An effective similarity determination model for case-based reasoning in support of low-carbon product design

Shedong Ren1,2, Fangzhi Gui1, Yanwei Zhao1, Min Zhan1 and Wanliang Wang2

Abstract
In the initial stage of low-carbon product design, design information is always uncertain and incomplete, as well as the coupling properties between design attributes, thus it requires retrospective coordination for design conflicts resulting from the inclusion of low-carbon requirements. Reusing the prior design knowledge can promote design efficiency, however, the acquisition of similar cases knowledge not only needs to consider the similarity of design problems, but also the adaptability of candidate cases. This study presents an effective similarity determination model to support low-carbon product design, and targets of the proposed model are (1) to reasonably determine design ranges of attribute values for product cases retrieval by representing the uncertain design attributes with fuzzy set theory; (2) to construct an efficient indexing structure to generate the index set of similar cases based on the improved discretized highest similarity method by proposing two effective strategies; and, (3) to establish similarity estimation models for different types of attributes, and it calculates the information content of each attribute to evaluate the adaptability of cases based on the Information Axiom. The applicability of the proposed model is demonstrated through a case study of similar cases retrieval for the vacuum pump low-carbon design.

Keywords
Low-carbon design, case-based reasoning, fuzzy requirement information, indexing structure, adaptability estimation

Date received: 6 August 2020; accepted: 23 September 2020

Handling Editor: James Baldwin

Introduction
In the 21st century, human society confronts with significant challenges from climate change. The IPCC’s fourth report of environmental assessment pointed out that climate change and global warming are mainly caused by the large amount of greenhouse gases generated by human activities, where the industrial production and manufacturing account for a large proportion.1 The International Academy for Production Engineering (CIRP) established the low-carbon manufacturing working group (CWG) in Pairs, France, in January 2008, it aims to guide enterprises to take measures to reduce carbon emissions throughout product life cycle, including energy saving and emission reduction at manufacturing stage, green supply chain, services at use stage, and components recycling at the end of life stage.2 The preliminary design stage of the product influences about 80% economic cost, environmental and social impacts in its whole life cycle.3 Low-
carbon product design takes account of the environmental indicators during product planning and conceptual design stages, and puts forward the strategies and ideas for the generation of new design schemes with the purpose of carbon emission reduction, and it has attracted a heated research in both academia and industry.4–6

The research of low-carbon product design mainly focuses on the carbon footprint estimation,7–9 low-carbon design theory and methodology,10,11 and low-carbon optimization.12–14 Carbon footprint, the sum of greenhouse gases emissions at each life cycle stage, is a wildly accepted quantitative metric to evaluate the environmental impact of the product. Song and Lee15 estimated the life cycle carbon footprint of parts based on the design information obtained from the bill of material (BOM), and established the greenhouse gases-BOM (g-BOM) to identify parts with high carbon emissions, which indicates the direction for subsequent improvement and optimization of the parts. Devanathan et al.16 combined the quality function deployment (QFD) and life cycle assessment (LCA) method to construct a function-environment impact matrix, which enables designers to evaluate the environmental impact of the product function at design stage. However, in the low-carbon product design, since the coupling properties of design quality characteristics, it will inevitably break the original design equilibrium system when low-carbon requirements are incorporated, which results in the complex design conflicting problems. For instance, modifying the component to reduce its carbon emission, but the economic cost is increased and the mechanical property is deteriorated. Therefore, it is necessary to employ the conflicting problem-solving method to coordinate the design contradictions and generate design strategies to make a trade-off between carbon footprint, cost and function performance. The common used contradictory problems solving methods include the theory of inventive problems solving (TRIZ, a Russian acronym),17,18 Extenics,19–21 constraint satisfaction problems (CSP),22 the negotiation method.23,24

From the abovementioned studies in low-carbon product design, it indicates that whether it is the carbon footprint estimation, the generation of low-carbon design schemes through solving the contradictory problems, and the low-carbon optimization method, they are all dependent on the preceding design knowledge. That is, it is a reasoning innovation on the basis of the product case knowledge. Thus, integrating the case-based reasoning (CBR) technique and design conflicting problem-solving methods can promote the design efficiency,25 and to accurately grasp the similar cases knowledge corresponding to the low-carbon requirements is the prerequisite for low-carbon design. In our research work, it presents an effective similarity determination model to obtain similar product cases in consideration of the similarity and adaptability for knowledge reuse and adaptation for the low-carbon product design.

The CBR method reuses the prior case knowledge to solve the similar design problems, it stores the product cases with the solution to problems previously solved in the case base, rather than extracting the complex rule knowledge, such as the rule-based method.26,27 Whereas CBR method requires less knowledge extraction in the knowledge acquisition task, it needs to collect a significant number of product cases to wildly cover the design problem and solution spaces in order to handle the new emergent design problems, which contributes to the retrieval utility problem of the case base. Retrieval utility problem refers to the opposite property between the accuracy and the efficiency,28 that is, when the number of product cases reaches the saturation point, the retrieval accuracy is basically stable, and if it continually adds new cases, the retrieval efficiency decreases. On the other hand, it is unreasonable to select the candidate cases by the single metric, similarity. Because the most similar candidate case may not be the most appropriate for the adaptation purpose. Consequently, it requires additional knowledge to construct the similarity determination model for the candidate cases in consideration of the similarity and adaptability.29,30

The most commonly used cases retrieval method is the nearest neighbor algorithm and its various improved variants.31,32 nearest neighbor method is wildly adopted for the flat case base as its simplicity and practicability. However, it needs to calculate the distance between each case in the case base (memory) and the query case, and then makes the final decision, thus, it will consume extensive searching time when facing the large size case base. Compared with the flat case base, the structure tree indexing method is more suitable for the cases retrieval in a hierarchical structure case base; it hierarchically divides the case base into multiple subsets based on the attribute nodes, thus it only needs to search the subset, in which the attribute values of the cases falling within the value range of the node attributes. Hjaltason and Samet33 summarized various structure tree indexing methods, and Stephane et al.30 pointed out the complexity in case base maintenance caused by the structural expansion under dynamic cases addition, and revealed the limitation in processing the overlapping cases. In addition, some useful machine learning methods are applied to the case retrieval, such as the neural network classification method,34 genetic algorithm,35 the clustering method.

Aiming at addressing the retrieval utility problem of the case base, the following three measures are generally adopted:28 (1) construct the indexing structure for
the case base, in this approach, the number of cases is not limited, and the constructed indexing structure can improve the retrieval efficiency without reducing the accuracy; however, it is not easy to construct an appropriate indexing structure according to the attribute properties of the cases; (2) formulate the principle of cases addition, that is, control the number of cases in the case base by assessing whether the new case satisfies the addition principle; (3) formulate the principle of cases deletion, namely, control the number of cases by evaluating whether the existing cases should be removed from the case base based on the deletion principle. The difficulty of the latter two measures lies in the formulation of evaluation criteria for the principles. In terms of the adaptability of the similar cases, Smyth and Keane\textsuperscript{29} pointed out that it requires additional knowledge to select the candidate case in addition to the similarity, and they proposed the concept of adaptation-guided retrieval scheme. Leake et al.\textsuperscript{38} further studied the relationship between the similarity and adaptability. Rousu and Aarts\textsuperscript{39} incorporated the adaptation cost into the evaluation indicators of the similarity and adaptability. Roldan Reyes et al.\textsuperscript{40} reported that the CSP and CBR depend on the general knowledge and contextual knowledge respectively, and they integrated these two methods to address adaptation problems in the process of engineering design. Qi et al.\textsuperscript{41} incorporated the support vector machine (SVM) and adaptability-related knowledge to obtain adaptable cases, the knowledge of adaptability contains the adaptability characteristics of the old cases and the guideline to the construction of the adaptability-involving SVM adaptation model with respect to retrieved cases.

In addition, in the earlier low-carbon design stage for products, facing the fuzzy low-carbon requirements, the incomplete design information, and the potential complex design conflicting problems, even if the designers are experienced domain experts, they cannot grasp the deep knowledge to determine the precise values for the design attributes. But they roughly know the possible design ranges, such as the lower and upper boundaries of the attribute values to guarantee that the decision making for the design parameters is still satisfying the design targets. Giachetti et al.\textsuperscript{42} revealed this uncertainty in the expression of the design information at each design stage, from the linguistic variables, fuzzy numbers, to the final real numbers. It also exactly demonstrates that the essence of product design is a complex backtracking process of clarifying the “ill-defined” target problems.\textsuperscript{43} The rest of this article is organized as follows: in the next section, low-carbon requirements are detailed analyzed, and the fuzzy set theory is used to represent the uncertain design attributes; the discretized highest similarity indexing method is improved with two proposed strategies to construct the effective indexing structure for the low-carbon product case base; the comprehensive estimation model by integrating the similarity and adaptability is presented to evaluate the candidate cases; a case study about the similar cases retrieval for the vacuum pump low-carbon design is provided to demonstrate the effectiveness of our proposed method; discussion is given, and conclusion is finally drawn along with the recommendations for the further research.

### Requirement analysis and representation for low-carbon product design

For low-carbon product design, it is essential to obtain similar cases from the case base according to the low-carbon requirements, and reuse the prior knowledge and conduct the adaptation to generate a new design scheme. Thus, it needs to extract the retrieval attributes from the requirement information. In addition, as the design information is incomplete and unascertained at earlier design stage, in this paper, different kinds of descriptions of the requirement information are converted into the standardized fuzzy numbers.

