Comput Syst Sci & Eng (2018) 5: 397–404
© 2018 CRL Publishing Ltd

International Journal of
Computer Systems
Science & Engineering

Research on the clustering analysis and similarity in factor space

Sha-Sha Li[1,2]*, **Tie-Jun Cui**[1,2,3]† **and Jian Liu**[1,2]‡

[1] *College of Safety Science and Engineering, Liaoning Technical Univ., 123000 Fuxin, China*
[2] *Key Laboratory of Mine Thermodynamic Disasters and Control of Ministry of Education, 123000 Fuxin, China*
[3] *Tunnel & Underground Structure Engineering Center of Liaoning, Liaoning Dalian Jiaotong University, 116028 Dalian, China*

In this paper, we study the influence of multiple domain attributes on the clustering analysis of object based on factor space. The representation method of graphical domain attribute is proposed for the object, which is called attribute circle. An attribute circle can represent infinite domain attributes. The similarity analysis of objects is first based on the concept of attribute circle, and the definition of graphical similarity is transformed into the definition of numerical similarity, and then the clustering analysis method of object set is studied and improved. Considering three kinds of graphical overlap, the analytic solution of similarity is obtained for numerical calculation. The clustering rules: strictly obey the similarity division and dissimilarity division, and refer to fuzzy similarity division. The reliability evaluation semantics of the actual electrical system are listed as the study object set, and the clustering analysis method and its improvement are carried out. The results show that the relation between decision set D and object set U means that the division of U is nonsingular and accurate for D. Although the system reliability is evaluated in different environments, these evaluation semantics are relatively objective, and can support each other. The two methods of similarity calculation have the same conclusion, but the improved method is more accurate and complex

Keywords: Factor space; Multiple domain attribute; Attribute circle; Object classification; Similarity

1. INTRODUCTION

The theory of factor space has been created by Mr. Wang Peizhuang and has been developed to some extent. The authors have been studied issues in the field of safety system engineering, and used traditional methods too difficult to handle [1-8]. For example, in an investigation of the safety of an electrical system, an answer from an operator of system on system safety is when the system is 12°C below, there are many faults; After working seventy or eighty days, there are many faults, and the system is seriously unstable. The example has some characteristics. The example is a multi-factor decision system; The expression of a factor is a domain value, that is, a factor is a range value; The basic data comes from the experience of some operators, and different working time and environments make the experience different in their evaluation of the system; Basic data is a evaluations of things, with fuzziness; How do you know the confidence of these evaluations and whether these evaluations can support each other.

There are, of course, some methods can deal with the evaluation semantics of things.

The following researches have been carried out in the world on evaluation semantic analysis and similarity. Similarity measurements was used on multi-scale qualitative locations [9]. Semantic trajectories was measured by multidimensional similarity [10]. A new recommender framework was put forward, which was combining semantic similarity fusion and bicluster collaborative filtering [11]. Noise suppression for dual-energy CT via penalized weighted least-square optimization with similarity-based regularization [12]. An adaptive color similarity function was obtained; it was suitable for image segmentation and its numerical evaluation [13]. Linguistic Vector Similarity Measures was given, and applied to linguistic information classification [14].

The authors have made some researches on reasoning method based on data and evaluation semantics. System reliability assessment method was given based on space fault tree [15]. Connectivity reliability of directed acyclic network was studied con-

*917045352@qq.com
†ctj.159@163.com
‡LJ1961@vip.sina.com

sidering nodes and lines [16]. Decision criterion discovery of system reliability was studied [4]. The factor importance distribution was definite in Continuous Space Fault Tree [17]. System safety classification decision rules were obtained considering the scope attribute [18]. Accident chain model was studied based on CBM and human error [19]. Above methods are based on the given qualitative and fuzzy information to separate the relevant knowledge in the information, simplify, inference and relevance. The knowledge is used to determine the safety of the system being evaluated.

It is very difficult that these methods deal with the example with above characteristics, and difficult to satisfy the problem analysis with these characteristics. Because the above methods cannot analyze the influence of multiple domain attributes on clustering analysis under the evaluation semantics.

In order to solve the problem of evaluation semantics of multiple domain attributes and their clustering analysis, the attribute representation method of object is modified in factor space. The method can represent infinite domain attributes within a unit attribute circle. Then we analyze the similarity of objects and transform them into numerical expressions of similarity. We obtain the rules of clustering divisions among objects. The improved method of similarity calculation is proposed considering graphical overlap.

2. PRELIMINARY KNOWLEDGE

In order to adapt to the information revolution and the needs of big data age, the early Chinese scholar Professor P. Z. Wang in 1983 proposed the Factor Space (FS) mathematical theory [20-23] for big data analysis and processing, and laid the mathematical foundation. At present, FS has been extensively studied and acknowledged, the related studies include: background information compression [24], FS processing unstructured data [25], factors granular space nested structure and data cognitive ecosystem [26- 29], evaluation and decision theory [30-34], FS and public safety [35], algebra, topology, differential geometry, category theory of comprehensive research [36], etc.

FS theory has made a great breakthrough in the field of intelligent science and big data, but not applied to the specific science and technology. The combination of Space Fault Tree(SFT) and FS provides the development space for FS in the reliability and fault analysis of safety science. Combined with SFT and FS theory, some researches have been done, including: the study of system safety classification decision rules considering range attributes [37], coal mine disaster safety analysis based on factor analysis [38], coal mine safety situation distinction method [39], and also carried out some system reliability analysis [40].

Factor is the element that analyzes the thing attributes and causal relationships. Factor space is the coordinate space named by factors. It is the universal mathematical framework for describing things. It is the basic mathematical theory of artificial intelligence, especially intelligent data science.

A factor is mathematically defined as a mapping. It maps an object to an attribute value, called a quantitative mapping. At the same time, it is mapped into an attribute, called qualitative mapping. For example, Jack's height is a mapping that maps an object(person) to a qualitative attribute of 'high'; and it is mapped into a quantitative attribute value 1.8m, as shown in Fig.1. Everything has two kinds of mapping states, namely quantitative and qualitative. From quantitative change to qualitative change, quantity determines quality.

