Advanced Engineering Informatics 57 (2023) 102089

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

A method for product appearance design evaluation based on heterogeneous data

Han Lai [a,b,c], Zheng Wu [a], Xiaokai Zhang [c], Huchang Liao [a,*], Edmundas Kazimieras Zavadskas [d]

[a] Business School, Sichuan University, Chengdu 610064, China
[b] Chongqing Key Laboratory of Intelligent Perception and BlockChain Technology, Chongqing Technology and Business University, Chongqing 400067, China
[c] School of Artificial Intelligence, Chongqing Technology and Business University, Chongqing 400067, China
[d] Institute of Sustainable Construction, Vilnius Gediminas Technical University, Sauletekio al. 11, Vilnius LT-10223, Lithuania

ARTICLE INFO

Keywords:
Product appearance design evaluation
Probabilistic linguistic term set with interval uncertainty
MARCOS
Eye-tracking data
Electroencephalogram data

ABSTRACT

The eye-tracking and electroencephalogram data, as physiological information, have been viewed as effective supplements to subjective reporting for guiding the product appearance design. In this context, how to combine heterogeneous information is a challenging question. This study proposes different methods to determine subjective and objective weights of criteria regarding the self-reporting, eye-tracking, and electroencephalogram data for the evaluation of product appearance design. We introduce the probabilistic linguistic term set with interval uncertainty (IUPLTS) to represent complex self-reporting data, and develop a method to aggregate IUPLTSs. An algorithm is proposed to fuse physiological data on the data layer and feature layer. To combine the obtained heterogeneous information, we define an objective weighting method that examines the differences in indicator data and the correlation between indicators, and then use a level difference maximization model to fuse subjective and objective weights. To ensure the stability of decision-making results for the problem involving a large number of indicators, we use the Measurement of Alternatives and Ranking according to the Compromise Solution (MARCOS) to rank alternatives. An example regarding the evaluation of automobile appearance design schemes is presented to show the validity and practicality of the proposed method. The prototype support system of the proposed method has been developed and is freely available at https://github.com/BitSecret/DAQQSO.

1. Introduction

The appearance design of a product has an important influence on customers' perceptions and purchasing behaviors. Given the fierce market competition, in addition to the functional characteristics of products, enterprises pay more and more attention to the appearance of their products [6]. Focusing on the aesthetic perceptions of product appearance, scholars used a variety of methods to measure the visual aesthetics of products. Due to the accessibility and intuitive benefits, subjective measurements were often used to measure visual aesthetics [29]. However, it is difficult to get the subjective appearance evaluation information in real time without disturbing the process of appreciation. Additionally, how to express uncertain evaluation information is a thorny problem for evaluators in appearance evaluation [31]. In terms of objective measurements in product appearance evaluation, scholars have explored from the perspective of Kansei engineering combined with physiological indices such as the eye-tracking metrics [6,13] and electroencephalogram (EEG) measurement [4,12]. The eye-tracking and EEG data belong to physiological data that can objectively display an user's attention distribution when observing the appearance of a product and the user's response to different appearance elements. Collecting these data from evaluators relies on devices and instruments with sensors such as the headwear eye tracker and electroencephalograph, which will not affect subjective evaluations. The eye-tracking metrics and EEG data have been viewed as effective supplements to subjective evaluations for guiding the evaluation and optimization of product appearance design [11].

Existing studies have investigated the evaluation of product appearance design considering evaluators' self-reported information, the obtained heterogeneous eye-tracking metrics, and EEG data (see Table A.1 in Appendix A for details). Crisp numbers [37,28] and linguistic terms [19,25,8,9,1] have been used to express evaluators' subjective reports. Since it is hard for evaluators to give exact ratings about the visual aesthetics of products in practice, some studies applied

* Corresponding author.
E-mail address: liaohuchang@163.com (H. Liao).

https://doi.org/10.1016/j.aei.2023.102089
Received 16 December 2022; Received in revised form 10 May 2023; Accepted 30 June 2023
Available online 12 July 2023

membership values [7], triangular fuzzy numbers [21], and continuous interval-valued linguistic term sets [31] to express evaluation information flexibly. However, these information representations cannot depict the evaluation information consisting of interval linguistic terms and interval probabilities. This is the first research gap that motivates this study.

Heterogeneous data fusion can be divided into three categories: data-level fusion, feature-level fusion and decision-level fusion [5]. Data layer fusion has the characteristics of high communication load, high processing complexity and no loss of information and performance, while feature level fusion has the characteristics of medium communication load, medium processing complexity and loss of information and performance [5]. Existing studies on design evaluation integrating eye-tracking and EEG data was mainly based on the data layer [37] and feature layer [14,19,20], separately. There is a lack of study to combine these two layers. This is the second research gap that motivates our work.

To combine the obtained heterogeneous eye-tracking metrics and EEG data with evaluators' self-reported information, it is essential to assign weights to the measured indexes. In existing studies on product appearance design evaluation, subjective weight assignment methods such as the analytic hierarchy process (AHP) [7] and best worst method (BWM) [21,31], and objective weighting methods such as entropy-based methods [1,25] and variance-based methods [2] have been used. However, the application of hybrid weight assignment involving both subjective and objective aspects in product appearance design evaluation has been rarely explored. In addition, existing technologies for product design evaluation mainly applied the BWM [31], TOPSIS [37,1,2,28], VIKOR [25], and WSM [7,14,19] to rank candidate products. However, the BWM method cannot adapt to the decision-making scenario with a large number of criteria and alternatives[1] due to time-consuming pairwise comparisons. The calculation process of TOPSIS, VIKOR, and WSM are simple, but they are easy to be affected by the change in criterion weight distribution, which means that their reliability and stability are limited [30]. Therefore, it is necessary to introduce a method to aggregate evaluators' self-reporting, eye-tracking and EEG data, which can adapt to a large number of criteria and alternatives and yield a stable result. This is the third research gap that motivates our work.

To effectively support evaluators to carry out complex qualitative evaluation, Wu et al. [32] introduced the probabilistic linguistic term set with interval uncertainty (IUPLTS). Compared with probabilistic linguistic term sets for depicting uncertain linguistic evaluation regarding quality improvement [34] and the hesitant fuzzy linguistic term sets for depicting the cognitive hesitation in the evaluation of service providers [30], IUPLTS characterized by interval probability and interval linguistic terms can cope with information from consumer historical data or demand analysis at a finer granularity. In this study, we apply the IUPLTSs to represent evaluators' complex subjective reports on product appearance design. Regarding the second research gap, we introduce an algorithm to fuse eye movement data and EEG data on the data layer and feature layer. As for the third research gap, we note that the method called Measurement of Alternatives and Ranking according to the Compromise Solution (MARCOS) [27] is free from complex calculations and pre-set parameters when solving multi-criteria decision-making (MCDM) problems with a large number of criteria and alternatives. However, to the best of our knowledge, there is no study using the MARCOS method for product appearance design evaluation considering self-reporting, eye-tracking and EEG data simultaneously.

In summary, this study aims to develop an MARCOS-based method to support product appearance design evaluation in which the subjective reports denoted by IUPLTSs and objective measurements in terms of eye tracking and EEG data are integrated. The method can reflect the actual process of product design evaluation, and the decision result is objective and robust. The main contributions of this paper are summarized as follows:

1) We introduce the IUPLTSs to represent evaluators' complex subjective reports on product appearance design. We transfer an IUPLTS to a crisp number to facilitate the comparison of IUPLTSs. Moreover, starting from the idea of reducing the weight of biased data and increasing the weight of public perception data, we give an aggregation method for the qualitative data expressed in IUPLTSs.
2) We propose an algorithm to fuse eye-tracking data and EEG data on the data layer and feature layer, such that the fusion of objective physiological index data considers not only the information of the data itself, but also the feature information contained in the data.
3) We develop an objective weight-calculation method, which examines both the differences of index data and the correlation degrees between indexes simultaneously, and is in line with the joint test of subjective feedback and objective physiological index data. We use a level-difference maximization method to realize subjective and objective weight fusion. Then, we develop an MARCOS-based method to support product appearance design evaluation in which the subjective reports denoted by IUPLTSs and objective eye-tracking and EEG data are integrated concerning subjective and objective weights.

The rest of this paper is organized as follows: Section 2 introduces preliminaries regarding IUPLTSs. Section 3 describes the proposed method in detail. Section 4 illustrates an application example to demonstrate the practicability of the method. Sensitivity analysis and comparison with other methods are also included in this section. Section 5 presents the conclusions.

2. Preliminary

This section introduces preliminaries regarding IUPLTSs to facilitate further presentation.

Let a linguistic term set (LTS) be denoted as $S = \{s_0, s_1, \cdots, s_q\}$, where s_α $(\alpha \in \{0, 1, \cdots, q\})$ represent people's subjective evaluations with $s_\alpha > s_\beta$, if $\alpha > \beta$. q is a positive integer that indicates the number of linguistic terms in the LTS. In this study, without specific justification, we suppose an LTS is predetermined as S={s_0(very poor), s_1(moderately poor), s_2(slightly poor), s_3(slightly good), s_4(moderately good), s_5 (very good)}. Using the linguistic scale function [38] like Eq. (1), we can obtain the corresponding semantic value of a linguistic term:

$$E(s_\alpha) = \frac{\alpha}{q}, \alpha = 0, 1, \cdots, q. \tag{1}$$

For complex decision-making problems, a decision-maker may not be able to give evaluation information in single linguistic terms, but in multiple linguistic terms associated with probabilities like "60 % likely to be moderately good, 30 % likely to be very good, and 10 % be uncertainty". To represent such evaluation information, Pang et al. [24] defined the PLTS on S as $h_S(p) = \{s_\alpha(p_\alpha) | s_\alpha \in S, \alpha = 0, 1, \cdots, q\}$, where $s_\alpha(p_\alpha)$ is the linguistic term s_α associated with the probability p_α, and q is the number of linguistic terms in the PLTS. Wu et al. [32] denoted the ignorance part of $h_S(p)$ as $s_S(p_S)$, and $p_S = 1 - \sum_{\alpha=0}^{q} p_\alpha$. For example, the aforementioned expression can be expressed in a PLTS as $\{s_4(0.6), s_5(0.3), s_S(0.1)\}$.

[1] Combined evaluators' self-reporting, eye-tracking and electroencephalogram data for product appearance design evaluation usually involve a large number of measurement indicators, such as Zhou et al. [39] used 1 qualitative evaluation index, 3 eye movement indexes and 11 electroencephalogram indexes, a total of 15 evaluation indexes to test 4 medical nursing bed design proposals; Tang et al. [29] used 1 qualitative evaluation index, 2 eye movement indexes and 10 electroencephalogram indexes, a total of 13 evaluation indexes to test 4 automotive industry design schemes.

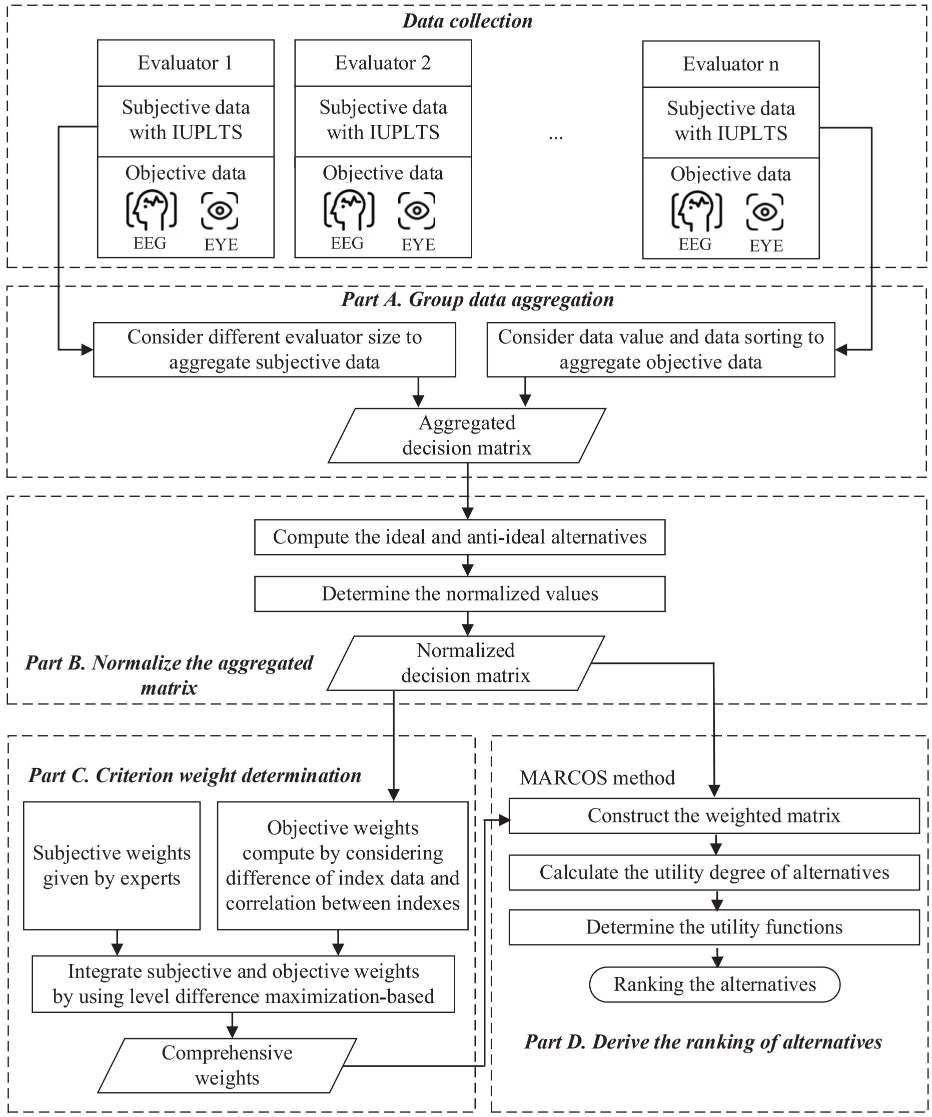

Fig. 1. Framework of the decision-making process.