#### Low-carbon requirement analysis

Low-carbon requirement comprises the traditional customer preference, that is the appearance attributes of the product, the general quality performance, etc., which is named the voice of customers (VOC), and the low-carbon demand, such as less material usage, low energy consumption, less carbon emissions. Mausi et al.\textsuperscript{44} summarized the customer requirements for the environment, defined as voice of environment (VOE), and the quality characteristics (QCs) for the product environmental conscious design (Eco-design), as listed in Table 1. Low-carbon design is a branch of the Eco-design, while the difference in VOC with different products, the VOE and corresponding QCs involved in the design requirements for different products are basically similar. Thus, this study incorporates those low-carbon related VOE and QCs into the design requirement analysis.

In addition, QFD method is a useful design tool to collect vaguely described needs and deploy them to the actual design work. The relationship of low-carbon requirements and quality characteristics can be established based on the house of quality (HoQ), as shown in Figure 1. According to the deployment for the low-carbon requirements, designers can identify the important quality characteristics with high relative weight, and extract them as the retrieval attributes to search the similar cases knowledge for low-carbon design.
In the life cycle-oriented low-carbon design process, the requirement information not only comes from the product users, but also includes the feedback information from the workers who are in charge of the product manufacturing, assembly, sales, maintenance, and recycling. However, it should be noted that some of this requirement information can be quantitatively expressed, for instance, specify the numerical attribute value with a crisp number, or designate it in a required range; and some requirement information is quantitative immeasurable, it should be expressed with a linguistic term. For example, it adopts the linguistic term to describe the low-carbon and environmental impact (LCEI) of the product as “LCEI is poor”, “LCEI is fair”, “LCEI is good”.

Thus, the different types of descriptions of the requirement information, including the precise numerical value, a range of the numerical value, and the linguistic term, make it difficult to construct the evaluation criteria of the similarity and adaptability for the cases. Fortunately, the above requirement information can be quantitative measured based on the fuzzy set theory. The fuzzy set in a domain $U$ is defined by a membership function $m$($x$), which has the values in range $[0, 1]$. $m$($x$) indicates the degree to which $x$ is the possible value in a fuzzy subset. In this paper, the uncertain attribute values of the requirement information are transformed into the standardized triangular fuzzy number, named STFN (Triangle), or the standardized trapezoidal fuzzy number, named STFN (Trapezoid), as illustrated in Figure 2.

Here, it takes the STFN (Trapezoid) as an example to describe its transformation effect. Suppose the attribute value $A$, the transformation result of $A$ based on the fuzzy set theory is represented as STFN ($A$) = ($a_l$, $a_m$, $a_n$, $a_u$), $d_l$ ≤ $a_m$ = $d_r$ ≤ $a_u$, as shown in Figure 2(b), where the domain $U$ of the fuzzy set is
expressed in Table 3.

and the fuzzy numerical value can also be converted
as described in Figure 2(a) and Table 2, respectively expressed by STFN (Triangle) or STFN (Trapezoid) is equal to the STFN (Triangle).

When \( a' = a'' = a'' = a'' \), \( A \) is a numerical value and it is transformed into a STFN(Trapezoid); when \( a' = a'' \), \( a'' = a'' \), \( A \) is a range value; and when \( a'' = a'' \), the STFN (Trapezoid) is equal to the STFN (Triangle).

Set \( U = [0, 1] \), the linguistic term can be quantitatively expressed by STFN (Triangle) or STFN (Trapezoid) as described in Figure 2(a) and Table 2, and the fuzzy numerical value can also be converted into the STFN (Triangle) or STFN (Trapezoid) as expressed in Table 3.

### The improved D-HS indexing method for the index set of similar cases

Indexing method is used to gather a group of similar cases for similarity calculation, and thus it can avoid the extensive searching for the whole memory. In this paper, an efficient indexing method named Discretized Highest Similarity (D-HS)\(^5\) is introduced, however, this general approach presents its weakness in application, two effective strategies are proposed to construct the improved indexing structure of the case base for low-carbon design.

### Principle of the D-HS indexing method

With the principle of D-HS method, it distributes the intervals of each attribute value of the product cases, and designates the spatial coordinates for the cases in the memory. Cases with the same coordinates as that of the query case constitute the similar group. The simple indexing structure of D-HS method facilitates the maintenance for the case base, and it can effectively address the utility problem.

Let \( C \) represent the low-carbon case base, it is composed of \( N \) product cases \((c_i)\), and each case consists of \( M \) attributes \((c_i)\) in a vector, as expressed in equations (2) and (3).

\[
C = \{c_i, i = 1, 2, \ldots, N\} \quad (2)
\]

\[
c = [c_i], j = 1, 2, \ldots, M \quad (3)
\]

where, it defines the expression \( c_c(c_i) \) as the value of attribute \( c_i \) of the \( c_c \).

Suppose the low-carbon query case \( q \), that is the target case, which consists of \( M \) attributes \( q_x \), and thus the query case can be described as below.

\[
q = [q_i], i = 1, 2, \ldots, M \quad (4)
\]

where, it defines the expression \( q_x(q) \) as the value of the attribute \( q_x \) of the query case \( q \).

Let \( S \) represent the index set of similar cases, generated by the indexing rule \( R \) based on the D-HS principle as expressed in equation (5).

\[
S = R(q) = \bigcap_{i=1}^{M} C^i,
\]

\[
C^i = \left\{ c_j \in C : c_{c_j}(c_i) \in \left[ q_x(q) - \frac{\Delta \delta_i}{2}, q_x(q) + \frac{\Delta \delta_i}{2} \right] \right\} \quad (5)
\]

Where, \( \Delta \delta_i \) denotes the interval width of the \( i \)-th attribute value; \( C^i \) is the subset of cases, in which the
attribute value of each case falls within the same interval as that of the query case for the \(i\)-th attribute.

Based on the principle, the indexing rule \(R\) can be described in detail as equation (6).

\[
S = R(q) = \bigcap_{i=1}^{M} C^R_i
\]

\[
C^R_i = \{ c_i \in \mathbb{C} : c_i(q_i) \in (l_i - 1)\Delta\delta_i, l_i, l_i + \Delta\delta_i ] \}
\]

\[
\Delta\delta_i = \frac{1}{L_i}, L_i = 1, ..., L, i = 1, 2, ..., M
\]

(6)

Where, \(L_i\) denotes the number of intervals of the \(i\)-th attribute, \(L\) is the index number of \(L_i\).

In equation (6), each attribute is discretized by designating the intervals. The number of intervals of the nominal attributes equals to the number of discrete attribute values, while the number of intervals of the numerical attributes is predefined by the designer. Take the two dimensional indexing space as an example, \(M = 2\), \(c = [c_x, c_y]\), define \(L_x = L_y = 4\), then the discrete intervals of \(c_x\) are described as: \([0, 0.25], (0.25, 0.5], (0.5, 0.75], (0.75, 1]\); the corresponding subsets of cases are \(\{C_{x1}, C_{x2}, C_{x3}, C_{x4}\}\). The discrete intervals of \(c_y\) are described as: \([0, 0.25], (0.25, 0.5], (0.5, 0.75], (0.75, 1]\); and the corresponding subsets of cases are \(\{C_{y1}, C_{y2}, C_{y3}, C_{y4}\}\). The two dimensional indexing structure is given as in Figure 3. Suppose the query case falls in the interval with index 3 of the attribute \(c_x\), and the interval with the index 3 of the attribute \(c_y\), then the cases falling within the intersection space constitute the index set \(S\), that is \(S = C_{x3} \cap C_{y3} = \{c_4, c_7, c_12, c_15\}\).

D-HS method makes the similarity evaluation limited to the cases in the index set \(S\), thus it greatly promotes the retrieval efficiency. Besides, it is easy to keep the maintenance for the case base when new cases are added or removed. However, the indexing structure of D-HS in Figure 3 has the limitations in retrieval accuracy, which mainly come from two aspects: (1) while the interval width of each attribute is equally designated, the attribute values of cases cannot be evenly distributed. (2) when the location of the query case is close to the boundaries of intervals, the similar cases are easily neglected.

**Strategies to improve the D-HS method**

**Strategy that keep approximately equal number of cases for attribute intervals division.** From the equation (6), it can be seen that the widths of intervals of one certain attribute are equally divided, that is, each interval width of the \(i\)-th attribute is \(1/L_i\). However, it cannot make sure the even distribution of the \(i\)-th attribute values of cases in each interval, which significantly impacts the retrieval result. For instance, the most cases fall within one certain interval, while few cases are sparsely distributed in other intervals. In this context, the index set \(S\) may gather most of the cases, and thus it cannot achieve the effect of reducing the similarity calculation for searching the whole memory by the indexing approach. On the contrary, \(S\) may contain no one case.

Fayyad and Irani\(^{48}\) proposed a multi-interval discretized method for continuous numerical attributes based on the information entropy, and constructed the MDLP interval division criteria. Galushka and Patterson\(^{47}\) applied the above method to construct a new indexing structure, name D-HSE, and proposed a selection mechanism between the D-HSM and the D-HSE according to the distribution of the cases, the former has better retrieval efficiency, and the latter has higher accuracy. Based on the principle of D-HS
In equation (8), when \( q \) does not fall within the center of the intervals, as shown in Figure 3, the location of \( q \) is near to the boundaries of the intersection. In this scenario, cases \( c_1, c_3, c_5, \) and \( c_{19} \) are not included in the index set \( S \), however they are more closer to the query case \( q \) than some cases in \( S \), which results in the retrieval inaccuracy.