Based on this philosophy, we set these two mappings together. The value of factor f is mapped into a (one-dimensional or higher dimensional) coordinate axis X_f. The attribute(means factor in attribute circle) values obtained by its quantitative mapping are ordinary set or fuzzy subsets in X_f. The method of fuzzy subset formation has already been solved by fuzzy set theory.

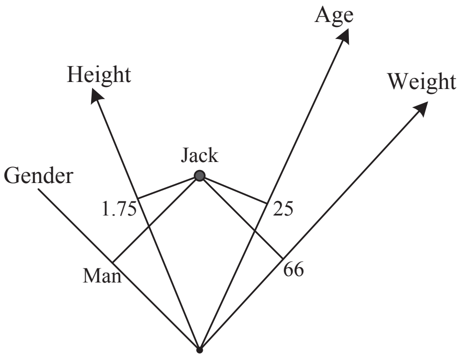

Figure 1 Person attributes in factor space.

From Fig. 1, the multiple factor axes are combined to obtain the coordinate frame named by factors, which is called factor space. Anything can be regarded as the point in the factor space. Its mathematical definition is a set that factor set F is the index set. Here, F is a collection of many factors. Since Boolean operations exist between factors and factors, F is a Boolean algebra too. Therefore, the factor space is defined as a set of families with Boolean algebra as the index set. Factor library is a new database, which is the data realization of factor space theory. It uses a series of basic table forms to process data.

3. THE CONCEPTS AND PROPERTIES OF ATTRIBUTE CIRCLE

Fig. 1 is the representation of human factors in factor space. The figure can represent the basic idea of factor space, that is, the relationship between an object (person) and its attributes(gender, height, age, weight). If these attributes are determined, then an instantiated person is determined. But in practical problems, the object has often more attributes. Using the form of Fig. 1, it is difficult to describe the value and state of the attributes and the relation between objects and their attributes, which is not conducive to further analysis. So the concept of attribute circle is put forward in this paper. For the expression, first, we give the attribute circle of x_1 in the example, as shown in Fig. 2. In attribute circle, the factor in factor space is called attribute.

Definition 1: Set system $T = (U, A, C, D)$ as decision table, $U = \{x_1, x_2, \ldots, x_m\}$ is object set, m is the number of objects; $C = \{a_1, a_2, \ldots, a_n\}$ is condition attribute set, n is the number of conditions; $a_q = [a_q^e, a_q^f]$ is a continuous range of attribute value, $i, j \in \{1, \ldots, m\}$, $q, p \in \{1, \ldots, n\}$ $D = \{d_1, d_2, \ldots, d_k\}$, k is the number of decisions. To distinguish the variable concept among objects, add x_i below the

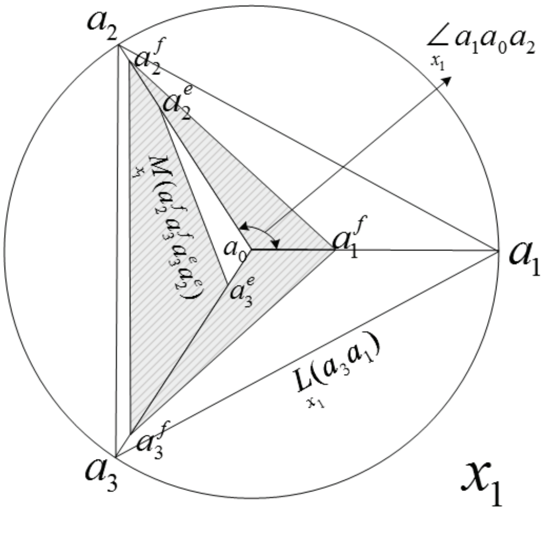

Figure 2 Attribute circle of the object x_1.

variable, such as $a_{1_{x_1}}$ represents the attribute a_1 of the object x_1.

Definition 2: Construct the basic information decision table $\Psi(T)$ to represent system T. The header of $\Psi(T)$ is $\{U, C, D\}$, where, the attribute a_q in C must be normalized. The actual domain of $a_{q_{x_i}}$ is $[A, B]$, the study domain is $[LL, UL]$, $LL \leq A$, $UL \geq B$, $a^e_{q_{x_i}} = (A - LL)/(UL - LL)$, $a^f_{q_{x_i}} = (B - LL)/(UL - LL)$.

From the above definition, we know that the data in $\Psi(T)$ is normalized, that is, $0 \leq a_q^e \leq 1$, $0 \leq a_q^f \leq 1$, $a_q^e \leq a_q^f$, which provides the basis for the establishment of attribute circle.

Definition 3: The attribute circle is a unit circle in the coordinate system, its radius is 1. In this coordinate system, the attribute circle can represent all objects in the object set. The connection line segment between a point a_q on the circumference of attribute circle and the center a_0 of the attribute circle is called attribute domain line segment (short for domain line). It represents the value ranges of attributes (normalization) of all objects in the domain, and the line segment length is 1. a_q^e, a_q^f are the endpoints of the domain line. a_q^e represents the start point of the attribute domain value, and a_q^f represents the end point. The domain line segment in the attribute circle is denoted by $L(\mathcal{K}_1, \mathcal{K}_2)$, and $\mathcal{K}_1$, $\mathcal{K}_2$ are any two points in the attribute circle, such as the a_q domain line denoted by $L(a_q, a_0)$. Attribute angle $\angle_{x_1} a_1 a_0 a_2$ is the angle between the domain line $L(a_q, a_0)$ and $L(a_{q+1}, a_0)$. The area in the attribute circle is denoted by $M(\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_O)$, and $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_O$ represent any number of points in the attribute circle, which can form a convex polygon in the order of occurrence. Attribute circle define rules can be summarized, as shown in Eq. (1). The parameters are defined by Def. 1 and 2.

$$\left\{\begin{array}{c} L(a_1, a_0) = L(a_2, a_0) = \cdots L(a_n, a_0) = 1 \\ \angle a_1 a_0 a_2 = \angle a_2 a_0 a_3 = \cdots = \angle a_{n-1} a_0 a_n \\ \angle a_1 a_0 a_2 + \angle a_2 a_0 a_3 + \cdots + \angle a_{n-1} a_0 a_n = 360^\circ \\ 0 \leq a_q^e \leq 10 \leq a_q^f \leq 1 a_q^e \leq a_q^f [a_q^e, a_q^f] \subseteq L(a_q, a_0) \end{array}\right. \tag{1}$$

Definition 4: $L(a^e_{q_{x_i}}, a^f_{q_{x_i}})$ or $L_{x_i}(a_q^e, a_q^f)$ represents the value domain of attribute a_q of object x_i. The larger the $L_{x_i}(a_q^e, a_q^f)$, it is the less influence of the attribute a_q on the object x_i; the smaller the $L_{x_i}(a_q^e, a_q^f)$, the greater the influence of the a_q on the object x_i.

