Technological and Economic Development of Economy
ISSN: 2029-4913 / eISSN: 2029-4921
2023 Volume 29 Issue 2: 411–437
https://doi.org/10.3846/tede.2022.17972

CUSTOMER PREFERENCE ANALYSIS FROM ONLINE REVIEWS BY A 2-ADDITIVE CHOQUET INTEGRAL-BASED PREFERENCE DISAGGREGATION MODEL

Huchang LIAO, Qian YANG*, Xingli WU

Business School, Sichuan University, Chengdu, China

Received 28 February 2022; accepted 17 October 2022; first published online 22 November 2022

Abstract. Online reviews have become an important data source for analyzing consumers' preferences. Consumer preference analysis assists product managers to understand consumers' propensity for different product attributes and make consumer-oriented market strategies. Existing studies on consumer preference analysis used simple additive algorithms to represent the relationship between overall ratings and attribute ratings, but ignored the interactions between attributes. In addition, not all attribute ratings were given by consumers when calculating the overall ratings of a product. To fill these gaps, a preference model based on the extended 2-additive Choquet integral is constructed. The 2-additive Choquet integral can reflect the importance of attributes and the interactions between pairs of attributes when integrating attribute ratings. In cases where consumers choose only a subset of product attributes to rate a product, we introduce the scale parameter into the 2-additive Choquet integral to characterize the relationship between different attribute subsets. Afterwards, a preference disaggregation paradigm based on nonlinear programming is provided to solve the preference model. Finally, the proposed method is validated by experimental analysis using the dataset collected from TripAdvisor.com. Experimental outcomes indicate that our approach can deduce consumers' preferences and approximate the evaluation behavior of consumers efficiently.

Keywords: consumer preferences, online reviews, preference disaggregation, multiple attribute decision aiding, 2-additive Choquet integral.

JEL Classification: C51, C61, D81, L00.

Introduction

Online reviews are primary means for consumers to describe their satisfaction with the overall performance and attribute characteristics of products (Zhang et al., 2021; Zhu et al., 2022). Consumers' preferences can be viewed as how much consumers like a product or product portfolio: the more important an attribute is perceived by consumers, the more a product's performance on that attribute can influence consumers' judgment. Scholars have proved the

*Corresponding author. E-mail: *yangqianscu@163.com*

effectiveness of online reviews in consumer preference analysis (Xiang et al., 2017; Tsai et al., 2020). By conducting consumer preference analysis based on online reviews, platform owners and product managers can know the reason why a consumer prefers one product to another and get an insight to design marketing strategies to maintain consumer loyalty and attract new consumers (Li et al., 2020; Guo et al., 2020).

Existing approaches for consumer preferences analysis based on online reviews can be classified into three categories, including conjoint analysis (Hung et al., 2019), recommender system-based approaches (Lee et al., 2010), and multiple attribute decision aiding (MADA) approaches (Guo et al., 2020). The conjoint analysis aims to analyze a group of consumers' preferences over product attributes. The recommender system-based approaches, such as collaborative filtering-based techniques (Shi et al., 2014) and content-based techniques (Pazzani & Billsus, 2007; Chung & Rao, 2012), are used to predict the overall ratings of products that are not used or purchased by consumers. These methods reflect the propensity degree of consumers for different products and make accurate recommendations, but did not analyze the consumers' preferences over product attributes. The MADA approaches can help product managers to approximately describe the decision-making behavior of consumers when evaluating products under multiple attributes through preference models, which provide insights on how consumers evaluate products and how much they attach importance to product attributes (Guo et al., 2020). Thus, to capture consumers' preferences for product attributes hidden in online reviews, in this study, we focus on building a consumer preference model from the perspective of MADA.

There were few studies on consumer preference analysis using MADA approaches under the background of online comments. Guo et al. (2020) proposed a data-driven MADA approach in which an additive preference model was used to integrate online information on attributes given by consumers. Zhu et al. (2022) developed an optimization model based on an additive preference model to model the decision process of an individual consumer based on individual consumers' online reviews and acquire the consumer's preferences for different attributes. These two studies both modeled consumers' preferences regarding the importance of attributes based on an additive preference model that was subject to a strict precondition of the mutual preference independence among attributes. Nevertheless, ignoring the interactions among attributes may limit the potential usage of these additive models in reflecting the real decision process. In an evaluation process, there may be some correlations between product attributes. When negative interactions exist, the combined effect of attributes on product performance is lower than the addition of the effects of these attributes on product performance. Conversely, when positive interactions appear, the joint impact of attributes on product performance should be larger than a simple summation of these impacts viewed individually (Guo et al., 2019). In this sense, it is necessary to consider consumers' preferences for interactions between attributes in a preference model for consumer preference analysis. In practical decision problems, it is impractical to describe the interactions between a large subset of attributes, which would place a great burden on a consumer to perform parameter identification in a preference model. In particular, the number of parameters to be confirmed by consumers will increase exponentially with the increase of the number of attributes. The 2-additive Choquet integral (Grabisch, 1997), a representative method to reflect interactions among pairs of attributes, offers a good compromise between the low complexity and richness of a preference model. Because of this feature, the 2-additive Choquet integral has been widely used to solve different problems, such as the global mental workload evaluation (Pelegrina et al., 2020), the identification of the best hospitals in weight loss surgery (Mayag & Bouyssou, 2020), and the sovereign ratings of European countries (Arcidiacono et al., 2021). However, it has not been used as a preference model to extract consumers' preferences within the context of online reviews. To fill this gap, this paper applies the 2-additive Choquet integral, as a preference model, to analyze consumers' preferences for attributes, especially interactions between attributes based on the ratings of online products.

In addition, when applying preference models, there is an assumption that the type and number of attributes used to evaluate alternatives are fixed. However, in online reviews, the relevant attributes used to evaluate products are usually different, even for the same category of products. Due to consumers' willingness and limited cognition, only part of a series of product attributes will be selected by consumers when evaluating products. The same consumer may also use different attributes when evaluating different products. In this sense, the second goal of this study is to improve the preference model to handle the problem that the attributes used to evaluate products are not fixed.

Based on the above analysis, this study aims to conduct the following innovative work: 1) We use the 2-additive Choquet integral as a preference model to portray a consumer's evaluation process, which can consider consumers' preferences for interactions between attributes. 2) The concept of scale parameter is introduced into the 2-additive Choquet integral to address the problem that relevant product attributes are not fully used when evaluating products. By portraying the relationships between the whole attribute set and attribute subsets, the scale parameter extends the application of the preference model in MADA. 3) A nonlinear programming model is proposed to learn unknown parameters in the preference model from online reviews and analyze consumers' preferences for product attributes. Finally, a real case study on Tripadvisor.com is proposed to verify the applicability of our proposed method.

The remainder of this paper is organized as follows. The next section provides related work and preliminaries of the 2-additive Choquet integral. In Section 2, the extended 2-additive Choquet integral is developed and the detailed learning procedures concerning the parameters in the preference model are presented. Section 3 implements several experiments with practical data from e-commerce platforms to test the applicability of the proposed method. The last section concludes this paper.

1. Related work and literature review

1.1. The consumer preference model based on MADA approaches

In the MADA framework, preference models can be classified into three types, including preference models based on multiple attribute utility/value theory (Dyer & Smith, 2021), outranking relation models (Govindan & Jepsen, 2016), and decision rule-oriented (if ..., then ...) models (Greco et al. 2016). The multiple attribute value theory refers to obtaining the marginal value of a product attribute through a marginal value function when the value dimensions of attributes are inconsistent, thus integrating the marginal value of the product under different attributes through a global value function to obtain the product's comprehensive performance. The basic idea of the outranking relation-based approach is to calculate the preference strength between products based on the evaluated values under each attribute, and to construct the consistency and inconsistency matrices of products. Besides, the overall preference strength of products is decided by integrating the predominance and non-predominance relationships. The decision rule-based approach is to explain the decision strategy using a preference model containing a decision rule form. Compared with the latter two approaches, the methodology of multiple attribute utility/value theory represented by value functions can explain more clearly how a consumer evaluates a set of products under different attributes. For example, whether consumers perceive different attributes to have different degrees of influence on product performance.

Because its treatment of consumers' preferences for different attributes is closer to the actual behavior of consumers, value functions have become the most commonly used preference model to analyze consumer preference in the field of MADA. Value functions can have different complexity, ranging from the additive value function (Kadzinski & Tervonen, 2013; Ghaderi et al., 2017; Guo et al., 2019; Grigoroudis et al., 2021) to the non-additive value functions (Branke et al., 2016; Aggarwal & Tehrani, 2019; Pereira et al., 2020). Many studies have employed the additive value function as an underlying preference model to represent consumers' preferences. For instance, Guo, Liao, and Liu (2019) employed an additive piecewise-linear value function as the basic preference model and augmented some components to handle the interactions among attributes. Grigoroudis et al. (2021) considered additive value functions as a consumer's preference model to analyze consumer behavior and study how different factors affect the price of artistic goods in the Art market. However, the assumption of mutual independence among attributes on which the additive value function depends was not always consistent with reality. Subsequently, non-additive value functions, represented by the family of Choquet integral (Choquet, 1954), were used as preference models to characterize complex decision-making behaviors of consumers.

Although value functions (Keeney & Raiffa, 1976) have been used as preference models and applied in different practical contexts such as alternative evaluation (Aggarwal & Tehrani, 2019) and performance assessment (Pereira et al., 2020), few studies have used value functions to extract consumers' preferences hidden in online reviews. Only few scholars (Guo et al., 2020; Zhu et al., 2022) have considered additive value functions as a preference model for consumer preference analysis. There is a lack of research on extracting complex consumer preferences based on non-additive value functions within the context of online reviews.

1.2. The 2-additive Choquet integral

The multiple attribute utility theory represents the preference relations of a set of alternatives $A = \{a_1, a_2, \cdots, a_m\}$ regarding a set of attributes $C = \{c_1, c_2, \cdots, c_n\}$ by the transitive decomposable model (Grabisch et al., 2008):

$$U(a_i) = F(u_1(a_i), \cdots, u_n(a_i)), \ \forall a_i \in A \tag{1}$$

with

$$a_i \succeq a_t \Leftrightarrow U(a_i) \geq U(a_t), \ \forall a_i, a_t \in A, \tag{2}$$

where $U(\cdot)$ indicates the global value of alternative a_i ($a_i \in A$), which is expected to represent a consumer's practical preferences as much as possible. $u(\cdot)$ is the marginal value of alternative a_i ($a_i \in A$) on the attribute set *C*, representing the consumer's cognition of the performance of an alternative under different attributes. The non-decreasing function $F(\cdot)$ is a value function to make trade-offs between the marginal values of an alternative over different attributes and output a global value. The preference relation "$\succeq$" on *A* is a separable weak order including the preference relation and indifference relation, which is determined by the global value *U*: when $U(a_i) \geq U(a_t)$, a_i is at least as good as a_t. The most frequently-encountered value function is based on the averaging strategy such as the arithmetic, geometric or power weighted sum. In these value functions, different attributes are treated independently.