Brochu\textsuperscript{49} proposed a spherical indexing method for neighboring cases, such as the two dimensional attributes in Figure 3(b), it takes the location of query case \( q \) as the center to draw a circle, then cases in the circle constitute the index set \( S \). However, this query sphere algorithm is not applicable to the hyper polyhedron in case that the lower and upper boundary values of one or more attribute intervals are asymmetric with respect to the recommended points. In this paper, a simple and effective strategy of similarity matching for attributes is proposed to obtain the similar edge candidate cases.

Table 4. Attribute values matching between the candidate cases and query case.
In Table 4, each attribute value of case $c_3$ is matched with that of the query case $q$; cases $c_4$, $c_7$, $c_{10}$ all have three attribute values falling within the same intervals as that of the query case $q$, respectively; for case $c_6$, two attribute values are matched, that is, the values of attributes $A_2$ and $A_4$; cases $c_1$ and $c_8$ only have single attribute value matched with that of the query case $q$. It concludes that cases having more matched attribute values may close to the query case $q$, such as the cases $c_3$, $c_4$, $c_7$, and $c_{10}$. Based on this principle, the strategy of similarity matching for attributes can be quantitatively modeled as equation (10).

$$SMA(c_i, q) = \sum_{j=1}^{M} \text{Match}(c_j(c_i), q_j(q))$$  

(10)

Where, $SMA(c_i, q)$ indicates the number of matched attributes between case $c_i$ and the query case $q$; $\text{Match}()$ is a matching function, based on the strategy that keep approximately equal number of cases for attribute intervals division as expressed in equation (7), then the matching function is constructed as below.

If $c_j(c_i), q_j(q) \in [c_j(c((L_1-1)\Delta_0)), c_j(c(L_1\Delta_0))] \cup [c_j(c((L_2-1)\Delta_0)), 1],$

then $\text{Match} (c_j(c_i), q_j(q)) = 1;$

If $\{ [c_j(c_i) \in [c_j(c((L_1-1)\Delta_0)), c_j(c(L_1\Delta_0))] \cap [c_j(c_i) \notin [c_j(c((L_1-1)\Delta_0)), 1]]\} \cap \{ [q_j(q) \in [c_j(c((L_2-1)\Delta_0)), c_j(c(L_2\Delta_0))] \cap [q_j(q) \notin [c_j(c((L_2-1)\Delta_0)), 1]]\}^{-1}$

and $l_{j1} \neq l_{j2},$

then $\text{Match} (c_j(c_i), q_j(q)) = 0.$

Based on the strategy of similarity matching for attributes, a new index set of similar cases $S_b$ is constructed as in equation (11), which is a supplement to the index set $S$ obtained from equation (7), and it can effectively avoids the neglect of similar cases when the query case $q$ locates near the boundaries of the intervals by setting the threshold $\text{Const}$ of the number of matched attributes. In final, the index set of similar cases $S_0$ in equation (12) is obtained.

$$S_b = \{ c_i : SMA(c_i, q) \geq \text{Const}, \text{ Const} \leq M \}$$  

(11)

$$S_0 = \{ c_i : c_i \in S \cup S_b \}$$  

(12)

**Construction of the indexing structure for the low-carbon product case base**

Based on the improved D-HS method, the indexing structure of the product case base for low-carbon design is established as illustrated in Figure 4.

Step 1: sequence the cases for $M$ dimensional attributes, respectively; designate the number of intervals of each attribute; and set up the attributes coordinate system for the product case base, labeled as ((0, $L_1$), (0, $L_2$), (0, $L_3$), ...).

Step 2: designate the coordinates for each product case according to the values of $M$ dimensional attributes, labeled as ($l_i(c_i)$, $l_j(c_i)$, $l_k(c_i)$, ...).

Step 3: Input the requirement information, namely the values of $M$ dimensional attributes of the query case $q$, and designate the attribute coordinates for the query case, labeled as ($l_i(q)$, $l_j(q)$, $l_k(q)$, ...).

Step 4: generate the index set $S$, in which the coordinates of cases are same as that of the query case based on the indexing rule $S = R(q)$.

![Figure 4. The procedure of the construction of the indexing structure for the case base.](image-url)
Step 5: define the threshold Const and gather the cases which satisfy the condition $SIMA(c_i, q) \geq Const$, these cases constitute the index set $S_0$, and it obtains the final index set $S_0 = S \cup S_b$.

Step 6: evaluate the similarity and adaptability of each product case in $S_0$, and output the acceptable cases for knowledge reuse and adaptation. The evaluation models for the similarity and adaptability are described in the following sections.

Step 7: if the output case is successfully adapted to satisfy the design requirements, then it is considered as a new product case added to the case base with updated attribute coordinates.

Step 8: the total number of cases will increase when new adapted cases are continuously added into the case base, thus, if the current number of cases $N_1$ satisfies the condition $N_1 \geq (1 + 10\%) N$, then the coordinate system of the case base should be updated.

### Similarity and adaptability evaluation models for cases retrieval

#### Similarity evaluation

In this section, it establishes the similarity evaluation models for different kinds of attributes in consideration of the fuzzy requirement information for the cases in the index set $S_0$.

**Similarity evaluation model for the precise numerical attribute.** When the requirement information for one certain numerical attribute is ascertained, namely, the numerical attribute value of the query case $q$ is a crisp number. For the precise numerical attribute, the classical distance measurement model is used to calculate the local similarity as expressed in equation (13).

$$sim_1(c_j, q_j) = 1 - \frac{|c_j(c) - q_j(q)|}{D_j} \quad (13)$$

Where, $D_j = \max$ value – minimum value of the $j$-th attribute.

**Similarity evaluation model for the fuzzy numerical attribute.** When the requirement information for the numerical attribute is unascertained, that is, the numerical attribute value is a fuzzy number, and usually it is a range number. In this paper, it sets up the feasible interval and desired interval for each fuzzy numerical attribute to make the decision making of the parameters more flexible for the designer; and the dependent function presented in Extenics$^{19,50}$ is adopted to establish the similarity evaluation model for fuzzy numerical attributes, as expressed in equation (14).

$$sim_2(c_k, q_k) = \begin{cases} 1, & K(c_k(c)) \geq 1 \\ e^{K(c_k(c)) - 1}, & K(c_k(c)) < 1 \end{cases} \quad (14)$$

Where, $K(c_k(c))$ is the dependent function for the value of the $k$-th attribute of case $c_i$.

Suppose the feasible interval $X = [a, b]$, and the desired interval $X_0 = [c, d]$, satisfying $X_0 \subseteq X$; given the ideal point $x_0$ of interval $X_0$, then the dependent function of an arbitrary point $x$ is calculated as in equation (15), here, $x$ can be taken as the attribute value of case $c_i$, namely $x = e_K(c)$.

$$K(x) = \frac{\rho(x, x_0, X)}{\rho(x, x_0, X) - \rho(x, x_0, X_0)} \quad (15)$$

Where, $\rho(x, x_0, X)$ is named extension distance, which describes the distance of the point $x$ to interval $X$; and $\rho(x, x_0, X_0)$ is the distance of the point $x$ to interval $X_0$, accordingly. However, the calculation of the extension distance depends on the location of the ideal point $x_0$. Here, it takes the feasible interval $X$ as the example, when $x_0$ locates at the left side of the middle point of interval $X$, then it should calculate left side extension distance $\rho_l(x, x_0, X)$, otherwise, it should calculate the right side extension distance $\rho_r(x, x_0, X)$, as expressed in equation (16).

$$\begin{align*}
\rho_l(x, x_0, X) &= \begin{cases} a - x, & x \leq a \\
\frac{b - x_0}{a - x_0} (x - a), & x \in (a, x_0) \\
x - b, & x \geq x_0 \end{cases} \\
\rho_r(x, x_0, X) &= \begin{cases} a - x, & x \leq x_0 \\
a - x_0, & x \in (x_0, b) \\
\frac{b - x_0}{b - x} (b - x), & x \in (x_0, b) \\
x - b, & x \geq b \end{cases}
\end{align*} \quad (16)$$

The dependent function of each attribute value of cases has the properties as below:

- when $x \in X_0$, then $K(x) \geq 1$, it reveals that the attribute value completely meets the requirement;
- when $x \in X - X_0$, then $0 \leq K(x) < 1$, it indicates that the attribute value is acceptable;
- and when $x \notin X$, then $K(x) < 0$, it reveals that attribute value cannot satisfy the requirement.

**Similarity evaluation model for the nominal attribute.** The $l$-th nominal attribute value of case $c_i$ and query case $q$ are firstly transformed into the standardized triangular fuzzy number $STFN(c_i(c))$ and $STFN(q(q))$, respectively; and then the local similarity of the $l$-th nominal attribute is estimated based on the ratio of area of the common range $A_{\text{CR}}$ to the area of design range $A_{\text{STFN(DR)}}$, as expressed below.
then the information content \( I \) of the design system is the sum of all probabilities. If one or more probabilities are zero, then \( I \) approaches to the infinity; in this scenario, the design system will not work although other design attributes are completely satisfied. Conversely, if all the probabilities are equal to one, then \( I \) approaches to zero, and this is an ideal design system.