The score function of a PLTS was defined as [33]:

$$S_{PLTS}(h_S(p)) = \sum_{\alpha=0}^{q} E(s_\alpha)p_\alpha \Big/ \sum_{\alpha=0}^{q} p_\alpha \tag{2}$$

Although the PLTS can express people's subjective evaluation values, in practical application, the evaluation opinions given by decision-makers may be more complex, such as "the probability between moderately good and very good is 70 %–90 %, and the probability of slightly poor is 10 %–20 %". To express such complex evaluation information, Wu et al. [32] defined the IUPLTS as $H_S(p)=\{[s_k^-,s_k^+][p_k^-,p_k^+]\,|k=1,2,\cdots K,\ s_k^-,s_k^+\in S,\ s_k^-\leqslant s_k^+,\ 0\leqslant p_k^-\leqslant p_k^+\leqslant 1\}$, where K is the number of different linguistic intervals in $H_S(p)$ with $S=\{s_0,s_1,\cdots,s_q\}$, and $[p_k^-,p_k^+]$ is the interval probability of linguistic interval $[s_k^-,s_k^+]$. p_k^+ and p_k^- are the upper and lower probability of the linguistic interval $[s_k^-,s_k^+]$, respectively. For example, the aforementioned expression can be expressed in an IUPLTS as $\{[s_2,s_2][0.1,0.2],[s_4,s_5][0.7,0.9]\}$, which can be further abbreviated as $\{s_2[0.1,0.2],[s_4,s_5][0.7,0.9]\}$.

It should be noticed that if $s_k^-=s_k^+$, $\forall k\in\{1,\cdots,K\}$, we obtain a specific form of IUPLTS, $H_S^{'}(p)=\{s_\alpha[p_\alpha^-,p_\alpha^+]|s_\alpha\in S,\alpha=0,1,\cdots,q\}$, which is called the probabilistic linguistic term set with interval probabilities (IPPLTS)[32]. When $p_\alpha^+>0$, we call s_α the focal linguistic term of $H_S^{'}(p)$. The non-focal linguistic terms do not need to be listed. $H_S(p)$ is more flexible than $H_S^{'}(p)$ in describing uncertainty, but its computational complexity is also greatly increased. An IUPLTS$H_S(p)=\{[s_k^-,s_k^+][p_k^-,p_k^+]\}$can be transformed into an IPPLTS $H_S^{'}(p)=\{s_\alpha[p_\alpha^-,p_\alpha^+]|s_\alpha\in S,\alpha=0,1,\cdots,q\}$, where[2] [32]:

$$p_\alpha^- = \sum_{s_\alpha\in[s_k^-,s_k^+] and s_k^-<s_k^+} \frac{p_k^-}{k^+-k^-+1} + \sum_{s_\alpha=s_k} p_k^-, \quad p_\alpha^+ = \sum_{s_\alpha\in[s_k^-,s_k^+] and s_k^-<s_k^+} \frac{p_k^+}{k^+-k^-+1} + \sum_{s_\alpha=s_k} p_k^+, \tag{3}$$

For $\alpha=S$, define $p_\alpha^-=p_S^-$, $p_\alpha^+=p_S^+$. For an IUPLTS containing interval terms and interval probabilities, we can use Eq. (3) to convert it into an IPPLTS. For example, the general form of the above evaluation opinion $\{s_2[0.1,0.2],[s_4,s_5][0.7,0.9]\}$ can be converted to $\{s_2[0.1,0.2],$ $s_4[0.35,0.45],s_5[0.35,0.45]\}$ by applying Eq. (3).

Let t IPPLTSs be $H_S^{'(i)}(p)=\{s_\alpha[p_\alpha^{(i)-},p_\alpha^{(i)+}]|s_\alpha\in S,\alpha=0,1,\cdots,q\}$

[2] In Eq. (3), k^+ and k^- are the subscripts of the linguistic terms s_k^+ and s_k^-.

$(i = 1,2,\cdots,t)$, and their corresponding weight vector be $W = (w_1, w_2, \cdots, w_t)^T$ with $w_i \geqslant 0$ $(i = 1,2,\cdots,t)$, and $\sum_{i=1}^{t} w_i = 1$. The weighted average operator of IPPLTSs was defined as [32]:

$$IPPLTS - WA\left(H_S^{'(1)}, H_S^{'(2)}, \cdots, H_S^{'(t)}\right) = \left\{ s_\alpha \left[\sum_{i=1}^{t} w_i p_\alpha^{(i)-}, \sum_{i=1}^{t} w_i p_\alpha^{(i)+} \right] | s_\alpha \in S, \alpha = 0, 1, \cdots, q \right\} \quad (4)$$

The score of $H_S^{'}(P)$ is an interval value, which can be denoted as $S_{IPPLTS}\left(H_S^{'}(p)\right) = [\phi^-, \phi^+]$ and obtained by Eqs. (5) and (6) [32].

$$\begin{aligned} \text{Min} \quad & \phi^- = \sum_{\alpha=0}^{q} E(s_\alpha) p_\alpha^o \\ s.t. \quad & p_\alpha^o \in \left[p_\alpha^-, p_\alpha^+\right] \\ & \sum_{\alpha=0}^{q} p_\alpha^o = 1 \end{aligned} \quad (5)$$

$$\begin{aligned} \text{Max} \quad & \phi^+ = \sum_{\alpha=0}^{q} E(s_\alpha) p_\alpha^* \\ s.t. \quad & p_\alpha^* \in \left[p_\alpha^-, p_\alpha^+\right] \\ & \sum_{\alpha=0}^{q} p_\alpha^* = 1 \end{aligned} \quad (6)$$

Although Wu et al. [32] proposed that the comparison of two IPPLTSs can be completed by the probability dominance method, it involves complex calculations. In Section 3.1, we give a conversion method based on the single-valued neutrosophic set (SVNS) theory, which converse the interval value to an SVNS, and then use the score function of SVNSs to further obtain the accurate score, so as to simplify the comparison process.

3. Proposed methodology

Let $E = \{e_1, e_2, \cdots, e_T\}$ be a set of evaluators, $A = \{a_1, a_2, \cdots, a_m\}$ be a set of alternatives, $C = \{c_1, c_2, \cdots, c_n\}$ be a set of criteria containing γ qualitative criteria and $n-\gamma$ quantitative criteria, $W = \{w_1, w_2, \cdots, w_n\}$ be a set of criteria weights. An LTS $S = \{s_0, s_1, \cdots, s_q\}$ is presented to evaluators for their self-reporting. The subjective evaluation data is given by the evaluator and expressed in IUPLTSs. The objective physiological eye-tracking and EEG data are collected by eye tracker and EEG device respectively, and are expressed in accurate values. The original data matrix can be expressed as $\widetilde{D}_t = \left[D_t^1, D_t^2\right]$, where t denotes the t-th evaluator for $t = 1, 2, \cdots, T$, $D_t^1 = \begin{bmatrix} h_{11}^t & h_{12}^t & \cdots & h_{1\gamma}^t \\ h_{21}^t & h_{22}^t & \cdots & h_{2\gamma}^t \\ \vdots & \vdots & \ddots & \vdots \\ h_{m1}^t & h_{m2}^t & \cdots & h_{m\gamma}^t \end{bmatrix}$ and $D_t^2 = \begin{bmatrix} x_{1\gamma+1}^t & x_{1\gamma+2}^t & \cdots & x_{1n}^t \\ x_{2\gamma+1}^t & x_{2\gamma+2}^t & \cdots & x_{2n}^t \\ \vdots & \vdots & \ddots & \vdots \\ x_{m\gamma+1}^t & x_{m\gamma+2}^t & \cdots & x_{mn}^t \end{bmatrix}$ denote the subjective evaluation data matrix and objective evaluation data matrix, respectively. $h_{ij}^t = \left\{ \left[s_{tk}^{ij-}, s_{tk}^{ij+}\right] \left[p_{tk}^{ij-}, p_{tk}^{ij+}\right] | k = 1, 2, \cdots K, s_{tk}^{ij-}, s_{tk}^{ij+} \in S, s_{tk}^{ij-} \leqslant s_{tk}^{ij+}, 0 \leqslant p_{tk}^{ij-} \leqslant p_{tk}^{ij+} \leqslant 1 \right\}$ $(i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, \gamma;\ t = 1, 2, \cdots, T)$ represents the evaluation in an IUPLTS of the i-th alternative on the j-th criterion given by the t-th evaluator.

In this section, we propose a method for product appearance design scheme evaluation, which has the following characteristics: (1) It considers the subjective criteria related to the self-reports of evaluators, as well as the objective criteria related to the eye-tracking and EEG data of the evaluators; (2) It uses IUPLTSs to express evaluators' opinions flexibly; (3) The subjective and objective weights of criteria are synthesized by considering the differences and correlations between various criteria simultaneously; (4) As for the prioritization, the MARCOS method [27] is applied to derive the final ranking of alternatives after combining the data in data layer and feature layer. The decision process can be intuitively demonstrated in Fig. 1.

3.1. Subjective data collection

Before aggregating subjective data, the evaluation values expressed in IUPLTSs should be transformed into IPPLTSs by Eq. (3), which unifies the expression form and simplifies the operation. After transformation, the evaluation values can be expressed as $h_{ij}^{'t} = \left\{ s_{t\alpha}^{ij} \left[p_{t\alpha}^{ij-}, p_{t\alpha}^{ij+}\right] \middle| s_{t\alpha}^{ij} \in S, \alpha = 0, 1, \cdots, q \right\}$ $(i = 1,2,\cdots,m;\ j = 1,2,\cdots,\gamma;\ t = 1,2,\cdots,T)$.

The general idea of subjective data aggregation is to reduce the weight of biased data and increase the weight of public cognitive data [17]. Based on this idea, the aggregation function can be given in different forms according to different sample size.[3]

(i) In the case of small evaluator size ($t = 2$), the bias cannot be identified without additional information. At this time, the primary task of aggregation is to avoid the loss of information. Thus, we use the union operation, such that

$$h_{ij}^{'} = g\left(h_{ij}^{'1}, h_{ij}^{'2}\right) = \left\{ s_\alpha^{ij} \left[\min\left(p_{1\alpha}^{ij-}, p_{2\alpha}^{ij-}\right), \max\left(p_{1\alpha}^{ij-}, p_{2\alpha}^{ij-}\right)\right] | s_\alpha^{ij} \in S, \alpha = 0, 1, \cdots, q \right\} \quad (7)$$

(ii) In the case of medium sample size ($3 \leq t \leq 30$), it is easy to distinguish bias information. Assume the evaluators' weight vector is $W_e = (w_1^e, w_2^e, \cdots, w_T^e)^T$. For a few data with great deviation, we can assign the weight to 0, which means the data is discarded. For the evaluators with different knowledge background and experience, different weights are given. Then, based on Eq. (4), we can aggregate evaluators' data by

$$h_{ij}^{'} = g\left(h_{ij}^{'1}, h_{ij}^{'2}, \cdots, h_{ij}^{'T}\right) = IPPLTS - WA\left(h_{ij}^{'1}, h_{ij}^{'2}, \cdots, h_{ij}^{'T}\right) = \left\{ s_\alpha \left[\sum_{t=1}^{T} w_t^e p_{t\alpha}^{ij-}, \sum_{t=1}^{T} w_t^e p_{t\alpha}^{ij+} \right] | s_\alpha \in S, \alpha = 0, 1, \cdots, q \right\} \quad (8)$$

(iii) In the case of large sample size ($t > 30$), most evaluators hold similar opinions, while a small number of evaluators hold opposite extreme opinions. Considering a common case that the factors affecting the cognitive ability of evaluators are random, we use a normal distribution to describe the group evaluations. The probability density function of normal distribution can be used to measure the weights of information from evaluators in the aggregation.