However, the attributes of an alternative are not always fully independent of each other in reality. The Choquet integral is a representative value function considering interactions among attributes. For an alternative $a_i \in A$, the Choquet integral with respect to the capacity μ is an aggregation operator presented as:

$$U_{\mu}(a_i) = F_{\mu}(u_1(a_i), u_2(a_i), \cdots, u_n(a_i)) = \sum_{j=1}^{n} (u_{(j)}(a_i) - u_{(j-1)}(a_i))\mu(\{c_{(j)}, \cdots, c_{(n)}\}), \qquad (3)$$

where $\{u_{(1)}(a_i), u_{(2)}(a_i), \cdots, u_{(n)}(a_i)\}$ is a permutation of $\{u_1(a_i), u_2(a_i), \cdots, u_n(a_i)\}$, such that $u_{(1)}(a_i) \leq u_{(2)}(a_i) \leq \cdots \leq u_{(n)}(a_i)$, and $u_{(0)}(a_i) = 0$. $\mu(\{c_{(j)}, \cdots, c_{(n)}\})$ is the capacity of the whole attribute set *C*, ($j = 1,..,n$), which assigns a weight to each subset of attributes. A capacity on *N* is a set function $\mu : 2^C \to [0,1]$, such that $\mu(\varnothing) = 0$; $\mu(C) = 1$; $\forall A, B \in 2^C$, $\left[A \subseteq B \Rightarrow \mu(A) \leq \mu(B)\right]$, where 2^C is the power set of *C* and $\mu(A)$ (or $\mu(B)$) represents the importance of attribute set *A* (or *B*). If the attributes are independent, then, $\mu(\{c_{(j)}, \cdots, c_{(n)}\}) = \mu(c_{(1)}) + \mu(c_{(2)}) + \cdots + \mu(c_{(n)})$ and Eq. (3) degenerates to the weighted arithmetic aggregation operator that $F_{\mu}(u_1(a_i), u_2(a_i), ... u_n(a_i)) = \sum_{j=1}^{n} \mu(c_{(j)}) u_j(a_i)$, where $\sum_{j=1}^{n} \mu(c_{(j)}) = 1$.

The Möbius transform $m^{\mu} : 2^C \to R$ of a capacity μ was introduced by Chateauneuf and Jaffray (1989) to simplify the calculation of the capacity, which was given as $m^{\mu}(X) := \sum_{Y \subseteq X} (-1)^{|X|-|Y|} \mu(Y)$, for all $X \in 2^C$. Conversely, we have $\mu(X) = \sum_{Y \subseteq X} m^{\mu}(Y)$, for all $X \in 2^C$, where $|X|$ (or $|Y|$) is the number of attributes in the attribute set *X* (or *Y*). According to the definition of capacity, the Möbius transform should satisfy $m^{\mu}(\varnothing) = 0$; $\sum_{X \in C} m^{\mu}(X) = 1; \forall c_j \in C$ and $\forall X \in C \setminus \{c_j\}$, $m\left(\{c_j\}\right) + \sum_{Y \subseteq X} m(Y \cup \{c_j\}) \geq 0$. Then, the Choquet integral can be expressed in terms of the Möbius transform m^{μ} of the capacity μ as:

$$U_{\mu}(a_i) = \sum_{X \in C} m^{\mu}(X) \times \min_{j=1,..,n} \left\{u_j(a_i)\right\}. \qquad (4)$$

To represent the capacity of the attribute set *C* with $\mu(\varnothing) = 0$ and $\mu(C) = 1$, the values $\mu(X)$ need to be assigned by the capacity μ to all other $2^{|C|} - 2$ subsets of *C*. In other words, we need to determine 2^n variables. The difficulty to measure a capacity increases exponentially with the number of attributes increasing. To control the complexity, Grabisch (1997) investigated the *k*-additive capacity to describe the capacity of an attribute subset whose number is equal to or smaller than *k*. The *k* -additive measure needs to define $\sum_{j\ 1}^{k} C_{n,j}$ variables. Especially, the 2-additive capacity considers the interactions only for pairs of attributes and reduces the complexity of capacity measurement to $n(n+1)/2$ variables.

Definition 1 (Grabisch, 1997)**.** μ is a 2-additive capacity if and only if its Möbius transform satisfies: 1) $\forall X \in 2^C$ with $|X|>2$, such that $m^{\mu}(X)=0$; 2) $\exists Y \in 2^C$ with $|Y|=2$, such that $m^{\mu}(Y)\neq 0$.

The 2-additive capacity of any subset of an attribute set C can be simplified as:

$$\mu(X)=\sum_{\{c_j\}\subseteq X\subseteq C} m^{\mu}(\{c_j\})+\sum_{\{c_j,c_k\}\subseteq X\subseteq C} m^{\mu}(\{c_j,c_k\}), \tag{5}$$

where $m^{\mu}(\{c_j\})$ is the Möbius transform of the capacity of the attribute c_j, and $m^{\mu}(\{c_j,c_k\})$ is the interaction degree between c_j and c_k (in this paper, $j\neq k$ unless otherwise states). According to the definition of Möbius transform, we have $m^{\mu}(\{c_j\})=\mu(\{c_j\})$, and $m^{\mu}(\{c_j,c_k\})=\mu(\{c_j,c_k\})-\mu(\{c_j\})-\mu(\{c_k\})$. For simplicity, we will use the following shorthand: $\mu_j=\mu(\{c_j\})$, $\mu_{jk}=\mu(\{c_j,c_k\})$, $m_j^{\mu}=m^{\mu}(\{c_j\})$ and $m_{jk}^{\mu}=m^{\mu}(\{c_j,c_k\})$ for all $j,k\in C$.

According to Eq. (5), Eq. (4) can be transformed as:

$$U_{\mu}(a_i)=\sum_{j\in C} m_j^{\mu}u_j(a_i)+\sum_{j,k\in C} m_{jk}^{\mu}\min(u_j(a_i),u_k(a_i)). \tag{6}$$

To further simplify the calculation of the capacity μ and its Möbius transform m^{μ}, the equivalent form of Eq. (6) for a 2-additive capacity was defined as follows (Grabisch, 1997):

$$U_{\mu}(a_i)=\sum_{j=1}^{n} v_j^{\mu}u_j(a_i)-\frac{1}{2}\sum_{j,k\in C} I_{jk}^{\mu}\left|u_j(a_i)-u_k(a_i)\right|, \tag{7}$$

where $v_j^{\mu}=\mu_j+\frac{1}{2}\sum_{k\in C\setminus\{j\}}\left[\mu_{jk}-\mu_j-\mu_k\right]$ is the importance of attribute c_j represented by the Shapley value of the capacity μ, and I_{jk}^{μ} is the interaction index between attributes c_j and c_k as defined in Grabisch (1997). When μ is a 2-additive capacity, $I_{jk}^{\mu}=m_{jk}^{\mu}=\mu_{jk}-\mu_j-\mu_k$. A conversion formula between v_j^{μ} and I_{jk}^{μ} is:

$$v_j^{\mu}=\mu_j+\frac{1}{2}\sum_{k\in C\setminus j} I_{jk}^{\mu}. \tag{8}$$

2. The proposed method

This section aims to take the 2-additive Choquet integral as the preference model of individual consumers and then extract customer preferences based on online ratings.

2.1. The extended 2-additive Choquet integral

The classical 2-additive Choquet integral can only be utilized to measure different alternatives on the premise that the alternatives have marginal values under all given attributes. However, within the context of online reviews, consumers do not always use all preset attributes but may choose some of them to evaluate products[1], which can be seen in Figure 1 and Figure 2

[1] In the remaining papers, we regard the phenomenon that a product has not been evaluated under all attributes as an incomplete evaluation of the product.

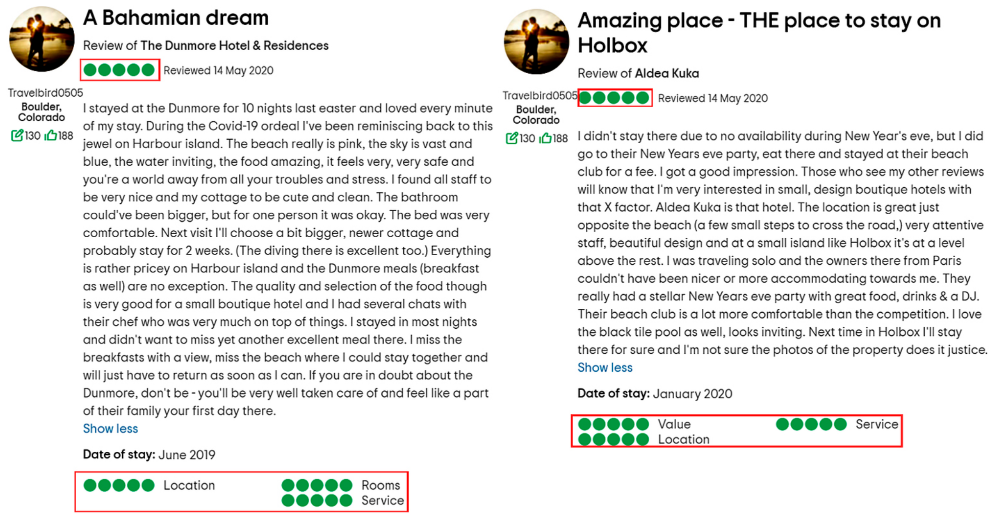


Figure 1. The online reviews of two hotels evaluated by the consumer “Travelbird0505” based on different types of attributes

