

A case retrieval method combined with similarity measurement and DEA model for alternative generation

Jing ZHENG^{1,2,*}, Ying-Ming WANG², Kai ZHANG³

¹ *College of Electronics and Information Science, Fujian Jiangxia University, Fujian 350108, P. R. China;*
E-mail: zhengjing80@fjxu.edu.cn

² *Decision Sciences Institute, Fuzhou University, Fujian 350116, P. R. China;*
E-mail: msymwang@hotmail.com

³ *Department of Information Engineering, Fujian Chuanzheng Communications College, Fuzhou 350007, PR China.*
E-mail: k7920@qq.com

Received 10 December 2017

Accepted 5 May 2018

Abstract

In alternative generation, reusing past experience is a potential methodology and case retrieval is a primary step. In order to improve the performance of case retrieval process, many applications have used different similarity measurements and the selection method for the most suitable historical case to solve problems. Many investigations have shown that human beings are usually bounded rational and their psychological behavior has certain influence on decision making. However, such behavior is neglected in similarity measurements and the selection method can only deal with the evaluation given by one decision maker (DM). This paper proposes a new case retrieval method that combines similarity measurement and data envelopment analysis (DEA) model. A similarity measurement based on cumulative prospect theory is proposed to consider the DM's psychological behavior. A hybridization of four similarity measurements is used to generate a set of similar historical cases. The DM evaluates the similar historical case set by a pairwise comparison matrix. A DEA model is constructed to get the priority vector. The most suitable historical case can then be picked out through the case similarity and the case priority. A case study is finally introduced to illustrate the use of the proposed method.

Keywords: Case-based reasoning; DEA model; multiple criteria decision analysis; prospect theory; similarity measurement

*Corresponding author.

1. Introduction

Case-based reasoning (CBR) is good at solving new problems by referring to the solution of similar past experience¹. It can help decision maker (DM) to generate alternatives quickly. Therefore, it has been widely used in many fields, such as environment preparedness system², emergency decision making³, business failure prediction⁴, medicine⁵, fault diagnosis⁶, wastewater treatment⁷. CBR usually includes four steps¹, i.e., retrieval, adaptation, revision and retaining. Among the four steps, retrieval is regarded as the first and core step. If the retrieved historical case(s) is the most desirable case, the solution will be effective; otherwise, the result would not be good. Hence, it is essential to study case retrieval methods.

Up to now, a number of case retrieval methods have been proposed in the practical CBR applications. There are mainly two kinds. One kind is to propose a similarity function to retrieve the similar historical case(s) and directly use it in the applications. For example, Kwong et al.⁸ proposed a similarity measure based on Euclidean distance to concurrent design of low power transformers. Yu et al.⁹ developed a hybridization of both symbolic and numeric reasoning techniques for mining of scarce construction databases. Li et al.⁴ proposed a similarity computation method, which transferred the attribute distance into Gaussian distance and solved the nonlinear data, to improve the prediction accuracy in business failure prediction. Sun et al.¹⁰ developed a similarity function using grey theory to improve the ability of similar case retrieval and prediction accuracy. Li et al.¹¹ proposed a similarity measure by combining four independent CBR models to amplify advantages of individual techniques and minimize their limitations. Apparently, it is useful to mix several similarity measure methods. The other kind of methods proposed case similarity measurements and selected the appropriate historical case(s) according to the evaluation of historical cases. To select the most

desirable historical case, some studies attempted to introduce multi-criteria decision making (MCDM) to select the most effective historical case. For example, Qi et al.¹² proposed a case retrieval method which used the algorithm of order preference by similarity to an ideal solution (TOPSIS) to evaluate the most similar cases in terms of product criteria to pick out the most suitable case. Li et al.¹³ developed a CBR forecasting method based on the similarities to positive and negative ideal cases. Fan et al.³ generated the desirable response alternative by evaluating the retrieved historical case(s).

The existing studies have made significant contributions to decision making based on CBR. These studies provided various retrieval methods for DMs to solve problems. However, among the existing studies, the similarity calculation methods have a premise, that is, the DM is perfectly rational. In fact, the DM has some emotions such as rejoicing regret, or dislike, in decision making¹⁴, and they would affect the DM's decision. In other words, people are often limitedly rational rather than perfectly rational when facing decision making. Therefore, it is necessary to investigate the case retrieval methods considering human behavior for the purpose of providing effective decision support to DMs. Furthermore, the existing case selection methods can only deal with the evaluation given by one DM, yet the process of decision making may have several formats such as in group decision-making. Therefore, case retrieval methods should consider the psychological behavior in the similarity calculation and several evaluation formats in the most suitable historical case selection.

Since Tversky and Kahneman proposed cumulative prospect theory (CPT)¹⁵, which is a descriptive model of decision making under conditions of risk, many scholars have employed it to solve various decision-making problems considering DM's behavior, such as, emergency decision making¹⁶, MCDM¹⁷. This is because CPT describes the DM's

behavior characteristics well and gives the calculation formulas on values and weights of potential outcome. In the case retrieval step, the DM usually has some emotions, i.e., loss aversion, diminishing sensitivity, and reference dependence. For instance, the DM has a thought that if the attribute value of one historical case is very different from the target case, even the case similarity is higher than the other historical cases, the historical case is not similar with the target case. Therefore, how to incorporate CPT into case retrieval deserves more attention.

Extensive studies of MCDM techniques have been undertaken over the past decades, where methods of analytic hierarchy process (AHP), elimination and choice translating reality (ELECTRE), and technique for TOPSIS have been proved to be effective approaches. TOPSIS method has been integrated into CBR to generate a more suitable alternative. But TOPSIS has a defect that cannot handle large set of alternatives and criteria¹⁸. AHP is a simple and effective MCDM aid tool, and it can evaluate several similar cases and identify a suitable design alternative. Rammanathan¹⁹ developed a method which combined data envelopment analysis (DEA) with AHP to form a DEAHP method, but it has a significant drawback that there is no guarantee that the DEAHP method can produce rational weight vectors for inconsistent pairwise comparison matrices (PCM). Afterwards, Wang et al.²⁰ proposed a new DEA model to determine the priority, which can derive logical priorities for PCM. Hence, how to derive the priority vector in the AHP for the similar case set is worth attention.

The objective of this paper is to develop a case retrieval method based on similarity measurement and DEA model for generating a desirable alternative. In similarity measurement stage, a similarity measurement based on CPT is proposed, which considers DM's psychological behavioral. It is more consistent with decision-making process. Meanwhile, we mix three classic similarity methods with the similarity

measurement based on CPT to get proper similar historical cases set. The mix can expand the advantages of the similarity measurements and make the decision result more effective. In alternative generation stage, a DEA model is constructed to get the priority vector in AHP, and generate a proper alternative. The DEA model can deal with various forms of evaluations and make the method more applicable.