In a design project, the probability of success for the design task is given by the tolerance specified by the designer, that is, the design range (DR), and the tolerance that the product system is capable of satisfying, that is, the system range (SR). Suppose the functional requirement \( FR_k \) follows the uniform distribution, and the system probability density function (pdf) is clear, as shown in Figure 5, it illustrates that the intersection between the design range and system range is the region where the acceptable solution exists. Therefore, \( P_k \) can be expressed as equation (20).

\[
P_k = \frac{A_{CR}}{A_{SR}} = \frac{L_{CR}}{L_{SR}}
\]

Where, \( A_{CR} \), \( A_{SR} \) are the area of common range and system range respectively; and \( L_{CR} \), \( L_{SR} \) are the interval width of common range and system range; accordingly, the information content for the \( FR_k \) is written as below.

\[
I_k = \log_2 \left( \frac{L_{SR}}{L_{CR}} \right)
\]

When \( FR_k \) follows the random distribution, as shown in Figure 6, the probability of achieving the \( FR_k \) in the common range (CR) is expressed as equation (22).

\[
P_k = \int_{CR}^{CR'} P_s(FR_k) dFR_k
\]
Where, $CR_l$, $CR_u$ are the lower boundary and upper boundary of $CR$, respectively; $Ps$ denotes the system probability density function (pdf). In this context, the $A_{SR}$ is equal to one, thus, the information content $I_k$ for $FR_k$ is calculated as in equation (23).

$$I_k = \log_2 \left( \frac{1}{A_{CR}} \right)$$  \hspace{1cm} (23)

However, in the earlier design stage, the information of design system is always incomplete and unascertained, thus it is hard to obtain the precise probability density function of the design system. In this paper, we adopt the membership function to transform the DR and SR to the STFN(DR) and STFN(SR) with crisp numbers as illustrated in Figure 7, and thus, the information content $I_k$ is expressed as in equation (24).

$$I_k = \log_2 \left( \frac{A_{STFN(SR)}}{A_{CR}} \right)$$  \hspace{1cm} (24)

Where, $A_{STFN(SR)}$ is the area of STFN(SR).

According to the evaluation model, it is crucial to obtain the DR, SR, and $A_{CR}$ of each attribute for the information content calculation. Here, we also specify the rules for Benefit type and Cost type attributes to construct the design range.

For Benefit type attributes, that is, the attribute value is like “at least a”, or “more than a”, the design range can be constructed as in the Figure 8(a). where, $a' = DR_l$, $a'' = a^* = \max(U(x))$, $\mu(a') = 0$, $\mu(a^*) = 1$; $DR_l$ denotes the lower boundary of the design range, $U(x)$ denotes the whole design domain.

For Cost type attributes, that is, the attribute value is like “at most a”, or “less than a”, then the design range can be constructed as in the Figure 8(b). where, $a' = a'' = \min(U(x))$, $a^* = DR_u$, $\mu(a') = 1$, $\mu(a^*) = 0$; $DR_u$ denotes the upper boundary of the design range.

In addition, when the system range of the design attribute is a crisp value, that is, $x_j = c_j(c)$, in this scenario, the system range is constructed as in Figure 9, here, the Benefit type attribute is taken as the example; and the information content is calculated as in equation (25).52

$$I_k = \log_2 \left( \frac{A_{STFN(SR)}}{A_{CR}} \right)$$  \hspace{1cm} (24)
\[ I_j = \log_2 \left( \frac{1}{P(c_j(c_i))} \right), \quad P(c_j(c_i)) = u(x_j) \]

\[
I_j = \begin{cases} 
\frac{x_j - a_j}{a^u - a_j}, & \text{Benefit type attributes} \\
\frac{a^u - x_j}{a^u - a_j}, & \text{Cost type attributes}
\end{cases} \tag{25}
\]

**A case study**

**Low-carbon requirement analysis for the vacuum pump**

The CNC cutting machine is widely used to automatically cut the leather and fabric materials, however, unlike the rigid material processing, the materials should be fixed to the working table by the adsorption force with the pressure difference. In the design task, it needs to configure a suitable vacuum pump for the working requirement of the cutting machine in consideration of low-carbon impact for the environment throughout the product life cycle as the demand of the government policy.

For the task, we concludes twelve design requirements for the vacuum pump after detailed discussion with the staff from the cutting machine industry, which include traditional performance needs and low-carbon demand, and related quality characteristics, as illustrated in Figure 10. Based on the HoQ for the requirements analysis, designers can clearly reveals the important QCs and pay more attentions to these attributes during the redesign process, and this study takes the QCs with high relative weight as the retrieval attributes to obtain the similar cases knowledge.

There are 151 product cases in the case base, to obtain the similar cases for knowledge reuse and adaptation, it extracts nine QCs as the retrieval attributes, the total sum of their relative weight are more than 80%. The nine QCs are labeled as \( A_1-A_9 \): rate of suction and exhaust \((A_1)\), ultimate pressure \((A_2)\), rated power \((A_3)\), noise \((A_4)\), cost \((A_5)\), low-carbon and environmental impact \((A_6)\), failure rate \((A_7)\), maintenance convenience \((A_8)\), and material recycling and reuse \((A_9)\); \( A_1-A_5 \) are numerical attributes, \( A_6-A_9 \) are nominal attributes, and each attribute value is provided with the feasible interval, desired interval, and the recommended value according to the fuzzy low-carbon requirements, as shown in Table 5.

For example, the determination of the attribute value of the rate of suction and exhaust \( A_1 \), during the cutting process, the air in the working chamber needs to be exhausted within 5–10 s, and thus the value of \( A_1 \) can be derived from the equation (26).

\[
t = 2.3K \frac{V}{S} \log \left( \frac{P_0}{P} \right) \tag{26}
\]

Where, \( t \) is the time to exhaust the air from the working chamber, \( P_0 \) is the initial pressure of the working chamber, and \( P \) is the final pressure, \( V \) is the working chamber volume, \( S \) is the value of \( A_1 \), \( K \) is the coefficient. Based on the calculation result, the rate of the suction and exhaust varies in range [50, 250] L/s. However, for the practical situation, factors such as pumping efficiency, air leakage, cost, and the demand of energy saving need to be considered, thus, in addition to the feasible interval \( X = [50, 250] \) L/s, the desired interval is also set as \( X_0 = [120, 180] \) L/s, and the recommended value is provided \( x = 150 \) L/s.

The ultimate pressure \((A_2)\) is determined by the required pressure difference to produce the adsorption
force to guarantee that the leather and fabric materials are stably fixed during the cutting process. Suggestion knowledge from the cutting machine customers is that when the pressure difference is more than 10,000 Pa, there is no relative motion between the cutting materials. Thus the working pressure of the vacuum pump is less than 90,000 Pa (set 1 atm = 100 kPa); and according to the design principle that the ultimate pressure is one order of magnitude smaller than the working pressure, thus the required ultimate pressure should be less than 9000 Pa. It also should take account of the air leakage factor and the harmful friction in the cutting process caused by the excessive pressure difference; thus, in addition to the feasible interval \( X = [0, 9000] \) Pa, the desired interval is set as \( X_0 = [3000, 7000] \) Pa, and the recommended value is provided \( x = 5000 \) Pa.

For the nominal attributes, the vacuum pump is required with a good low-carbon and environmental performance, low failure rate, good property in maintenance, and high rate of material recycling and reuse. These linguistic terms for the requirements are transformed into the standardized triangular fuzzy numbers, and according to the actual needs from the working situation of the cutting machine, the feasible and desired intervals for attributes can also be designated along with the recommended values.

**Figure 10.** The mapping analysis for low-carbon requirements and QCs based on the HoQ.

**Generation of the index set of similar cases**

The indexing structure of the product case base is constructed based on the strategy that keep approximately equal number of cases for attribute intervals division. Values of each attribute are normalized and sequenced from smallest to largest, respectively. For attributes \( A_1 \) to \( A_6 \), the number of intervals are predefined as 5, and for attributes \( A_7 \) to \( A_9 \), the number of the intervals are predefined as 3, as illustrated in Figures 11 and 12.

Take \( A_1 \) as an example, attribute \( A_1 \) is split into five intervals, in the interval with index 1, there are 33 cases, the attribute value varies in a range \((0, 0.019)\); in the interval with index 2, there are 30 cases, the attribute value varies in a range \((0.019, 0.063)\); in the interval with index 3, there are 32 cases, the attribute value
varies in a range (0.063, 0.125); in the interval with index 4, there are 30 cases, the attribute value varies in a range (0.125, 0.25); and in the interval with index 5, there are 26 cases, the attribute value varies in a range (0.25, 1). The distribution of cases for each attribute is estimated based on the criterion in equation (8). It indicates the ratio of $s_i$ to $s_{\text{max}(i)}$, $\sigma^*$ is smaller than the threshold $\sigma^*$ of each attribute, especially for $A_7$-$A_9$, although many attribute values are overlapping, the splitting result still make sure these attributes belong to the “good” attributes with basically uniform distribution.