Based on Eqs. (5) and (6), the lowest score $h_{ij}^{'t\min} = \left\{ s_\alpha\left(p_{t\alpha}^{ijo}\right) | s_\alpha \in S, \alpha = 0, 1, \cdots, q \right\} (t = 1, 2, \cdots, T)$ and the highest score $h_{ij}^{'t\max} = \left\{ s_\alpha\left(p_{t\alpha}^{ij*}\right) | s_\alpha \in S, \alpha = 0, 1, \cdots, q \right\}$ can be obtained from $h_{ij}^{'1}, h_{ij}^{'2}, \cdots, h_{ij}^{'T}$, where $p_{t\alpha}^{ijo}$ denotes the probability of linguistic term s_α in $h_{ij}^{t\min}$ and $p_{t\alpha}^{ij*}$ denotes the probability of linguistic term s_α in $h_{ij}^{t\max}$, for $t = 1,2,\cdots,T$. For $h_{ij}^{1\min}, h_{ij}^{2\min}, \cdots, h_{ij}^{T\min}$, Eqs. (9)–(14) are used to aggregate to get the lowest

[3] In this study, the critical size is given with reference to existing literature [17]. The main principle is whether it is easy to distinguish biased and public evaluation data. Then, different aggregation formulas suitable for specific objectives can be given. t = 2 and t = 30 are compatible with the group evaluation based on interval-valued data. For different data forms and practical backgrounds, the critical values can be identified through simulation analysis or empirical research that can examine the distribution of group opinions.

score $h_{ij}^{'\min}$.

$$p_{\alpha}^{ij^{o}} = \sum_{t=1}^{T} p_{t\alpha}^{ij^{o}} \tag{9}$$

$$\mu_{ij}^{\min} = \frac{\sum_{\alpha=0}^{q} \alpha p_{\alpha}^{ij^{o}}}{\sum_{\alpha=0}^{q} p_{\alpha}^{ij^{o}}} \tag{10}$$

$$\left(\sigma_{ij}^{\min}\right)^2 = \frac{\sum_{\alpha=0}^{q} \left(\alpha - \mu_{ij}^{\min}\right)^2 p_{\alpha}^{ij^{o}}}{\sum_{\alpha=0}^{q} p_{\alpha}^{ij^{o}} - 1} \tag{11}$$

$$\varphi_{ij}^{\min}(\alpha) = \frac{1}{\sigma_{ij}^{\min}\sqrt{2\pi}} e^{-\left(\alpha-\mu_{ij}^{\min}\right)^2 / 2\left(\sigma_{ij}^{\min}\right)^2} \tag{12}$$

$$w_{\alpha}^{\min} = \varphi_{ij}^{\min}(\alpha) \Big/ \sum_{\alpha=0}^{q} \varphi_{ij}^{\min}(\alpha) \tag{13}$$

$$h_{ij}^{'\min} = \left\{ s_{\alpha}^{ij}\left(w_{\alpha}^{\min}\right) | \alpha = 0, 1, \cdots, q \right\} \tag{14}$$

where $p_{\alpha}^{ij^{o}}$ denotes the frequency of occurrence of the linguistic term s_α in the lowest score when considering the group evaluation. Then, $\sum_{\alpha=0}^{q} p_{\alpha}^{ij^{o}}$ is the sample size to examine the value of α. $\mu_{ij}^{\min}$ denotes the sample mean. $\left(\sigma_{ij}^{\min}\right)^2$ denotes the sample variance, which is an unbiased estimate and takes $\sum_{\alpha=0}^{q} p_{\alpha}^{ij^{o}} - 1$ as the degree of freedom. $\varphi_{ij}^{\min}(\alpha)$ denotes the probability density of each linguistic term s_α in $h_{ij}^{'\min}$. Since not all continuous values of α constitute the domain of discourse of LTSs and only discrete integer values are examined, a normalization formula Eq. (13) is necessary. Finally, $w_{\alpha}^{\min}$ denotes the probability of S_α in the final aggregated result.

Similarly, we can get the highest score $h_{ij}^{'\max} = \left\{ s_{\alpha}^{ij}\left(w_{\alpha}^{\max}\right) | \alpha = 0, 1, \cdots q \right\}$. Then, according to Eq. (2), the aggregated results $h_{ij}^{\min}$ and $h_{ij}^{\max}$ are transformed into interval values. An application example is described in Appendix B.

$$\overline{u}_{ij} = \left[S_{PLTS}\left(h_{ij}^{'\min}\right), \quad S_{PLTS}\left(h_{ij}^{'\max}\right) \right] = \left[\phi_{ij}^{-}, \phi_{ij}^{+}\right] \tag{15}$$

By the above aggregation method, we can obtain the aggregated qualitative evaluation matrix $\overline{D}^1 = \begin{bmatrix} \overline{u}_{11} & \overline{u}_{12} & \cdots & \overline{u}_{1\gamma} \\ \overline{u}_{21} & \overline{u}_{22} & \cdots & \overline{u}_{2\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{u}_{m1} & \overline{u}_{m2} & \cdots & \overline{u}_{m\gamma} \end{bmatrix}$, which is expressed in the form of interval values.

Next, we define a conversion rule to further convert interval numbers into precise numbers for subsequent comparison. The idea is to convert the interval values into SVNSs, and then use the score function of SVNS to get precise values. The calculation steps are described below.

For an interval number $\overline{u}_{ij} = \left[u_{ij}^{-}, u_{ij}^{+}\right]$ $(0 \leqslant u_{ij}^{-} \leqslant u_{ij}^{+} \leqslant 1)$, we can correspond it with the perfect value "1", and regard the upper and lower bounds of the interval number as the degree of perfection. Based on this idea, we can define $\theta_{ij} = u_{ij}^{-}$, which means the degree of "being sure to be perfect"; $\eta_{ij} = u_{ij}^{+} - u_{ij}^{-}$, which means the degree of "not sure if it is perfect"; $\psi_{ij} = 1 - u_{ij}^{+}$, which means the degree of "certainty cannot reach perfection". Therefore, we can get a single valued neutrosophic number (SVNN) $\left(\theta_{ij}, \eta_{ij}, \psi_{ij}\right)$. Furthermore, through the score function of SVNNs [26], i.e., Eq. (16), we can obtain the precise value of $\overline{u}_{ij}$ and form a subjective evaluation score matrix $D^1 = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1\gamma} \\ u_{21} & u_{22} & \cdots & u_{2\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{m\gamma} \end{bmatrix}$.

$$u_{ij} = \frac{3 + \theta_{ij} - 2\eta_{ij} - \psi_{ij}}{4} \tag{16}$$

Using the above conversion rules, the conversion from IUPLTSs to crisp numbers can be realized, which can be used for subsequent comparison.

3.2. Objective data collection

The aggregation function in the double normalization-based multi-aggregation (DNMA) method [16] takes into account the coordinated utility value and rank of each scheme, thus improving the reliability of the ranking results. Inspired by this method, we aggregate the objective eye-tracking and EEG data from evaluators in data layer and feature layer. The calculation steps are described as follows:

Firstly, we normalize each evaluator's original physiological data to eliminate dimensional effects:

$$u_{ij}^{t} = \begin{cases} \dfrac{x_{ij}^{t}}{\max\limits_{i} x_{ij}^{t}}, \text{ for benifit criterion } c_j \\ \dfrac{\min\limits_{i} x_{ij}^{t}}{x_{ij}^{t}}, \text{ for cost criterion } c_j \end{cases} \quad \text{for } i = 1, 2, \cdots, m; j = \gamma + 1, \gamma + 2, \cdots, n; t = 1, 2, \cdots, T \tag{17}$$

According to the descending order of u_{ij}^{t}, the rank r_{ij}^{t} of each alternative corresponding to each evaluator through physiological reactions is assigned.[4] Then, we aggregate the objective data and corresponding ranks by

$$u_{ij} = \sum_{t=1}^{T} w_{t}^{e} \sqrt{\varphi\left(u_{ij}^{t}\right)^2 + (1-\varphi)\left(\left(m - r_{ij}^{t} + 1\right)/m\right)^2}, \quad \text{for } i = 1, 2, \cdots, m; \quad j = \gamma + 1, \gamma + 2, \cdots, n \tag{18}$$

where w_{t}^{e} is the weight of each evaluator, and φ is the weight of the original values compared with ranks.

In this way, the aggregation matrix $D^2 = \begin{bmatrix} u_{1\gamma+1} & u_{1\gamma+2} & \cdots & u_{1n} \\ u_{2\gamma+1} & u_{2\gamma+2} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m\gamma+1} & u_{m\gamma+2} & \cdots & u_{mn} \end{bmatrix}$ is constructed, and the comprehensive decision matrix incorporating subjective and objective criteria is $D = \left[D^1, D^2\right]$.

3.3. Normalize the decision matrix

From the decision matrix D, we can obtain the ideal alternative with the best characteristic $[\mho_1, \mho_2, \cdots, \mho_n]$ and the anti-ideal alternative with the worst characteristic $[\Omega_1, \Omega_2, \cdots, \Omega_n]$, where

$$\mho_j = \begin{cases} \max\limits_{i} u_{ij}, j \in B \wedge j \in Su \\ \min\limits_{i} u_{ij}, j \in C \wedge j \in Su \\ \max\limits_{i} u_{ij}, j \in Ob \end{cases}, \quad \Omega_j = \begin{cases} \min\limits_{i} u_{ij}, j \in B \wedge j \in Su \\ \max\limits_{i} u_{ij}, j \in C \wedge j \in Su \\ \min\limits_{i} u_{ij}, j \in Ob \end{cases} \tag{19}$$

where B denotes the benefit criteria, while C denotes the cost criteria. Su denotes the subjective criteria, while Ob denotes the objective type criteria.[5]

Then, we can obtain the normalized decision matrix $D' = \left[u_{ij}'\right]_{m \times n}$ where

[4] In this study, we use Besson's mean ranks [35] to sort alternatives. The rule is that if an object a_i ranks at the ρ-th position, then its rank is ρ; if both a_i and $a_{i'}$ $(i, i' \in (1, 2, \cdots, m)$ and $i \neq i')$ rank at the ρ-th position, then the ranking levels of a_i and $a_{i'}$ are the same, which can be calculated as: $(\rho + (\rho + 1))/2 = \rho + 0.5$. For example, for a set of data $\{2.1, 3.2, 3.2, 1.1\}$, the orders of them are $\{3, 1.5, 1.5, 4\}$.

[5] Since the data of objective criteria have completed the type conversion in matrix D^2, they are all treated as benefit type data.

Table 2
Results of qualitative data aggregation.

Scheme	Qualitative data aggregation results			
	PLTS with interval probabilities	Minimum PLTS and Maximum PLTS	Interval number	Precise value
A_1	$\{s_1[0.1305,0.1355], s_2[0.4000,0.4175], s_3[0.2545,0.2760], s_4[0.1905,0.2005]\}$	Min: $\{s_1(0.1355), s_2(0.4175), s_3(0.2565), s_4(0.1905)\}$ Max: $\{s_1(0.1305), s_2(0.4), s_3(0.269), s_4(0.2005)\}$	[0.5004,0.5079]	0.75
A_2	$\{s_3[0.0105,0.0155], s_4[0.4740,0.4850], s_5[0.4895,0.5055]\}$	Min: $\{s_3(0.0155), s_4(0.485), s_5(0.4995)\}$ Max: $\{s_3(0.0105), s_4(0.484), s_5(0.5055)\}$	[0.8968,0.899]	0.95
A_3	$\{s_1[0.0645,0.072], s_2[0.143,0.158], s_3[0.391,0.3985], s_4[0.316,0.3305], s_5(0.0635)\}$	Min: $\{s_1(0.072), s_2(0.1575)], s_3(0.391), s_4(0.316), s_5(0.0635)\}$ Max: $\{s_1(0.0645), s_2(0.143)], s_3(0.3985), s_4(0.3305), s_5(0.0635)\}$	[0.6283,0.6371]	0.81
A_4	$\{s_2[0.0315,0.0335], s_3[0.149,0.154], s_4[0.5935,0.6065], s_5[0.2175,0.22]\}$	Min: $\{s_2(0.0335), s_3(0.154), s_4(0.595), s_5(0.2175)\}$ Max: $\{s_2(0.0315), s_3(0.149), s_4(0.5995), s_5(0.22)\}$	[0.7993,0.8016]	0.90

Table 3
Subjective, objective and comprehensive weights of criteria.