The rest of the paper is organized as follows. In section 2, we give a brief review of the classic similarity measurements and CPT. In section 3, we develop a case retrieval method combined with similarity measurement and DEA model for alternative generation. In section 4, numerical examples are provided to illustrate the use of the proposed method. In section 5, the discussion of this study is presented. In section 6, conclusions of this study are provided.

2. Preliminaries

This section provides a brief introduction about concepts related to similarity measurement and CPT that are used in the later proposed method.

2.1. Similarity measurement

Case retrieval is the core step of the CBR. Similarity measurement between target case and historical cases has great influence on retrieval quality. The similarity assessment based on distance function is the typical measurement, such as, Euclidean distance¹¹, Manhattan distance¹¹, Gaussian distance⁴. We assume the attribute distance between historical case C_g and target case

C_0 concerning the attribute P_l is d_{0gl} , $l \in \{1, 2, \dots, m\}$, $g \in \{1, 2, \dots, h\}$ and the formula for the case similarity based on Euclidean distance Sim_g^1 is shown by the following way:

$$d_{0gl} = \frac{|x_{0l} - x_{gl}|}{\max_{1 \leq g \leq h} \{|x_{0l} - x_{gl}|\}} \quad (1)$$

$$Sim_g^1 = \frac{1}{1 + \sqrt{\sum_{l=1}^m (w_l^p d_{0gl})^2}} \quad (2)$$

where x_{0l} denotes the feature values of the target case C_0 concerning the attribute P_l , x_{gl} denotes the feature values of the target case C_0 concerning the attribute P_l , w_l^p denotes the weight of attribute P_l , such that $\sum_{l=1}^m w_l^p = 1$ and $w_l^p \geq 0$. The CBR model built on this Euclidean distance is called as ECBR.

We assume the attribute distance between historical case C_g and target case C_0 concerning the attribute P_l is g_{0gl} , and the formula for the case similarity based on Gaussian distance Sim_g^2 is shown by the following way:

$$g_{0gl} = \exp\left[-\left(\frac{d_{0gl}}{\sqrt{2} \times \sigma_l}\right)^2\right] \quad (3)$$

$$Sim_g^2 = \sqrt{\sum_{l=1}^m (w_l^p g_{0gl})^2} \quad (4)$$

where $\delta_l = \delta \times (\max_{1 \leq g \leq h} (d_{0gl}) - \min_{1 \leq g \leq h} (d_{0gl}))$, $\delta \in [0,1]$, and it indicates the error deviation degree. The CBR model built on this Gaussian distance is called GCBR.

Then, in order to improve the performance of retrieval, grey coefficient degree¹¹ is used to calculate

the similarity. Assume the attribute distance between historical case C_g and target case C_0 concerning the attribute P_l is r_{0gl} , and the formula for the case similarity based on grey coefficient Sim_g^3 is shown by the following way:

$$r_{0gl} = \frac{\min_{0 \leq g \leq h} |x_{gl} - x_{0l}| + 0.5 \times \max_{0 \leq g \leq h} |x_{gl} - x_{0l}|}{d_{0gl} + 0.5 \times \max_{0 \leq g \leq h} |x_{gl} - x_{0l}|} \quad (5)$$

$$Sim_g^3 = \sqrt{\sum_{l=1}^m (w_l^p r_{0gl})^2} \quad (6)$$

The CBR model built on grey coefficient degree is called RCBR.

2.2. CPT method

A lot of psychological studies have shown that there are several psychological characteristics of human behavior under risk and uncertainty, such as reference dependence, loss aversion, and judgmental distortion of likelihood of almost impossible and certain outcomes^{16, 21, 22}. The decision-making problem is uncertain and risk sometimes, so it is necessary to consider DMs' psychological behavior. Since Tversky and Kahneman proposed CPT, many scholars have employed it to solve various decision-making problems considering DM's behavior, such as, emergency decision making¹⁶, MCMD¹⁷. CPT includes two steps. Firstly, the outcomes of gains and losses are calculated by a reference point and the prospect value is evaluated by a value function. The value function is expressed in the form of a power law according to the following expression²²

$$v(x) = \begin{cases} x^\alpha, & x \geq 0 \\ -\lambda(-x)^\beta, & x < 0 \end{cases} \quad (7)$$

where x denotes the gains or losses; $x \geq 0$ represents the gains and $x < 0$ represents the losses. α and β are exponent parameters related to gains and losses, respectively, $0 \leq \alpha, \beta \leq 1$. λ is the risk aversion parameter, which represents the characteristic of steeper for losses than for gains, $\lambda > 1$ represents a prospect value function with convex and concave S-shapes for losses and gains, respectively. The values of α , β , and λ are determined by experiments^{23,24}.

3. The proposed method

In this section, we present a method for case retrieval

which combines similarity measurement and DEA model for alternative generation as Fig. 1. Firstly, a hybrid similarity measurement is proposed, which mixes the similarity measurement based on CPT with three classic similarity measurements. Then alternative evaluation is got using PCM. Furthermore, a DEA model is constructed to gain the priority of the similar historical cases. Finally, we gain the ranking order of the similar historical case and the desirable historical case will be determined. The method is introduced as follows.