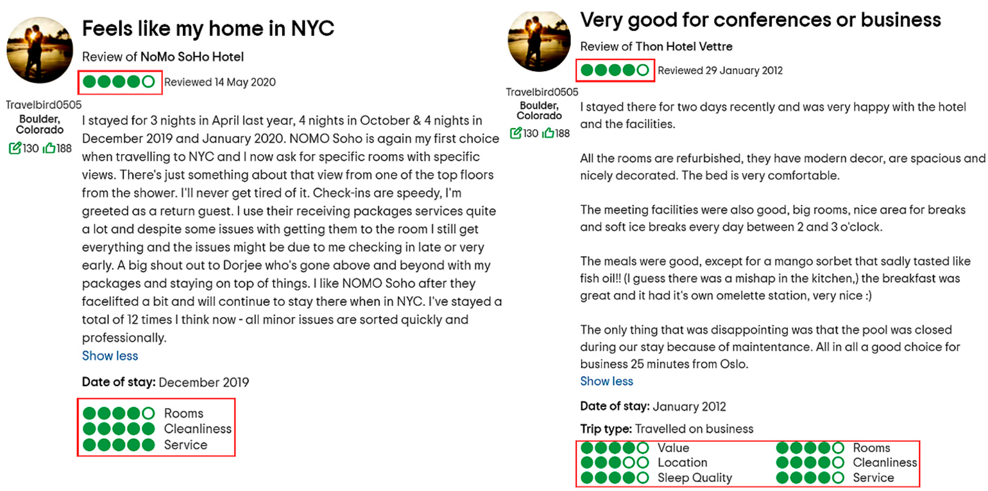


Figure 2. The online reviews of two hotels given by the consumer “Travelbird0505” based on different numbers of attributes

as examples. Figure 1 shows online reviews provided by the consumer “Travelbird0505” on the hotel “Aldea Kuka” and “The Dunmore Hotel & Residences”, respectively. Each review consists of the overall star rating of the hotel and the star rating on each attribute, which are both based on the scale from 1 (“very unsatisfied”) to 5 (“very satisfied”). This consumer gave the two hotels the same overall rating (five points) considering the same number (three) but different types of attributes. Three attributes including location, service, and value are chosen to evaluate the hotel “Aldea Kuka”, while other three attributes including location, rooms, and service are selected to evaluate the hotel “The Dunmore Hotel & Residences”. In addition, the same consumer “Travelbird0505” used different numbers of attributes to determine the satisfactory degrees of the other two hotels that can be seen in Figure 2. Although both hotels received the same overall rating (four points), the chosen attributes cover different aspects and numbers. It verifies that different products may be evaluated from different attributes both in types and numbers by the same consumer, even though these products fall into the same category.

For a whole attribute set $C=\{c_1,c_2,\cdots,c_n\}$, let $C_i=\{c_{1_i,j},c_{2_i,j},\cdots,c_{n_i,j}\}$ be a subset of C for evaluating alternative a_i, where $c_{n_i,j}$ indicates that this attribute is the nth attribute of a_i (in subset C_i) and it also exists in the attribute set C, $n_i \le n$. If the whole attribute set C has m subsets, then, $C_1 \cup C_2 \cup \cdots \cup C_m = C$. As long as $C_i \subset C$ exist, the predefined attributes are not all evaluated in online reviews. In this case, the 2-additive Choquet integral model cannot be directly used as a value function to analyze consumer preferences, especially to obtain the interactions between attributes. To solve this problem, we extend the 2-additive Choquet integral according to the characteristics of online reviews.

Generally speaking, if a consumer is concerned about specific attributes of a product, the consumer will actively express his/her true opinion in online reviews, no matter the overall evaluation is positive or negative. Since consumers are not willing to evaluate or not interested in some attributes of a product, or the product has a general performance on some attributes, consumers may only be willing to disclose the evaluations of the attributes that are more important to them, or the attributes that have extreme performance (very good or very poor) in online reviews. Therefore, this study assumes that, for consumers, the attributes mentioned in online reviews are more important than those not mentioned. Specifically, we prefer to allocate the weights of non-evaluated attributes to evaluated attributes in proportion. In this sense, we introduce a self-defined parameter called the scale parameter into the 2-additive Choquet integral. In this regard, we need to define the following hypotheses.

Hypothesis 1: For a decision maker, the relative scale between the capacity of attributes c_j and c_k does not change from C to any of its subset that includes these two attributes.

Let the capacity of c_j in the whole attribute set C be μ_j, and the capacity of this attribute, $c_{l_i,j}$, in the subset C_i be $\mu_{l_i,j}$. Similarly, μ_k and $\mu_{t_i,k}$, $l_i, t_i \in n_i$, are derived. According to Hypotheses 1, the relative scale of the capacity of two attributes is consistent in the whole attribute set C and its subset C_i, which means:

$$\frac{\mu_j}{\mu_k}=\frac{\mu_{l_i,j}}{\mu_{t_i,k}}. \tag{9}$$

According to the formula of alternando in mathematics, we can deduce that $\frac{\mu_j}{\mu_{l_i,j}}=\frac{\mu_k}{\mu_{t_i,k}}$. Then, we can introduce a scale parameter η_i, $i = 1, \ldots, m$, to model the scale between μ_j (or μ_k) and $\mu_{l_i,j}$ (or $\mu_{t_i,k}$), such that:

$$\frac{\mu_j}{\mu_{l_i,j}}=\frac{\mu_k}{\mu_{t_i,k}}=\eta_i, \tag{10}$$

where the scale parameter η_i indicates the extent to which a consumer assigns the capacity to the attributes in subsets of attributes compared with in the whole attribute set. In other words, when the evaluations of alternatives under all given attributes is incomplete, a consumer assigns the capacity of these attributes to other attributes on which the evaluation of the alternatives is not missing. For a consumer, each alternative a_i has the scale parameter $\mu_{l_i,j} = \eta_i \mu_j$, $\eta_i \geq 1$, $i = 1,\ldots,m$, $j = 1,\ldots,n$, $l_i = 1_i,\ldots,n_i$. $\eta_i = 1$ indicates that the consumer evaluates an alternative with complete preset attributes.

Hypothesis 2: When evaluating the same type of products, a consumer has a relatively fixed cognition regarding the interactions between attributes, which does not go away but changes in proportion with the change of subsets of evaluation attributes. To be specific, the interaction index I^{μ}_{jk} between attributes c_j and c_k changes in proportion from C to any of its subset that includes these two attributes.

Let I^{μ}_{jk} be the interaction index between attributes c_j and c_k in the whole attribute set, which satisfies:

$$I^{\mu}_{jk} = \mu_{jk} - \mu_j - \mu_k, \tag{11}$$

where $I^{\mu}_{jk} > 0$ means that c_j and c_k are complementary, $I^{\mu}_{jk} < 0$ indicates the redundancy relation, and $I^{\mu}_{jk} = 0$ implies the independency relation. Its positive or negative value depends on the complementary or redundancy relationship between attributes. The greater its absolute value is, the greater the interaction strength is.

The interaction index $I^{\mu}_{jk(i)}$ between $c_{l_i,j}$ and $c_{t_i,k}$ with respect to C_i can be defined as:

$$I^{\mu}_{jk(i)} = \eta_i I^{\mu}_{jk}. \tag{12}$$

That is to say,

$$I^{\mu}_{jk(i)} = \eta_i (\mu_{jk} - \mu_j - \mu_k). \tag{13}$$

According to the definition of capacity, the boundary and monotonicity of the capacity of C_i can be expressed as:

$$1)\ \sum_{C_i \in C} m^{\mu}(C_i) = \sum_{l_i=1}^{n_i} \mu_{l_i,j} + \sum I^{\mu}_{jk(i)} = \sum_{j=1}^{n} \eta_i \mu_j + \sum_{k \in C \setminus \{j\}} \eta_i I^{\mu}_{jk} = 1; \tag{14}$$

$$2)\ \mu_{l_i,j} + \sum_{t_i \in n_i \setminus l_i} I^{\mu}_{jk(i)} \geq 0,\ \forall c_j, c_k \in C,\ \forall c_{l_i}, c_{t_i} \in C_i. \tag{15}$$

Based on the above analysis, the capacities and interactions of attributes in C_i can be represented by the capacities and interactions of attributes in C with the help of the scale parameter η_i. Meanwhile, η_i can be deduced according to the boundary of the capacity of C_i when we obtain the capacities and interactions of attributes in C. In this sense, the extended 2-additive Choquet integral of a_i can be defined as:

$$\begin{aligned} U_{\mu}(a_i) &= \sum_{l_i=1}^{n_i} v^{\mu}_{l_i,j} u_j(a_i) - \frac{1}{2} \sum_{l_i,t_i \in n_i} I^{\mu}_{jk(i)} \left| u_j(a_i) - u_k(a_i) \right| \\ &= \sum_{l_i=1}^{n_i} (\mu_{l_i,j} + \frac{1}{2} \sum_{t_i \in n_i \setminus l_i} I^{\mu}_{jk(i)}) u_j(a_i) - \frac{1}{2} \sum_{l_i,t_i \in n_i} I^{\mu}_{jk(i)} \left| u_j(a_i) - u_k(a_i) \right| \\ &= \sum_{j=1}^{n} \eta_i (\mu_j + \frac{1}{2} \sum_{k \in C \setminus j} (\mu_{jk} - \mu_j - \mu_k)) u_j(a_i) - \frac{1}{2} \sum_{\{j,k\} \in C} \eta_i (\mu_{jk} - \mu_j - \mu_k) \left| u_j(a_i) - u_k(a_i) \right|, \end{aligned} \tag{16}$$

where $\mu_{l_i,j} = \eta_i \mu_j$, $\mu_{t_i,k} = \eta_i \mu_k$, $i = 1,2,\cdots,m$, $j = 1,2,\cdots,n$, $k = 1,2,\cdots,n$.

In the following, we take two examples to illustrate how the extended 2-additive Choquet integral can be used to evaluate alternatives when the attribute evaluation information is incomplete.

Example 1. Let us compare the comprehensive ability of two talents (a_1, a_2) in the field of basic research and applied technology, respectively. The collective attribute set is defined as {published papers (c_1), application effect (c_2), recognition by peers (c_3), ownership of intellectual property (c_4)}. Suppose that the capacities of these attributes are $\mu_1 = 0.4$, $\mu_2 = 0.2$, $\mu_3 = 0.1$ and $\mu_4 = 0.2$, and the interaction parameters of pairs of attributes are $\mu_{12} = 0.5$, $\mu_{13} = 0.6$, $\mu_{14} = 0.6$, $\mu_{23} = 0.4$, $\mu_{24} = 0.4$ and $\mu_{34} = 0.3$. The attribute subset $\{c_1, c_3, c_4\}$ is applied to measure the performance of a_1, and its attribute values are assumed to be $\{0.6, 0.8, 0.5\}$; Similarly, the attribute subset $\{c_2, c_3, c_4\}$ is applied to measure the performance of a_2, and its attribute values are given as $\{0.8, 0.6, 0.3\}$. By Eq. (11), we can calculate the interactions between attributes c_1 and c_2, namely, $I_{12} = \mu_{12} - \mu_1 - \mu_2 = 0.5 - 0.4 - 0.2 = -0.1$. Similarly, we can obtain $I_{13} = I_{23} = 0.1$ and $I_{14} = I_{24} = I_{34} = 0$. By Eq. (14), we have $(\mu_1 + \mu_3 + \mu_4)\eta_1 + \eta_1 \times I_{13} + \eta_1 \times I_{14} + \eta_1 \times I_{34} = 1$. The solution is $\eta_1 = 1.25$. Similarly, we have $\eta_2 = 1.67$. Meanwhile, the overall value of a_i can be calculated by Eq. (16). That is, $U_\mu(a_1) = 0.48$, $U_\mu(a_2) = 0.34$. Therefore, $U_\mu(a_1) > U_\mu(a_2)$ and $a_1 \succ a_2$.