Indicator	Subjective weight w_j^s	Objective weight			Comprehensive weight w_j
		G_j	R_j	w_j^o	
C_1	0.0769	0.7480	12.2982	0.0798	0.0769
C_2	0.0769	0.7280	12.3868	0.0778	0.0769
C_3	0.0769	0.7348	12.3151	0.0785	0.0769
C_4	0.0769	0.7183	11.5352	0.0766	0.0769
C_5	0.0769	0.7265	12.1925	0.0776	0.0770
C_6	0.0769	0.7341	10.0573	0.0775	0.0769
C_7	0.0769	0.7212	12.2967	0.0771	0.0771
C_8	0.0769	0.6962	12.2460	0.0746	0.0769
C_9	0.0769	0.7033	11.3007	0.0750	0.0769
C_{10}	0.0769	0.7196	12.0984	0.0769	0.0769
C_{11}	0.0769	0.7216	11.4893	0.0769	0.0769
C_{12}	0.0769	0.7018	12.3835	0.0752	0.0769
C_{13}	0.0769	0.7152	12.2681	0.0765	0.0769

$$u'_{ij} = \begin{cases} \dfrac{u_{ij}}{\mho_j}, j \in B \wedge j \in Su \\ \dfrac{\mho_j}{u_{ij}}, j \in C \wedge j \in Su \\ \dfrac{u_{ij}}{\mho_j}, j \in Ob \end{cases} \tag{20}$$

3.4. Criterion weight determination

The determination of criterion weights is an important part of MCDM. To simplify the presentation, this study supposes that the decision-maker gives subjective weight vector $W^s = (w_1^s, w_2^s, \cdots, w_n^s)^T$ to criteria. This section first presents an objective weight calculation method which not only considers the differences of values on each criterion, but also considers the correlation between criteria. Then, we adopt a level difference maximization method to determine the final weights of criteria considering the subjective and objective weights of criteria.

We use the Gini index [18] that describes the variations in the assessments of multiple alternatives over a specific criterion. The Gini index measures the purity/reliability level of each criterion. The larger the Gini index of a criterion is, the higher the reliability of the criterion is and thus a higher weight should be assigned to the criterion. It should be noted that our aim is to use subjective evaluations and objective physiological indicators for joint testing. For the horizontal correlations of criteria, the higher the correlation between one criterion and other criteria, the higher the consistency of the data on that criterion is, and thus the criterion should assign a higher weight. In addition, the correlation coefficient may have negative values in the calculation process. The absolute value of the correlation coefficient is taken to simplify the numerical calculation.

Firstly, we adopt the linear normalization to normalize the decision matrix D':

$$\xi_{ij} = \frac{u'_{ij}}{\sum_{i=1}^{m} u'_{ij}}, \quad \text{for} \quad i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n \tag{21}$$

Compute the Gini index G_j for each criterion, which represents the purity/reliability of the criterion:

$$G_j = 1 - \sum_{i=1}^{m} \xi_{ij}^2, \quad \text{for} \quad i = 1, 2, \cdots, m; j = 1, 2, \cdots, n, \tag{22}$$

The correlation coefficient between the j-th and k-th criteria is determined by

$$\varsigma_{jk} = \frac{\sum_{i=1}^{m} \left(u'_{ij} - \overline{u'_j}\right)\left(u'_{ik} - \overline{u'_k}\right)}{\sqrt{\sum_{i=1}^{m} \left(u'_{ij} - \overline{u'_j}\right)^2 \sum_{i=1}^{m} \left(u'_{ik} - \overline{u'_k}\right)^2}} \tag{23}$$

where $\overline{u'_j}$ is the mean value of the j-th criterion, such that

$$\overline{u'_j} = \sum_{i=1}^{m} u'_{ij} \Big/ m, j = 1, 2, \cdots, n. \tag{24}$$

Assume that the objective weight vector of criteria is $W^o = (w_1^o, w_2^o, \cdots, w_n^o)^T$. Since the harmonic average method [35] can reduce the influence of outliers and emphasize the consistency between Gini index and correlation coefficient when aggregating the two parts of information, we use this method to calculate the comprehensive information of each criterion and normalize it to get the objective weight:

$$w_j^o = \frac{2G_jR_j/(G_j + R_j)}{\sum_{j=1}^{n} 2G_jR_j/(G_j + R_j)}, \quad \text{where} \quad R_j = \sum_{k=1}^{n} |\varsigma_{jk}|, \quad \text{for} \quad j = 1, 2, \cdots, n \tag{25}$$

Since the level difference maximization method [15] can not only consider the subjective and objective weights, but also has a good interpretability, we use this method to determine the final weights of criteria. Thus, the final weights of criteria considering the subjective and objective weights of criteria are obtained by solving the following optimization model:

$$\max \quad s^2 = \frac{1}{m-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(w_j \left(u'_{ij} - \overline{u'_j} \right) \right)^2 \tag{26}$$

Table 4
The utility values of alternatives and their ranks.

Alternatives	Φ_i	κ_i^+	κ_i^-	$f(\kappa_i^-)$	$f(\kappa_i^+)$	$f(\kappa_i)$	Rank
A_1	0.453	1.078	0.456	0.297	0.703	0.405	4
A_2	1.000	2.382	1.008	0.297	0.703	0.895	1
A_3	0.548	1.306	0.553	0.297	0.703	0.491	3
A_4	0.804	1.914	0.810	0.297	0.703	0.720	2

Table 5
The final utility values with respect to different values of φ.

Alternatives	$\varphi = 0$	$\varphi = 0.1$	$\varphi = 0.2$	$\varphi = 0.3$	$\varphi = 0.4$	$\varphi = 0.5$	$\varphi = 0.6$	$\varphi = 0.7$	$\varphi = 0.8$	$\varphi = 0.9$	$\varphi = 1$
A_1	0.3428	0.3596	0.3735	0.3854	0.3959	0.4053	0.4138	0.4214	0.4284	0.4347	0.4405
A_2	0.9372	0.9270	0.9180	0.9099	0.9024	0.8954	0.8889	0.8827	0.8769	0.8715	0.8664
A_3	0.4504	0.4600	0.4689	0.4769	0.4842	0.4910	0.4972	0.5030	0.5083	0.5133	0.5179
A_4	0.7107	0.7122	0.7139	0.7158	0.7177	0.7196	0.7216	0.7235	0.7255	0.7274	0.7293

$$s.t.\quad \begin{cases} \sum_{j=1}^{n} w_j = 1 \\ \min\left(w_j^s, w_j^o\right) \leqslant w_j \leqslant \max\left(w_j^s, w_j^o\right) \end{cases}$$

3.5. Derive the ranking of alternatives

The MARCOS method first proposed by Stević et al. [27] can improve the robustness of MCDM. The core idea of the method is to determine the value function of alternatives by defining the relationships between alternatives and reference objects, and to achieve a compromise ranking associated with the ideal and anti-ideal alternatives. The MARCOS method is free from complex calculations and pre-set parameters. In view of these merits, we apply the MARCOS method to aggregate the data in the product appearance design evaluation problem.

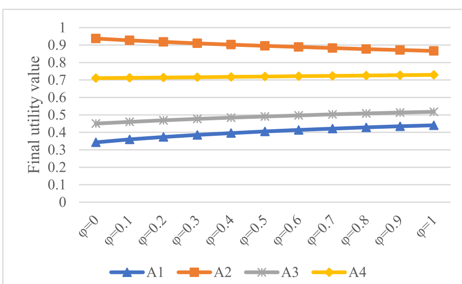

Fig. 2. Influence of change of φ on change of scheme utility value.

We have obtained the normalized decision matrix D', ideal alternative and anti-ideal alternative. Then, the weighted average of each alternative, and those of the ideal alternative and anti-ideal alternative are respectively calculated by

$$\Phi_i = \sum_{j=1}^{n} w_j u_{ij}^{'}, i = 1, 2, \cdots, m. \tag{27}$$

$$\Phi_{ai} = \sum_{j=1}^{n} w_j \mho_j \tag{28}$$

$$\Phi_{aai} = \sum_{j=1}^{n} w_j \Omega_j \tag{29}$$

Then, for each value, we have $\kappa_i^- = \frac{\Phi_i}{\Phi_{aai}}$, $\kappa_i^+ = \frac{\Phi_i}{\Phi_{ai}}$. Thus, the final feature value of each alternative can be calculated by

$$f(\kappa_i) = \frac{\kappa_i^+ + \kappa_i^-}{1 + \frac{1 - f\left(\kappa_i^+\right)}{f\left(\kappa_i^+\right)} + \frac{1 - f\left(\kappa_i^-\right)}{f\left(\kappa_i^-\right)}} \tag{30}$$

where $f\left(\kappa_i^+\right) = \frac{\kappa_i^-}{\kappa_i^+ + \kappa_i^-}$, $f\left(\kappa_i^-\right) = \frac{\kappa_i^+}{\kappa_i^+ + \kappa_i^-}$. Alternatives can be ranked in the

Table 6
The advantages of our method compared with existing methods.

	1. Consider subjective and objective physiological indexes	2. Ability to express and process fuzzy and uncertain information	3. Consider sample size	4. Aggerate physiological indicators from data and feature layers	5. Integrate subjective and objective weights	6. Consider difference and correlation of indicator data for objective weight	7. Suit to MCDM under a number of indexes
Our method	√	H	√	√	√	√	√
Hsiao and Ko [7]	×	L	×	×	×	×	×
Li, Wang, et al. [14]	×	L	×	×	×	×	×
Yang, Chen, et al. [1]	×	H	×	×	×	×	×
Lou et al. [19]	√	M	×	×	×	×	×
Qi et al. [25]	×	M	×	×	×	×	×
Maghsoodi et al. [21]	×	M	×	×	×	×	×
Lou et al. [20]	√	M	×	×	×	×	×
Jing et al. [8]	×	M	×	×	×	×	×
Zhou et al. [37]	√	L	×	×	×	×	×
Tang et al. [28]	√	L	×	×	×	×	×
Chen et al. [1]	×	M	×	×	×	×	×
Deng and Wang [2]	×	L	×	×	×	×	×
Wu and Liao [31]	×	H	×	×	×	×	×
Jing et al. [9]	×	H	×	×	×	×	×

Note: L-Low, M–Medium, H-High, √-support, ×-not support.

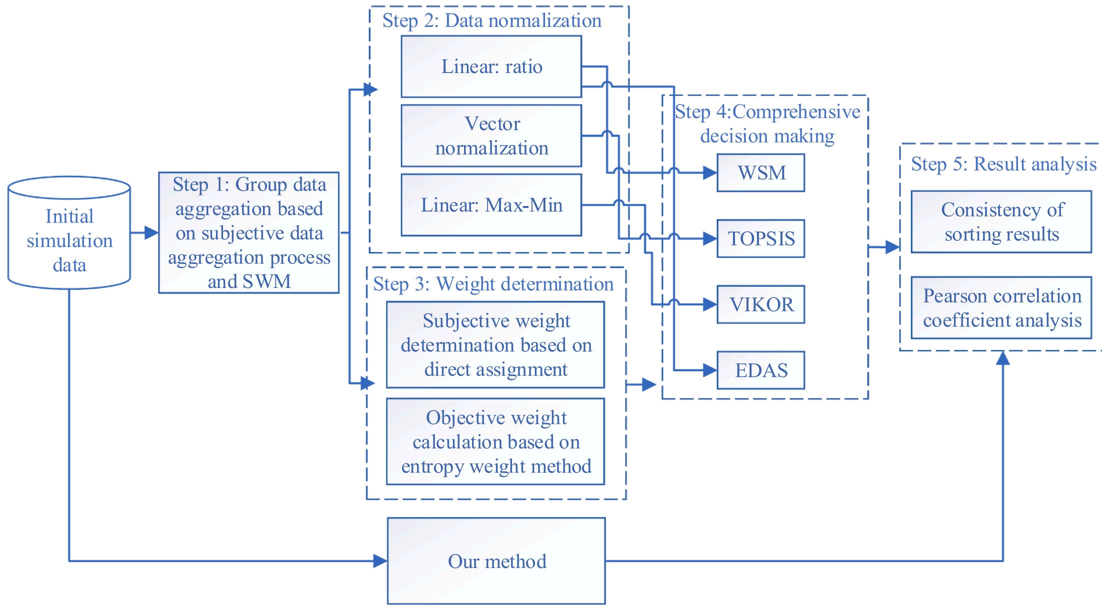

Fig. 3. The implement framework of empirical comparisons.

Table 7
Ranking of the alternatives by using different MCDM methods.