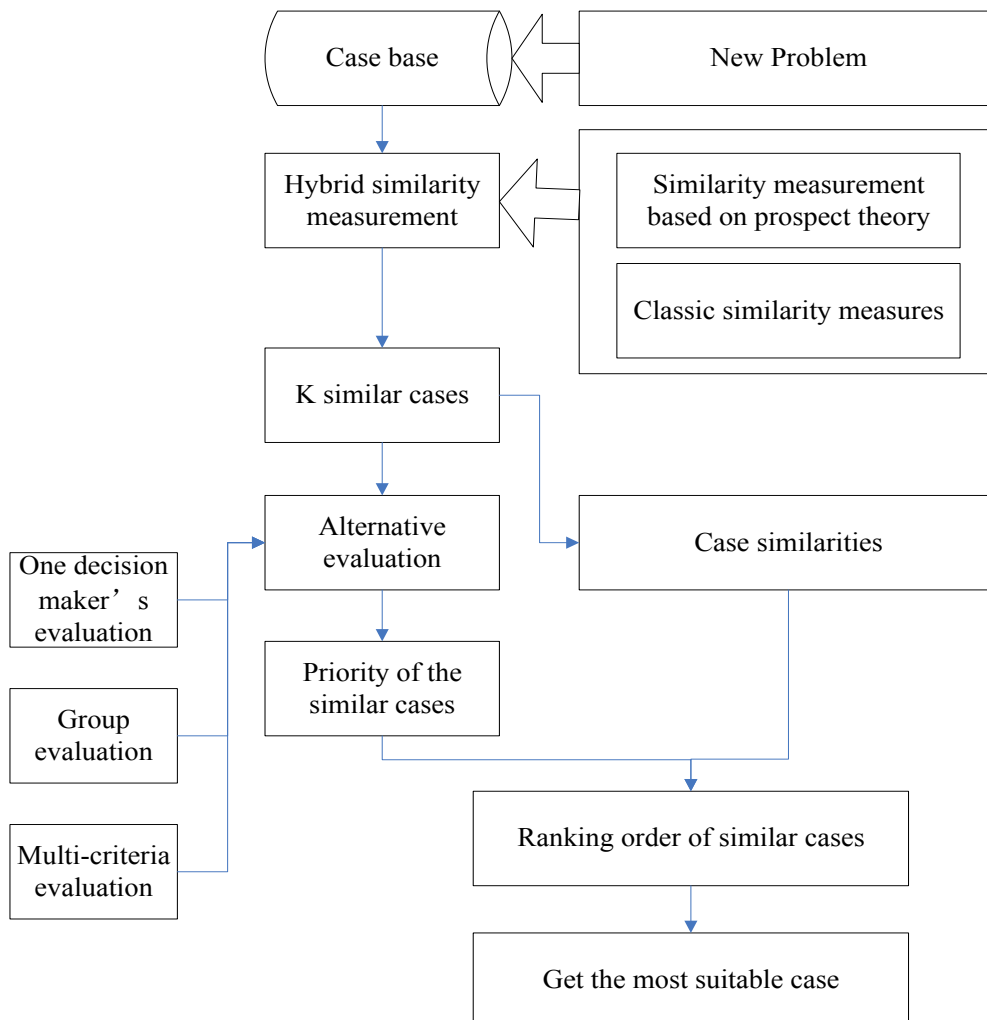


Fig. 1. Framework of the case retrieval model

3.1. Similarity measurement

In order to amplify the advantages of similarity measurements and minimize their limitations¹¹, the combination of several similarity measurements is employed to gain the similarities between historical

cases and target case. The case similarity measurement includes two aspects and is shown in Fig. 2. The first aspect is to calculate the case similarity based on prospect theory, the other aspect is to calculate the case similarity based on three classic similarity measurements.

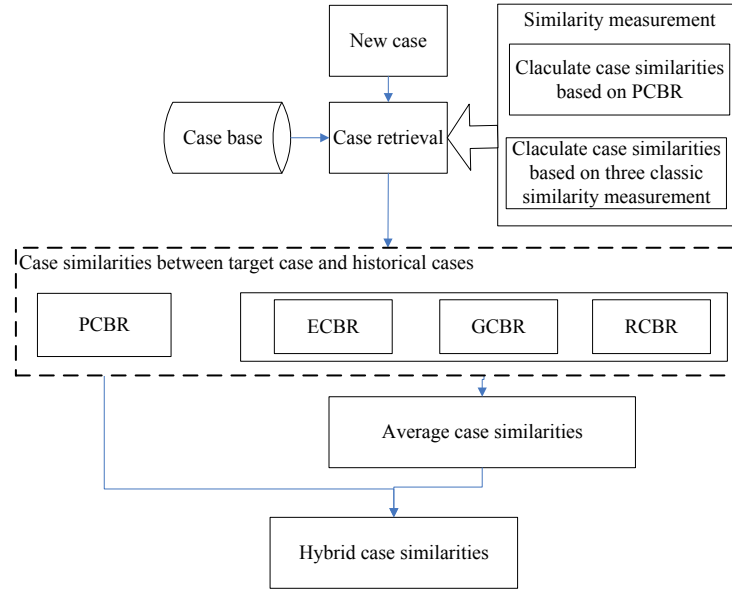


Fig. 2. The process of the hybrid similarity calculation

3.1.1. Similarity measurement based on cumulative prospect theory

Suppose there are g historical cases denoted by $C_g (g=1, \dots, h)$ and one target case denoted by C_0 . Let $P = \{P_1, P_2, \dots, P_m\}$ represent the vector of m attributes with regard to the problem of both the historical cases and the target case, where P_l denotes the l th attribute, $l \in \{1, 2, \dots, m\}$. Let $W^P = \{w_1^P, w_2^P, \dots, w_m^P\}$ be a vector of attribute weights, where w_l^P denotes the weight of

the attribute P_l , such that $\sum_{l=1}^m w_l^P = 1$ and $0 \leq w_l^P \leq 1$.

Let $X_g = \{x_{g1}, x_{g2}, \dots, x_{gm}\}$ represent a vector of attribute value with regard to the historical case C_g , where x_{gl} denotes the attribute value concerning attribute P_l with regard to the historical case C_g . Let $X_0 = \{x_{01}, x_{02}, \dots, x_{0m}\}$ represent a vector of attribute value with regard to the target case C_0 , where x_{0l} denotes the attribute value concerning attribute P_l

with regard to the target case C_0 . Then we can define the attribute distance d_{0gl} between historical case C_g and target case C_0 concerning the attribute P_l as follows:

$$d_{0gl} = \frac{|x_{0l} - x_{gl}|}{\max_{0 \leq g \leq h} \{x_{gl}\} - \min_{0 \leq g \leq h} \{x_{gl}\}} \quad (8)$$

where $0 \leq d_{0gl} \leq 1$. Thus, the case similarity can be expressed as $1 - d_{0gl}^{25}$.

Let Sim_{0gl} denote the attribute similarity between historical case C_g and target case C_0 concerning the attribute P_l and d_l denote a reference point with regard to the l th attribute distance. Because the DM's preference for the attribute distance is d_l , it means that if the attribute distance is less than d_l , the DM would feel it is "gain", otherwise, if the attribute distance is more than d_l , the DM would feel it is "loss". Based on CPT, we can define the attribute similarity Sim_{0gl} between historical case C_g and target case C_0 concerning the attribute P_l as follows:

$$sim_{0gl} = \begin{cases} (1 - d_{0gl})^\alpha, & d_{0gl} \leq d_l \\ -\lambda(1 - d_{0gl})^\beta, & d_{0gl} > d_l \end{cases} \quad (9)$$

where $(1 - d_{0gl})^\alpha$ represents the attribute similarity when the DM feels "gain", $-\lambda(1 - d_{0gl})^\beta$ represents

the attribute similarity when the DM feels "loss". According to the [21], let $\alpha = 0.89$, $\beta = 0.92$, $\lambda = 2.25$.

Since different attribute similarities are usually incommensurate, Sim_{0gl} needs to be normalized as

\overline{Sim}_{0gl} by using the following formula:

$$\overline{Sim}_{0gl} = \frac{Sim_{0gl}}{|Sim_{0g}|^{\max}} \quad (10)$$

where $|Sim_{0g}|^{\max} = \max\{|Sim_{0g1}|, |Sim_{0g2}|, \dots, |Sim_{0gm}|\}$,

and it is obvious that $0 \leq \overline{Sim}_{0gl} \leq 1$.

Finally, by using the simple additive weighting (SAW) method, the overall prospect of each historical case similarity Sim_g'' can be calculated as follows:

$$Sim_g'' = \sum_{l=1}^m w_l \overline{Sim}_{0gl} \quad (11)$$

The CBR model built on prospect theory is called PCBR.