4. OBJECT CLUSTERING ANALYSIS METHOD BASED ON ATTRIBUTE CIRCLE

For the description of the classification method, the graphical similarity definition between x_1 and x_6 is given, as shown in Fig. 3.

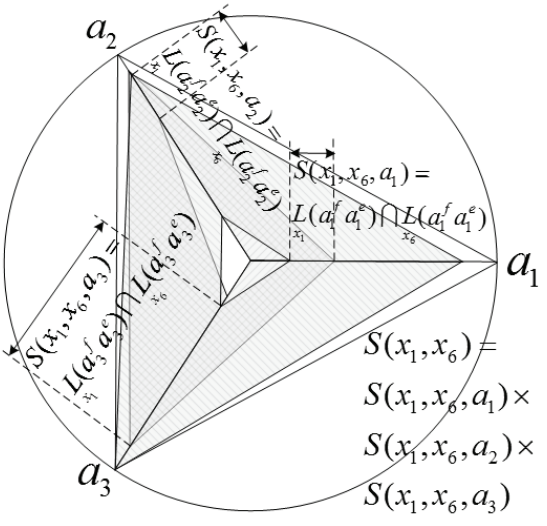
a_2
$S(x_1, x_6, a_2) =$
$L_{x_1}(a_2^f a_2^e) \bigcap L_{x_6}(a_2^f a_2^e)$
$S(x_1, x_6, a_1) =$
$L_{x_1}(a_1^f a_1^e) \bigcap L_{x_6}(a_1^f a_1^e)$
a_1
$S(x_1, x_6, a_3) =$
$L_{x_1}(a_3^f a_3^e) \bigcap L_{x_6}(a_3^f a_3^e)$
$S(x_1, x_6) =$
$S(x_1, x_6, a_1) \times$
$S(x_1, x_6, a_2) \times$
$S(x_1, x_6, a_3)$
a_3

Figure 3 Similarity definition between x_1 and x_6.

Firstly, the concept of object similarity is given from the geometric meaning, which is graphical similarity considering graphical overlap. As shown in Fig. 2, $M_{x_1}(a_2^f a_3^f a_3^e a_2^e)$ represents a convex polygon for representing both the attribute values of the object x_1 on the attribute a_2, a_3. Fig. 3 shows graphical overlap of the attribute circles of x_1 and x_6, then the overlap area of $M_{x_1}(a_2^f a_3^f a_3^e a_2^e)$ and $M_{x_6}(a_2^f a_3)$ can greatly reflect the similarity between x_1 and x_6 about the attribute a_2, a_3. Please note that the shading states of x_1 and x_6 are different.

However, it is difficult to determine the similarity between x_1 and x_6 using the above methods. The overlap area of $M_{x_1}(a_2^f a_3^f a_3^e a_2^e)$ and $M_{x_6}(a_2^f a_3^f a_3^e a_2^e)$ in the above method reflects the similarity between x_1 and x_6 about the two attribute a_2, a_3, and the single attribute is determined difficulty. On the other hand, the overlap area between $M_{x_1}(a_2^f a_3^f a_3^e a_2^e)$ and $M_{x_6}(a_2^f a_3^f a_3^e a_2^e)$ needs to be calculated by complex analytical methods, which is not satisfactory for engineering applications. Therefore, graphical similarity is transformed into numerical calculation method to define and use.

From Fig. 3, there is the overlap line segment between $L_{x_1}(a_3^e, a_3^f)$ and $L_{x_6}(a_3^e, a_3^f)$ on the attribute a_3. This overlap line segment shows that the attribute a_3 has an line segment $(L_{x_1}(a_3^e, a_3^f) \bigcap L_{x_6}(a_3^e, a_3^f))$ that has the same effect on x_1 and x_6, that is, x_1 and x_6 are similar in that line segment. Based on this idea, the similarity is defined considering overlap line segment, not overlap area.

Definition 5: In system T, $xi, x_j \in U$, define $S(x_i, x_j, a_q)$ as the similarity between x_i and x_j about attribute a_q, $S(x_i, x_j, a_q)$ is determined as follows.

When $i = j$, $S(x_i, x_j, a_q) = 1$, an object is compared to itself, the similarity is 1.

When $i \neq j$, Compare the relative overlap line segment of $a_q = [a_q^e, a_q^f]$ and $a_q = [a_q^e, a_q^f]$.

When there is no overlap line segment between $a_{q_{x_i}}$ and $a_{q_{x_j}}$ on $L(a_q, a_0)$, $S(x_i, x_j, a_q) = 0$, which indicates that two objects are not related to a_q.

When there is the overlap line segment between $a_{q_{x_i}}$ and $a_{q_{x_j}}$ on $L(a_q, a_0)$, $S(x_i, x_j, a_q) = 0$, according to the overlap, the $S(x_i, x_j, a_q)$ is obtained, as shown in Eq. (2).

$$\begin{aligned} S(x_i,x_j, a_q) &= \underset{x_i}{L}(a_q^f a_q^e) \cap \underset{x_j}{L}(a_q^f a_q^e) \\ &= MIN\left(\left|\underset{x_j}{a_q^f} - \underset{x_j}{a_q^e}\right|, \left|\underset{x_j}{a_q^f} - \underset{x_i}{a_q^e}\right|, \left|\underset{x_i}{a_q^f} - \underset{x_j}{a_q^e}\right|, \left|\underset{x_i}{a_q^f} - \underset{x_i}{a_q^e}\right|\right) \\ &\quad \div MAX\left(\left|\underset{x_j}{a_q^f} - \underset{x_j}{a_q^e}\right|, \left|\underset{x_j}{a_q^f} - \underset{x_i}{a_q^e}\right|, \left|\underset{x_i}{a_q^f} - \underset{x_j}{a_q^e}\right|, \left|\underset{x_i}{a_q^f} - \underset{x_i}{a_q^e}\right|\right) \end{aligned} \tag{2}$$

Where, $0 \leq S(x_i, x_j, a_q) \leq 1$.

