R.A. Carrasco, M.F. Blasco, E. Herrera-Viedma, A 2-tuple fuzzy linguistic RFM model and its implementation, Procedia Comput. Sci. 55 (2015) 1340–1347.

[46] A. Graf, P. Maas, Customer value from a customer perspective: a comprehensive review, J. Betriebswirtschaft 58 (1) (2008) 1–20.

[47] V.A. Zeithaml, Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence, J. Mark. 52 (3) (1988) 2–22.

[48] B. Gale, B.T. Gale, R.C. Wood, Managing Customer Value: Creating Quality and Service That Customers Can See, Simon and Schuster, 1994.

[49] M.B. Holbrook, The nature of customer value: An axiology of services in the consumption experience, Serv. Qual.: New Directions in Theory Pract. 21 (1) (1994) 21–71.

[50] R.B. Woodruff, Customer value: the next source for competitive advantage, J. Acad. Mark. Sci. 25 (2) (1997) 139.

[51] Y.-H. Hu, K. Chen, P.-J. Lee, The effect of user-controllable filters on the prediction of online hotel reviews, Inf. Manag. 54 (6) (2017) 728–744.

[52] D. Schuff, S. Mudambi, What makes a helpful online review? A study of customer reviews on amazon. com, Soc. Sci. Electron. Publ. 34 (1) (2012) 185–200.

[53] C. Forman, A. Ghose, B. Wiesenfeld, Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets, Inf. Syst. Res. 19 (3) (2008) 291–313.

[54] A.M. Weiss, N.H. Lurie, D.J. MacInnis, Listening to strangers: whose responses are valuable, how valuable are they, and why? J. Mark. Res. 45 (4) (2008) 425–436.

[55] Z. Liu, S. Park, What makes a useful online review? Implication for travel product websites, Tour. Manag. 47 (2015) 140–151.