The following example illustrates the feasibility and effectiveness for the similarity measurement based on CPT.

Example 1 A high-rise building fire took place in City F. The emergency decision center considers four attributes mainly, which are fire rating (P_1), fire area (P_2 , unit: m^2), casualties (P_3 , unit: person) and economic losses (P_4 , unit: ten thousand RMB). Table 1 shows the attribute values with regard to the target case C_0 and historical

cases $C_i (i=1, 2, 3)$. The attribute weight vector provided by the emergency management center is $W = (0.25, 0.25, 0.25, 0.25)$, the DM preference attribute distance vector is $(0.4, 0.3, 0.3, 0.3)$.

Table 1. Attribute values with regard to target case and historical cases

	P ₁	P ₂	P ₃	P ₄
C ₁	1	16	18	7
C ₂	3	19	20	10
C ₃	4	25	25	11
C ₀	2	18	20	8

According to Eqs.(1)-(2), the attribute distance d_{0gl} and case similarity Sim_g are gained, and the results can be found in Table 2.

Table 2. The computation result of attribute distance and case similarity using Euclidean distance

	d_{0g1}	d_{0g2}	d_{0g3}	d_{0g4}	Sim_g
C ₁	0.3333	0.2222	0.2857	0.2500	0.7272
C ₂	0.3333	0.1111	0.0000	0.5000	0.7639
C ₃	0.6667	0.7778	0.7143	0.7500	0.3562

According to Eqs.(8)-(11), the attribute similarity sim_{0gl} and case similarity Sim_g'' are gained, and the results can be found in Table 3.

Table 3. The computation result of attribute similarity and case similarity using similarity measurement based on CPT

	sim_{0g1}	sim_{0g2}	sim_{0g3}	sim_{0g4}	Sim_g''
C ₁	0.6971	0.7996	0.7412	0.7741	0.7530
C ₂	0.6971	0.9005	1.0000	-1.1891	0.3521
C ₃	-0.8189	-0.5639	-0.7106	-0.6285	-0.6805

In Table 2, historical case C_2 is the most similar case with the target case C_0 , but the attribute distance d_{024} is very large. It means that the target case C_0

and the historical case C_2 have significant differences on attribute P_4 and it would lead to different alternatives for the historical case C_2 and the target case C_0 . So, DMs have their preference in the attribute distance. DMs prefer to choose the historical case with a lower similarity and lower attribute distance as the most similar case. The similarity measurement based on CPT can well express this kind of psychological behavior of DMs. The attribute similarity sim_{024} in the third row and fifth column of Table 3 is very small, because its attribute distance is larger than the DM preference attribute distance 0.3. The most similar case is C_1 . Therefore, the similarity measurement based on CPT is in accordance with the DMs' behavior.

3.1.2. Classic similarity measurement

Classic similarity measurements are proved to be useful for the case retrieval, so we consider their effectiveness. We assemble ECBR, GCBR and RCBR to get the classic case similarity. When the ECBR is got using Eqs.(1)-(2), GCBR is got using Eqs.(3)-(4), RCBR is got using Eqs.(5)-(6), we use weighted mean for aggregating three classic similarity measurements. Let Sim_g' be the average similarity of the above three classic similarities. The aggregation formula is given by

$$Sim_g' = \frac{Sim_g^1 + Sim_g^2 + Sim_g^3}{3} \quad (12)$$

In order to aggregate the advantages of these similarity measurements, it is necessary to mix the similarities based on the CPT and the average similarity based on three classic similarities measurements. Let

Sim_g denote the hybrid similarity between the historical case C_g and the target case C_0 . The calculation of Sim_g is given by

$$Sim_g = \theta Sim'_g + \gamma Sim''_g \quad (13)$$

where θ and γ represent the preference of the two similarities, and $0 \leq \theta, \gamma \leq 1$, $\theta + \gamma = 1$.

Let ξ be the threshold for the similarity between target case and historical cases, such that $\xi \in [\min\{Sim_g | g \in \{1, 2, \dots, h\}\}, \max\{Sim_g | g \in \{1, 2, \dots, h\}\}]$. The bigger the value of ξ is, the higher requirement of the case similarity the DM has. The value of ξ usually is given by the DM according to his (her) experience and the realistic data. When $Sim_g \geq \xi$, the historical case C_g would be extracted, and it would constitute the similar historical cases set Z^{Sim} , i.e., $Z^{Sim} = \{C_k | k \in M^{Sim}\}$, where $M^{Sim} = \{g | Sim_g \geq \xi, g = 1, 2, \dots, h\} = \{1, 2, \dots, s\}$. If the similar case set has only one case, the case would be selected as the most suitable case. Otherwise, we would select the most suitable case by a DEA model.

3.2. Multi-criteria decision making

The DEA model proposed by Wang et al.²⁰ can produce true weights for PCM. Here, we use this DEA model to evaluate the similar cases.

3.2.1. DEA model for alternatives priority

Based on the similar case set, the DM evaluates the

alternatives by using a PCM denoted as $A^1 = (a_{ij})_{s \times s}$,

which satisfies $a_{ii} = 1$ and $a_{ji} = 1/a_{ij}$ for $j \neq i$.

Let $W = (w_1, w_2, \dots, w_s)$ be the priority vector of the matrix A^1 .

We use a DEA model to get the priority vector of the matrix A^1 and get the most suitable historical case.

According to [20], we view each row of the matrix A^1 as a decision making unit (DMU), each column as an output and assume a dummy input value of one for all the DUMs. Each DMU has s outputs and one dummy constant input, based on which the DEA model for relative score can be formulates as

$$\begin{aligned} & \text{Maximize } w_0 = \sum_{j=1}^s a_{0j} z_j \\ & \text{Subject to } \begin{cases} \sum_{j=1}^n (\sum_{i=1}^n a_{ij}) z_j = 1, \\ \sum_{j=1}^n a_{ij} z_j \geq n z_i, \quad i = 1, \dots, s, \\ z_j \geq 0, \quad j = 1, \dots, s. \end{cases} \quad (14) \end{aligned}$$

where DMU_0 represents the similar historical case under evaluation, w_0^* represents the DEA efficiency of DMU_0 and is used as its priority. Linear programming model (14) is solved for all the similar historical cases in Z^{Sim} to obtain the priority vector $W^* = (w_1^*, w_2^*, \dots, w_s^*)$ of the PCM A^1 .

3.2.2. DEA model for alternatives priority concerning group decision making

In decision analysis, group decision making is usually used to assemble several experts' wisdom. In order to select the most effective alternative, sometimes DMs use group decision making. We consider several DMs to give their preference on the similar case set Z^{Sim} by PCM. Let $A^{(k)} = (a_{ij}^{(k)})_{s \times s}$ be a PCM provided by the k th DM _{k} ($k = 1, \dots, f$), $r_k > 0$ be its relative

importance weight satisfying $\sum_{k=1}^f r_k = 1$. The selection steps for the most suitable historical case are introduced as follows.