Definition 6: The whole similarity of x_i, x_j is $S(x_i, x_j)$, for $C = \{a_1, a_2, \ldots, a_n\}$, then $S(x_i,x_j) = \prod_{q=1}^{n} S(x_i,x_j, a_q), a_q \in C$.

The specific processes of the above definitions are shown in Fig. 3.

Definition 7: Whole similarity $S(x_i, x_j)$ of x_i, x_j is classified by the following rules. Set λ_{a_q} as the similarity threshold of x_i, x_j for a single attribute a_q, generally $0.4 \leq \lambda_{a_q} \leq 0.6$. $1 \geq S(x_i, x_j, a_q) \geq \lambda_{a_q}$ means similarity, $S(x_i, x_j, a_q) = 0$ means dissimilarity, the values between them means fuzzy similarity. So for $S(x_i, x_j)$, $1 \geq S(x_i,x_j) = \prod_{q=1}^{n} S(x_i,x_j, a_q) \geq \prod_{q=1}^{n} \lambda_{a_q}$ means similarity, $S(x_i, x_j) = 0$ means dissimilarity, the values between them means fuzzy similarity.

5. IMPROVEMENTS OF CLUSTERING ANALYSIS METHOD

The above method simplifies geometric similarity and obtains a simple similarity algorithm. But the original geometric similarity is more accurate, so this section gives the concept of similarity algorithm based on geometric similarity. In Fig. 2, $M_{x_1}(a_2^f a_3^f a_3^e a_2^e)$ represents a convex polygon with representing both the attribute values of the object x_1 on the attribute a_2, a_3. The overlap area of $M_{x_1}(a_2^f a_3^f a_3^e a_2^e)$ and $M_{x_6}(a_2^f a_3^f a_3^e a_2^e)$ can greatly reflect the similarity between x_1 and x_6 about the attribute a_2, a_3.

Calculates the similarity between x_i and x_j on attributes a_q and a_{q+1}, and that is the overlap area of $M_{x_i}(a_q^f a_{q+1}^f a_{q+1}^s a_q^s)$ and $M_{x_1}(a_q^f a_{q+1}^f a_{q+1}^e a_q^e)$.

Definition 8: In system T, x_i, $x_j \in U$, define $S(x_i, x_j, (a_q, a_{q+1}))$ as the similarity between x_i and x_j about attribute a_q, a_{q+1} as shown in Eq. (3).

$$\begin{aligned} &S(x_i, x_j, (a_q, a_{q+1})) \\ &\quad = \frac{M_{x_i}(a_q^f a_{q+1}^f a_{q+1}^e a_q^e) \cap M_{x_j}(a_q^f a_{q+1}^f a_{q+1}^e a_q^e)}{M_{x_i}(a_q^f a_{q+1}^f a_{q+1}^e a_q^e) \cup M_{x_j}(a_q^f a_{q+1}^f a_{q+1}^e a_q^e)} \end{aligned} \tag{3}$$

The different states are described as followed.

1) When there is no overlap line segment between $a_{q_{x_i}}$ and $a_{q_{x_j}}$ on $L(a_q, a_0)$, or $a_{q_{x_i}}$ and $a_{q_{x_j}}$ on $L(a_{q+1}, a_0)$, $S(x_i, x_j, (a_q, a_{q+1})) = 0$, which indicates that two objects are not related to a_q, a_{q+1}.

2) When there is the overlap line segment between $a_{q_{x_i}}$ and $a_{q_{x_j}}$ on $L(a_q, a_0)$, in addition $a_{q+1_{x_i}}$ and a_{q+1} on $L(a_{q+1}, a_0)$; at the same time, the line segment $(a_{q_{x_j}}^e, a_{q+1_{x_j}}^e)$ does not intersect the line segment (a_q^e, a_{q+1}^e), in addition the line segment (a_q^f, a_{q+1}^f) does not intersect the line segment (a_q^f, a_{q+1}^f), then the overlap area is defined as shown in Eqs. (4) and (5). This is the second overlap state as shown in Fig. 4.

$$\begin{aligned} &\underset{x_i}{M}(a_q^f a_{q+1}^f a_{q+1}^e a_q^e) \cap \underset{x_j}{M}(a_q^f a_{q+1}^f a_{q+1}^e a_q^e) = \tfrac{1}{2}(bd - ac)\sin\theta \\ &\theta = \angle a_q a_0 a_{q+1}, d = \min\{\underset{x_i}{a_q^f}, \underset{x_j}{a_q^f}\}, b = \min\{\underset{x_i}{a_{q+1}^f}, \underset{x_j}{a_{q+1}^f}\}, \\ &a = \max\{\underset{x_i}{a_q^e}, \underset{x_j}{a_q^e}\}, c = \max\{\underset{x_i}{a_{q+1}^e}, \underset{x_j}{a_{q+1}^e}\} \end{aligned} \tag{4}$$

$$\begin{aligned} &\underset{x_i}{M}(a_q^f a_{q+1}^f a_{q+1}^e a_q^e) \cup \underset{x_j}{M}(a_q^f a_{q+1}^f a_{q+1}^e a_q^e) = \tfrac{1}{2}(bd - ac)\sin\theta \\ &\theta = \angle a_q a_0 a_{q+1}, d = \max\{\underset{x_i}{a_q^f}, \underset{x_j}{a_q^f}\}, b = \max\{\underset{x_i}{a_{q+1}^f}, \underset{x_j}{a_{q+1}^f}\}, \\ &a = \min\{\underset{x_i}{a_q^e}, \underset{x_j}{a_q^e}\}, c = \min\{\underset{x_i}{a_{q+1}^e}, \underset{x_j}{a_{q+1}^e}\} \end{aligned} \tag{5}$$

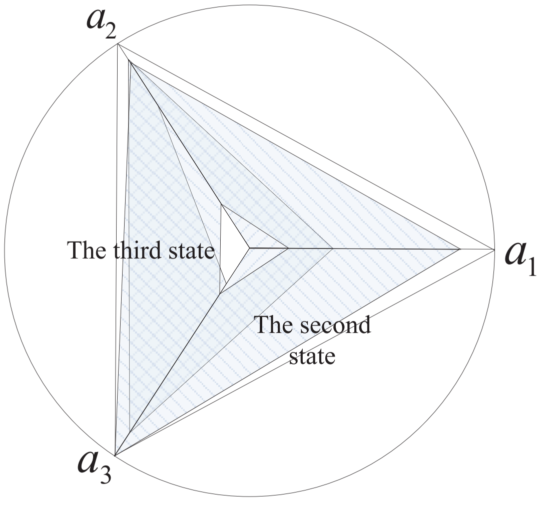

Figure 4 Overlap states of similarity definition between x_1 and x_6.