Example 2. We consider an extension of the example of evaluating students in Grabisch (1996). The attributes of the evaluation of students in high school are the scores of four subjects: Mathematics (c_1), Physics (c_2), Literature (c_3) and English c_4 (the scores are given on the scale ranging from 0–10). Suppose that the capacities of the four attributes are $\mu_1 = 0.4$, $\mu_2 = 0.3$, $\mu_3 = 0.3$ and $\mu_4 = 0.2$. In order to favor well-equilibrated students, the redundant interactions between scientific subjects need to be reduced. From this point, the capacities attributed to different subsets of attributes are set as $\mu_{12} = 0.4$, $\mu_{13} = 0.7$, $\mu_{14} = 0.7$, $\mu_{23} = 0.7$, $\mu_{24} = 0.5$, $\mu_{34} = 0.4$. In the evaluation process, four students are evaluated by different subsets of four given attributes. The first student is evaluated by the whole attributes and the other three students are evaluated by a subset of attributes $\{c_1, c_2, c_3, c_4\}$ of which the incomplete evaluations exist. Then, four students (a_1, a_2, a_3 and a_4) are ranked by the overall scores derived by the extended 2-additive Choquet integral. The overall score of the student with complete evaluation (a_1) is aggregated directly by the classical 2-additive Choquet integral. Following Eq. (11), $I_{12} = -0.3$, $_{13} = 0$, $I_{14} = 0.1$, $I_{23} = 0.1$, $I_{24} = 0$, $I_{34} = -0.1$. Then, according to Eq. (7) and Eq. (8), $U_\mu(a_1) = 7.20$. In this case, the value of the scale parameter $\eta_1 = 1$ due to the complete evaluation of the student a_1 under four attributes. We also take the computation of student a_2 with incomplete evaluations as an illustration for addressing missing values. By Eq. (14), we have $(\mu_1 + \mu_3 + \mu_4)\eta_2 + \eta_2 \times I_{13} + \eta_2 \times I_{14} + \eta_2 \times I_{34} = 1$. Thus, $\eta_2 = 1.1$. Similarly, the value of η_3 and η_4 can be calculated by $(\mu_1 + \mu_2 + \mu_3)\eta_3 + \eta_3 \times I_{12} + \eta_3 \times I_{13} + \eta_3 \times I_{23} = 1$ and $(\mu_2 + \mu_3 + \mu_4)\eta_4 + \eta_4 \times I_{23} + \eta_4 \times I_{24} + \eta_4 \times I_{34} = 1$. Then, according to Eq. (16), the overall score of three students (a_2, a_3 and a_4) can be measured. The results are shown in Table 1.

According to Eq. (16), we can get an overall scores of the students a_2, a_3 and a_4 by the extended 2-additive Choquet integral, which does not add additional information but puts more focus on the information that already exists. The best choice can be gained if this decision maker wants to use the proposed method to provide an optimal decision. From Table 1,

Table 1. The results of the overall evaluation of four students based on the proposed method

Student	Score under four attributes				η_i	Overall score
	c_1	c_2	c_3	c_4		
a_1	10	8	6	4	1.00	7.20
a_2	8	–	10	6	1.10	8.14
a_3	6	8	8	–	1.25	7.75
a_4	–	6	7	8	1.25	6.75

we can get the ranking of four students ($a_2 \succ a_3 \succ a_1 \succ a_4$). Considering the comprehensive development of students, the school may give more approval to a_2 or a_3. The good performance in scientific subjects of student a_1 is offset by the bad performance in English, so this student does not get a high ranking. The extended 2-additive Choquet integral gives more capacity to the known attributes without adding additional information.

2.2. Application to consumer preference analysis based on online ratings

From Examples 1 and 2, it is easy to use the extended 2-additive Choquet integral if we know the capacities and interaction parameters of attributes. However, when using online reviews for consumer preference analysis, it is difficult to interact with consumers to get information about the importance of product attributes and the interactions among attributes. The technique of preference disaggregation would be helpful for indirectly extracting the preference information over the importance of product attributes and the interactions among attributes. This section aims to extract consumers' preference on product attributes from online reviews, so as to help the managers of online platform know consumers' purchase and evaluation behaviors. We apply the extended 2-additive Choquet integral as the preference model of different individual consumers.

2.2.1. The determination of attribute values based on online ratings

Ratings in online reviews can be a critical heuristic tool to perceive the evaluation of online consumer information (Li et al., 2020). Generally, a consumer evaluates products by ratings on a 1-5 scale concerning the attributes designed by platforms and provides an overall rating to express his/her overall opinion on a product. Since overall ratings and separate ratings on attributes can be seen as a visual reflection of a consumer's preferences for different products, this paper mainly takes online ratings of an individual consumer as the input of the proposed method.

Different individual consumers may not have a uniform rating scale (Wu & Liao, 2021). For different consumers, the meaning of a 5-rating is different, and a 5-rating does not mean that consumers are most satisfied with the overall performance of a product or the performance of this product under a certain attribute. In this sense, we consider learning the preferences of an individual consumer from his/her numerous historical online ratings. Here we mainly extract the preferences of individual consumers who have made a series of online ratings on the same type of products. Considering that product attributes directly provided by the online platform for a consumer to evaluate products in the form of ratings are representative and general, the set of all attributes used by a consumer to evaluate products can be seen as the attribute set. It should be emphasized that a consumer may use different attributes to evaluate various products.

We select m products, $A=\{a_1,a_2,\cdots,a_m\}$, in the same type of product set evaluated by a consumer. Suppose that there are n attributes, $C=\{c_1,c_2,\cdots,c_n\}$, which constitute the attribute set. Each product a_i is evaluated by a part of the attributes in C, i.e., an attribute subset $C_i=\{c_{1_i},c_{2_i},\cdots,c_{n_i}\}$, $n_i \le n$. To build a preference learning model, we set the evaluation scale $S=\{s_\phi \mid \phi=1,2,3,4,5\}$ as the notation of the rating {1-rating, 2-rating, 3-rating, 4-rating, 5-rating}. The overall rating S_i provided by a consumer to product a_i indicates its overall evaluation scale. The higher the rating given by the individual consumer is, the higher the consumer's satisfaction with the product.

How to characterize the marginal value of ratings is a widely studied issue. A relatively complete summary has been achieved by Li, Liu, and Zhu (2020). Considering the uniform distribution of the scale of ratings, we simplify the form of marginal value function as a linear form, referring to the method of Wu and Liao (2021). We normalize the ratings ranging from one to five into the interval [0, 1], and consider the normalized ratings as the global value of a product and its marginal values under different attribute. The normalization function is a monotonically increasing function $g(\cdot)$ and the global value of the overall rating can be defined as $U_S(a_i)=g(s_\phi)$, $g(s_\phi)\in[0,1]$, and $g(s_\phi)\le g(s_\varphi)$ if $\phi<\varphi$. Meanwhile, the marginal value function with respect to the attribute c_j can be also defined as $u_j(a_i)\quad g(s\ \ )$, $g(s_\phi)\in[0,1]$. When consumers do not rate products under some attributes, incomplete product attribute evaluations appear. At this time, we can determine the global value of products by Eq. (16).

2.2.2. Preference analysis in the paradigm of preference disaggregation

The main objective of this part is to learn the values of unknown parameters, which reflect the importance of product attributes and the interactions among them, in the established preference model from historical information given by the same consumer. The posted ratings are composed of the overall ratings of products and the ratings on separate product attributes. We can describe the input and output of preference analysis as follows:

Input. The marginal value of the online rating $u_j(a_i)$ provided by a consumer with respect to different attribute subsets $C_i \subseteq C$, and the global value of the overall rating $U_S(a_i)$ for a set of products $a_i \in A$ in the same category.

Output. The unknown parameters related to consumer preference in the extended 2-additive Choquet integral, so as to describe the importance a consumer attaches to the attributes and whether to consider interactions between attributes.

The aggregation-disaggregation paradigm in the preference disaggregation method refers to eliciting preference information and inferring a decision model from a set of decision examples on some reference alternatives (Fürnkranz & Hüllermeier, 2010; Doumpos & Zopounidis, 2011; Liu et al., 2019). In this regard, if the obtained model's parameters are in line with the actual preferential propensity of a consumer, then the decision model deduced from historical data is consistent with the real evaluation process of the alternatives by this consumer and thus can be applied to make new decisions for this consumer. Inspired by this idea of preference disaggregation, in this paper, the part of aggregation means the process of aggregating the performance values of products under all involved attributes based on the extended 2-additive Choquet integral, while the part of disaggregation refers to constructing nonlinear constraints shown in Model 1 for determining unknown parameters compatible with the individual consumer preferences. We assume that the global value of a product's overall rating is equal to the integrated value of the marginal values of the product under all attributes. When the differences between the global value of the overall rating and the integrated value under the attributes are minimum, the parameters learned from historical reviews can be seen as practicable for representing individual consumer preferences. In this sense, the objective function of such a preference extraction model is to minimize the differences between the true and integrated global values of products. Formally, we establish the following nonlinear programming to learn the preference information of a consumer:

Model 1. $$\min d_{\mu}(S)=\frac{1}{m}\sum_{i=1}^{m}(\xi_i^{+}+\xi_i^{-}); \tag{17}$$

$$\text{s.t.: } U_{\mu}(a_i)-U_S(a_i)-\xi_i^{+}+\xi_i^{-}=0,\ \forall a_i\in A; \tag{18}$$

$$U_{\mu}(a_i)=\sum_{j=1}^{n}\eta_i(\mu_j+\frac{1}{2}\sum_{k\in C\setminus j}I_{jk}^{\mu})u_j(a_i)-\frac{1}{2}\sum_{\{j,k\}\in C}\eta_i I_{jk}^{\mu}\left|u_j(a_i)-u_k(a_i)\right|; \tag{19}$$

$$\mu(C_i)=\sum_{j=1}^{n}\eta_i\mu_j+\sum_{k\in C\setminus\{j\}}\eta_i I_{jk}^{\mu}=1,\ \forall i\in\{1,2,\cdots,m\}; \tag{20}$$

$$\mu_{l_i,j}+\sum_{t_i\in n_i\setminus l_i}I_{jk(i)}^{\mu}\geq 0,\ \forall X\in 2^{C_i},\ \forall c_j,c_k\in C,\ \forall c_{l_i},c_{t_i}\in C_i; \tag{21}$$

$$\mu_{jk}\geq\mu_j,\mu_{jk}\geq\mu_k\ \ \forall j,k\in\{1,2,\cdots,n\},\ j\neq k; \tag{22}$$

$$\eta_i\geq 1,\ \forall i\in\{1,2,\cdots,m\}; \tag{23}$$

$$\mu_j,\mu_k,\mu_{jk}>0,\ \forall j,k\in\{1,2,\cdots,n\}; \tag{24}$$

$$v_j^{\mu}=\mu_j+\frac{1}{2}\sum_{k\in C\setminus j}I_{jk}^{\mu},\ \forall j,k\in C; \tag{25}$$

$$\sum_{j\in C}v_j^{\mu}=1; \tag{26}$$

$$\xi_i^{+}\geq 0,\ \xi_i^{-}\geq 0,\ \forall i\in\{1,2,\cdots,m\}, \tag{27}$$

where ξ_i^{+} and ξ_i^{-} are two non-negative slack variables, $\mu_{l_i,j}=\eta_i\mu_j$, $I_{jk(i)}^{\mu}=\eta_i I_{jk}^{\mu}$, $I_{jk}^{\mu}=\mu_{jk}-\mu_j-\mu_k$, $\eta_i\geq 1$ ($\forall j,k\in\{1,2,\cdots,n\},\forall i\in\{1,2,\cdots,m\}$). $U_{\mu}(a_i)$ is the integrated value of product a_i under all attributes based on the extended 2-additive Choquet integral, and $U_S(a_i)$ is the true global value of the overall rating of product a_i. The normalization, monotonicity and nonnegativity of the capacities of attributes are specified by Eqs. (20)–(24), respectively.

The importance of each attribute can be calculated by Eqs. (25)–(26). The constraints of slack variables in Eq. (27) are utilized to limit the slack variables. In fact, Eq. (18) can be satisfied and the feasible solutions can be derived when Eq. (20) is used.

The monotonicity condition of capacities is the main reason to produce the computational complexity. Each attribute in C_i is compared with other attributes in the same subset. Therefore, the monotonicity generates $n(n-1)$ constraints in maximum. Given that $\eta_i \geq 1$, Eq. (21) can be simplified as $\mu_j + \sum_{k \in n \setminus j} I_{jk}^{\mu} \geq 0$, $\forall j,k \in \{1,2,\cdots,n\}$, $j \neq k$. Furthermore, the unknown variables in this nonlinear programming include: the capacities μ_j $(j=1,2,\cdots,n)$ of n attributes, the interaction index I_{jk}^{μ} $(j,k=1,2,\quad,n, j \neq k)$ between attributes, and the scale parameter μ_i (i = 1, 2, …, m). μ_j and I_{jk}^{μ} reflect the preferences of a consumer on different attributes of products, while μ_i describes the relationship between the capacity of an attribute in the whole attribute set and the capacity of this attribute in the attribute subset. Since the scale parameters μ_i (i = 1, 2, …, m) can be deduced by Eq. (20) when μ_j $(j=1,2,\cdots,n)$ and I_{jk}^{μ} are known, Model 1 mainly focuses on deducing the values of μ_j (j = 1, 2, …, n) and I_{jk}^{μ}, $\forall j,k \in \{1,2,\cdots,n\}$, $j \neq k$.

Even though Model 1 is a non-linear programming model, it still can be easily solved using the optimization packages in Lingo or MATLAB. If we have deduced the unknown parameters, including the capacities of attributes and the interaction index in the proposed model, then we can utilize the extended 2-additive Choquet integral with inferred parameters to predict the global values of other products that the consumer will be included in the decision-making process.

2.3. Outline of the proposed consumer preference analysis method

The framework of preference analysis for individual consumer is shown in Figure 3, which includes the following steps:

Step 1. Data collection and processing. Collect all online reviews of an individual consumer on products in the same category, including overall ratings of products and ratings on separate product attributes. During data preprocessing, duplicate comments and inconsistent comments should be deleted. The inconsistent review implies that the overall rating of a product is inconsistent with the possible aggregation value of the ratings on separate attribute given by consumers.

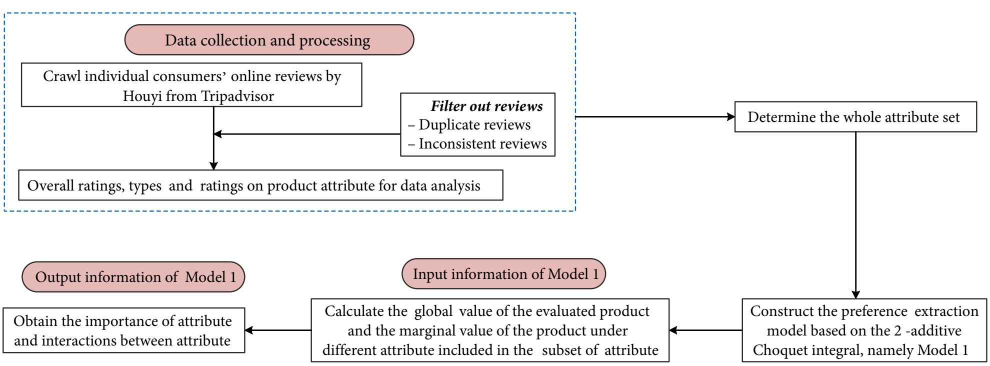


Figure 3. Framework of the proposed consumer preference analysis method

Step 2. Determine the whole attribute set. Identify all attribute types from the collected information about product attribute types in Step 1. The set of all attributes used by a consumer to evaluate products is seen as the attribute set for subsequent preference analysis.

Step 3. Construct the preference model based on the extended 2-additive Choquet integral. Since the type and number of attributes used to evaluate different products may differ, $U_{\mu}(a_i)$ for different products can be obtained according to Eq. (16). Based on the assumption that the global value of a product's overall rating is equal to the integrated value of the marginal values of the product under all separate attributes, Model 1 can be constructed to elicit the individual consumer's preferences.

Step 4. Determine the input information for preference model. Through the normalization function, the global values $U_S(a_i)$ of the overall ratings of products and the marginal values $u_j(a_i)$ of these products under different attributes can be calculated and further served as the input information of the preference model.

Step 5. Derive the individual consumer's preferences, including the importance of diverse attributes and interactions between attributes, by solving Model 1.

3. Numerical example

This section conducts experimental analyses to illustrate the feasibility of the proposed method for consumer preference analysis.

3.1. Data collection

TripAdvisor.com (https://www.tripadvisor.com) is a popular travel review website, bringing together more than 200 million real-world reviews of hotels, restaurants, attractions, airlines, and cruises from global travelers. To demonstrate the feasibility of our proposed model, we crawled the different amounts of experimental data from Tripadvisor.com using a crawling software Houyi crawler (http://www.houyicaiji.com).

Two cases with different data volumes were used for consumer preference analysis. The first case considers the website platform itself as a consumer who uses four product attributes including value (c_1), service (c_2), cleanliness (c_3), location (c_4) as attributes to evaluate different hotels on the platform. The historical online ratings of 73 hotels regarding the overall ratings and the ratings on four attributes are collected. These two types of online ratings are normalized into [0, 1] as the true overall value $U_s(a_i)$ and the marginal values $u_j(a_i)$, $j = 1,2,3,4$, respectively. Parts of the processed data of the first case are shown in Table 2. The data of the second case is 200 online comments on different hotels published by consumers named "Khaven" on the platform. These 200 data also include the overall ratings and the ratings of these hotels under five attributes (service, cleanliness, location, room and food), but the attributes used to evaluate products in this case are different in number and type, unlike in the first case where the attributes are fixed and identical. Further, these ratings are also normalized into [0, 1]. Due to the limited space, the collected data is not presented here.

Table 2. The evaluations of parts of 73 hotels given by the platform itself based on the same four attributes

	Hotel	Overall Score	Attributes			
			value	service	cleanliness	location
Hotel 1	Hotel Bradford Elysees – Astotel	1.0	0.9	1.0	1.0	1.0
Hotel 2	Hotel Ekta	0.9	0.9	0.9	1.0	1.0
Hotel 3	Hotel La Manufacture	0.9	0.9	0.9	0.9	0.9
Hotel 4	Hotel Eiffel Blomet	0.9	0.9	0.9	1.0	0.9
Hotel 5	Secret de Paris – Hotel & Spa	0.9	0.9	1.0	1.0	0.9
Hotel 6	Hotel Rose Bourbon	0.9	0.9	0.9	0.9	0.9
Hotel 7	Hotel des Grands Hommes	0.9	0.9	0.9	1.0	1.0
Hotel 8	Hotel Terminus Lyon	0.8	0.8	0.9	0.9	0.9
Hotel 9	B Montmartre Hotel	0.9	0.9	1.0	1.0	0.9
Hotel 10	Hotel CitizenM Paris Gare de Lyon	0.9	0.9	0.9	0.9	0.9
Hotel 11	Hotel Maxim Folies	0.9	0.9	1.0	0.9	1.0
Hotel 12	Hotel Apollon Montparnasse	0.9	0.9	0.9	0.9	0.9
Hotel 13	Hotel 30.8B – Astotel	1.0	0.9	0.9	1.0	1.0
Hotel 14	Hotel Darcet	0.9	0.9	0.9	1.0	0.9
Hotel 15	Hotel Da Vinci & Spa	1.0	0.9	1.0	1.0	1.0
Hotel 16	Pullman Paris Eiffel Tower Hotel	0.8	0.7	0.8	0.8	0.9
Hotel 17	Hotel L'interlude	0.8	0.8	0.8	0.9	0.8
Hotel 18	Le Relais Saint Charles	0.8	0.8	0.8	0.9	0.9
Hotel 19	Hotel Mademoiselle	0.9	0.9	0.9	0.9	0.9
Hotel 20	Hyatt Regency Paris Etoile	0.8	0.8	0.8	0.9	0.9
…	…	…	…	…	…	…

3.2. Solutions

According to the procedure of consumer preference analysis, we need to substitute the overall values and the marginal values under different attributes of all similar products into Model 1 to determine preference parameters. Once these preference parameters are determined, the extended 2-additive value function can approximate the consumer preference model. Specifically, when the error between the overall value calculated by the inferred model and the overall value directly calculated by normalized in Section 2.2.1 is relatively small, the inferred preference parameters can be regarded as compatible with the actual preferences of consumers.