First, we use SAW method to aggregate several DMs' preferences, which means that we integrate the k $A^{(k)}$ into a PCM $B = (b_{ij})_{s \times s}$. The formula is given as follows:

$$b_{ij} = \sum_{k=1}^f r_k a_{ij}^{(k)} \quad (15)$$

Second, we use the following DEA model²⁰ to gain the priority vector:

$$\begin{aligned} & \text{Maximize } w_0 = \sum_{j=1}^s b_{ij} z_j, \\ & \text{Subject to } \begin{cases} \sum_{j=1}^s (\sum_{i=1}^s b_{ij}) z_j = 1, \\ \sum_{j=1}^s b_{ij} z_j \geq sz_i, \quad i = 1, \dots, s, \\ z_j \geq 0, \quad j = 1, \dots, s. \end{cases} \end{aligned} \quad (16)$$

By solving the above model (16) for each w_i ($i = 1, \dots, s$), we will get the best priority vector

W^* .

3.2.3. DEA model for alternative priority concerning multi-criteria evaluation

In decision making, the DM cannot evaluate the alternative directly sometimes, and what she/he can do is to evaluate the alternative from a few criteria. In this situation, a hierarchical structure often exists as shown in Fig. 3. In order to select the most suitable historical case, we can evaluate the similar historical cases set from different criteria. Between the historical case selection level and the p decision criteria level, we can get a priority vector $\{v_1, v_2, \dots, v_p\}$ by solving the model (14) and between the p criteria level and s similar historical cases level, we can get p priority vectors $\{v_{1j}, v_{2j}, \dots, v_{sj}\}$ ($j = 1, \dots, p$) by solving the model (14). Based on the above priority vectors, we use the SAW to aggregate them and get a global priority vector $\{w_1, w_2, \dots, w_s\}$ as follows:

$$w_i = \sum_{j=1}^p v_{ij} v_j \quad (17)$$

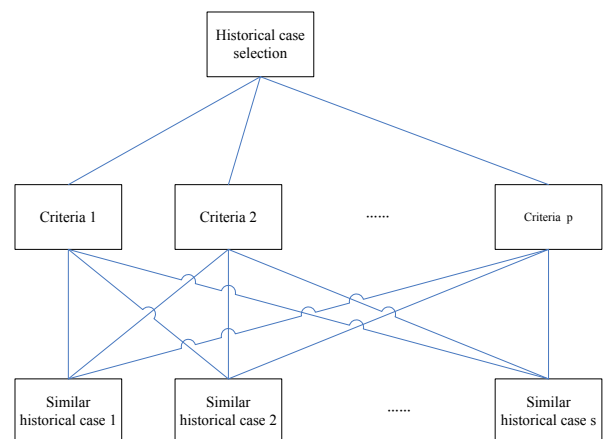


Fig. 3. A hierarchical structure for multi-criteria historical case selection

3.3. Comprehensive coefficient

A comprehensive coefficient (CC) is defined to determine the ranking order of the similar historical cases when w_i of each similar historical case has been calculated. The case similarity is a very important indicator for selecting the most suitable historical case. So, we should consider the case similarity and the alternative evaluation simultaneously. Let D_i denote the CC of the historical case C_i . The calculation formula of D_i is given by:

$$D_i = Sim_i \times w_i \quad (18)$$

where Sim_i is the case similarity between the historical case C_i and the target case C_0 , $Sim_i \in Z^{Sim}$, $i \in M^{Sim}$. w_i is the priority weight about the similar historical case. Obviously, $D_i \in [0,1]$ and the greater the value of D_i is, the more suitable the historical case C_i will be to the target case C_0 .

According to D_i , we can determine the ranking order of all similar historical cases and select the best cases from the similar historical cases set to generate the alternative of the target case.

In summary, the steps of the proposed method for case retrieval are given as follows:

Step 1. For attribute P_l , calculate the case similarities Sim_g^1 based on Euclidean distance using

Eqs.(1)-(2), Sim_g^2 based on Gaussian distance using

Eqs.(3)-(4), Sim_g^3 based on grey coefficient degree

using Eqs.(5)-(6), and Sim_g'' based on CPT using Eqs.(8)-(11).

Step 2. Calculate the average similarity of the three classic similarities, Sim_g' , using Eq. (12), and the hybrid similarity, Sim_g , using Eq. (13).

Step 3. When there is only one DM who gives the evaluation for the similar historical cases, the alternative priority vector W^* is determined by model (14); otherwise, by using Eq. (15) and model (16). When the DM evaluates the alternatives from a few criteria, the alternative priority vector W^* is determined by model (17).

Step 4. Calculate the comprehensive coefficient D_i using Eq. (18), based on which all the similar historical cases can be ranked and the best historical case(s) can be selected.

4. Example

In this section, we provide a numerical example from three aspects to illustrate the advantages of the proposed method and its potential applications in case retrieval. Consider an application of the proposed method to parametric car design. A car design company E constructs a data base which includes 15 historical cases $(C_1, C_2, \dots, C_{15})$ concerning parametric car design.

According to parametric car design, the car design company considers four attributes mainly, namely, hundred kilometers acceleration (P_1 , unit: second), braking distance (P_2 , unit: m), horsepower (P_3 , unit: hp) and hundred kilometers fuel consumption (P_4 , unit: l).

Now, a new kind of car needs to be designed, and it is regarded as the target case C_0 . Table 4 shows the values of the attributes with regard to the historical cases C_i and the target case C_0 . Then we use the case retrieval

method combined with similarity measurement and DEA model to generate an alternative. The computation processes and results are presented as follows.

Table 4. The attribute values of the historical cases and the target case

Cases	P ₁	P ₂	P ₃	P ₄
C ₁	12	42	105	12
C ₂	11	43	101	12
C ₃	8	39	110	18
C ₄	11	41	110	16
C ₅	10	40	123	17
C ₆	7	38	128	20
C ₇	9	43	125	18
C ₈	6	37	130	20
C ₉	5	36	130	25
C ₁₀	12	45	125	13
C ₁₁	10	41	123	14
C ₁₂	11	43	121	13
C ₁₃	9	40	117	14
C ₁₄	7	39	120	17
C ₁₅	6	37	117	16
C ₀	10	42	120	12

"Step 1: The case similarity based on CPT Sim_g " can be calculated by using Eqs.(8)-(11), and the computation results are shown in Table 5.

"Step 2: The case similarities by using three classic similarity measurements are calculated. The case similarity Sim_g^1 is calculated by using Eq. (1) and Eq. (2), and the results are shown in Table 5. The case similarity Sim_g^2 is calculated by using Eq. (3) and Eq. (4), and the results are shown in Table 5. The case

similarity Sim_g^3 is calculated by using Eq. (5) and Eq. (6), and the results are shown in Table 5. Based on these case similarities, we use Eq.(12) to get the average similarity Sim_g' , and the results are shown in Table 5. Considering the importance of the two kinds of similarity measurements is equal, let $\gamma = \theta = 0.5$. Then, the hybrid similarity Sim_g is calculated by using Eq. (13), the results are shown in Table 5.