[56] R.A. King, P. Racherla, V.D. Bush, What we know and don't know about online word-of-mouth: A review and synthesis of the literature, J. Interact. Mark. 28 (3) (2014) 167–183.
[57] R. Filieri, S. Alguezaui, F. McLeay, Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth, Tour. Manag. 51 (2015) 174–185.
[58] P. Wang, Understanding the influence of electronic word-of-mouth on outbound tourists' visit intention, in: Conference on E-Business, E-Services and E-Society, Springer, 2014, pp. 33–45.
[59] S. Pourfakhimi, T. Duncan, W.J. Coetzee, Electronic word of mouth in tourism and hospitality consumer behaviour: state of the art, Tour. Rev. (2020).
[60] S. Zhou, Q. Yan, M. Yan, C. Shen, Tourists' emotional changes and eWOM behavior on social media and integrated tourism websites, Int. J. Tour. Res. 22 (3) (2020) 336–350.
[61] L.-Y. Leong, T.-S. Hew, K.-B. Ooi, B. Lin, Do electronic word-of-mouth and elaboration likelihood model influence hotel booking? J. Comput. Inf. Syst. 59 (2) (2019) 146–160.
[62] K. Kitcharoen, The effect of E-word of mouth (E-WOM) on various factors influencing customers' hotel booking intention, ABAC ODI J. Vis. Action. Outcome 6 (1) (2019) 62.
[63] U. Chakraborty, Perceived credibility of online hotel reviews and its impact on hotel booking intentions, Int. J. Contemp. Hosp. Manag. (2019).
[64] S.W. Sussman, W.S. Siegal, Informational influence in organizations: An integrated approach to knowledge adoption, Inf. Syst. Res. 14 (1) (2003) 47–65.
[65] Y. Luo, X. Xu, Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: A case study of yelp, Sustainability 11 (19) (2019) 5254.
[66] R. Filieri, F. McLeay, B. Tsui, Z. Lin, Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services, Inf. Manag. 55 (8) (2018) 956–970.
[67] R. Rietsche, D. Frei, E. Stöckli, M. Söllner, Not all reviews are equal-a literature review on online review helpfulness, 2019.
[68] S. Chatterjee, Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach, Int. J. Hosp. Manag. 85 (2020) 102356.
[69] T. Ghosh, Predicting hotel book intention: The influential role of helpfulness and advocacy of online reviews, J. Hosp. Mark. Manag. 27 (3) (2018) 299–322.
[70] S. Panigrahi, N.A. Azizan, M. Waris, Investigating the empirical relationship between service quality, trust, satisfaction, and intention of customers purchasing life insurance products, Indian J. Mark. January (2018).
[71] N. Zarifah Dhabitah Mahat, M. Hafiz Hanafiah, Help me TripAdvisor! Examining the relationship between TripAdvisor e-WOM attributes, trusts towards online reviews and travellers behavioural intentions, J. Inf. Organ. Sci. 44 (1) (2020) 83–112.
[72] J.A. Yeap, J. Ignatius, T. Ramayah, Determining consumers' most preferred eWOM platform for movie reviews: A fuzzy analytic hierarchy process approach, Comput. Hum. Behav. 31 (2014) 250–258.
[73] R. Ramli, A.Z.M. Ibrahim, A.A. Bakar, A.H. Mohamad, Evaluating customer trust towards online SNSs sellers using fuzzy AHP, in: Journal of Physics: Conference Series, Vol. 1529, IOP Publishing, 2020, 052016.
[74] K. Palanivel, R. Sivakumar, A study on collaborative recommender system using fuzzy-multicriteria approaches, Int. J. Bus. Inf. Syst. 7 (4) (2011) 419–439.
[75] M. Nilashi, E. Yadegaridehkordi, O. Ibrahim, S. Samad, A. Ahani, L. Sanzogni, Analysis of travellers' online reviews in social networking sites using fuzzy logic approach, Int. J. Fuzzy Syst. 21 (5) (2019) 1367–1378.
[76] S. Dogan, M.A. Basaran, K. Kantarci, Determination of attributes affecting price-performance using fuzzy rule-based systems: online ratings of hotels by travel 2.0 users, J. Hosp. Tour. Technol. 11 (2) (2020) 291–311.
[77] C.-H. Cheng, Y.-S. Chen, Classifying the segmentation of customer value via RFM model and RS theory, Expert Syst. Appl. 36 (3) (2009) 4176–4184.
[78] M. Khajvand, M.J. Tarokh, Estimating customer future value of different customer segments based on adapted RFM model in retail banking context, Procedia Comput. Sci. 3 (2011) 1327–1332.
[79] Y. Zhang, E.T. Bradlow, D.S. Small, Predicting customer value using clumpiness: From RFM to RFMC, Mark. Sci. 34 (2) (2015) 195–208.
[80] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, in: Learning Systems and Intelligent Robots, Springer, 1974, pp. 1–10.
[81] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
[82] F. Herrera, L. Martínez, A 2-tuple fuzzy linguistic representation model for computing with words, IEEE Trans. Fuzzy Syst. 8 (6) (2000) 746–752.
[83] F. Herrera, L. Martínez, A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision-making, IEEE Trans. Syst. Man Cybern. B 31 (2) (2001) 227–234.
[84] A.M. Hughes, Strategic database marketing: The masterplan for starting and managing a profitable, customer-based marketing program, 1994.
[85] T.L. Saaty, Axiomatic foundation of the analytic hierarchy process, Manage. Sci. 32 (7) (1986) 841–855.
[86] T.L. Saaty, How to make a decision - The analytic hierarchy process, Interfaces 24 (6) (1994) 19–43.
[87] T.L. Saaty, Decision making with the analytic hierarchy process, Int. J. Serv. Sci. 1 (1) (2008) 83–97.
[88] G. Kou, C. Lin, A cosine maximization method for the priority vector derivation in ahp, Eur. J. Oper. Res. 235 (1) (2014) 225–232.