We substitute all the data of the two cases into Model 1, and get the corresponding results. For the first case, the deduced results of preference parameter are shown in Table 3 and Table 4. Because the platform uses four attributes to evaluate 73 hotels simultaneously, all scale parameters (η_i, i =1...73) are taken as $\eta_i = 1$. At this time, the extended 2-additive Choquet integral is equal to the 2-additive Choquet integral. According to the results in Table 3, the interactions between attributes exist. For example, there is a positive interaction between "service" and "cleanliness" ($I_{23} = 0.8$). That is to say, a hotel with high "service" and high "cleanliness" is much appreciated by the platform, since the joint impact of such a pair of attributes is higher than a simple addition of the two impacts viewed separately. In addition, there are also a small degree of negative interactions among three pairwise attributes of "value" and "service", "value" and "cleanliness", "service" and "location". According to the property of a 2-additive capacity and Eq. (8), the importance of the attributes concerned by the platform can be derived. The results ($v_2^{\mu} = 0.45$, $v_3^{\mu} = 0.45$) show that when the platform ratings these hotels, it pays more attention to two attributes, namely, "service" and "cleanliness". That means that if a hotel wants to get a high ranking or high overall rating on the platform, it is necessary to improve the comprehensive performance in two attributes "service" and "cleanliness".

In the second case, the consumer "Khaven" evaluated 200 hotels with different types and number of attributes, so the scale parameters in Model 1 will take different values to assign the capacity to the subsets of attributes. The preference parameters for the consumer "Khaven" are deduced, as shown in Table 5 and Table 6. According to the experimental results, the consumer"Khaven" does not show more attention to attributes "food" in the comprehensive evaluation process, but she/he may consider the other four attributes. In addition, for the consumer "Khaven", the interactions between attributes "service" and "location", "cleanliness" and "room", "cleanliness" and "food", "location" and "food" are negative, which means that the impact of these pairs of attributes on products' performance are redundant. For example, the consumer "Khaven" may believe that the food and rooms provided by a hotel should be inherently clean. Thus, when considering these pairs of attributes ("cleanliness" and "room", or "cleanliness" and "food") jointly in the evaluation process, she/he will think that the impact of the pair of attributes on the comprehensive performance of this hotel should be lesser than a simple addition of the impacts generated by each of the two attributes separately.

Table 3. The results of capacity of attributes in the first case

Attributes	The capacity of single attributes	The capacity of two attributes
Value (c_1)	$\mu_1 = 0.099$	$\mu_{12} = 0.1, \mu_{13} = 0.1, \mu_{14} = 0.127$
Service (c_2)	$\mu_2 = 0.099$	$\mu_{12} = 0.1, \mu_{23} = 0.909, \mu_{24} = 0.1$
Cleanliness (c_3)	$\mu_3 = 0.001$	$\mu_{13} = 0.1, \mu_{23} = 0.909, \mu_{34} = 0.1$
Location (c_4)	$\mu_4 = 0.001$	$\mu_{14} = 0.127, \mu_{24} = 0.1, \mu_{34} = 0.1$

Table 4. The results of interaction index and importance of attributes in the first case

Attributes	The interaction index between attributes ($I_{jk}, j,k \in 1,...,n, j \neq k$)	The importance of attributes (v_j^{μ})
Value (c_1)	$I_{12} = -0.098, I_{13} = -0.009, I_{14} = 0.018$	$v_1^{\mu} = 0.05$
Service (c_2)	$I_{12} = -0.098, I_{23} = 0.8, I_{24} = -0.009$	$v_2^{\mu} = 0.45$
Cleanliness (c_3)	$I_{13} = -0.009, I_{23} = 0.8, I_{34} = 0.08$	$v_3^{\mu} = 0.45$
Location (c_4)	$I_{14} = 0.018, I_{24} = -0.009, I_{34} = 0.08$	$v_4^{\mu} = 0.05$

Table 5. The capacity of attributes for consumer "Khaven" in the second case

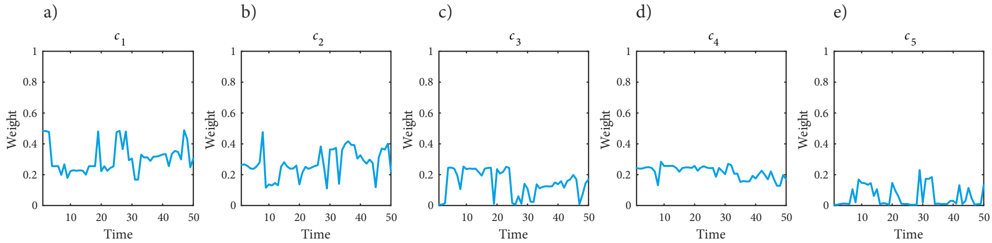
Attributes	The capacity of single attributes	The capacity of two attributes
Service (c_1)	$\mu_1 = 0.01$	$\mu_{12} = 0.325, \mu_{13} = 0.014, \mu_{14} = 0.152, \mu_{15} = 0.155$
Cleanliness (c_2)	$\mu_2 = 0.013$	$\mu_{12} = 0.325, \mu_{23} = 0.133, \mu_{24} = 0.014, \mu_{25} = 0.014$
Location (c_3)	$\mu_3 = 0.01$	$\mu_{13} = 0.014, \mu_{23} = 0.133, \mu_{34} = 0.323, \mu_{35} = 0.011$
Room (c_4)	$\mu_4 = 0.001$	$\mu_{14} = 0.152, \mu_{24} = 0.014, \mu_{34} = 0.323, \mu_{45} = 0.028$
Food (c_5)	$\mu_5 = 0.001$	$\mu_{15} = 0.155, \mu_{25} = 0.014, \mu_{35} = 0.011, \mu_{45} = 0.028$

Table 6. The results of interaction index and importance of attributes in the second case

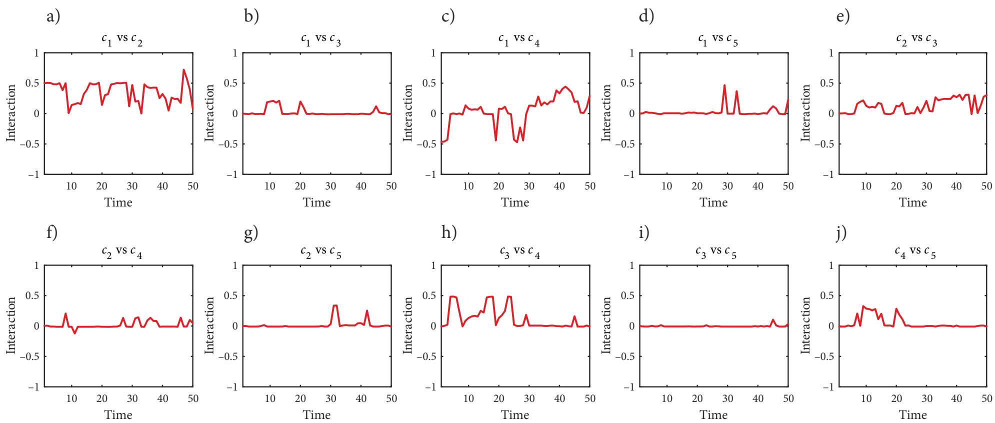
Attributes	The interaction index between attributes ($I_{jk}, j,k \in 1,...,n, j \neq k$)	The importance of attributes (v_j^{μ})
Service (c_1)	$I_{12} = 0.302$, $I_{13} = -0.009$, $I_{14} = 0.129$, $I_{15} = 0.135$	$v_1^{\mu} = 0.289$
Cleanliness (c_2)	$I_{12} = 0.302$, $I_{23} = 0.110$, $I_{24} = -0.012$, $I_{25} = -0.009$	$v_2^{\mu} = 0.209$
Location (c_3)	$I_{13} = -0.009$, $I_{23} = 0.110$, $I_{34} = 0.300$, $I_{35} = -0.009$	$v_3^{\mu} = 0.207$
Room (c_4)	$I_{14} = 0.129$, $I_{24} = -0.012$, $I_{34} = 0.300$, $I_{45} = 0.005$	$v_4^{\mu} = 0.224$
Food (c_5)	$I_{15} = 0.135$, $I_{25} = -0.009$, $I_{35} = -0.009$, $I_{45} = 0.005$	$v_5^{\mu} = 0.071$

3.3. Robustness analysis

We use the data of the second case as the basis to verify the robustness of the proposed preference model. The 100 sets of data are randomly selected 50 times from the 200 sets of data, and then they are input into Model 1 separately. Finally, the change trend of the importance of five attributes and their interaction index results obtained from 50 calculations can be seen in Figure 4 and Figure 5. The results show that the calculation results are relatively stable without large-scale fluctuations. The average values of the importance of five attributes obtained from 50 calculations are 0.305, 0.271, 0.147, 0.219, and 0.058, which are not significantly different from the importance obtained from the 200 sets of data that we calculated in the previous section. Both two types of variances based on the 50 calculations are small. The variance of the importance of five attributes is less than 0.009, while the variance of the eight groups of interaction indexes is less than 0.05. Subsequently, we also count the proportion of positive and negative interaction indices between pairs of attributes shown in Table 7, and find that the positive and negative distributions of attribute interactions are also stable, which is consistent with the results obtained by one-time calculation of 200 sets of data. The results of the robustness analysis show that our method can be used to approximately express the process of integrating ratings for consumers and extract consumer preferences for attributes such as the importance of attributes and interactions among attributes.