Methods	Utility Value and Rank	Calculation results			
		A_1	A_2	A_3	A_4
SW-WSM	Utility value	0.519	1.000	0.607	0.847
	Rank	4	1	3	2
OW-WSM	Utility value	0.430	1.000	0.484	0.819
	Rank	4	1	3	2
SW-TOPSIS	Utility value	0.089	1.000	0.249	0.682
	Rank	4	1	3	2
OW-TOPSIS	Utility value	0.086	1.000	0.140	0.660
	Rank	4	1	3	2
SW-VIKOR	Utility value	1.000	0.000	0.896	0.460
	Rank	4	1	3	2
OW-VIKOR	Utility value	1.000	0.000	0.977	0.372
	Rank	4	1	3	2
SW-EDAS	Utility value	0.000	1.000	0.173	0.703
	Rank	4	1	3	2
OW-EDAS	Utility value	0.000	1.000	0.087	0.711
	Rank	4	1	3	2
Our method	Utility value	0.405	0.895	0.491	0.720
	Rank	4	1	3	2

Note: SW-subjective weight; OW-objective weight.

descending order of their feature values.

4. Application example: Evaluation of automobile appearance design scheme

In this section, an example of appearance design scheme evaluation in the automotive industry is illustrated and solved by the proposed method.

Automobile appearance design is an important part of automobile research and development. It has become an important factor for the survival, profitability and development of relevant enterprises in automobile industry. Tang et al. [28] applied the Trinity research framework of self-report, eye movement and EEG to the field of automobile appearance design evaluation. They obtained subjective evaluation data by subjects scoring, collected eye movement data by Dikablis glasses 2.0 headwear eye tracker, and obtained EEG data by Brain Amp64 Electroencephalograph. The eye movement data can reveal which part of a product captures the attention of evaluators when observing the appearance of products. Thus, it can help designers understand how users perceive products, and how to adjust the design to improve the usability and attractiveness of products. The EEG data can reveal how evaluators respond to different appearance elements, such as color and shape, and reveal how to optimize product design to improve the visual aesthetics of products.

In our example, Monte Carlo simulation method is used to generate simulated data of 20 samples of four schemes $\{A_1, A_2, A_3, A_4\}$. The original data obtained by simulation are shown in Table A.2. The evaluation involves 12 indicators. C_1 denotes self-report data, which is represented by the IUPLTSs and can be collected by questionnaire. The used LTS is $S = \{s_0$(very poor), s_1(moderately poor), s_2(slightly poor), s_3(slightly good), s_4(moderately good), s_5(very good)$\}$. C_2 and C_3 represent the average fixation time (seconds) and the number of fixations (times), respectively. The average fixation time means the time when participants see the effect drawing of automobile appearance design scheme until they make decisions [37]. The number of fixations means that the times of gaze within the time before participants made the decision after watching the effective drawing of automobile appearance design scheme [37]. C_4-C_{13} are the electrode data of EEG, which can be collected from the brainwave topography. The indicators C_4-C_{13} focus the eleven electrodes used for the sense of beauty [22], including F3, F4, F7, F8, FZ, FC1, FC2, FC5, FC6, FT9, FT10. The data represents the "amplitude" of the corresponding electrode with unit "μV". All indicators belong to benefit type. Lingo 17.0 is used to solve all objective functions in this study.

4.1. Decision-making process

Since there are IUPLTSs in qualitative evaluation data (see the second column of Table A.2), we need to transform these data into IPPLTSs based on Eq. (3). The subjective evaluation data regarding 20 samples are aggregated by Eq. (8). When aggregating subjective evaluation data, we claim that each evaluator has the same position, that is, each evaluator is assigned to the same weight $w_t^e = 0.05(t=1,2,\cdots,20)$. Then, we can use Eqs. (5) and (6) to obtain the data in the form of PLTSs. Next, the data can be transformed into intervals by Eq. (2). We can further translate the interval values into SVNSs by the conversion rule in Section 3.1 and get the precise values based on Eq. (16). The intermediate process of qualitative data aggregation and transformation is shown in Table 2. For quantitative data, since all indicators are benefit types, the

Table 8
Pearson correlation detection of output results from different MCDM methods.

	M1	M2	M3	M4	M5	M6	M7	M8	Our method
M1	1	0.996	1	0.993	−0.989	−0.989	1	0.995	0.999
M2	0.996	1	0.996	0.998	−0.989	−0.997	0.997	1	0.995
M3	1	0.996	1	0.994	−0.993	−0.99	0.999	0.994	1
M4	0.993	0.998	0.994	1	−0.994	−0.999	0.993	0.997	0.994
M5	−0.989	−0.989	−0.993	−0.994	1	0.991	−0.986	−0.984	−0.994
M6	−0.989	−0.997	−0.99	−0.999	0.991	1	−0.99	−0.996	−0.99
M7	1	0.997	0.999	0.993	−0.986	−0.99	1	0.997	0.998
M8	0.995	1	0.994	0.997	−0.984	−0.996	0.997	1	0.993
Our method	0.999	0.995	1	0.994	−0.994	−0.99	0.998	0.993	1

Table A1
Representative studies on the different decision-making methods for product design evaluation.

Ref.	Subjective Information representation	Objective Information representation	Weight determination	Criteria correlation	Data aggregation	Evaluation object
Hsiao and Ko [7]	Membership value	×	(S)AHP	×	WSM	Bicycle design
Li, Wang [14]	/	Eye-tracking and EEG data > Feature value	(S)Direct rating	√	WSM	Women's shirts
Yang, Chen, et al. [1]	IVQROFSs	×	(S)Direct rating	×	WSM	SmartWatch appearance
Lou et al. [19]	SLTs → Cloud model	EEG data → Feature value	(S)Direct rating	×	WSM	Elevator
Qi et al. [25]	SLTs → Rough number	Crisp value → Rough number	(O)Entropy	×	VIKOR	Emergency cutting off device
Maghsoodi et al. [21]	TFNs	×	(S)BWM	×	HBWFAD method	Loudspeaker prototype
Lou et al. [20]	SLTs → ILNs → Trapezium Cloud model	EEG data → Feature value	×	√	Choquet integral	Elevator
Jing et al. [8]	SLTs → IFS → IVIFS	Crisp value → IFS → IVIFS	(O)demand preference strategy	×	Rough set technology + WSM	Form design of mobile phone
Zhou et al. [37]	Crisp score	Eye-tracking and EEG data	×	×	TOPSIS	Nursing Bed
Tang et al. [28]	Crisp score	Eye-tracking and EEG data	×	×	TOPSIS	Automobile industry design
Chen et al. [1]	SLTs → TFNs	×	(O)Entropy	×	TOPSIS	Color design in aircraft cockpits
Deng and Wang [2]	×	Crisp value	(O)Variance	×	TOPSIS	Human-machine interaction interface layout
Wu and Liao [31]	CIVLTSs:	×	(S)BWM + WSM	√	WSM	Product and service design
Jing et al. [9]	SLTs → Rough number	×	×	√	Rough set technology + Cooperative game	Blood pressure meter

Note: The full names of the abbreviations in Table A.1 are: Singleton linguistic terms-SLTs; AHP-analytic hierarchy process; WSM-Weighted Sum Model; IVQROFSs-interval-valued q-rung orthopair fuzzy set; VIKOR- Višekriterijumska Optimizacija i kompromisno Rešenje; TFNs-Triangular fuzzy numbers; BWM-best worst method; ILNs-intuitionistic linguistic numbers; IVIFS-interval-valued intuitionistic fuzzy set; IFS-Intuitionistic fuzzy set; TOPSIS-The technique for order of preference by similarity to ideal solution; CIVLTSs-continuous interval-valued linguistic term set; QFD-Quality function deployment; FAD- Fuzzy Axiomatic Design; HBWFAD-Hierarchal Best-Worst Fuzzy Axiomatic Design; (O)-Objective weight; (S)-Subjective weight. Symbol definitions in Table 1: "→": convert to; "√": support; "×": not support.

eye movement and EEG data of each evaluator are normalized and aggregated by Eqs. (17) and (18), respectively. When applying Eq. (18) to obtain the comprehensive decision value of eye movement and EEG data, we think that the data itself and data order have the same status, that is, $\varphi = 0.5$. The aggregated data is shown in the decision matrix D.

$$D = \begin{bmatrix} 0.75 & 0.48 & 0.51 & 0.32 & 0.41 & 0.48 & 0.37 & 0.34 & 0.48 & 0.47 & 0.54 & 0.33 & 0.36 \\ 0.95 & 1.00 & 1.00 & 0.97 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 0.98 \\ 0.81 & 0.53 & 0.64 & 0.66 & 0.64 & 0.28 & 0.59 & 0.34 & 0.28 & 0.43 & 0.40 & 0.39 & 0.52 \\ 0.90 & 0.79 & 0.78 & 0.89 & 0.82 & 0.76 & 0.78 & 0.74 & 0.76 & 0.77 & 0.80 & 0.75 & 0.86 \end{bmatrix}$$

After normalizing the data by Eqs. (19) and (20), the normalized decision matrix $D^{'}$ is obtained:

$$D^{'} = \begin{bmatrix} 0.79 & 0.48 & 0.51 & 0.33 & 0.41 & 0.48 & 0.37 & 0.34 & 0.48 & 0.47 & 0.54 & 0.33 & 0.36 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.86 & 0.53 & 0.64 & 0.69 & 0.64 & 0.28 & 0.59 & 0.34 & 0.28 & 0.43 & 0.40 & 0.39 & 0.52 \\ 0.95 & 0.79 & 0.78 & 0.92 & 0.82 & 0.76 & 0.78 & 0.74 & 0.76 & 0.77 & 0.80 & 0.75 & 0.88 \end{bmatrix}$$

For the subjective weight, we assume that the 13 indicators have the same subjective weight, *i.e.*, $w_j^S = 0.0769$ $(j = 1, 2, \cdots, 13)$. The objective weights are calculated by Eqs. (21)–(25). Solving the level difference maximization model of Eq. (26), we can get the comprehensive weights. The results are shown in Table 3.

By Eqs. (27)–(30), we get the utility value $f(\kappa_i)$ for each alternative. The calculation process is shown in Table 4. The final utility values of alternatives are 0.405, 0.895, 0.491 and 0.720, and the ranking is $A_2 \succ A_4 \succ A_3 \succ A_1$. Therefore, A_2 is the optimal one.

4.2. Sensitivity analysis

In this section, sensitivity analysis is carried out on the parameters of data aggregation, that is, the parameter φ in Eq. (18). We make it change in the range of 0 to 1 and take 0.1 as the increment. The change of φ causes the change of the decision matrix D, and then affects the objective weights, and finally causes the change of utility values. The final utility values with respect to different values of φ are shown in Table 5.

Fig. 2 shows the change of utility values of four alternatives when φ takes different values. We find that when φ is close to 0, the greater the difference between the utility values of each alternative is; when φ is

Table A2
Simulated raw data.