Table 5. The case similarities by five methods

Cases	Sim_g^1	Sim_g^2	Sim_g^3	Sim_g'	Sim_g''	Sim_g
C_1	0.8328	0.7238	0.7358	0.7642	0.2345	0.4993
C_2	0.8148	0.6967	0.6994	0.7370	0.3063	0.5216
C_3	0.7373	0.5142	0.5157	0.5890	0.2215	0.4053
C_4	0.8152	0.6665	0.6426	0.7081	0.2330	0.4705
C_5	0.8492	0.7511	0.7313	0.7772	0.3112	0.5442
C_6	0.6939	0.4469	0.4686	0.5365	-0.4827	0.0269
C_7	0.8183	0.6884	0.6599	0.7222	0.3190	0.5206
C_8	0.6571	0.3811	0.4238	0.4873	-0.3920	0.0477
C_9	0.5947	0.3011	0.3718	0.4225	-0.1394	0.1416
C_{10}	0.8216	0.6618	0.6443	0.7093	0.5994	0.6543
C_{11}	0.9157	0.8486	0.8187	0.8610	0.6891	0.7750
C_{12}	0.9163	0.8422	0.8089	0.8558	0.6895	0.7727
C_{13}	0.8654	0.7455	0.7097	0.7735	0.6460	0.7098
C_{14}	0.7772	0.6254	0.6299	0.6775	-0.1242	0.2767
C_{15}	0.7223	0.5193	0.5347	0.5921	-0.4484	0.0718

Step 3: The DMs give the threshold of the case similarity according to his (her) experience, i.e., $\xi = 0.5$, and we gain the similar cases set, i.e.,

$$C^{Sim} = (C_2, C_5, C_7, C_{10}, C_{11}, C_{12}, C_{13}) .$$

Step 4: The DM gives a PCM A concerning the similar cases set according to his (her) experience as follows:

$$A = \begin{bmatrix} 1 & 1 & 1/3 & 1/4 & 1/5 & 1/6 & 1/4 \\ 1 & 1 & 1/5 & 1/6 & 1/5 & 1/7 & 1/6 \\ 3 & 5 & 1 & 1/2 & 1/4 & 1/3 & 1/4 \\ 4 & 6 & 2 & 1 & 1/2 & 1/3 & 1/2 \\ 5 & 5 & 4 & 2 & 1 & 3 & 4 \\ 6 & 7 & 3 & 3 & 1/3 & 1 & 2 \\ 4 & 6 & 4 & 2 & 1/4 & 2 & 1 \end{bmatrix}$$

Step 5: The priority of the similar cases set can be gained by solving model (14), the result is

$$w = (0.0304, 0.0261, 0.0609, 0.0609, 0.5478, 0.1826, 0.0913) .$$

Step 6: Finally, the comprehensive coefficient CC of the similar cases set can be got by using Eq. (18), the result is $Sim = (0.0159, 0.0142, 0.0317, 0.0398, 0.4246, 0.1411, 0.0648)$. The most suitable historical case is got by ranking CCs and is determined to be C_{11} .

It is indicated from the computational results obtained by using the proposed case retrieval method that the retrieved historical case C_{11} is the most suitable one to the target case C_0 . Thus, the design plan for the case C_{11} can be considered as that for the target case C_0 . Further, designer of company E can make improvements based on the design plan of C_{11} .

In order to further illustrate the proposed method, alternative selection based on group decision making is described below. Three experts from three different departments are invited to make comparisons about the eight similar historical cases, and the three PCMs provided by them are as follows:

$$A^1 = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/2 & 1/7 & 1/6 & 1/5 \\ 2 & 1 & 1/3 & 1/2 & 1/5 & 1/6 & 1/4 \\ 3 & 3 & 1 & 2 & 1/3 & 1/4 & 1/2 \\ 2 & 2 & 1/2 & 1 & 1/5 & 1/3 & 1/3 \\ 7 & 5 & 3 & 5 & 1 & 5 & 4 \\ 6 & 6 & 4 & 3 & 1/5 & 1 & 2 \\ 5 & 4 & 2 & 3 & 1/4 & 1/2 & 1 \end{bmatrix}$$

$$A^2 = \begin{bmatrix} 1 & 1 & 1/3 & 1/2 & 1/6 & 1/7 & 1/5 \\ 1 & 1 & 1/3 & 1 & 1/6 & 1/5 & 1/3 \\ 3 & 3 & 1 & 2 & 1/5 & 1/4 & 1/3 \\ 2 & 1 & 1/2 & 1 & 1/5 & 1/4 & 1/3 \\ 6 & 6 & 5 & 5 & 1 & 3 & 4 \\ 7 & 5 & 4 & 4 & 1/3 & 1 & 2 \\ 5 & 3 & 3 & 3 & 1/4 & 1/2 & 1 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 1 & 1/2 & 1/2 & 1/3 & 1/2 & 1/8 & 1/5 \\ 2 & 1 & 1 & 1/3 & 1/2 & 1/3 & 1/3 \\ 3 & 3 & 1 & 2 & 1/5 & 1/4 & 1/3 \\ 2 & 2 & 1/2 & 1 & 1/6 & 1/5 & 1/5 \\ 7 & 4 & 5 & 6 & 1 & 5 & 4 \\ 8 & 3 & 4 & 5 & 1/5 & 1 & 3 \\ 5 & 3 & 3 & 5 & 1/4 & 1/3 & 1 \end{bmatrix}$$

Suppose the relative importance weights of the three DMs are $(0.5, 0.3, 0.2)$, and we use Eq. (15) to transform the three matrices into a matrix B. Then, by solving model (16) for each similar historical case, we get the best priorities shown in the first row of Table 6. If the weights of the three DMs are different, it would get a different matrix B. Then a different priority vector can be derived by model (16). Table 6 shows the best priority vectors derived under different weights of the three DMs. Furthermore, we get the comprehensive coefficient CC of the similar cases set by using Eq. (18), and the results are shown in Table 7, from which we can see that the most suitable historical case is C_{11} . So we select the

alternative of C_{11} as the finally alternative.