3) The third state, when there is the overlap line segment between $a_{q_{x_i}}$ and $a_{q_{x_j}}$ on $L(a_q, a_0)$, in addition a_{q+1} and $a_{q+1_{x_j}}$ on

$L(a_{q+1}, a_0)$; at the same time, the line segment $(a^e_{q_{x_j}}, a^e_{q+1_{x_j}})$ intersect the line segment $(a^e_{q_{x_i}}, a^e_{q+1_{x_i}})$, or the line segment $(a^f_{q_{x_j}}, a^f_{q+1_{x_j}})$ intersect the line segment $(a^f_{q_{x_i}}, a^f_{q+1_{x_i}})$, then the overlap area is defined as shown in Eqs. (6) and (7). This is the third overlap state as shown in Fig. 4.

$$\begin{aligned}&\underset{x_i}{M}(a^f_q a^f_{q+1} a^e_{q+1} a^e_q) \cap \underset{x_j}{M}(a^f_q a^f_{q+1} a^e_{q+1} a^e_q) = \\ &\tfrac{1}{2}hc\sin\theta - \tfrac{1}{2}(h-g)^2 \times \tfrac{\cos\theta}{h/c-g/d} - \tfrac{1}{2}eb\sin\theta \\ &-\tfrac{1}{2}(f-e)^2 \times \tfrac{\cos\theta}{f/a-e/b}\end{aligned} \tag{6}$$

$$\begin{aligned}&\underset{x_i}{M}(a^f_q a^f_{q+1} a^e_{q+1} a^e_q) \cup \underset{x_j}{M}(a^f_q a^f_{q+1} a^e_{q+1} a^e_q) = \\ &\tfrac{1}{2}dg\sin\theta - \tfrac{1}{2}(h-g)^2 \times \tfrac{\cos\theta}{h/c-g/d} - (\tfrac{1}{2}fa\sin\theta \\ &-\tfrac{1}{2}(f-e)^2 \times \tfrac{\cos\theta}{f/a-e/b})\end{aligned} \tag{7}$$

Where, $\theta = \angle a_q a_0 a_{q+1}$, $a = a^e_{q+1_{x_j}}$, $b = a^e_{q+1_{x_i}}$, $c = a^f_{q+1_{x_j}}$, $d = a^f_{q+1_{x_i}}$, $e = a^e_{q_{x_i}}$, $f = a^e_{q_{x_j}}$, $g = a^f_{q_{x_i}}$, $h = a^f_{q_{x_j}}$.

To ensure that Eqs. (6) (7) are clean and tidy; letters are substituted for numerical markers.

The results in 2) and 3) are the analytic solution obtained by graphical calculation, because of the limitation of space, the analytic procedure here is abbreviated.

Definition 9: The whole similarity of x_i, x_j is $S(x_i, x_j)$, for $C = \{a_1, a_2, \ldots, a_n\}$, then $S(x_i,x_j) = \prod_{q=1}^{n-1} S(x_i,x_j,(a_q, a_{q+1})) \times S(x_i,x_j,(a_n, a_1))$, $a_q \in C$.

Definition 10: Whole similarity $S(x_i, x_j)$ of x_i, x_j is classified by the following rules. Set $\lambda_{a_q,a_{q+1}}$ as the similarity threshold of x_i, x_j for two attribute a_q, a_{q+1}, generally $0.4 \leq \lambda_{a_q,a_{q+1}}$. $1 \geq S(x_i, x_j, (a_q, a_{q+1})) \geq \lambda_{a_q}$ means similarity, $S(x_i, x_j, (a_q, a_{q+1})) = 0$ means dissimilarity, the values between them means fuzzy similarity. So for $S(x_i, x_j(a_a, a_{q+1})) = 0$, means similarity, means dissimilarity, the values between them means fuzzy similarity.

6. EXAMPLE ANALYSES

For reliability analysis of an electrical system, 7 operators who have used the system have been investigated. They give some descriptions for evaluation of the system reliability. Because of their work, scheduling and other reasons, the environment of their operating systems are different. In fact, there are many factors affecting the probability of component fault in the system. For example, the diode fault in the electrical system has a direct relationship with working time, temperature, current and voltage. If the system is analyzed, the adaptive working time and temperature of each component may be different. With the change of the whole system working time and environment temperature, the system reliability is different too[41]. Therefore, the basic environment of the system reliability evaluation is different for the above operators.

Proposed methods are used to classify the reliability evaluation of these operators. If the object set classification (semantic description sets) is the same of decision set classification (semantic result sets), then these operators are objective and can support each other for system reliability evaluation. If the classification does not correspond, the evaluation description of the more operators should be added to further determine the accuracy of the descriptions.

Based on the survey, an operator answer for system reliability is: system faults was more below 12°C; there are more faults after working seventy or eighty days, and serious instability (due to limited space, the other 6 are not listed). The system is generally overhauled every 100 days, we can set the time domain of [0, 100] day. The temperature domain is [0,40]°C; The humidity is generally determined by the seasonal climate during the work period.