Subject	Criteria data of alternative 1												
	X1	Y1	Y2	FZ	FC1	F1	F2	F3	F4	F5	F6	F7	F8
e1	{s_1(0.70,0.80),s_2(0.25)}	14.79	50.17	1.20	2.39	2.02	1.91	0.92	0.92	1.56	1.11	0.99	0.78
e2	{s_1(0.56),s_2(0.44)}	14.93	51.56	1.32	2.65	2.30	2.08	1.00	1.02	1.84	1.12	1.01	0.79
e3	{s_1(0.43),s_2(0.57)}	15.03	51.96	1.35	2.76	2.75	2.08	1.01	1.43	1.86	1.22	1.02	0.84
e4	{s_1(0.29),s_2(0.71)}	16.87	53.12	1.41	2.91	2.95	2.16	1.03	1.73	1.91	1.43	1.04	0.86
e5	{s_1(0.25),s_2(0.75)}	17.52	54.12	1.42	3.19	3.00	2.60	1.04	1.73	1.98	1.58	1.06	0.92
e6	{s_1(0.15),s_2(0.85)}	17.67	54.30	1.64	3.31	3.02	2.61	1.04	1.96	2.04	1.75	1.11	0.92
e7	{s_1(0.10),s_2(0.80,1.0)}	17.86	55.51	1.66	3.37	3.18	2.69	1.05	1.99	2.65	1.87	1.13	0.96
e8	{s_1(0.07),s_2(0.93)}	18.21	56.63	1.76	3.47	3.26	2.83	1.07	2.13	3.04	1.93	1.14	1.04
e9	{s_1(0.06),s_2(0.94)}	18.40	56.72	1.81	3.71	3.37	3.10	1.14	2.39	3.27	1.95	1.16	1.09
e10	{s_2(0.48),s_3(0.52)}	18.48	56.79	2.08	4.08	3.45	3.18	1.15	2.43	3.27	1.96	1.92	1.21
e11	{s_2(0.47),s_3(0.53)}	19.24	57.12	2.17	4.16	3.46	3.38	1.19	2.64	3.36	2.08	1.94	1.32
e12	{s_2(0.38),s_3(0.62)}	19.35	58.21	2.20	4.44	3.76	3.61	1.36	2.65	3.38	2.11	1.99	1.60
e13	{[s_2,s_3](0.30,0.60),[s_3,s_4](0.40,0.80)}	20.50	58.34	2.53	4.48	3.79	3.78	1.59	2.82	3.59	2.14	2.07	1.73
e14	{s_2(0.14),s_3(0.86)}	21.41	58.53	2.61	4.69	3.80	4.11	2.12	3.00	3.93	2.31	2.08	1.74
e15	{s_2(0.14),s_3(0.86)}	21.70	59.21	3.08	4.74	3.85	4.27	2.46	3.23	3.97	2.32	2.66	1.79
e16	{s_3(0.50,0.58),s_4(0.46)}	25.01	62.79	3.21	4.82	4.23	4.30	2.55	3.49	4.18	2.61	2.74	2.10
e17	{s_3(0.37),s_4(0.63)}	25.04	63.52	3.87	5.11	4.27	4.33	2.77	3.87	4.20	2.65	2.78	2.31
e18	{s_3(0.32),s_4(0.68)}	25.11	65.05	3.93	5.58	4.33	4.64	3.03	4.15	4.23	2.65	2.93	2.53
e19	{s_3(0.10),s_4(0.90)}	25.20	67.63	4.20	5.86	5.10	4.73	3.51	4.20	4.30	2.65	3.08	2.62
e20	{s_3(0.06),s_4(0.94)}	26.55	68.71	4.55	6.27	6.11	5.62	4.98	4.25	5.44	4.56	4.12	2.86
	Criteria data of alternative 2												
	X1	Y1	Y2	FZ	FC1	F1	F2	F3	F4	F5	F6	F7	F8
e1	{s_3(0.08),s_4(0.92)}	25.70	75.56	4.23	5.70	3.90	5.14	3.93	2.55	3.70	2.33	3.50	1.74
e2	{s_3(0.03),s_4(0.97)}	27.78	76.84	4.29	5.70	4.47	5.23	4.22	3.42	5.01	2.43	4.22	2.13
e3	{s_4(0.99),s_5(0.01)}	28.67	80.12	4.52	6.71	4.67	5.40	4.78	4.02	5.03	2.50	4.41	2.31
e4	{s_4(0.91),s_5(0.09)}	30.08	81.41	4.62	6.79	4.79	5.83	4.81	4.82	5.08	2.78	4.44	2.60
e5	{s_4(0.79),s_5(0.10,0.32)}	30.37	81.43	5.40	6.83	5.02	6.05	4.95	4.89	5.16	3.03	4.59	2.82
e6	{s_4(0.74),s_5(0.26)}	31.49	82.47	5.72	7.37	5.19	6.36	4.98	4.91	5.21	3.12	4.70	3.04
e7	{s_4(0.74),s_5(0.26)}	31.65	83.29	5.73	7.56	5.27	6.83	5.10	5.06	5.26	3.17	4.87	3.12
e8	{s_4(0.69),s_5(0.31)}	31.77	83.37	5.76	7.58	5.53	6.91	5.65	5.53	5.39	3.29	4.91	3.19
e9	{s_4(0.64),s_5(0.36)}	33.87	83.43	6.04	7.61	5.64	7.02	5.76	5.59	5.72	3.36	5.00	3.25
e10	{[s3,s4](0.20,0.40),[s4,s5](0.40,0.60)}	33.91	84.60	6.05	7.86	5.72	7.21	5.84	5.60	5.75	3.45	5.01	3.31
e11	{s_4(0.40),s_5(0.60)}	34.50	84.74	6.07	7.95	5.73	7.32	5.97	5.68	5.85	3.56	5.40	3.49
e12	{s_4(0.30,0.32),s_5(0.69)}	34.94	85.75	6.15	7.96	5.91	7.60	6.04	5.73	6.23	3.59	5.42	3.74
e13	{s_4(0.29),s_5(0.71)}	35.27	86.82	6.46	8.10	6.12	7.73	6.36	5.82	6.43	3.88	5.77	3.83
e14	{s_4(0.27),s_5(0.73)}	35.47	87.07	6.46	8.21	6.14	7.83	6.42	6.44	6.57	3.99	5.90	3.98
e15	{s_4(0.14),s_5(0.86)}	36.15	87.36	6.84	8.21	6.15	7.92	7.04	6.47	6.69	4.06	6.08	3.98
e16	{s_4(0.13),s_5(0.87)}	36.31	87.87	7.48	8.33	6.26	8.08	7.34	6.47	6.89	4.20	6.38	4.00
e17	{s_4(0.09),s_5(0.91)}	36.54	91.34	7.62	8.44	6.40	8.08	7.39	6.58	7.09	4.33	6.54	4.00
e18	{s_4(0.09),s_5(0.91)}	37.04	91.69	7.72	9.19	6.78	8.51	8.20	6.82	7.29	4.92	6.90	4.17
e19	{s_4(0.07),s_5(0.93)}	37.44	92.36	8.11	9.77	7.12	9.29	8.39	7.50	7.59	5.00	7.76	4.59
e20	{s_4(0.01),s_5(0.99)}	38.43	92.47	8.73	10.13	7.19	9.66	8.82	8.07	8.05	5.01	8.21	4.73
	Criteria data of alternative 3												
	X1	Y1	Y2	FZ	FC1	F1	F2	F3	F4	F5	F6	F7	F8
e1	{s_1(0.53),s_2(0.47)}	13.77	56.36	2.53	2.91	2.42	2.35	0.93	0.91	1.58	0.93	1.00	0.77
e2	{s_1(0.43),s_2(0.57)}	15.32	56.96	3.26	4.35	4.10	2.98	0.93	0.96	1.69	1.01	1.01	0.86
e3	{s_1(0.23),s_2(0.77)}	17.00	57.80	3.60	4.40	4.57	3.61	0.94	0.99	1.90	1.01	1.05	0.87
e4	{s_2(0.47),s_3(0.53)}	17.97	58.40	3.81	4.75	4.58	3.70	1.01	1.08	2.03	1.09	1.12	0.91
e5	{s_2(0.15),s_3(0.85)}	18.07	58.88	4.06	5.11	4.63	3.81	1.01	1.09	2.20	1.09	1.13	1.23
e6	{[s1,s2](0.2,0.5),[s2,s3](0.5,0.8)}	18.07	58.91	4.36	5.22	4.71	4.44	1.02	1.10	2.24	1.16	1.15	1.31
e7	{s_2(0.08),s_3(0.92)}	18.22	59.00	4.37	5.51	4.75	4.44	1.03	1.12	2.34	1.25	1.51	1.32
e8	{s_3(0.98),s_4(0.02)}	18.38	59.79	4.52	5.62	4.91	4.46	1.12	1.53	2.61	1.31	1.59	1.35
e9	{s_3(0.72),s_4(0.28)}	18.88	60.35	4.60	5.64	4.93	4.49	1.13	1.63	2.67	1.57	1.75	1.61
e10	{s_3(0.64),s_4(0.30,0.42)}	21.16	62.06	4.83	5.83	4.94	4.58	1.14	1.67	2.76	1.63	1.79	2.06
e11	{s_3(0.49),s_4(0.51)}	21.42	62.08	5.20	5.85	4.98	4.78	1.80	1.76	2.89	1.77	1.89	2.06
e12	{s_3(0.47),s_4(0.53)}	22.16	62.46	5.41	6.15	5.18	4.84	2.00	1.78	2.91	2.11	2.00	2.09
e13	{s_3(0.44),s_4(0.56)}	22.94	63.11	5.44	6.43	5.33	5.35	2.16	2.17	3.22	2.11	2.09	2.25
e14	{s_3(0.40),s_4(0.60)}	23.25	63.83	5.71	6.49	5.60	5.62	2.17	2.23	3.50	2.18	2.11	2.40
e15	{s_3(0.40),s_4(0.60)}	23.41	65.42	5.76	6.52	5.75	5.78	2.37	2.24	3.53	2.22	2.13	2.46
e16	{s_3(0.39),s_4(0.61)}	23.70	66.33	6.13	6.65	5.78	5.78	2.89	2.27	3.55	2.37	2.31	2.52
e17	{s_3(0.31),s_4(0.61,0.78)}	24.13	68.69	6.17	7.18	5.86	6.07	3.36	2.30	4.18	2.37	2.55	2.88
e18	{s_3(0.03),s_4(0.97)}	24.26	68.72	6.25	7.19	6.04	6.13	3.47	2.38	4.32	2.64	3.58	2.91
e19	{s_4(0.70),s_5(0.30)}	25.16	75.31	6.48	7.76	6.29	6.28	3.64	3.40	4.42	2.74	3.88	2.99
e20	{s_4(0.03),s_5(0.97)}	25.21	75.53	7.53	8.44	6.65	6.52	3.87	3.41	5.44	3.43	4.38	3.14
	Criteria data of alternative 4												
	X1	Y1	Y2	FZ	FC1	F1	F2	F3	F4	F5	F6	F7	F8
e1	{s_2(0.33),s_3(0.67)}	20.52	55.17	3.61	4.35	3.20	3.74	2.07	1.56	2.47	0.92	1.95	1.11
e2	{s_2(0.30,0.34),s_3(0.68)}	22.19	61.17	4.47	5.70	3.45	4.19	2.70	2.63	3.38	1.76	2.15	1.82

(*continued on next page*)

Table A2 (*continued*)

	Criteria data of alternative 4												
	X1	Y1	Y2	FZ	FC1	F1	F2	F3	F4	F5	F6	F7	F8
e3	{s_3(0.48),s_4(0.52)}	23.66	62.11	4.50	5.71	3.46	4.48	2.90	2.70	3.44	2.49	2.56	1.94
e4	{s_3(0.33),s_4(0.67)}	23.70	63.16	4.84	5.75	4.43	4.84	3.47	2.77	3.65	2.61	2.73	2.53
e5	{s_3(0.26),s_4(0.74)}	24.66	65.78	5.00	5.93	4.58	5.20	3.49	2.85	3.89	2.65	2.83	2.55
e6	{s_3(0.20),s_4(0.80)}	25.20	66.79	5.17	6.52	4.64	5.28	3.84	4.03	4.32	2.83	3.04	2.89
e7	{[s3,s4](0.60,0.80),[s4,s5](0.30,0.40)}	25.38	67.22	5.25	6.54	4.65	5.32	3.99	4.06	4.50	2.87	3.17	2.95
e8	{s_3(0.03),s_4(0.97)}	25.52	68.52	5.49	6.65	4.66	5.41	4.01	4.28	4.56	2.97	3.43	2.96
e9	{s_3(0.03),s_4(0.97)}	25.64	68.61	5.58	6.86	4.78	5.47	4.24	4.28	4.58	3.05	4.45	3.00
e10	{s_4(0.94),s_5(0.06)}	26.45	69.11	5.73	6.88	4.82	5.72	4.33	4.30	4.72	3.07	4.54	3.16
e11	{s_4(0.91),s_5(0.09)}	27.85	70.09	5.87	6.94	5.00	5.80	4.42	4.39	4.91	3.13	4.56	3.39
e12	{s_4(0.82,0.93),s_5(0.13)}	28.59	70.62	6.10	6.98	5.04	5.96	4.62	4.84	4.99	3.47	4.95	3.43
e13	{s_4(0.84),s_5(0.16)}	29.77	71.11	6.27	7.02	5.04	6.05	4.67	5.05	5.35	3.77	5.15	3.65
e14	{s_4(0.83),s_5(0.17)}	30.39	71.25	6.58	7.04	5.25	6.08	4.95	5.17	5.44	3.87	5.15	3.68
e15	{s_4(0.64),s_5(0.36)}	30.67	72.48	6.92	7.12	5.39	6.22	5.17	5.26	5.55	3.94	5.32	3.99
e16	{s_4(0.63),s_5(0.37)}	31.18	72.92	6.97	7.68	5.65	6.27	5.18	5.31	5.58	4.00	5.37	4.08
e17	{s_4(0.48),s_5(0.52)}	31.64	72.99	7.20	8.07	5.69	6.48	5.68	5.44	5.80	4.02	5.39	4.10
e18	{s_4(0.25),s_5(0.75)}	31.65	74.99	7.45	8.36	5.91	6.57	6.06	5.89	5.81	4.07	5.51	4.15
e19	{s_4(0.22),s_5(0.78)}	32.79	76.81	7.85	8.43	5.98	8.18	6.63	6.52	5.82	4.13	5.56	4.23
e20	{s_4(0.19),s_5(0.81)}	33.24	79.08	9.15	9.47	6.37	8.72	7.58	6.68	7.27	4.37	6.21	4.40