In order to make a clear comparison with the original method for intervals division in the D-HS, that is each interval width is equal to $1/L_i$, $L_i$ is the interval number of the $i$-th attribute. The result of the cases distribution is illustrated in Figure 13, where the interval number of each numerical attribute $A_1$-$A_5$ is predefined as 5, and the interval number of each nominal attribute $A_6$-$A_9$ is predefined as 10. It indicates that most of cases are falling within the interval with index 1 for $A_1$-$A_5$, and the uniformity of cases distribution of attributes $A_6$-$A_9$ is also not ideal, which will badly impact the accuracy of the similar cases retrieval.

The distribution of cases in Figures 11 and 12 demonstrates that the predefined interval number of each attribute is acceptable, and accordingly, the
coordinates system of the case base can be established as shown in Table 6; and then the coordinates of each case should be designated according to the attribute values. Take the product case c7 as the example, the coordinates are given as described in Table 7.

The query case is the target case in the retrieval algorithm, the coordinates of query case are also designated as shown in Table 8 based on the recommended value of each attribute provided in Table 5.

The cases labeled with the same coordinates as the query case are extracted to constitute the index set \( S \), and the number of attributes which are matched with the query case \( q \) for each case \( c_i \), \( S.M.A(c_i, q) \) is estimated, as shown in Table 9. From Table 9, it indicates that there is no one case can completely match with the query case in each attribute, that is, the index set \( S \) is an empty set. In this situation, the threshold Const is set, \( Const = 4 \), and to obtain the index set \( S_b = \{ c \} \):
Table 8. The coordinates of \( l(q) \) of the query case \( q \).

| Attributes | \( A_1 \) | \( A_2 \) | \( A_3 \) | \( A_4 \) | \( A_5 \) | \( A_6 \) | \( A_7 \) | \( A_8 \) | \( A_9 \) |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Values     | 150      | 5000     | 15       | 75       | 8000     | 0.75     | 0.30     | 0.75     | 0.75     |
| Normalization | 0.188    | 0.25     | 0.188    | 0.370    | 0.075    | 0.75     | 0.30     | 0.75     | 0.75     |
| \( l(q) \)  | 4        | 5        | 4        | 3        | 3        | 4        | 2        | 3        | 3        |

**Estimation of the similarity and adaptability**

When the index set \( S_0 \) is generated, cases in the \( S_0 \) still need to be estimated from two perspectives, the similarity and adaptability according to the low-carbon requirements. For the estimation of similarity, \( A_1 \)–\( A_5 \) belong to fuzzy numerical attributes, thus it adopts the dependent function to calculate the local similarity of each attribute; and for the nominal attributes \( A_6 \)–\( A_9 \), the estimation model based on the membership function as constructed in equation (17) is used to calculate the local similarity of each attribute. The result of the similarity estimation for each case is listed in Table 11.

Estimation of the adaptability for cases is based on the information content of each attribute, which is determined by the design range and system range. The design range of each attribute is obtained from the requirement analysis, as explained in Table 5, and the range values of numerical attributes need to be transformed into the standardized triangular fuzzy number. The system range of each attribute reveals the maximum range for adaptation of the series product cases under the existing technical conditions. For example, the ultimate pressure of the SK series vacuum pump is generally at the range 6000–8000 Pa, but it can reach 3000 Pa or lower by installing a fore-stage nozzle. Here, it takes case \( c_{25} \) as the example to describe the estimation of the adaptability for each attribute.

In Table 12, the normalized attribute values of design range and system range of the case \( c_{25} \) are provided. For attributes \( A_1 \) (rate of suction and exhaust), \( A_2 \) (ultimate pressure), \( A_3 \) (rated power), they follow the fuzzy triangular distribution; \( A_2 \) belongs to the Cost type attribute, and the system range is a crisp value, thus the system area of \( A_2 \) is constructed based on the principle in Figure 9, here with the Cost type attribute. In Figure 14, the calculation models of information content \( I_{A_j} \) with respect to \( A_1, A_2, A_3 \) of case \( c_{25} \) are illustrated, respectively:

\[
I_{A_1} = \log_2(1/ (A_{CR} / A_{STFN(SR)})) = \log_2(1/0.28) = 1.836;
\]
### Table 10. Cases in the index set $S_0$.

| Cases/ci | $A_1/(L/s)$ | $A_2/Pa$ | $A_3/kW$ | $A_4/dB$ | $A_5/Yuan$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ |
|----------|-------------|-----------|-----------|-----------|-------------|-------|-------|-------|-------|
| C7       | 150         | 6666      | 15        | 74        | 3200        | 0.75* | 0.3*  | 0.75* | 0.75* |
| C6       | 100         | 6666      | 11        | 72        | 2500        | 0.75* | 0.3*  | 0.75* | 0.75* |
| C8       | 200         | 6666      | 18.5      | 75        | 4900        | 0.75* | 0.3*  | 0.75* | 0.75* |
| C39      | 138         | 3300      | 15        | 74        | 6800        | 0.75* | 0.3*  | 0.65* | 0.75* |
| C43      | 111         | 3300      | 11        | 73        | 5500        | 0.75* | 0.3*  | 0.65* | 0.75* |
| C44      | 138         | 3300      | 15        | 74        | 7200        | 0.75* | 0.3*  | 0.65* | 0.75* |
| C38      | 111         | 3300      | 11        | 73        | 4050        | 0.75* | 0.3*  | 0.65* | 0.75* |
| C9       | 250         | 6666      | 30        | 78        | 6000        | 0.7*  | 0.3*  | 0.75* | 0.75* |
| C17      | 100         | 3333      | 15        | 74        | 2750        | 0.65* | 0.3*  | 0.75* | 0.75* |
| C23      | 95          | 3300      | 11        | 76        | 3800        | 0.75* | 0.25* | 0.75* | 0.75* |
| C134     | 70          | 2600      | 7.5       | 75        | 7550        | 0.5*  | 0.3*  | 0.75* | 0.75* |
| C5       | 50          | 6666      | 5.5       | 67        | 1950        | 0.75* | 0.25* | 0.75* | 0.75* |
| C18      | 200         | 3333      | 22        | 78        | 5400        | 0.65* | 0.3*  | 0.75* | 0.75* |
| C22      | 65          | 3300      | 7.5       | 72        | 3600        | 0.75* | 0.25* | 0.75* | 0.75* |
| C24      | 113         | 3300      | 15        | 80        | 4500        | 0.75* | 0.25* | 0.7*  | 0.75* |
| C25      | 130         | 3300      | 15        | 80        | 4850        | 0.75* | 0.25* | 0.7*  | 0.75* |
| C47      | 191         | 8106      | 22        | 82        | 9400        | 0.65* | 0.3*  | 0.7*  | 0.65* |
| C87      | 150         | 1        | 15        | 78        | 85000       | 0.85* | 0.3*  | 0.65* | 0.75* |
| C114     | 120         | 10        | 15        | 78        | 85000       | 0.85* | 0.3*  | 0.7*  | 0.75* |
| C123     | 150         | 0.05     | 2.2       | 80        | 6200        | 0.7*  | 0.3*  | 0.75* | 0.7*  |
| C130     | 166         | 5489.9    | 15        | 82        | 7150        | 0.7*  | 0.3*  | 0.7*  | 0.7*  |
| C133     | 50          | 2600      | 5.5       | 72        | 5200        | 0.5*  | 0.3*  | 0.75* | 0.75* |
| C144     | 150         | 2600      | 11        | 80        | 14000       | 0.7*  | 0.3*  | 0.7*  | 0.7*  |
| C145     | 200         | 2600      | 15        | 80        | 18300       | 0.7*  | 0.3*  | 0.7*  | 0.75* |
| C150     | 50          | 300       | 7.5       | 76        | 8200        | 0.7*  | 0.3*  | 0.7*  | 0.75* |

Note: $0.25^* = (0.1, 0.25, 0.4); 0.3^* = (0.2, 0.3, 0.4); 0.35^* = (0.2, 0.35, 0.5); 0.5^* = (0.4, 0.5, 0.6); 0.65^* = (0.5, 0.65, 0.8); 0.7^* = (0.6, 0.7, 0.8); 0.75^* = (0.65, 0.75, 0.85); 0.85^* = (0.75, 0.85, 0.95).

### Table 11. The result of similarity estimation for cases in $S_0$.

| $c_i$ | $\text{sim}_3(A_i, q_i), i = 1, 2, 3, 4, 5$ | $\text{sim}_3(A_i, q_i), i = 6, 7, 8, 9$ | $\text{SIM}(c_i, q)$ |
|-------|------------------------------------------|------------------------------------------|---------------------|
| C25   | 1.000                                   | 1.000                                    | 1.000               |
| C23   | 0.700                                   | 1.000                                    | 1.000               |
| C39   | 1.000                                   | 1.000                                    | 1.000               |
| C44   | 1.000                                   | 1.000                                    | 1.000               |
| C24   | 0.912                                   | 1.000                                    | 1.000               |
| C7    | 1.000                                   | 1.000                                    | 1.000               |
| C43   | 0.882                                   | 1.000                                    | 1.000               |
| C38   | 0.882                                   | 1.000                                    | 1.000               |
| C8    | 0.753                                   | 1.000                                    | 1.000               |
| C114  | 1.000                                   | 0.369                                    | 1.000               |
| C6    | 0.753                                   | 1.000                                    | 1.000               |
| C18   | 0.753                                   | 1.000                                    | 1.000               |
| C17   | 0.751                                   | 1.000                                    | 1.000               |
| C144  | 1.000                                   | 0.875                                    | 1.000               |
| C130  | 1.000                                   | 1.000                                    | 1.000               |
| C22   | 0.452                                   | 1.000                                    | 1.000               |
| C145  | 0.753                                   | 0.875                                    | 1.000               |
| C9    | 0.368                                   | 1.000                                    | 1.000               |
| C47   | 0.844                                   | 0.641                                    | 0.827               |
| C134  | 0.490                                   | 0.875                                    | 0.368               |
| C87   | 1.000                                   | 0.368                                    | 0.223               |
| C133  | 0.368                                   | 0.875                                    | 0.208               |
| C123  | 1.000                                   | 0.368                                    | 0.082               |
| C150  | 0.368                                   | 0.407                                    | 0.368               |
| C5    | 0.368                                   | 1.000                                    | 0.208               |
$\log_2\left(\frac{1}{A_{CR}/A_{STFN(SR)}}\right) = \log_2(0.747) = 0.421$;
$\log_2\left(\frac{1}{A_{CR}/A_{STFN(SR)}}\right) = \log_2(0.679) = 0.558$.