Define system $T = (U, A, C, D)$, the descriptions of the 7 operators as object set $U = \{x_1, x_2, \ldots, x_7\}$, and x_i is the description of the ith operator, $i \in \{1, \ldots, 7\}$. The working time, temperature, humidity of system make up condition attribute set $C = \{a_1, a_2, a_3\}$, a_1 is temperature, a_2 is time, a_3 is humidity. a_1, a_2 and a_3 are the continuous domain value, and are normalized according to the descriptions of the 7 operators. If the operator's answer is x_1, the domains of evaluation semantics are working temperature of [0,12] C, working time of [70,95]d, and humidity. They are normalized as followed: $a_{1_{x_1}} = [a^e_1, a^f_1]$, $a^e_{1_{x_1}} = (0-0)/(40-0) = 0$, $a^f_{1_{x_1}} = (12-0)/(40-0) = 0.3$, similarly, $a_{2_{x_1}} = [0.7, 0.95]$, $a_{3_{x_1}} = [0.2, 0.9]$. The decision set $D = \{d_1, d_2, d_3\}$ represents the safety levels from one to three, which are "unreliable", "generally reliable" and "very reliable". Get the basic information decision table $\Psi(T)$, as shown in Tab. 1. Without considering the decision set D, we study the attribute circle representation of the object set and attribute set. The attribute circle x_1 is given. The attribute circles $x_2 \sim x_7$ are shown in Fig. 5. According to Tab. 1 and Defs. 5 and 6, we obtain the object similarity table, as shown in table 2.

Table 1 Basic information decision table $\Psi(T)$.

U	a$_1$ temperature	a$_2$ time	a$_3$ humidity	D safety levels
x$_1$	[0,0.3]	[0.7,0.95]	[0.2,0.9]	1 low
x$_2$	[0,0.4]	[0.8,1]	[0.3,0.9]	1
x$_3$	[0.2,0.7]	[0.5,0.8]	[0.1,0.8]	2
x$_4$	[0.6,0.9]	[0.2,0.6]	[0.2,0.8]	3
x$_5$	[0.25,0.7]	[0.5,0.9]	[0.1,0.9]	2
x$_6$	[0.2,0.8]	[0.3,0.9]	[0.3,1]	2
x$_7$	[0.6,0.9]	[0.2,0.4]	[0,0.9]	3 high

For object set classification, first define $\lambda_{a_1} = 0.5$, $\lambda_{a_2} = 0.5$, $\lambda_{a_3} = 0.5$, and similarity division of $S(x_i, x_j)$ is {similarity, fuzzy similarity, dissimilarity}={[1,0.125], (0.125,0), 0}. The similar objects obtained with Tab. 2 are classified as follows:

Similarity: $S(x_2, x_1) = 0.3214$, $S(x_5, x_3) = 0.5906$, $S(x_6, x_3) = 0.2315$, $S(x_6, x_5) = 0.2632$, $S(x_7, x_4) = 0.2592$.

Fuzzy similarity: $S(x_3, x_1) = 0.0238$, $S(x_4, x_3) = 0.0204$, $S(x_3, x_1) = 0.0278$, $S(x_5, x_2) = 0.0321$, $S(x_5, x_4) = 0.0165$, $S(x_6, x_1) = 0.0288$, $S(x_6, x_2) = 0.0306$, $S(x_6, x_4) = 0.0765$, $S(x_7, x_6) = 0.0245$.

Dissimilarity: $S(x_3, x_2)$, $S(x_4, x_1) = 0$, $S(x_4, x_2) = 0$, $S(x_7, x_1) = 0$, $S(x_7, x_2) = 0$, $S(x_7, x_3) = 0$, $S(x_7, x_5) = 0$.

The clustering analysis rules: strictly obey the similarity division and dissimilarity division, and refer to fuzzy similarity division. For example, $S(x_2, x_1) = 0.3214$ means x_2, x_1 should

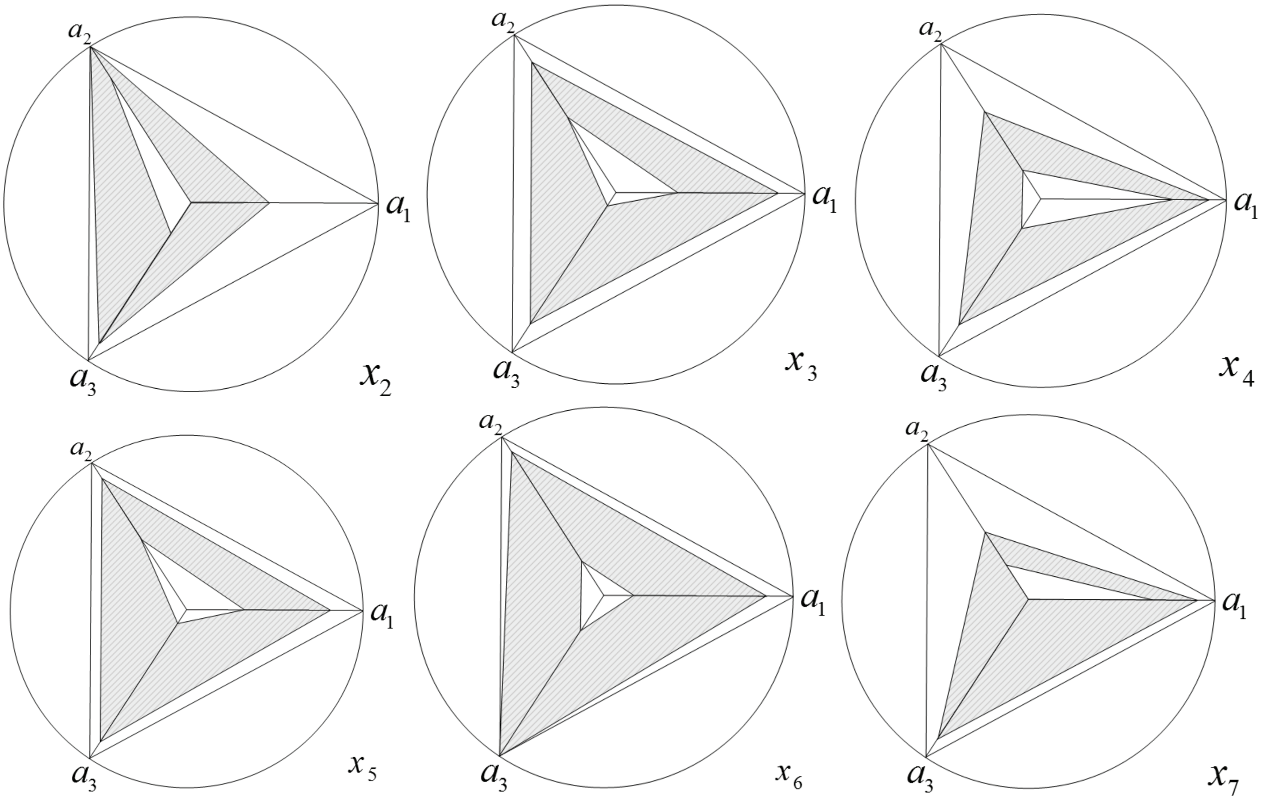

Figure 5 Attribute circle of the object $x_2 \sim x_7$.