Table 6. The best priorities of the eight similar historical cases in a group decision making problem

	W_1	W_2	W_3	W_4	W_5	W_6	W_7
$(0.5, 0.3, 0.2)$	0.0332	0.0468	0.0924	0.0591	0.4188	0.2293	0.1526
$(0.5, 0.25, 0.25)$	0.0329	0.0471	0.0922	0.0590	0.4207	0.2294	0.1529
$(0.33, 0.33, 0.33)$	0.0330	0.0477	0.0894	0.0573	0.4184	0.2302	0.1528
$(1, 0, 0)$	0.0327	0.0445	0.0991	0.0634	0.4223	0.2227	0.1488
$(0, 1, 0)$	0.0371	0.0458	0.0880	0.0561	0.4003	0.2362	0.1541
$(0, 0, 1)$	0.0300	0.0536	0.0851	0.0549	0.4358	0.2364	0.1587

Table 7. The comprehensive coefficients of the eight similar historical cases

	Sim_2	Sim_5	Sim_7	Sim_{10}	Sim_{11}	Sim_{12}	Sim_{13}
$(0.5, 0.3, 0.2)$	0.0173	0.0254	0.0481	0.0382	0.3246	0.1772	0.1083
$(0.5, 0.25, 0.25)$	0.0171	0.0256	0.0480	0.0380	0.3260	0.1772	0.1085

(0.33,0.33,0.33)	0.0172	0.0260	0.0466	0.0371	0.3242	0.1779	0.1085
(1,0,0)	0.0171	0.0242	0.0516	0.0409	0.3273	0.1721	0.1056
(0,1,0)	0.0194	0.0249	0.0458	0.0371	0.3102	0.1825	0.1094
(0,0,1)	0.0157	0.0292	0.0443	0.0341	0.3377	0.1827	0.1126

According to the study of [26], we evaluate the alternatives of the similar historical cases from three criteria, such as safety (G_1), cost (G_2) and accident loss (G_3). In order to improve the requirement of similarity, we set the threshold of the similarity as $\xi = 0.54$, and the similar historical cases set is $(C_5, C_{10}, C_{11}, C_{12}, C_{13})$. Fig 4 shows the hierarchical structure for this alternative selection problem. Table 8 shows the PCM for three criteria. We get the priority weights by solving Eq. (14), the results are shown in

Table 8. Table 9 shows the PCM for the five alternatives with respect to the three criteria. The priority weights are gained by solving Eq. (14), respectively, the results are shown in Table 9. Based on the above results, we get the global priority using Eq.(17), that is $w^g = (0.0973, 0.0639, 0.4072, 0.2695, 0.1700)$. Finally, we use Eq.(18) to get the comprehensive coefficient, and the result is that $Sim = (0.0530, 0.0418, 0.3156, 0.2083, 0.1207)$, based on which, we select alternative C_{11} as the response alternative.

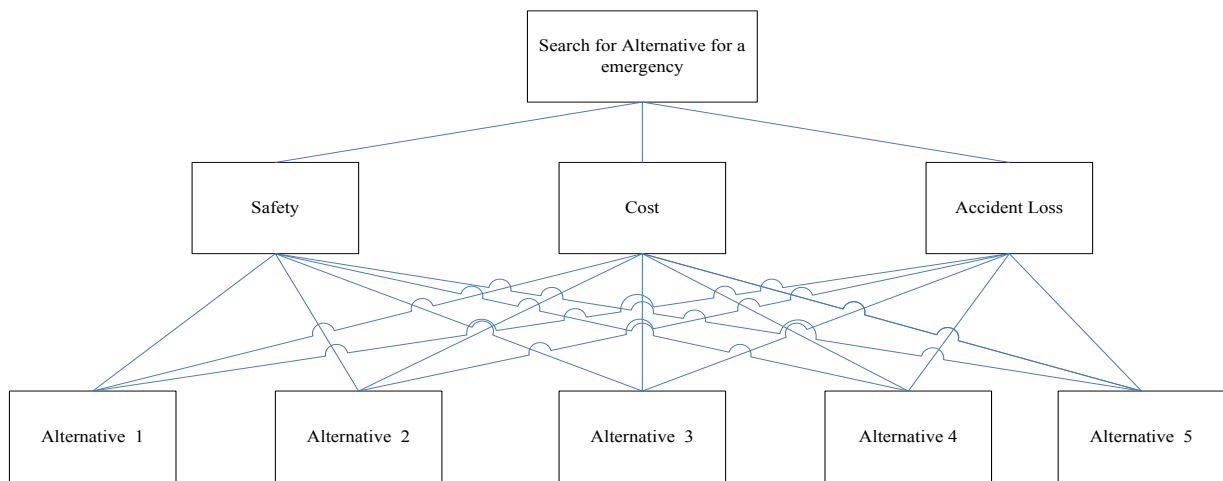


Fig.4. Hierarchy of selection for the most suitable historical case

Table 8. PCM for three criteria and its priorities

	G_1	G_2	G_3	Priority
G_1	1	3	2	0.5294
G_2	1/3	1	1/3	0.1404
G_3	1/2	3	1	0.3333

Table 9. PCM s for five historical cases with regard to three criteria and their priority

	C_5	C_{10}	C_{11}	C_{12}	C_{13}	Priority
Pairwise comparisons of five historical cases with respect to safety						
C_5	1	2	1/5	1/4	1/3	0.0792
C_{10}	0.5	1	1/5	1/5	1/3	0.0574
C_{11}	5	5	1	2	3	0.4230
C_{12}	4	5	0.5	1	2	0.2785
C_{13}	3	3	1/3	0.5	1	0.1675
Pairwise comparisons of five historical cases with respect to cost						
C_5	1	2	1/3	0.5	1	0.1294
C_{10}	1/2	1	1/7	1/6	1/4	0.0506
C_{11}	3	7	1	2	3	0.4204
C_{12}	2	6	1/2	1	1	0.2314
C_{13}	1	4	1/3	1	1	0.1719
Pairwise comparisons of five historical cases with respect to accident loss						
C_5	1	2	1/3	1/3	1/2	0.1118
C_{10}	1/2	1	1/4	1/3	1/2	0.0793
C_{11}	3	4	1	2	2	0.3729
C_{12}	3	3	1/2	1	2	0.2689
C_{13}	2	2	1/2	1/2	1	0.1717

5. Comparative analysis and discussion

To further verify the validity of the proposed method, the results derived from the proposed method are further compared with other existing case retrieval methods.

First, PCBR is compared with the classic case similarity measurements, such as, ECBR, GCBR, RCBR. As can be seen from Example 1, the similarity measurement based on CPT can express the psychological behavior of DM's preference on distance well. From Table 5, the case similarities and the historical ranking based on CPT is a little different from the three classic case similarity measurements, because it

considers the DM's psychological behavior and can distinguish the DM's preference strictly. This is more in line with the actual decision situations.

Furthermore, from Table 5, the classic case similarity measurements and PCBR are different, because ECBR considers the distance between two points, GCBR uses the Gaussian function to change the Euclidean distance into nonlinear form, RCBR represents the case similarity using the case correlation, PCBR considers the DM's psychological behavior. The proposed method aggregates all the advantages of these methods.

The case retrieval method combined with similarity measurement and MCDM has been proposed by Qi et

al.¹² and Li et al.¹³. They used the technique for TOPSIS to evaluate the similar historical case set. But [13] pointed out that the TOPSIS has a drawback that the number of cases under evaluation is limited. However, DEA has no limit on the number of DMU and the DEA model has no such restriction. Furthermore, the DEA model not only can solve the situation where one DM gives his/her evaluation for the alternatives by using PCM, but also can solve group decision making and hierarchical structure in the AHP. From Tables 6, 8 and 9, it can be seen that the DEA model can give a best priority vector from a PCM and can help DM to make more accurate decision making.