For attributes $A_4$ (noise), $A_5$ (cost), they follow the uniform distribution, the calculation models of the information content for $A_4$ and $A_5$ are illustrated in Figure 15.

\begin{align*}
I_{A_4} &= \log_2\left(\frac{1}{L_{CR}/L_{SR}}\right) = \log_2(1/0.20) = 2.322;
I_{A_5} &= \log_2\left(\frac{1}{L_{CR}/L_{SR}}\right) = \log_2(1/0.727) = 0.460.
\end{align*}

For nominal attributes $A_6$-$A_9$, the linguistic term of each attribute value is transformed into the standardized triangular fuzzy number, and the calculation models of the information content for $A_6$-$A_9$ are illustrated in Figure 16.

\begin{align*}
I_{A_6} &= \log_2\left(\frac{1}{A_{CR}/A_{STFN(SR)}}\right) = \log_2(1/0.143) = 2.806;
I_{A_7} &= \log_2\left(\frac{1}{A_{CR}/A_{STFN(SR)}}\right) = \log_2(1/0.296) = 1.756;
I_{A_8} &= \log_2\left(\frac{1}{A_{CR}/A_{STFN(SR)}}\right) = \log_2(1/0.036) = 4.796;
I_{A_9} &= \log_2\left(\frac{1}{A_{CR}/A_{STFN(SR)}}\right) = \log_2(1/0.143) = 2.806.
\end{align*}

Then the information content of case $c_{25}$ is obtained:

\begin{align*}
I_{c_{25}} &= v_{A_1}, v_{A_2}, v_{A_3}, v_{A_4}, v_{A_5}, v_{A_6}, v_{A_7}, v_{A_8}, v_{A_9}(\frac{1}{C_{15}})IA_{1}, IA_{2}, IA_{3}, IA_{4}, IA_{5}, IA_{6}, IA_{7}, IA_{8}, IA_{9}(\frac{T}{C_{15}}) = 1.836, 0.421, 0.558, 2.322, 0.460, 2.806, 1.756, 4.796, 2.806^T = 1.890.
\end{align*}

The information content of the remaining cases in $S_0$ also can be calculated with the same method, the result is shown in Table 13. In addition, in order to consider the comprehensive influence of similarity and adaptability of the product cases, the comparison analysis histogram of the similarity and information content of each case in the index set $S_0$ is generated, as illustrated in Figure 17.

In Figure 17, it indicates that cases $c_{25}$, $c_{23}$, $c_{39}$, and $c_{44}$ have high similarity and small information content;
and for the case c7, although the similarity is not high, the information content is small; therefore, these cases are considered as the similar cases knowledge for reuse. For cases c43, c38, c8, although the similarity is not low, the information content is relative large; however, the information content of cases c114, c47, c87, c133, c123 are all approaching infinity, it indicates that there are one or more requirement attributes cannot be satisfied by adaptation under the current technical conditions.

Meanwhile, the local attribute similarity and information content can guide the designer to make targeted improvement for the candidate cases, and it also could indicate the direction of technology updates to some extent. For instance, the attributes $A_4$ (noise) and $A_8$ (maintenance convenience) of c25 cannot be completely satisfied, and the related information content of $A_4$ and $A_8$ are relative high. The designer maybe adopt new scheme to enhance the local similarity rather than make adaption on the original structure, such as install the noise reduction device, and replace the mechanical sealing component with a simple and effective packing sealing component to reduce the failure rate and facilitate the maintenance. Case c114 is a screw vacuum pump, which has good environmental performance and high work efficiency, but its cost is far beyond the scope of

Figure 15. Information content estimation for attributes $A_4$ and $A_6$: (a) information content $l_{A_4}$ and (b) information content $l_{A_6}$.

Figure 16. Information content estimation for attributes $A_6$-$A_9$: (a) information content $l_{A_6}$, (b) information content $l_{A_7}$, (c) information content $l_{A_8}$, and (d) information content $l_{A_9}$. 
the cost constraint; and thus reducing the cost with new technologies is a key issue. The information content of attribute $A_6$ (low-carbon and environmental impact) of cases $c_{87}, c_{133}, c_{134}$ are all approaching infinity, it reveals that the traditional technology can no longer meet the customer’s low-carbon and environmental needs, and it is urgent to implement the green manufacturing technologies to update the product cases to meet the market demand and the government environmental regulations.

**Table 13.** The result of information content estimation for cases in $S_0$.

| $c_i$ | $I_{A1}$ | $I_{A2}$ | $I_{A3}$ | $I_{A4}$ | $I_{A5}$ | $I_{A6}$ | $I_{A7}$ | $I_{A8}$ | $I_{A9}$ | $I_{A10}$ |
|------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| $c_{25}$ | 1.836 | 0.421 | 0.558 | 2.322 | 0.460 | 2.806 | 1.756 | 4.796 | 2.806 | 1.890 |
| $c_{23}$ | 2.988 | 0.421 | 1.910 | 1.114 | 0.460 | 2.806 | 1.756 | 2.806 | 2.806 | 1.986 |
| $c_{39}$ | 1.966 | 0.421 | 0.558 | 0.737 | 0.615 | 2.806 | 3.000 | 5.573 | 2.806 | 1.970 |
| $c_{44}$ | 1.966 | 0.421 | 0.558 | 0.737 | 0.474 | 2.806 | 3.000 | 5.573 | 2.806 | 1.956 |
| $c_{24}$ | 2.988 | 0.421 | 0.558 | 1.806 | 0.460 | 2.806 | 1.756 | 4.796 | 2.806 | 2.020 |
| $c_7$ | 1.221 | 2.556 | 0.558 | 1.586 | 0.480 | 2.806 | 3.000 | 2.806 | 2.806 | 1.913 |
| $c_{43}$ | 3.059 | 0.421 | 1.910 | 1.224 | 0.571 | 2.806 | 3.000 | 5.573 | 2.806 | 2.352 |
| $c_{38}$ | 3.059 | 0.421 | 1.910 | 1.000 | 0.648 | 2.806 | 3.000 | 5.573 | 2.806 | 2.342 |
| $c_8$ | 3.023 | 2.556 | 0.548 | 2.124 | 0.460 | 2.806 | 3.000 | 2.806 | 2.806 | 2.149 |
| $c_{114}$ | 2.857 | 1.977 | 0.737 | 0.806 | 3.000 | 4.796 | 2.806 | 2.806 | 2.806 | 2.419 |
| $c_6$ | 3.265 | 2.556 | 1.910 | 1.224 | 0.927 | 2.806 | 3.000 | 2.806 | 2.806 | 2.307 |
| $c_{18}$ | 3.023 | 0.421 | 1.514 | 1.737 | 0.460 | 4.158 | 3.000 | 2.806 | 2.806 | 2.286 |
| $c_{17}$ | 3.265 | 0.421 | 0.910 | 1.586 | 0.808 | 4.158 | 3.000 | 2.806 | 2.806 | 2.286 |
| $c_{144}$ | 1.224 | 0.816 | 1.910 | 1.486 | 1.552 | 4.796 | 3.000 | 4.796 | 2.806 | 2.493 |
| $c_{130}$ | 1.214 | 0.396 | 0.910 | 3.699 | 1.351 | 4.796 | 3.756 | 4.796 | 4.796 | 2.728 |
| $c_{22}$ | 3.427 | 0.421 | 2.434 | 1.415 | 0.619 | 2.806 | 1.756 | 2.806 | 2.806 | 2.163 |
| $c_{145}$ | 3.023 | 0.816 | 0.558 | 1.486 | 2.523 | 4.796 | 3.000 | 4.796 | 2.806 | 2.679 |
| $c_9$ | 3.857 | 2.556 | 3.878 | 1.224 | 0.514 | 4.796 | 3.000 | 2.806 | 2.806 | 3.044 |
| $c_{47}$ | 2.699 | $\alpha$ | 3.718 | 2.322 | 0.876 | 5.573 | 3.000 | 4.796 | 5.573 | $\alpha$ |
| $c_{134}$ | 3.251 | 0.816 | 2.434 | 1.221 | 0.533 | $\alpha$ | 3.000 | 2.806 | 2.806 | $\alpha$ |
| $c_{87}$ | 0.941 | 1.977 | 0.558 | 3.171 | 3.699 | $\alpha$ | 3.000 | 5.573 | 2.806 | $\alpha$ |
| $c_{133}$ | 3.836 | 0.816 | 0.558 | 0.460 | 2.806 | 3.000 | 4.796 | 2.806 | 2.806 | $\alpha$ |
| $c_{123}$ | 0.941 | 1.977 | $\alpha$ | 2.322 | 0.456 | 4.796 | 3.000 | 2.806 | 4.796 | $\alpha$ |
| $c_{150}$ | 3.836 | 1.862 | 1.565 | 1.000 | 0.626 | 4.796 | 3.000 | 4.796 | 2.806 | 2.810 |
| $c_5$ | 3.836 | 2.556 | 3.012 | 1.114 | 1.806 | 2.806 | 1.756 | 2.806 | 2.806 | 2.629 |

**Note:** $\alpha$ denotes that the corresponding information content approaches infinity.

**Figure 17.** Comparison analysis of the similarity and information content for cases in $S_0$.

**Discussion**

In this paper, an effective similarity determination model which integrates the representation of fuzzy requirement information, the efficient indexing approach, and the evaluation method for the similarity and adaptability into a holistic similar cases retrieval methodology is proposed to support low-carbon product design. A case study concerning the similar cases retrieval for the vacuum pump low-carbon design is conducted to verify the applicability of the proposed
methodology; and the theoretical and practical implications of our research work are discussed below.