Table B1
Initial sample data.

e_t	Score	e_t	Score	e_t	Score
1	{[s_2,s_3][0.25,0.33], s_3[0.60,0.70]}	13	{s_3(0.03), s_4(0.97)}	25	{s_4(0.84), s_5(0.16)}
2	{[s_2,s_3][0.25,0.33], s_3[0.60,0.70]}	14	{s_3(0.03), s_4(0.97)}	26	{s_4(0.84), s_5(0.16)}
3	{s_2(0.32), s_3(0.68)}	15	{s_3(0.03),s_4(0.97)}	27	{s_4(0.83), s_5(0.17)}
4	{s_3(0.48), s_4(0.52)}	16	{s_3(0.03), s_4(0.97)}	28	{s_4(0.64), s_5(0.36)}
5	{s_3(0.48), s_4(0.52)}	17	{s_4(0.94), s_5(0.06)}	29	{s_4(0.64), s_5(0.36)}
6	{s_3(0.33), s_4(0.67)}	18	{s_4(0.94), s_5(0.06)}	30	{s_4(0.63), s_5(0.37)}
7	{s_3(0.26), s_4(0.74)}	19	{s_4(0.91), s_5(0.09)}	31	{s_4(0.48), s_5(0.52)}
8	{s_3(0.26), s_4(0.74)}	20	{s_4(0.91), s_5(0.09)}	32	{s_4(0.48), s_5(0.52)}
9	{s_3(0.1,0.20), s_4(0.70,0.90)}	21	{s_4(0.91), s_5(0.09)}	33	{s_4(0.25), s_5(0.75)}
10	{s_3(0.1,0.20), s_4(0.70,0.90)}	22	{s_4[0.80,0.87], s_5[0.10,0.16]}	34	{s_4[0.1,0.22], s_5[0.70,0.8]}
11	{s_3(0.1,0.20), s_4(0.70,0.90)}	23	{s_4[0.80,0.87], s_5[0.10,0.16]}	35	{s_4[0.1,0.22], s_5[0.70,0.8]}
12	{s_3(0.03), s_4(0.97)}	24	{s_4(0.84), s_5(0.16)}	36	{s_4(0.19), s_5(0.81)}

close to 1, the smaller the difference between the utility values of each alternative is. This indicates that, instead of only focusing on the data itself (data layer), when more consideration is given to the data order (feature layer) in the aggregation of objective data, more cardinality differences can be reflected. The proposed algorithm for fusing physiological data can distinguish objective evaluations with slight difference but with ordinal significance, and thus meets practical needs of processing objective data, which are quite different from that of processing subjective evaluations.

When φ is greater than 0.5, the change of φ value has little influence on the final result, which shows a nonlinear response of the simulated data to the parameter φ. Additionally, no matter how the value of φ changes, the ranking of the four alternatives is consistent, and there is no crossover. It shows the rationality of taking $\varphi = 0.5$ when using the proposed method to process the simulated data in our numerical example. That is to say, for the simulated data, the parameter φ will not change the ranking result, and thus the verification of the validity and rationality of the proposed methods in Section 4.3 is not subject to the parameter φ.

Table B2
The lowest scores.

e_t	Score	e_t	Score	e_t	Score
1	{s_2(0.33),s_3(0.67)}	13	{s_3(0.03),s_4(0.97)}	25	{s_4(0.84),s_5(0.16)}
2	{s_2(0.33),s_3(0.67)}	14	{s_3(0.03),s_4(0.97)}	26	{s_4(0.84),s_5(0.16)}
3	{s_2(0.32),s_3(0.68)}	15	{s_3(0.03),s_4(0.97)}	27	{s_4(0.83),s_5(0.17)}
4	{s_3(0.48),s_4(0.52)}	16	{s_3(0.03),s_4(0.97)}	28	{s_4(0.64),s_5(0.36)}
5	{s_3(0.48),s_4(0.52)}	17	{s_4(0.94),s_5(0.06)}	29	{s_4(0.64),s_5(0.36)}
6	{s_3(0.33),s_4(0.67)}	18	{s_4(0.94),s_5(0.06)}	30	{s_4(0.63),s_5(0.37)}
7	{s_3(0.26),s_4(0.74)}	19	{s_4(0.91),s_5(0.09)}	31	{s_4(0.48),s_5(0.52)}
8	{s_3(0.26),s_4(0.74)}	20	{s_4(0.91),s_5(0.09)}	32	{s_4(0.48),s_5(0.52)}
9	{s_3(0.20),s_4(0.80)}	21	{s_4(0.91),s_5(0.09)}	33	{s_4(0.25),s_5(0.75)}
10	{s_3(0.20),s_4(0.80)}	22	{s_4(0.87),s_5(0.13)}	34	{s_4(0.22),s_5(0.78)}
11	{s_3(0.20),s_4(0.80)}	23	{s_4(0.87),s_5(0.13)}	35	{s_4(0.22),s_5(0.78)}
12	{s_3(0.03),s_4(0.97)}	24	{s_4(0.84),s_5(0.16)}	36	{s_4(0.19),s_5(0.81)}

Table B3
The highest scores.

e_t	Score	e_t	Score	e_t	Score
1	{s_2(0.30),s_3(0.70)}	13	{s_3(0.03),s_4(0.97)}	25	{s_4(0.84),s_5(0.16)}
2	{s_2(0.30),s_3(0.70)}	14	{s_3(0.03),s_4(0.97)}	26	{s_4(0.84),s_5(0.16)}
3	{s_2(0.32),s_3(0.68)}	15	{s_3(0.03),s_4(0.97)}	27	{s_4(0.83),s_5(0.17)}
4	{s_3(0.48),s_4(0.52)}	16	{s_3(0.03),s_4(0.97)}	28	{s_4(0.64),s_5(0.36)}
5	{s_3(0.48),s_4(0.52)}	17	{s_4(0.94),s_5(0.06)}	29	{s_4(0.64),s_5(0.36)}
6	{s_3(0.33),s_4(0.67)}	18	{s_4(0.94),s_5(0.06)}	30	{s_4(0.63),s_5(0.37)}
7	{s_3(0.26),s_4(0.74)}	19	{s_4(0.91),s_5(0.09)}	31	{s_4(0.48),s_5(0.52)}
8	{s_3(0.26),s_4(0.74)}	20	{s_4(0.91),s_5(0.09)}	32	{s_4(0.48),s_5(0.52)}
9	{s_3(0.10),s_4(0.90)}	21	{s_4(0.91),s_5(0.09)}	33	{s_4(0.25),s_5(0.75)}
10	{s_3(0.10),s_4(0.90)}	22	{s_4(0.84),s_5(0.16)}	34	{s_4(0.20),s_5(0.80)}
11	{s_3(0.10),s_4(0.90)}	23	{s_4(0.84),s_5(0.16)}	35	{s_4(0.20),s_5(0.80)}
12	{s_3(0.03),s_4(0.97)}	24	{s_4(0.84),s_5(0.16)}	36	{s_4(0.19),s_5(0.81)}

Table B4
The frequency of each linguistic term in the highest and lowest scores.

	s_0	s_1	s_2	s_3	s_4	s_5
Frequency (lowest score)	0	0	0.98	4.58	23.89	6.55
Frequency (highest score)	0	0	0.92	4.34	24.09	6.65

Table B5
Sample mean and sample variance of terms in the lowest and highest scores.

	Mean	Variance
Lowest score	4.0003	0.4300
Highest score	4.0131	0.4190

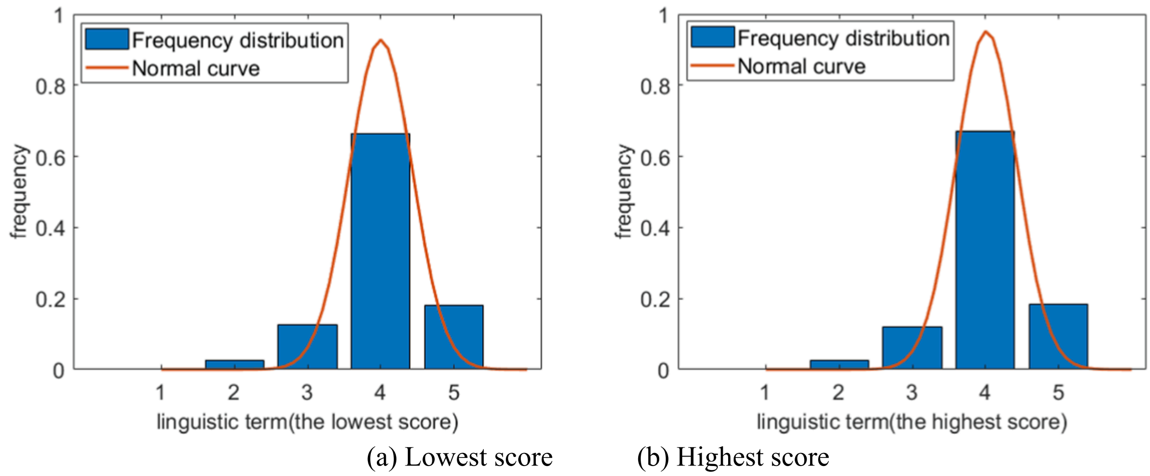

Fig. B1. Term frequency and normal distribution curve about the lowest score.

4.3. Comparative analysis with existing methods

This section aims to conduct a comparative analysis to test the validity and rationality of the methods. We perform a qualitative comparison and an empirical comparison based on the same simulated data in Table A.2.

The main advantages compared our proposed method with the existing methods are shown in Table 6.

From Table 6, we can see that in item 1, only our method and Lou et al. [19,20], Zhou et al. [37], and Tang et al. [28] considered the fusion of subjective evaluation and objective physiological data at the same time; in item 2, although interval-valued q-rung orthopair fuzzy set [1] and continuous interval-valued linguistic term set [31] were used in fuzzy evaluation of product design, our method is more flexible than interval-valued q-rung orthopair fuzzy set and continuous interval-valued linguistic term set in expressing and processing fuzzy and uncertain information by supporting the general form of IUPLTSs; in items 3–7, our method shows advantages, and supports the relevant benefits which is more in line with the actual decision context. Additionally, specific to the criterion weight determination, compared with the CRITIC method [3], our proposed method is not based on a standard deviation when computing the vertical differences of criteria, which can avoid the low accuracy and large error of the CRITIC method [36].

Fig. 3 displays the implement framework of empirical comparisons, including the aggregation methods of initial group data, weight determination methods, and MCDM methods. In Step 1, the initial data are aggregated by subjective data aggregation process described in Section 3.1 and the WSM method, and the aggregated matrix is

$$\widehat{D} = \begin{bmatrix} 0.75 & 19.943 & 58 & 2.4 & 4.1 & 3.6 & 3.4 & 1.8 & 2.6 & 3.2 & 2.1 & 1.9 & 1.5 \\ 0.95 & 33.369 & 85 & 6.2 & 7.8 & 5.7 & 7.2 & 6.1 & 5.6 & 6 & 3.6 & 5.5 & 3.4 \\ 0.81 & 20.624 & 63 & 5 & 5.9 & 5.1 & 4.8 & 1.9 & 1.8 & 3 & 1.8 & 2 & 1.9 \\ 0.9 & 27.534 & 69 & 6 & 6.9 & 4.9 & 5.8 & 4.5 & 4.4 & 4.8 & 3.2 & 4.2 & 3.2 \end{bmatrix}$$

Then, in order to promote the following objective weight calculation and the use of comprehensive decision-making method, the matrix $\widehat{D}$ is normalized by different technologies in Step 2, including linear ratio, linear max–min, and vector-based method, respectively. In Step 3, we give the subjective and objective weights of indicators. The subjective weights are assigned as the same, that is, $w_j^S = 0.0769$ ($j = 1, 2, \cdots, 13$); the objective weights are obtained by the entropy weight method, and the result isW^o = [0.0067, 0.0372, 0.0171, 0.0874, 0.0418, 0.0212, 0.0566, 0.2098, 0.1416, 0.0665, 0.0635, 0.1621, 0.0884].

In Step 4, we use the classical MCDM methods including WSM, TOPSIS [23], VIKOR[6] [23] and EDAS [10] to make final decisions. Based on the two weighting methods and four MCDM methods, we generated eight methods (2 × 4) to compare, i.e., SW-SWM (M1), OW-SWM (M2), SW-TOPSIS (M3), OW- TOPSIS (M4), SW-VIKOR (M5), OW-VIKOR (M6), SW-EDAS (M7), OW-EDAS (M8). Last, we observe the consistency of the ranking results of different methods with that of our proposed method, and calculate the Pearson correlation coefficient of the values of alternatives obtained by different methods. Table 7 shows the feature values calculated by different methods and the rankings of alternatives. We can find that A_2 is the best one among all methods. Likewise, as we can see in Table 8, the correlation coefficients between our method and other different methods are close to "1". Compared with existing methods, the proposed method ensures the consistency and reliability when considering heterogeneous data in the product design evaluation.