Note: $c_1 \rightarrow service$; $c_2 \rightarrow Cleanliness$; $c_3 \rightarrow location$; $c_4 \rightarrow room$; $c_5 \rightarrow food$.
Figure 4. The importance of attributes obtained in robustness analysis



Note: $c_1 \rightarrow service$; $c_2 \rightarrow Cleanliness$; $c_3 \rightarrow location$; $c_4 \rightarrow room$; $c_5 \rightarrow food$.
Figure 5. The interaction index obtained in robustness analysis

Table 7. The proportion of positive and negative interactions obtained from 50 calculations

Type of interaction	Interaction index									
	c_1 vs c_2	c_1 vs c_3	c_1 vs c_4	c_1 vs c_5	c_2 vs c_3	c_2 vs c_4	c_2 vs c_5	c_3 vs c_4	c_3 vs c_5	c_4 vs c_5
Positive	**100%**	32%	**66%**	**60%**	**82%**	28%	34%	**72%**	12%	**54%**
Negative	0%	**68%**	34%	40%	18%	**72%**	**66%**	28%	**88%**	46%

Note: Boldface represents the higher proportion of positive or negative interactions.

3.4. Comparative analysis

3.4.1. Comparisons with the additive function

In this part, we compare the proposed method with the additive value function to judge whether and how different individual consumers consider the interactions between attributes in the process of evaluating products. Considering that some predefined attributes are not used to evaluate a product, it is not possible to directly use the additive value function to analyze consumer preferences. We also introduce the scale parameter that represents the relationship between the whole attributes set and a subset of attributes to the additive value function, so that the additive value function can be used for consumer preference analysis based on online reviews when a product is not evaluated under some attributes.

Two representative additive value functions are utilized here for comparisons.

- The linear value function (LAU). This is to sum the weighted attribute values. Its general form can be expressed as (Beliakov et al., 2007):

$$LAU(a_i) = \sum_{j=1}^{n} w_j u_j(a_i). \tag{28}$$

- The quadratic utility function (QUF). Its main idea is to model the global value of each alternative by comparing with the idea solution which performs the best under all attributes. The mathematical form of the QUF can be expressed as (Aggarwal & Tehrani, 2019):

$$QUF(a_i) = 1 - \sqrt{\sum_{j=1}^{n} w_j^{\,2} (u_j(a_i) - u_j(a^*))^2}, \tag{29}$$

where a^* is an ideal solution. Here, we suppose that a^* is a product with a 5-rating under all given attributes. Thus, we have the marginal value of a^* under each attribute being $u_j(a^*) = 1$, $j = 1, 2, \cdots, n$. When the problem that certain predefined attributes are not used to evaluate a product exists, the scale parameter is introduced to these two additive value functions. The weight w_j of attributes c_j for evaluating product a_i, can be transformed as $w_{j(i)}$, satisfying:

$$w_{j(i)} = \begin{cases} \delta w_j, \text{if } c_j \in C_i \\ 0, \quad \text{otherwise} \end{cases}, \tag{30}$$

where δ is the introduced scale parameter, $\delta \geq 1$. Furthermore, two additive value function above can respectively be transformed as the extended linear value function (ELAU) and the extended quadratic utility function (EQUF):

$$ELAU(a_i) = \sum_{j=1}^{n} w_{j(i)} u_j(a_i); \tag{31}$$

$$EQUF(a_i) = 1 - \sqrt{\sum_{j=1}^{n} w_{j(i)}^{\,2} (u_j(a_i) - u_j(a^*))^2}. \tag{32}$$

According to the assumptions in Section 2.2.2, we establish Model 2 to elicit consumers' preferences. It is worth noting that Model 1 uses the extended 2-additive value function(Subsequently abbreviated as E-2CI) as the preference model to approximate the individual consumers' decision-making process, while in Model 2, we use two types of the additive value function to express their decision-making process separately. The parameters relevant to preferences obtained from Model 1 mainly include the importance of attributes and the possible interaction between attributes, while Model 2 can only infer the importance of attributes.

Model 2. $\min \frac{1}{m}\sum_{i=1}^{m}(\xi_i^+ + \xi_i^-)$

$$\text{s.t. } U_A(a_i) - U_S(a_i) - \xi_i^+ + \xi_i^- = 0,\ \forall a_i \in A; \tag{33}$$

$$\xi_i^+ \ge 0,\ \xi_i^- \ge 0,\ \forall a_i \in \{1,2,\cdots,m\}; \tag{34}$$

$$w_{j(i)} \ge 0,\ \sum_{j=1}^{n} w_{j(i)} = 1,\ \forall i \in \{1,2,\cdots,m\},\ \forall j \in \{1,2,\cdots,n\}; \tag{35}$$

$$w_j \ge 0,\ \sum_{j=1}^{n} w_j = 1,\ \forall i \in \{1,2,\cdots,m\}, \tag{36}$$

where $U_A(a_i)$ is the overall value of a_i aggregated by the additive value function. When the linear value function is considered as the preference model, there is $U_A(a_i) = ELAU(a_i)$. Similarly, when the quadratic utility function is chosen as the preference model, then $U_A(a_i) = EQUF(a_i)$. $U_S(a_i)$ is the overall value of the overall rating of product a_i calculated by normalization.

Next, we conduct consumer preference analysis using the ELAU, EQUF, E-2CI as the preference model respectively according to the procedure in Section 2.3. The online ratings of both two experimental cases are utilized as the experimental data to verify the differences of the aforementioned three value functions in extracting preferences of individual consumers. All the experimental outcomes are shown in Tables 8–10.

Table 8. The results of the first case based on the three extended methods

	The weight of four attributes learned experimentally				The interaction index between attributes $I_{jk}(j \ne k,\ j,k \in 1,\ldots,n)$
	Value	Service	Cleanliness	Location	
ELAU	0.05	0.85	0.05	0.05	–
EQUF	0.01	0.97	0.01	0.01	–
E-2CI	0.402	0.227	0.185	0.054	$I_{12} = -0.098, I_{13} = -0.009, I_{14} = 0.018, I_{23} = 0.8, I_{24} = -0.009, I_{34} = 0.08$

Table 9. The learning results of the consumer "Khaven" in second case based on the three extended methods

	The weight of five attributes learned experimentally					The interaction index between attributes $I_{jk}(j \ne k,\ j,k \in 1,\ldots,n)$
	Service	Cleanliness	Location	Room	Food	
ELAU	0.485	0.01	0.01	0.485	0.01	–
EQUF	0.96	0.01	0.01	0.01	0.01	–
E-2CI	0.289	0.209	0.207	0.224	0.071	$I_{12} = 0.302, I_{13} = -0.009, I_{14} = 0.129, I_{15} = 0.135, I_{23} = 0.110, I_{24} = -0.012, I_{25} = -0.009, I_{34} = 0.300, I_{35} = -0.009, I_{45} = 0.005$

Table 10. The average calculation error of the extended learning models in two experimental cases

	Average error of 73sets of data for the first case	Average error of 200 sets of data for the second case
ELAU	0.025	0.062
EQUF	0.024	0.074
E-2CI	0.018	0.052

After introducing the scale parameter to solve the problem that the product has not been evaluated by some attributes, both the additive value function (the ELAU and EQUF) and the non-additive E-2CI can be used to analyze consumers' preferences for product attributes. All these three value functions with inferred preference parameters can be regarded as the preference model, but from the perspective of the average fitting error, the E-2CI as a preference model is closer to the decision-making process of consumers because the calculated error is smaller. The above experimental results also show that the individual consumer will not only differently weight diverse attributes in a decision-making process, but also consider the possible interactions between attributes. In addition, different consumers' behaviors of giving weights to attributes and considering the interactions between attributes are different. For example, for the platform itself, the attributes "value" will be considered in the evaluation process, while for the consumer "Khaven", other attributes "room" and "food" will be included. Although the preference model we derived by the E-2CI may be only one of the models compatible with consumers' real preferences, it may also illustrate the need to consider the interactions between attributes when conducting consumer preferences analysis based on online reviews.

3.4.2. Comparisons with the existing algorithms for consumer preference analysis

Comparisons between our proposed method and the existing studies that were used for consumer preference analysis based on value functions in the context of online reviews are discussed in the following:

(1) The source of extracting product attribute information. The study of Guo et al. (2020) and Zhu et al. (2022) both extracted product attribute information based on textual comments. In Guo et al. (2020), the structure of attributes, attributes values and the relationship between the importance of attributes were determined by the calculation rules related to the frequency of attributes in textual comments. In Zhu et al. (2022), the attributes and their corresponding marginal values were extracted from textual comments. In this paper, we determine attributes and the marginal values of products under each attribute according to the online ratings related to product attributes.

(2) The value functions selected as the preference model. Guo et al. (2020) and Zhu et al. (2022) used additive value functions to model the decision-making process of consumers, without considering the interactions between attributes. Our proposed approach extracts consumers' preferences regarding the importance of attributes and interactions among them from online reviews based on a non-additive value function.

(3) The way to deal with the problem that attributes are not fully used by consumers to evaluate products. A hierarchical structure was utilized by Guo et al. (2020) to determine attributes, but this method only allowed different products to be evaluated by consumers with different product attributes, which are still summarized into the same fixed attributes. If a situation like that shown in Figure 1 and Figure 2 occurred, the problem that the attributes used to evaluate products are not fixed in number and type cannot be solved. Zhu et al. (2022) regarded this problem as that the evaluation value of a product under a certain attribute is the missing value, and further used an optimization model to extract consumer preferences. Our proposed method provides new ideas to solve this problem by assigning more weights to the attributes used to evaluate the product.

(4) Different perspectives of consumer preference analysis. Guo et al. (2020) used previous online reviews of all consumers to infer the preferences of a new consumer, so as to realize product recommendations. Both the method of Zhu et al. (2022) and the approach we proposed are from the perspective of an individual consumer, inferring his/her preferences from their historical comments respectively. In addition to the importance of the attribute concerned by Zhu et al. (2022), our proposed method also focuses on whether consumers consider the interactions between attributes in the evaluation process.

When considering the interactions between attributes, how will individual consumers' decision-making process be, how much attention will be paid to different attributes, which different attributes will have interactions, and whether these interactions are positive or negative? These issues can be better explained by our proposed consumer preference analysis approach without the participation of consumers.

3.5. Implications

We derive some implications about consumer preference analysis from the perspective of platform merchants or product managers:

(1) Using historical data to infer consumers' preferences enriches the application of value functions in the context of online comments. Previous MADA approaches based on value functions require consumers to evaluate products according to a set of preset attributes and provide relevant parameters, such as the relative importance of attributes. But these requirements are hard to be satisfied in practice. Using value functions as a consumer preference analysis model, and inferring preference parameters based on historical data and the aggregation-disaggregation framework can help to obtain real consumer preferences. It is helpful to reveal the personalized behavior of consumers.