6. Conclusion

This paper develops a new case retrieval method from a comprehensive view of combining the similarity measurement and DEA model. The primary contributions are summarized as follows:

- (1) A similarity measurement based on CPT is proposed. This method considers the DM's psychological behavior and makes the decision result more realistic.
- (2) The case similarity is gained through the aggregation of four case similarities. This can amplify the advantage of case similarity measurements and overcome the one-sidedness brought about by a single approach.
- (3) In the case selection, a DEA model is constructed to get the priority vector. It can produce true priority vector for PCM and the DEA model can deal with three formats of evaluations, such as a DM evaluates the alternatives, several DMs evaluate the alternatives and the DM evaluates the alternative from a few criteria. These evaluation formats are usually used very often. So, this DEA model has good usability.

An example about the high-rise fire has shown the necessity of the similarity measurement based on CPT, which considers the DM's psychological behavior. In addition, the example about the car design has shown the feasibility and validity of the proposed method, which

can help DM to find the suitable historical case and to make decision.

For future work, a promising research direction is the use of artificial intelligence method^{27, 28} to improve the validity of the retrieval results. In addition, the case adaptation seems a promising and fruitful research line.

Acknowledgments

We thank two anonymous reviewers for their valuable comments and suggestions which have helped to improve the paper. This work was partly supported by the National Natural Science Foundation of China under the Grant Nos. 71371053 and 61773123, Humanities and Social Science Foundation of Chinese Ministry of Education under the Grant No. 16YJC630008, Fujian Natural Science Foundation of China, No. 2017J01513.

References

1. Aamodt, Agnar, Plaza, and Enric, Case-based reasoning: foundational issues, methodological variations, and system approaches, *AI Communications*, 7(1) (1994) 39-59.
2. Z. Liao, X. Mao, P. M. Hannam and T. Zhao, Adaptation methodology of CBR for environmental emergency preparedness system based on an Improved Genetic Algorithm, *Expert Systems with Applications*, 39(8) (2012) 7029-7040.
3. Z. Fan, Y. Li and Y. Zhang, Generating project risk response strategies based on CBR: A case study, *Expert Systems with Applications*, 42(6) (2015) 2870-2883.
4. H. Li, and J. Sun, Gaussian case-based reasoning for business failure prediction with empirical data in China, *Information Sciences*, 179(1) (2009) 89-108.
5. D. Gu, C. Liang, and H. Zhao, A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis, *Artificial Intelligence in Medicine*, 77, (2017) 31-47.
6. H. Zhao, J. Liu, W. Dong, et al, An improved case-based reasoning method and its application on fault diagnosis of tennessee eastman process, *Neurocomputing*, 249 (C)

- (2017) 266-276.
7. A. Yan, H. Shao and P. Wang, A soft-sensing method of dissolved oxygen concentration by group genetic case-based reasoning with integrating group decision making, *Neurocomputing*, 169 (2015) 422-429.
8. C. K. Kwong and S. Tam, Case-based reasoning approach to concurrent design of low power transformers, *Journal of Materials Processing Technology*, 128(1) (2002) 136-141.
9. W. Yu, and Y. Liu, Hybridization of CBR and numeric soft computing techniques for mining of scarce construction databases, *Automation in Construction*, 15(1) (2006) 33-46.
10. J. Sun, and H. Li, Financial distress prediction using grey case-based reasoning optimized by genetic algorithm, *Science Research Management*, 30(2) (2009) 119-125. (Press in China)
11. J. Sun, and H. Li, Majority voting combination of multiple case-based reasoning for financial distress prediction, *Expert Systems with Applications*, 36(3) (2009) 4363-4373.
12. J. Qi, J. Hu, Y. H. Peng, W. Wang and Z. Zhang, A case retrieval method combined with similarity measurement and multi-criteria decision making for concurrent design, *Expert Systems with Applications*, 36(7) (2009) 10357-10366.
13. H. Li, H. Adeli, J. Sun and J. G. Han, Hybridizing principles of TOPSIS with case-based reasoning for business failure prediction, *Computers & Operations Research*, 38(2) (2011) 409-419.
14. Y. Lin, Y. M. Wang, and S. Q. Chen, Hesitant fuzzy multiattribute matching decision making based on regret theory with uncertain weights, *International Journal of Fuzzy Systems*, 19(4) (2016) 1-12.
15. A. Tversky, and D. Kahneman, Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5(4), (1992) 297-323.
16. Y. Liu, Z. P. Fan, Y. Zhang, Risk decision analysis in emergency response: A method based on cumulative prospect theory, *Computers & Operations Research*, 42 (2014) 75-82
17. Y. Qi. The same indifference interval multiple criteria matching decision method based on prospect theory, *Journal system science & mathematics science*, 33(12) (2013) 1447-1455. (Press in China)
18. P. C. Chang, C. H. Liu and R. K Lai, A fuzzy case-based reasoning model for sales forecasting in print circuit board industries. *Expert Systems with Applications*, 34(3) (2008) 2049-2058.
19. R. Ramanathan, Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process. *Computers & Operations Research*, 33(5), (2006) 1289-1307.
20. Y.M. Wang and K.S. Chin, A new data envelopment analysis method for priority determination and group decision making in the analytic hierarchy process. *European Journal of Operational Research*, 195(1) (2009) 239-250.
21. D. Kahneman, and A. Tversky, Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, (1979) 263-291.
22. A. Tversky and D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), (1992) 297-323.
23. M. Abdellaoui, H. Bleichrodt, and C. Paraschiv, Loss Aversion Under Prospect Theory: A Parameter-Free Measurement. *Management Science*, 53(10) (2007) 1659-1674.
24. H. Bleichrodt, U. Schmidt, and H. Zank, Additive Utility in Prospect Theory. *Management Science*, 55(5) (2009) 863-873.
25. J. Qi, J. Hu, Y. H. Peng, et al. AGFSM: An new FSM based on adapted Gaussian membership in case retrieval model for customer-driven design, *Expert Systems with Applications An International Journal*, 38(1) 2011 894-905.
26. Q. Y. Guan, X. L. Chen and Y. Z. Wang, Distance entropy based decision-making information fusion method. *Systems Engineering-Theory & Practice*, 35(1)

- (2015) 216-227. (Press in China)
27. M. Relich and P. Pawlewski, A case-based reasoning approach to cost estimation of new product development. *Neurocomputing*. (2017)
 28. G. N. Zhu, J. Hu, J. Qi, J. Ma, and Y. H. Peng, An integrated feature selection and cluster analysis techniques for case-based reasoning. *Engineering Applications of Artificial Intelligence*, 39, (2015) 14-22.