5. Conclusion

In this paper, we proposed an MCDM method for product appearance design scheme evaluation. We took into account the qualitative indicators of subjective feedback, as well as the physiological indicators of eye movement and EEG to measure an evaluator's preferences for alternatives. The method used IUPLTSs to represent the fuzzy and uncertain information of evaluators flexibly and naturally. We discussed the information measurement of IUPLTSs. We aggregated the quantitative physiological index data in data layer and feature layer simultaneously. We proposed an objective weight calculation method which takes into account the data differences of indicators and the correlations between indicators. Considering the stability of decision-making results when a large number of indicators are involved, we integrated the MARCOS method for comprehensive decision making. Through the simulation analysis regarding the evaluation of automobile appearance design scheme, we checked the effectiveness of our method by sensitivity analysis and comparative analysis.

There are still unsolved issues which could be investigated in the future. Due to experimental limitations, we used simulated data to verify the effectiveness of our method. An empirical analysis would be meaningful to further check the effectiveness of our method. It is also necessary to expand the number of indicators and schemes to meet the

[6] Note: we set the weight of the strategy of "the maximum group utility'' is 0.5 in VIKOR.

requirement of practice. In addition, the development of software support tools will contribute to the adoption of technology in industry. We have developed a Java programming language-based prototype tool (named: "DAQQ") to support our decision algorithm and reduce the computational burden of algorithm users. But the system functionality is limited. In the future, we plan to enhance the interactivity of the prototype system to make our system user-friendly.

Informed consent

All authors participated in this study with their consent.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

CRediT authorship contribution statement

Han Lai: Conceptualization, Methodology, Funding acquisition, Writing – original draft. **Zheng Wu:** Conceptualization, Methodology, Writing – review & editing. **Xiaokai Zhang:** Conceptualization, Methodology, Writing – original draft. **Huchang Liao:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing – review & editing. **Edmundas Kazimieras Zavadskas:** Methodology, Writing – review & editing.

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.

Data availability

All the data used in this study can be used in any research project by just citing this article.

Acknowledgements

This study was funded by the National Natural Science Foundation of China (71771156, 71971145, 72171158) and the National Training Program of Innovation and Entrepreneurship for Undergraduates (202011799002). Zheng Wu contributed equally to the first author.

Appendix A

See Table A.1 and Table A.2.

Appendix B

Example. Let $S = \{s_0, s_1, \cdots, s_5\}$ be an LTS. Suppose that there are thirty-six students $e_t(t = 1,2\cdots,36)$, who are invited to evaluate the teaching skill of a teacher. Suppose that the students' judgments are given in IUPLTSs shown in Table B.1.

We first convert the data to IPPLTSs using Eq. (3). Then, we can apply Eqs. (5) and (6) to obtain the lowest scores and highest scores as shown in Tables B.2 and B.3.

Using Eq. (10), we can get the frequency of linguistic terms in the lowest scores. Similarly, we can also get frequency of linguistic terms in the highest scores. The results are shown in Table B.4.

Using Eqs. (11) and (12), for the lowest and highest scores, the sample mean and sample variance of terms are obtained respectively. The results are shown in Table B.5.

Next, we can obtain the normal distribution function of evaluation term in the lowest and highest scores by Eq. (13). Fig. B.1 shows the term frequency and normal distribution curve about the lowest and highest scores, respectively.

Finally, we can obtain the lowest score and the highest score after aggregation by Eqs. (14) and (15).

$$h^{'\min} = \{s_3(0.0590), s_4(0.8819), s_5(0.0590)\}$$

$$h^{'\max} = \{s_3(0.0482), s_4(0.8959), s_5(0.0559)\}$$

References

[1] Y. Chen, S. Yu, J. Chu, M. Yu, D. Chen, Fuzzy emotional evaluation of color matching for aircraft cockpit design, J. Intell. Fuzzy Syst. 40 (3) (2021) 3899–3917.
[2] L. Deng, G.H. Wang, Quantitative evaluation of visual aesthetics of human-machine interaction interface layout, Comput. Intell. Neurosci. 2020 (2020), 9815937, https://doi.org/10.1155/2020/9815937.
[3] D. Diakoulaki, G. Mavrotas, L. Papayannakis, Determining objective weights in multiple criteria problems: the critic method, Comput. Oper. Res. 22 (7) (1995) 763–770.
[4] Y. Ding, Y. Cao, Q. Qu, V.G. Duffy, An exploratory study using Electroencephalography (EEG) to measure the smartphone user experience in the short term, Int. J. Hum.-Comput. Interact. 36 (11) (2020) 1008–1021.
[5] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80.
[6] F. Guo, M.M. Li, M.C. Hu, F.X. Li, B.Z. Lin, Distinguishing and quantifying the visual aesthetics of a product: an integrated approach of eye-tracking and EEG, Int. J. Ind. Ergon. 71 (2019) 47–56.
[7] S.W. Hsiao, Y.C. Ko, A study on bicycle appearance preference by using FCE and FAHP, Int. J. Ind. Ergon. 43 (4) (2013) 264–273.
[8] L.T. Jing, J. Yao, F. Gao, J.Q. Li, X. Peng, S.F. Jiang, A rough set-based interval-valued intuitionistic fuzzy conceptual design decision approach with considering diverse customer preference distribution, Adv. Eng. Inf. 48 (2021), 101284, https://doi.org/10.1016/j.aei.2021.101284.
[9] L.T. Jing, S.F. Jiang, J.Q. Li, X. Peng, J.F. Ma, A cooperative game theory based user-centered medical device design decision approach under uncertainty, Adv. Eng. Inf. 47 (2021), 101204, https://doi.org/10.1016/j.aei.2020.101204.
[10] M. Keshavarz Ghorabaee, E.K. Zavadskas, L. Olfat, Z. Turskis, Multi-criteria inventory classification using a new method of evaluation based on distance from average solution (EDAS), Informatica 26 (3) (2015) 435–451.
[11] M.S.A. Khannan, A.E. Tontowi, M.K. Herliansyah, A. Sudiarso, New product development method trends and future research: a systematic literature review, Jurnal Teknik Industri 23 (1) (2021) 11–24.
[12] S. Kumar, M. Yadava, P.P. Roy, Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction, Inf. Fusion 52 (2019) 41–52.
[13] J.Y. Kuo, C.H. Chen, S. Koyama, D.N. Chang, Investigating the relationship between users' eye movements and perceived product attributes in design concept evaluation, Appl. Ergon. 94 (2021), 103393, https://doi.org/10.1016/j.apergo.2021.103393.
[14] B.R. Li, Y. Wang, K.S. Wang, A novel method for the evaluation of fashion product design based on data mining, Adv. Manuf. 5 (4) (2017) 370–376.

[15] G. Li, J.P. Li, X.L. Sun, M. Zhao, Research on a combined method of subjective-objective weighing and the its rationality, Manage. Rev. 29 (12) (2017), 17–26+61 (in Chinese).
[16] H.C. Liao, X.L. Wu, DNMA: a double normalization-based multiple aggregation method for multi-expert multi-criteria decision making, Omega 94 (2020), 102058, https://doi.org/10.1016/j.omega.2019.04.001.
[17] H.C. Liao, X.L. Wu, X.D. Liang, J.B. Yang, D.L. Xu, F. Herrera, A continuous interval-valued linguistic ORESTE method for multi-criteria group decision making, Knowl.-Based Syst. 153 (2018) 65–77.
[18] H. Liu, M. Cocea, Induction of classification rules by gini-index based rule generation, Inf. Sci. 436 (2018) 227–246.
[19] S.H. Lou, Y.X. Feng, Z.W. Li, H. Zheng, J.R. Tan, An integrated decision-making method for product design scheme evaluation based on cloud model and EEG data, Adv. Eng. Inf. 43 (2020), 101028, https://doi.org/10.1016/j.aei.2019.101028.
[20] S.H. Lou, Y.X. Feng, Z.W. Li, H. Zheng, Y.C. Gao, J.R. Tan, An edge-based distributed decision-making method for product design scheme evaluation, IEEE Trans. Ind. Inf. 17 (2) (2021) 1375–1385.
[21] A.I. Maghsoodi, M. Mosavat, A. Hafezalkotob, A. Hafezalkotob, Hybrid hierarchical fuzzy group decision-making based on information axioms and BWM: prototype design selection, Comput. Ind. Eng. 127 (2019) 788–804.
[22] L.A. Medina, G.E.O. Kremer, R.A. Wysk, Supporting medical device development: a standard product design process model, J. Eng. Des. 24 (2) (2013) 83–119.
[23] S. Opricovic, G.H. Tzeng, Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS, Eur. J. Oper. Res. 156 (2) (2004) 445–455.
[24] Q. Pang, H. Wang, Z.S. Xu, Probabilistic linguistic term sets in multi-attribute group decision making, Inf. Sci. 369 (2016) 128–143.
[25] J. Qi, J. Hu, Y. Peng, Modified rough VIKOR based design concept evaluation method compatible with objective design and subjective preference factors, Appl. Soft Comput. 107 (2021), 107414, https://doi.org/10.1016/j.asoc.2021.107414.
[26] P. Rani, A.R. Mishra, Novel single-valued neutrosophic combined compromise solution approach for sustainable waste electrical and electronics equipment recycling partner selection, IEEE Trans. Eng. Manag. (2020), https://doi.org/10.1109/TEM.2020.3033121.
[27] Ž. Stević, D. Pamučar, A. Puška, P. Chatterjee, Sustainable supplier selection in healthcare industries using a new MCDM method: measurement of alternatives and ranking according to COmpromise solution (MARCOS), Comput. Ind. Eng. 140 (2020), 106231, https://doi.org/10.1016/j.cie.2019.106231.
[28] B.B. Tang, G. Guo, K. Wang, L. Lin, J. Zhou, X. Fan, N. Xu, X.Y. Guo, User experience evaluation and selection of automobile industry design with eye movement and electroencephalogram, Comput. Integr. Mauf. Syst. 21 (2015) 1449–1459, https://doi.org/10.13196/j.cims.2015.06.006 (in Chinese).
[29] V. Wagner, M. Scharinger, C.A. Knoop, W. Menninghaus, Effects of continuous self-reporting on aesthetic evaluation and emotional responses, Poetics (2020), 101497, https://doi.org/10.1016/j.poetic.2020.101497.
[30] Z. Wen, H.C. Liao, E. Kazimieras Zavadskas, A. Al-Barakati, Selection third-party logistics service providers in supply chain finance by a hesitant fuzzy linguistic combined compromise solution method, Econ. Res.-Ekonomska Istraživanja 32 (1) (2019) 4033–4058.
[31] X.L. Wu, H.C. Liao, Customer-oriented product and service design by a novel quality function deployment framework with complex linguistic evaluations, Inf. Process. Manag. 58 (2) (2021), 102469, https://doi.org/10.1016/j.ipm.2020.102469.
[32] X.L. Wu, H.C. Liao, W. Pedrycz, Probabilistic linguistic term set with interval uncertainty, IEEE Trans. Fuzzy Syst. (2021), https://doi.org/10.1109/TFUZZ.2020.3025699.
[33] X.L. Wu, H.C. Liao, A consensus-based probabilistic linguistic gained and lost dominance score method, Eur. J. Oper. Res. 272 (3) (2019) 1017–1027.
[34] X.L. Wu, H.C. Liao, An approach to quality function deployment based on probabilistic linguistic term sets and ORESTE method for multi-expert multi-criteria decision making, Inf. Fusion 43 (2018) 13–26.
[35] Z.S. Xu, Fuzzy harmonic mean operators, Int. J. Intell. Syst. 24 (2) (2009) 152–172.
[36] H. Yang, Y. Tan, X. Sun, X. Cheng, G. Liu, G. Zhou, Comprehensive evaluation of water resources carrying capacity and analysis of obstacle factors in Weifang City based on hierarchical cluster analysis-VIKOR method, Environ. Sci. Pollut. Res. (2021), https://doi.org/10.1007/s11356-021-14236-3.
[37] Z.Y. Zhou, J.X. Cheng, W.J. Wei, L. Lee, Validation of evaluation model and evaluation indicators comprised Kansei Engineering and eye movement with EEG: an example of medical nursing bed, Microsyst. Technol. 27 (2021) 1317–1333.
[38] H.G. Peng, J.Q. Wang, Hesitant uncertain linguistic z-numbers and their application in multi-criteria group decision-making problems, Int. J. Fuzzy Syst. 19 (5) (2021) 1300–1316.