Table 2 Object similar table.

	x1			x2			x3			x4			x5			x6			x7		
	a1	a2	a3	a1	a2	a3	a1	a2	a3	a1	a2	a3	a1	a2	a3	a1	a2	a3	a1	a2	a3
x1	1	1	1																		
	$S(x_1,x_1)=1$																				
x2	0.7500	0.5000	0.8571	1	1	1															
	$S(x_2,x_1)=0.3214$			$S(x_2,x_2)=1$																	
x3	0.1429	0.2222	0.7500	0	0	0.6250	1	1	1												
	$S(x_3,x_1)=0.0238$			$S(x_3,x_2)=0$			$S(x_3,x_3)=1$														
x4	0	0	0.8571	0.2222	0	0.7143	0.1429	0.1667	0.8571	1	1	1									
	$S(x_4,x_1)=0$			$S(x_4,x_2)=0$			$S(x_4,x_3)=0.0204$			$S(x_4,x_4)=1$											
x5	0.0714	0.4444	0.8750	0.2143	0.2	0.7500	0.9000	0.7500	0.8750	0.1538	0.1429	0.7500	1	1	1						
	$S(x_5,x_1)=0.0278$			$S(x_5,x_2)=0.0321$			$S(x_5,x_3)=0.5906$			$S(x_5,x_4)=0.0165$			$S(x_5,x_5)=1$								
x6	0.1250	0.3077	0.7500	0.2500	0.1429	0.8571	0.8333	0.5000	0.5556	0.2857	0.4286	0.6250	0.5921	0.6667	0.6667	1	1	1			
	$S(x_6,x_1)=0.0288$			$S(x_6,x_2)=0.0306$			$S(x_6,x_3)=0.2315$			$S(x_6,x_4)=0.0765$			$S(x_6,x_5)=0.2632$			$S(x_6,x_6)=1$					
x7	0	0	0.8889	0	0	0.6667	0.1429	0	0.7778	1	0.3333	0.7778	0.1538	0	0.8889	0.2857	0.1429	0.6000	1	1	1
	$S(x_7,x_1)=0$			$S(x_7,x_2)=0$			$S(x_7,x_3)=0$			$S(x_7,x_4)=0.2592$			$S(x_7,x_5)=0$			$S(x_7,x_6)=0.0245$			$S(x_7,x_7)=1$		

be divided into one group; $S(x_3, x_2) = 0$ means x_3, x_2 should not be divided into one group. So the final object set is divided into $U = \{\{x_2, x_1\}, \{x_7, x_4\}, \{x_5, x_3, x_6\}\}$. Considering the corresponding relation between the decision set D and the object set U in Tab. 1, we find $U \rightarrow D = \{\{x_2, x_1\} \rightarrow d_1, \{x_7, x_4\} \rightarrow d_3, \{x_5, x_3, x_6\} \rightarrow d_2\}$. This shows that the classification of object sets is non-singular and accurate for decision set.

Similarly, in accordance with the above settings, the improved method is used to obtain the object similarity, and the result is shown in Tab. 3.

Set $\lambda_{a_1} = 0.5$, $\lambda_{a_2} = 0.5$, $\lambda_{a_3} = 0.5$, and similarity division of $S(x_i, x_j)$ is {similarity, fuzzy similarity, dissimilarity} = {[1,0.125], (0.125,0), 0}. The similar objects obtained with Tab. 3 are classified as follows:

Similarity: $S(x_2, x_1) = 0.5138$, $S(x_4, x_3) = 0.1650$, $S(x_5, x_3) = 0.7706$, $S(x_6, x_3) = 0.4225$, $S(x_7, x_4) = 0.4069$, $S(x_6, x_5) = 0.5288$.

Fuzzy similarity: $S(x_3, x_1) = 0.0384$, $S(x_5, x_1) = 0.0197$, $S(x_6, x_1) = 0.0717$, $S(x_5, x_2) = 0.0486$, $S(x_6, x_2) = 0.0989$, $S(x_5, x_4) = 0.0372$, $S(x_6, x_4) = 0.0650$, $S(x_7, x_6) = 0.0709$.

Dissimilarity: $S(x_4, x_1)$, $S(x_7, x_1) = 0$, $S(x_3, x_2) = 0$, $S(x_4, x_2) = 0$, $S(x_7, x_2) = 0$, $S(x_7, x_3) = 0$, $S(x_7, x_5) = 0$.

Final object set division $U = \{\{x_2, x_1\}, \{x_7, x_4\}, \{x_5, x_3, x_6\}\}$. But this division is inconsistent with $S(x_4, x_3) = 0.1650$. Since

Table 3 Object similar table.