In the earlier design stage, designers can basically specify the accurate values for the very familiar requirement attributes. However, when taking account of the low-carbon requirements, it requires the reconfiguration for corresponding quality characteristics, which possibly contributes to complex design conflicting problems as the coupling property between the quality characteristics, namely, one design factor about the low-carbon requirements is improved, while the related factor is deteriorated. Consequently, even if there are experienced domain experts for the design task, sometimes they cannot grasp deep knowledge and precisely address the complex low-carbon design problems with incomplete design information, but they can roughly know the possible range of design attribute values, and the tolerable differences between the query case attribute values and that of the candidate cases to ensure the designated parameter ranges are still satisfying the design purpose. Therefore, for the fuzzy requirement attributes, this study adopts the fuzzy set theory to represent the incomplete and uncertain design information, and it defines the feasible interval and desired interval to make flexible for the designers in decision-making to the design parameters.

For the improved D-HS method, it adopts the recommended attribute values to assign the indexing coordinates for the query case rather than uses the design range of each attribute to construct the index set mainly due to the following considerations: if it uses design ranges of attributes to search the similar cases, then it must identify all cases in the memory whether they are falling within the index set every time with a new requirement, and thus it loses the meaning of the intervals division. Conversely, with the recommended values, the coordinates of the query case can be designated, and cases having the same coordinates with that of the query case are extracted to build the index set; in addition, when the coordinates of the current query case are same as that of the previous one, similar cases in the index set saved in the database can be immediately reused.

In our study, we propose the strategy that the number of cases is approximately equal for each interval of attributes to make sure the cases can be evenly distributed. In case study, it predefine the interval number of the i-th attribute \( L_i \), however, the determination of \( L_i \) is actually an iterative process to satisfy the criteria in equation (8). Besides, for two neighboring cluster cases, if the average values of the i-th attribute of the two clusters are quite different, then it is necessary to construct the attribute interval for two clusters separately, and re-designate the \( L_i \) even though only the sum of the number of cases of two clusters can satisfy the condition of uniform distribution of the cases. The strategy of similarity matching for attributes proposed in the paper provides a supplementary index set \( S_0 \) by defining the required number of matched attributes, namely the threshold Const, which effectively avoids the omission of the similar cases. For example, in the case study, cases \( c_{25}, c_{24}, c_{23} \) are included in the initial index set \( S_0 \), the number of the matched attributes of these three cases are not high, they can be seen as the edge cases as illustrated in Figure 3(b); although they do not fall within the same intervals with the query case, they have the high similarity, and luckily, they are not neglected.

The effective estimation models of the similarity and adaptability of the cases are respectively established in our study. For the similarity calculation of precise numerical attributes, it adopts the traditional distance measurement to calculate the local similarity; for the fuzzy numerical attributes, it uses the dependent function to construct the local similarity evaluation model, as it can better distinguish the degree to which the attribute values of the candidate case meet the design requirements. That is, when the value of dependent function of the attribute is equal or more than one, it indicates the attribute value is completely meet the demand; when the value of dependent function is between zero and one, then the attribute value is acceptable; while when the value of dependent function is less than zero, the attribute value cannot be satisfied. For the nominal attributes, the traditional method is to use the binary estimation model, 0 and 1. For instance, the linguistic term of the attribute low-carbon and environmental impact \( A_6 \) of the query case is required “good”, when the linguistic term of \( A_6 \) of the candidate case is “good”, then the similarity is equal to 1, otherwise is 0; however, when the linguistic term of \( A_6 \) of the candidate case is “very good”, “fair”, “poor”, or “very poor”, it cannot clearly distinguish the similarity of these values. In our study, the linguistic terms of the nominal attribute are transformed into the standardized fuzzy triangular numbers, and the similarity estimation model is constructed based on the ratio of area of the common range to the area of the design range.

The estimation model of the adaptability of the candidate case is proposed based on the information content presented in the Information Axiom. For one certain attribute of the case, the smaller the information content, the better the adaptability performance, because the system range of this attribute has a high degree of overlap with the design range, which indicates it has a high probability to meet the design requirements through modification; and the information content of the case is the sum of the information content of each attribute. To estimate the information content of each attribute, it is important to identify its system range that is based on the existing manufacturing technologies, design specification, related laws and regulations, etc. Thus it is hard to delineate a clear system
range of the attribute with the incomplete design information. In this study, for the numerical attribute, the upper and lower boundary values that the attribute can reach under the current technical conditions of the series product cases are defined as the system range. For example, the rate of suction and exhaust of the SK series vacuum pump can be promoted by enhancing the rated power or increasing the number of impellers, and thus to obtain the maximum and minimum value of the rate of suction and exhaust to construct the system range. For the nominal attribute, the system range is defined by evaluating the corresponding sub-factors. For instance, to define the system range of the attribute \( A_6 \), it needs to evaluate its sub-factors, such as the life cycle carbon footprint of the product, assemble and disassemble for the components reuse, the material recycling and disposal methods; the linguistic term of the attribute is finally obtained by scoring each sub-factor, and then the system range of the attribute is specified by comparing series product cases.

**Conclusion and further research**

In low-carbon product design, it takes account of the low-carbon requirements at preliminary design stage in consideration of the reduction of the carbon emissions throughout the product life cycle. However, as the coupling properties between the quality characteristics, it inevitably causes the design conflicting problems when incorporating the low-carbon indicators, which needs to go through multiple backtracking design processes to coordinate the contradictions. Thus, it is a good strategy for exploiting the prior design knowledge and experience to solve the similar design problems. In our study, we propose an effective similarity determination method to obtain acceptable cases knowledge for reuse, similarity here is not only referred to the similarity between the attribute values of the query case and that of the candidate case, but the similarity of the system range of the candidate case and the design range of the query case.

The incomplete and uncertain requirements information is represented based on the fuzzy set theory. The efficient D-HS indexing approach is improved with two valid strategies, the strategy that keep approximately equal number of the cases for intervals division, and the strategy of similarity matching for attributes; the improved D-HS method can effectively guarantee the basically even distribution of cases in each attribute interval, and avoid the omission of the similar cases when the location of the query case is close to the boundaries of the design ranges. In terms of the similarity estimation, it adopts the classical distance measurement and dependent function to calculate the local similarity for the precise numerical attribute and the fuzzy numerical attribute; and for the fuzzy nominal attribute, the similarity is based on the ratio of area of the common range to the area of the design range. For the adaptability estimation, it adopts the information content presented in the Information Axiom to construct the evaluation model for the adaptability of each attribute, the smaller the total information content, the better the adaptability of the candidate case. Thus, the cases having high similarity and small information content can be considered as the acceptable cases for knowledge reuse and adaptation.

In our proposed methodology, the information content of each attribute depends on the corresponding design range and system range, in the further research, it is necessary to use the sensitivity analysis and Monte Carlo analysis methods to more reasonably determine the boundaries of the design range and system range of each attribute, as well as the distribution of its fuzzy membership function. In addition, it is essential to adopt the contradictory problems solving methods, such as the TRIZ, Extensis, to address the design conflicts during the adaptation for the similar cases, which facilitate the generation of the low-carbon innovative design schemes.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this research work was supported by the National Natural Science Foundation of China (Grant No. 51875524 and No. 61873240) and the Zhejiang Province Postdoctoral Research Foundation (Grant No. zj2019130).

**ORCID iDs**

Shedong Ren https://orcid.org/0000-0002-4466-6192  
Fangzhi Gui https://orcid.org/0000-0002-8779-192X  
Yanwei Zhao https://orcid.org/0000-0002-0810-4563  
Min Zhan https://orcid.org/0000-0002-7064-283X  
Wanliang Wang https://orcid.org/0000-0002-3408-3321

**References**

1. Huisingh D, Zhang Z, Moore JC, et al. Recent advances in carbon emissions reduction: policies, technologies, monitoring, assessment and modeling. *J Clean Prod* 2015; 103: 1–12.

2. Kellens K, Dewulf W, Overcash M, et al. Methodology for systematic analysis and improvement of manufacturing unit process life cycle inventory (UPLCI) CO2PE! initiative (cooperative effort on process emissions in
manufacturing). Part 2: case studies. *Int J Life Cycle Ass* 2011; 17: 242–251.