[89] J. Zhang, G. Kou, Y. Peng, Y. Zhang, Estimating priorities from relative deviations in pairwise comparison matrices, Inf. Sci. 552 (2021) 310–327.
[90] G. Kou, D. Ergu, J. Shang, Enhancing data consistency in decision matrix: adapting hadamard model to mitigate judgment contradiction, Eur. J. Oper. Res. 236 (1) (2014) 261–271.
[91] C. Lin, G. Kou, Y. Peng, F.E. Alsaadi, Aggregation of the nearest consistency matrices with the acceptable consensus in ahp-gdm, Ann. Oper. Res. (2020) 1–17.
[92] M.M. Masrurul, Impact of service quality on customer satisfaction in Bangladesh tourism industry: An empirical study, Adv. In Manag. Vol 12 (2019) 1.
[93] K. Pervez, E. Uddin, A. Shah, F. Prodhan, M. Sheikh, Fuzzy-likert scale based assessment of marketing risk faced by the hybrid rice growers of Bangladesh, Ekonomika poljoprivrede 66 (1) (2019) 9–22.
[94] H.A. Megeirhi, K.M. Woosnam, M.A. Ribeiro, H. Ramkissoon, T.J. Denley, Employing a value-belief-norm framework to gauge Carthage residents' intentions to support sustainable cultural heritage tourism, J. Sustain. Tour. 28 (9) (2020) 1351–1370.
[95] D. Pan, J. Yuan, L. Li, D. Sheng, Deep neural network-based classification model for sentiment analysis, in: 2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC), IEEE, 2019, pp. 1–4.
[96] H. Sharma, A. Tandon, P. Kapur, A.G. Aggarwal, Ranking hotels using aspect ratings based sentiment classification and interval-valued neutrosophic TOPSIS, Int. J. Syst. Assur. Eng. Manag. 10 (5) (2019) 973–983.
[97] X. Liang, P. Liu, Z. Wang, Hotel selection utilizing online reviews: A novel decision support model based on sentiment analysis and DL-VIKOR method, Technol. Econ. Dev. Econ. 25 (6) (2019) 1139–1161.
[98] G.A. Miller, The magical number seven, plus or minus two: Some limits on our capacity for processing information, Psychological review 63 (2) (1956) 81.
[99] H.A. Lee, R. Law, J. Murphy, Helpful reviewers in TripAdvisor, an online travel community, J. Travel Tour. Mark. 28 (7) (2011) 675–688.
[100] A. Ganzaroli, I. De Noni, P. van Baalen, Vicious advice: Analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice, Tour. Manag. 61 (2017) 501–510.
[101] A. Valdivia, E. Hrabova, I. Chaturvedi, M.V. Luzón, L. Troiano, E. Cambria, F. Herrera, Inconsistencies on TripAdvisor reviews: A unified index between users and sentiment analysis methods, Neurocomputing 353 (2019) 3–16.
[102] W. Kasper, M. Vela, Sentiment analysis for hotel reviews, in: Computational Linguistics-Applications Conference, Vol. 231527, 2011, pp. 45–52.
[103] N.T. Duyen, N.X. Bach, T.M. Phuong, An empirical study on sentiment analysis for Vietnamese, in: 2014 International Conference on Advanced Technologies for Communications (ATC 2014), IEEE, 2014, pp. 309–314.
[104] A. Hasan, S. Moin, A. Karim, S. Shamshirband, Machine learning-based sentiment analysis for twitter accounts, Math. Comput. Appl. 23 (1) (2018) 11.
[105] Y.-C. Chang, C.-H. Ku, C.-H. Chen, Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor, Int. J. Inf. Manage. 48 (2019) 263–279.
[106] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, X. Wu, Chinese text sentiment analysis based on extended sentiment dictionary, IEEE Access 7 (2019) 43749–43762.
[107] C. Song, X.-K. Wang, P.-F. Cheng, J.-Q. Wang, L. Li, SACPC: A framework based on probabilistic linguistic terms for short text sentiment analysis, Knowl.-Based Syst. 194 (2020) 105572.
[108] A. Hermawan, et al., The hotel recommendation system using SAW (simple additive weighting) and TOPSIS (the technique for order of preference by similarity to ideal solution) method, bit-Tech 1 (3) (2019) 131–145.
[109] M. Leon Santiesteban, A.M. Larranaga Nunez, Integration of knowledge in restaurants through multicriteria analysis for decision making, Inquietud Empresarial 19 (2) (2019) 25–38.
[110] Z. Wei, X. Shan, RFM high-speed railway customer value classification model based on spark, in: IOP Conference Series: Materials Science and Engineering, Vol. 563, IOP Publishing, 2019, 052081.
[111] R.A. Carrasco, M.F. Blasco, J. García-Madariaga, E. Herrera-Viedma, Integrating a tourism service quality evaluation linguistic multi-criteria decision making model into a relational database management system, in: Advances in Fuzzy Logic and Technology 2017, Springer, 2017, pp. 408–418.

[112] H. Wang, Y. Lu, C. Zhai, Latent aspect rating analysis on review text data: a rating regression approach, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010, pp. 783–792.
[113] H. Wang, Y. Lu, C. Zhai, Latent aspect rating analysis without aspect keyword supervision, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011, pp. 618–626.
[114] P. Digkoglou, A. Dragoslis, J. Papathanasiou, V. Kostoglou, Using AHP and VIKOR to evaluate the hotel industry of eight European countries, in: Balkan Region Conference on Engineering and Business Education, Vol. 3, Sciendo, 2017, pp. 8–13.
[115] Y. Song, H. Wang, M. Zhu, Sustainable strategy for corporate governance based on the sentiment analysis of financial reports with CSR, Financial Innov. 4 (1) (2018) 2.
[116] W.N. Wan Ahmad, N.M. Ali, A study on persuasive technologies: The relationship between user emotions, trust and persuasion, Int. J. Interact. Multimedia Artif. Intell. 5 (1) (2018) 57–61.