(2) Through introducing scale parameters to model the relations of the capacity of attributes in the whole attribute set and its subsets, the additive or non-additive value function can be extended to aggregate the overall value of a consumer when the product has not been evaluated under some attributes. The application of the scale parameters provides a new perspective for solving consumer preference analysis problems based on MADA approaches.

(3) Individual consumers will make different decision-making behaviors for different products of the same type. The case study results show that, in addition to using different attributes to evaluate products, consumers also pay different attention to diverse attributes and consider the interactions between attributes. For different consumers, these interactions may be inconsistent. For example, for one consumer, there is a negative interaction between “service” and “location”, but for another consumer, the interaction between these two attributes may be positive. According to these experimental results, the platform merchants or product managers can understand which attributes a consumer prefers to and how these attributes affect the individual consumers’ evaluation of products. They can utilize the historical data of consumers who have a relatively large volume of reviews on the platform to extract the comprehensive consumers’ preferences and recommend products to different consumers in line with their preferences, which can further improve the product sales and marketing strategies of the platform.

Conclusions

In this paper, we proposed an individual consumer preferences analysis approach based on a value function considering the interactions between attributes in the context of online reviews. To solve the problem that the attribute relevant to products are not fully used as the attributes to evaluate products, the 2-additive Choquet integral was extended as the basic preference model by introducing the scale parameters. To obtain the preference of individual consumers over the attributes, we developed a nonlinear programming within the aggregation-disaggregation paradigm to learn these parameters related to consumers’ preferences. Numerical examples suggest that, for the individual consumers, the value function considering attribute interactions can describe their evaluation and decision-making behaviors more approximately than the additive value function. This means that consumers will consider the interactions between different product attributes when evaluating products, and these interactions can be positive or negative. For product managers or platform merchants, they can improve their marketing strategies and product recommendations according to the different individual consumers’ preference.

There are still some limitations in this paper. First, this paper only referred to online ratings for determining the information about attributes but did not consider textual reviews which also provide consumers’ evaluation of product attributes. Future studies should combine online ratings and textual comments to obtain more accurate consumer preferences. In addition, this study did not cover a large volume of historical data to support the feasibility of the experiment. The data-driven analytic approach such as Web crawling (e.g., text mining), machine learning and statistical methods should be concerned for consumer preference analysis in the future. Besides, there is a lack of discussion about the changes in consumer preferences over time, which could prevent us from obtaining more accurate consumer preferences. A framework needs to be established for analyzing consumers’ preferences in real-time and obtaining more value functions compatible with the consumer decision-making process.

Funding

The work was supported by the National Natural Science Foundation of China under Grant 71771156, 71971145, 72171158, the Sichuan Planning Project of Social Science under Grant SC22A021, the 2022 Key Project of the Key Research Institute of Humanities and Social Sciences in Sichuan Province under Grant CJZ22-04, and the Chengdu Philosophy and Social Sciences Planning Project under Grant 2022L09.

Author contributions

Authors contributed equally.

Disclosure statement

The authors have no competing financial, professional, or personal interests from other parties that are related to this paper.

References

Aggarwal, M., & Tehrani, A. F. (2019). Modelling human decision behaviour with preference learning. *INFORMS Journal on Computing*, *31*(2), 318–334. https://doi.org/10.1287/ijoc.2018.0823

Arcidiacono, S. G., Corrente, S., & Greco, S. (2021). Robust stochastic sorting with interacting criteria hierarchically structured. *European Journal of Operational Research*, *292*(2), 735–754. https://doi.org/10.1016/j.ejor.2020.11.024

Beliakov, G., Pradera, A., & Calvo, T. (2007). *Aggregation functions: A guide for practitioners*. Springer.

Branke, J., Corrente, S., Greco, S., Słowinski, R., & Zielniewicz, P. (2016). Using Choquet integral as preference model in interactive evolutionary multiobjective optimization. *European Journal of Operational Research*, *250*(3), 884–901. https://doi.org/10.1016/j.ejor.2015.10.027

Chateauneuf, A., & Jaffray, J. Y. (1989). Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, *17*(3), 263–283. https://doi.org/10.1016/0165-4896(89)90056-5

Choquet, G. (1954). Theory of capacities. *Annales de l'Institut Fourier*, *5*, 131–295. https://doi.org/10.5802/aif.53

Chung, J., & Rao, V. R. (2012). A general consumer preference model for experience products: Application to internet recommendation services. *Journal of Marketing Research*, *49*(3), 289–305. https://doi.org/10.1509/jmr.09.0467

Doumpos, M., & Zopounidis, C. (2011). Preference disaggregation and statistical learning for multicriteria decision support: A review. *European Journal of Operational Research*, *209*(3), 203–214. https://doi.org/10.1016/j.ejor.2010.05.029

Dyer, J. S., & Smith, J. E. (2021). Innovations in the science and practice of decision analysis: The role of management science. *Management Science*, *67*(9), 5364–5378. https://doi.org/10.1287/mnsc.2020.3652

Fürnkranz, J., & Hüllermeier, E. (2010). *Preference learning*. Springer. https://doi.org/10.1007/978-3-642-14125-6

Ghaderi, M., Ruiz, F., & Agell, N. (2017). A linear programming approach for learning non-monotonic additive value functions in multiple criteria decision aiding. *European Journal of Operational Research*, *259*(3), 1073–1084. https://doi.org/10.1016/j.ejor.2016.11.038

Govindan, K., & Jepsen, M. B. (2016). ELECTRE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, *250*(1), 1–29. https://doi.org/10.1016/j.ejor.2015.07.019

Grabisch, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, *89*(3), 445–456. https://doi.org/10.1016/0377-2217(95)00176-X

Grabisch, M. (1997). *k*-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, *92*(2), 167–189. https://doi.org/10.1016/S0165-0114(97)00168-1

Grabisch, M., Kojadinovic, I., & Meyer, P. (2008). A review of methods for capacity identification in Choquet integral based multi-attribute utility theory applications of the Kappalab R package. *European Journal of Operational Research*, *186*(2), 766–785. https://doi.org/10.1016/j.ejor.2007.02.025

Greco, S., Matarazzo, B., & Słowiński, R. (2016). Decision rule approach. In S. Greco, M. Ehrgott, & J. R. Figueira (Eds.), *International series in operations research & management science: Vol. 233. Multiple criteria decision analysis: State of the art surveys* (pp. 497–552). Springer-Verlag. https://doi.org/10.1007/978-1-4939-3094-4_13

Grigoroudis, E., Noel, L., Galariotis, E., & Zopounidis, C. (2021). An ordinal regression approach for analyzing consumer preferences in the art market. *European Journal of Operational Research*, *290*(2), 718–733. https://doi.org/10.1016/j.ejor.2020.08.031

Guo, M., Liao, X., & Liu, J. (2019). A progressive sorting approach for multiple criteria decision aiding in the presence of non-monotonic preferences. *Expert Systems with Applications*, *123*, 1–17. https://doi.org/10.1016/j.eswa.2019.01.033

Guo, M. Z., Liao, X. W., Liu, J. P., & Zhang, Q. P. (2020). Consumer preference analysis: A data-driven multiple criteria approach integrating online information. *Omega*, *96*, 102074. https://doi.org/10.1016/j.omega.2019.05.010

Hung, K., Guillet, B. D., & Zhang, H. Q. (2019). Understanding luxury shopping destination preference using conjoint analysis and traditional item-based measurement. *Journal of Travel Research*, *58*(3), 411–426. https://doi.org/10.1177/0047287518760259

Kadziński, M., & Tervonen, T. (2013). Robust multi-criteria ranking with additive value models and holistic pair-wise preference statements. *European Journal of Operational Research*, *228*(1), 169–180. https://doi.org/10.1016/j.ejor.2013.01.022

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. Wiley.

Lee, S. K., Cho, Y. H., & Kim, S. H. (2010). Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations. *Information Sciences*, *180*(11), 2142–2155. https://doi.org/10.1016/j.ins.2010.02.004

Li, X., Liu, H. F., & Zhu, B. (2020). Evolutive preference analysis with online consumer ratings. *Information Science*, *541*, 332–344. https://doi.org/10.1016/j.ins.2020.06.048

Liu, J. P., Liao, X. W., Kadziński, M., & Słowiński, R. (2019). Preference disaggregation within the regularization framework for sorting problems with multiple potentially non-monotonic criteria. *European Journal of Operational Research*, *276*(3), 1071–1089. https://doi.org/10.1016/j.ejor.2019.01.058

Mayag, B., & Bouyssou, D. (2020). Necessary and possible interaction between criteria in a 2-additive Choquet integral model. *European Journal of Operational Research*, *283*(1), 308–320. https://doi.org/10.1016/j.ejor.2019.10.036

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *Lecture notes in computer science: Vol. 4321. The adaptive web* (pp. 325–341). Springer. https://doi.org/10.1007/978-3-540-72079-9_10

Pelegrina, G. D., Duarte, L. T., Grabisch, M., & Romano, J. M. T. (2020). The multilinear model in multicriteria decision making: The case of 2-additive capacities and contributions to parameter identification. *European Journal of Operational Research*, *282*(3), 945–956. https://doi.org/10.1016/j.ejor.2019.10.005

Pereira, M. A., Figueira, J. R., & Marques, R. C. (2020). Using a Choquet integral-based approach for incorporating decision-maker's preference judgments in Data Envelopment Analysis model. *European Journal of Operational Research*, *284*(3), 1016–1030. https://doi.org/10.1016/j.ejor.2020.01.037

Shi, Y., Larson, M., & Hanjalic, A. (2014). Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys* (*CSUR*), *47*(1), 1–45. https://doi.org/10.1145/2556270

Tsai, C. F., Chen, K., Hu, Y. H., & Chen, W. K. (2020). Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tourism Management*, *80*, 104–122. https://doi.org/10.1016/j.tourman.2020.104122

Wu, X. L., & Liao, H. C. (2021). Modeling personalized cognition of customers in online shopping. *Omega*, *104*, 102471. https://doi.org/10.1016/j.omega.2021.102471

Xiang, Z., Du, Q. Z., Ma, Y. F., & Fan, W. G. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Zhang, C. X., Xu, Z. S., Gou, X. J., & Chen, S. X. (2021). An online reviews-driven method for the prioritization of improvements in hotel services. *Tourism Management*, *87*, 104382. https://doi.org/10.1016/j.tourman.2021.104382

Zhu, B., Guo, D. F., & Ren, L. (2022). Consumer preference analysis based on text comments and ratings: A multi-attribute decision-making perspective. *Information* & *Management*, *59*(3), 103626. https://doi.org/10.1016/j.im.2022.103626