3. Wang Q, Tang DB, Yin LL, et al. Bi-objective optimization for low-carbon product family design. *Robot CIM Int Manuf* 2016; 41: 53–65.

4. Kuo TC. The construction of a collaborative framework in support of low carbon product design. *Robot CIM Int Manuf* 2013; 29: 174–183.

5. He B, Zhang D, Gu ZC, et al. Skeleton model-based product low carbon design optimization. *J Clean Prod* Epub ahead of print 2020. DOI: 10.1016/j.jclepro.2020.121687.

6. Chen PK, Fortuny Santos J, Lujan I, et al. Sustainable manufacturing: exploring antecedents and influence of Total Productive Maintenance and lean manufacturing. *Adv Mech Eng* 2019; 11: 1–16.

7. BSI. Guide to PAS 2050 how to assess the carbon footprint of goods and services. British Standards Institution, London, 2008.

8. Jeswiet J and Kara S. Carbon emissions and CES™ in manufacturing. *CIRP Ann: Manuf Techn* 2008; 57: 17–20.

9. He B, Pan QJ and Deng ZQ. Product carbon footprint for product lifecycle under uncertainty. *J Clean Prod* 2018; 187: 459–472.

10. Sakao T. A QFD-centred design methodology for environmentally conscious product design. *Int J Prod Res* 2007; 45: 4143–4162.

11. Peng J, Li WQ, Li Y, et al. Innovative product design method for low-carbon footprint based on multi-layer carbon footprint information. *J Clean Prod* 2019; 228: 729–745.

12. Xu ZZ, Wang YS, Teng ZR, et al. Low-carbon product multi-objective optimization design for meeting requirements of enterprise, user and government. *J Clean Prod* 2015; 103: 747–758.

13. Zhang C, Huang HH, Zhang L, et al. Low-carbon design of structural components by integrating material and structural optimization. *Int J Adv Manuf Tech* 2018; 95: 4547–4560.

14. Gui FZ, Ren SD, Zhao YW, et al. Activity-based allocation and optimization for carbon footprint and cost in product lifecycle. *J Clean Prod* Epub ahead of print 2019. DOI: 10.1016/j.jclepro.2019.117627.

15. Song JS and Lee KM. Development of a low-carbon product design system based on embedded GHG emissions. *Resour Conserv Recy* 2010; 54: 547–556.

16. Devanathan S, Ramanujan D, Bernstein WZ, et al. Integration of subsystems into early design through the function impact matrix. *J Mech Design* 2010; 132: 1–8.

17. Chechurin L and Borgianni Y. Understanding TRIZ through the review of top cited publications. *Comput Ind* 2016; 82: 119–134.

18. Chou JR. An ARIZ-based life cycle engineering model for eco-design. *J Clean Prod* 2014; 66: 210–223.

19. Yang CY and Cai W. *Extenics: theory, method and application*. Beijing, China: Science Press, 2013.

20. Chen J, Zhao YW, Li FY, et al. Transforming bridge-based conflict resolution for product green design. *J Mech Eng* 2010; 46: 132–142 (in Chinese).

21. Li WJ, Song ZH, Mao ER, et al. Using Extenics to describe coupled solutions in Axiomatic design. *J Eng Design* 2018; 30: 1–31.

22. Popoff A and Millet D. Sustainable life cycle design using constraint satisfaction problems and quality function deployment. In: *24th CIRP conference on life cycle engineering*; Kamakura, Japan, 8–10 March 2017, pp.75–80. Amsterdam, Netherlands: Elsevier Ltd.

23. Chen L, Wang W and Huang B. A negotiation methodology for multidisciplinary collaborative product design. *Adv Eng Inform* 2014; 28: 469–478.

24. Dong YF, Wei XN and Liu FR. Novel collaborative optimization framework with a negotiation model for satellite system design. *Eng Optimiz* 2018; 50: 1395–1414.

25. Ren SD, Gui FZ, Zhao YW, et al. Accelerating preliminary low-carbon design for products by integrating TRIZ and Extenics methods. *Adv Mech Eng* 2017; 9: 1–18.

26. Aamodt A and Plaza E. Case-based reasoning - Foundational Issues, methodological variations, and system approaches. *AI Commun* 1994; 7: 39–59.

27. Zhang JH, Wang FL and Wang CK. Integrating case-based with rule-based reasoning in body-in-white fixture design. *Int J Adv Manuf Tech* 2015; 85: 1807–1824.

28. Patterson DW, Rooney N and Galushka M. Efficient similarity determination and case construction techniques for case based reasoning. In: *6th European conference on case-based reasoning*, Robert Gordon Univ, Aberdeen, Scotland, 4–7 September 2002, pp.292–305. Berlin: Springer-Verlag.

29. Smyth B and Keane MT. Adaptation-guided retrieval: questioning the similarity assumption in reasoning. *Artif Intell* 1999; 102: 249–293.

30. Stéphane N, Hector R and Marc LLJ. Effective retrieval and new indexing method for case based reasoning: application in chemical process design. *Eng Appl Artif Intell* 2010; 23: 880–894.

31. Cost S and Salzberg S. A weighted nearest neighbor algorithm for learning with symbolic features. *Mach Learn* 1993; 10: 57–78.

32. Diab DM and El Hindi K. Using differential evolution for improving distance measures of nominal values. *Appl Soft Comput* 2018; 64: 14–34.

33. Hjaltason GR and Samet H. Index-driven similarity for approximate similarity queries. *Data Knowl Eng* 1998; 102: 917–1013.

34. Polat K. Similarity-based attribute weighting methods via clustering algorithms in the classification of imbalanced medical datasets. *Neural Comput Appl* 2018; 28: 517–580.

35. Zhao YW, Ren SD, Chen WG, et al. Extension classifier construction based on improved BP neural network. *Comput Integr Manuf Syst* 2015; 21: 2807–2815 (in Chinese).

36. Bueno R, Traina AJM and Traina C. Genetic algorithms for approximate similarity queries. *Data Knowl Eng* 2007; 62: 459–482.

37. Polat K. Similarity-based attribute weighting methods via clustering algorithms in the classification of imbalanced medical datasets. *Neural Comput Appl* 2018; 30: 987–1013.

38. Liu W, Tan RH, Cao GZ, et al. Creative design through knowledge clustering and case-based reasoning. *Eng Comput* 2019; 36: 527–541.

39. Leake DB, Kinley A and David W. Case-based similarity assessment: estimating adaptability from experience. In: *Proceedings of the fourteenth national conference on artificial intelligence*. Menlo Park, CA: AAAI Press, 2001; 190–197.
artificial intelligence. Providence, Rhode Island, USA, 27–31 July 1997, pp.674–679. Menlo Park, California: AAAI Press.

39. Rousu J and Aarts RJ. Adaptation cost as a criterion for solution evaluation. In: 3rd European workshop on case-based reasoning, Lausanne, Switzerland, 14–16 November 1996, pp.354–361. Berlin: Springer-Verlag.

40. Roldan Reyes E, Negny S, Cortes Robles G, et al. Improvement of online adaptation knowledge acquisition and reuse in case-based reasoning: application to process engineering design. Eng Appl Artif Intel 2015; 41: 1–16.

41. Qi J, Hu J and Peng YH. Incorporating adaptability-related knowledge into support vector machine for case-based design adaptation. Eng Appl Artif Intel 2015; 37: 170–180.

42. Giachetti RE, Young RE, Roggatz A, et al. A methodology for the reduction of imprecision in the engineering process. Eur J Oper Res 1997; 100: 277–292.

43. Chandrasegaran SK, Ramani K, Sriram RD, et al. The evolution, challenges, and future of knowledge representation in product design systems. Comput-Aided Des 2013; 45: 204–228.

44. Masui K, Sakao T, Kobayashi M, et al. Applying quality function deployment to environmentally conscious design. Int J Qual Reli Manag 2003; 20: 90–106.

45. Zadeh L. Fuzzy sets. Inf Control 1965; 8: 338–353.

46. Zeng JH, An M and Smith NJ. Application of a fuzzy based decision making methodology to construction project risk assessment. Int J Proj Manag 2007; 25: 589–600.

47. Galushka M and Patterson D. Intelligent index selection for case-based reasoning. Knowl-Based Syst 2006; 19: 625–638.

48. Fayyad UM and Irani KB. On the handling of continuous-valued attributes in decision tree generation. Mach Learn 1992; 8: 87–102.

49. Brodu N. Query Sphere Indexing for neighborhood requests. J Graphics Tools 2008; 13: 35–51.

50. Florentin S. Extenics in higher dimensions. Columbus, Ohio: Education Publisher, 2012.

51. Kulak O and Kahraman C. Fuzzy multi-attribute selection among transportation companies using axiomatic design and analytic hierarchy process. Inf Sci 2005; 170: 191–210.

52. Kahraman C and Cebi S. A new multi-attribute decision making method: hierarchical fuzzy axiomatic design. Expert Syst Appl 2009; 36: 4848–4861.

53. Suh NP. Axiomatic design theory for systems. Res Eng Des 1998; 10: 189–209.

54. Xu CH, Lu GZ, Tan ZX, et al. Guide to vacuum equipment selection and procurement. Beijng, China: Chemical Industry Press, 2013 (in Chinese).