	x_1			x_2			x_3			x_4			x_5			x_6			x_7		
	a_1,a_2	a_2,a_3	a_3,a_1	a_1,a_2	a_2,a_3	a_3,a_1	a_1,a_2	a_2,a_3	a_3,a_1	a_1,a_2	a_2,a_3	a_3,a_1	a_1,a_2	a_2,a_3	a_3,a_1	a_1,a_2	a_2,a_3	a_3,a_1	a_1,a_2	a_2,a_3	a_3,a_1
x_1	1	1	1																		
	$S(x_1,x_1)=1$																				
x_2	0.8210	0.8549	0.7321	1	1	1															
	$S(x_2,x_1)=0.5138$			$S(x_2,x_2)=1$																	
x_3	0.1891	0.6323	0.3212	0	0	0.4550	1	1	1												
	$S(x_3,x_1)=0.0384$			$S(x_3,x_2)=0$			$S(x_3,x_3)=1$														
x_4	0	0	0	0	0	0	0.4623	0.6450	0.5532	1	1	1									
	$S(x_4,x_1)=0$			$S(x_4,x_2)=0$			$S(x_4,x_3)=0.1650$			$S(x_4,x_4)=1$											
x_5	0.1099	0.8344	0.2150	0.1756	0.6324	0.4375	0.8750	0.9200	0.9573	0.0988	0.5790	0.6500	1	1	1						
	$S(x_5,x_1)=0.0197$			$S(x_5,x_2)=0.0486$			$S(x_5,x_3)=0.7706$			$S(x_5,x_4)=0.0372$			$S(x_5,x_5)=1$								
x_6	0.3450	0.8312	0.2500	0.4500	0.6155	0.3571	0.6756	0.7690	0.8132	0.2290	0.4490	0.6322	0.7300	0.8121	0.8920	1	1	1			
	$S(x_6,x_1)=0.0717$			$S(x_6,x_2)=0.0989$			$S(x_6,x_3)=0.4225$			$S(x_6,x_4)=0.0650$			$S(x_6,x_5)=0.5288$			$S(x_6,x_6)=1$					
x_7	0	0	0.4859	0	0	0.4226	0	0	0.8320	0.8002	0.7060	0.7203	0	0	0.9590	0.2230	0.3454	0.9200	1	1	1
	$S(x_7,x_1)=0$			$S(x_7,x_2)=0$			$S(x_7,x_3)=0$			$S(x_7,x_4)=0.4069$			$S(x_7,x_5)=0$			$S(x_7,x_6)=0.0709$			$S(x_7,x_7)=1$		

the similarities among x_5, x_3, x_6 are higher than the similarity between x_4 and x_3, and x_7 can not be divided with x_5, x_3, and the similarity between x_7 and x_4 is very high, so the above division is reasonable. Considering the corresponding relation between the decision set D and the object set U in Tab. 1, we find $U \to D = \{\{x_2, x_1\} \to d_1, \{x_7, x_4\} \to d_3, \{x_5, x_3, x_6\} \to d_2\}$. This shows that the object set division is non-singular and accurate for decision set. This shows that the initial division of the decision set is correct and can be tested in practice.

The definition of attribute circle is put forward and the methods of object similarity calculation are constructed based on attribute circle. The first method reduces geometric similarity to facilitate computation, but at the expense of accuracy. The improved method is based on geometric similarity, more precise, but complex. In general, the results are the same, but the more attributes, the more accurate the improved method, but the more complex.

7. CONCLUSIONS

The attribute representation method of the factor space object is modified. In a unit attribute circle, the infinite multiple domain attributes can be indicated. Then we can analyze the similarity of objects and transform them into numerical expressions of similarity. The similarity calculation method is improved. In the attribute circle, the overlap area of different objects for the same attributes is defined as similarity, that is, the geometric overlap similarity calculation method. In general, the results of two methods are the same, the more attributes, the more accurate the improved method, but the more complex the calculation.

We obtain the rules of object clustering, which strictly follow the similarity and dissimilarity division, and refer to the fuzzy similarity division to classify the object set.

The results show that if the correspondence relation between object sets U and decision set D is nonsingular ($U \to D = \{\{x_2, x_1\} \to d_1, \{x_7, x_4\} \to d_3, \{x_5, x_3, x_6\} \to d_2\}$), although the system has different environmental factors, but a attribute evaluation semantic of the system is relatively objective for every object, and can support each other, then these evaluation semantic are correct. If it is strange, then needs to be added the semantic evaluation to describe, final further determine it.

Acknowledgements

The author wishes to thank all his friends for their valuable critics, comments and assistances on this paper. This study was partially supported by the grants (Grant Nos. 2017YFC1503100, 51704141, 51474121) from the Natural Science Foundation of China.

REFERENCES

1. Tie-Jun Cui, Pei-Zhuang Wang, Sha-Sha Li. The function structure analysis theory based on the factor space and space fault tree. Cluster Computing, 2017, 20(2): 1387-1398.
2. Tie-Jun Cui, Sha-Sha Li. Study on the Relationship between System Reliability and Influencing Factors under Big data and Multi-factors. Cluster Computing, https://doi.org/10.1007/s10586-017-1278-5.
3. Tie-Jun Cui, Sha-Sha Li. Study on the construction and application of Discrete Space Fault Tree modified by Fuzzy Structured Element. Cluster Computing, https://doi.org/10.1007/s10586-018-2342-5.
4. Tie-Jun Cui, Sha-Sha Li. Deep Learning of System Reliability under Multi-factor Influence Based on Space Fault Tree. Neural Computing and Applications, https://doi.org/10.1007/s00521-018-3416-2.
5. Sha-Sha Li, Tie-Jun Cui, Jian Liu. Study on the construction and application of Cloudization Space Fault Tree. Cluster Computing, https://doi.org/10.1007/s10586-017-1398-y.
6. CUI Tie-jun, MA Yun-dong. Discrete Space Fault Tree Construction and Failure Probability Space Distribution Determine. Systems Engineering-Theory & Practice, 2016, 36(4): 1081-1088
7. CUI Tie-jun, WANG Pei-zhuang, MA Yun-dong. Inward analysis of system factor structure in 01 space fault tree. Systems Engineering-Theory & Practice, 2016, 36(8): 2152-2160.
8. CUI Tiejun, MA Yundong. Inaccurate reason analysis of the factors projectionfitting method in DSFT. Systems Engineering - Theory & Practice, 2016, 36(5): 1340-1345.
9. Shihong Du, Luo Guo. Similarity measurements on multi-scale qualitative locations, Transactions in GIS, 20, 824-847.

10. Andre Salvaro Furtado, Despina Kopanaki, Luis Otavio Alvares et al., Multidimensional similarity measuring for semantic trajectories. Transactions in GIS, 2016, 20, 280-298.
11. Faezeh S. Gohari, Mohammad Jafar Tarokh, New recommender framework: combining semantic similarity fusion and bicluster collaborative filtering. Computational Intelligence, 2016, 32, 561-586.
12. Joseph Harms, Tonghe Wang, Michael Petrongolo, et al., Noise suppression for dual-energy CT via penalized weighted least-square optimization with similarity-based regularization. Medical Physics, 2016, 43, 2676-2686.
13. Rodolfo Alvarado Cervantes, Edgardo M. Felipe Riverón, Vladislav Khartchenko, et al. An adaptive color similarity function suitable for image segmentation and its numerical evaluation. Color Research & Application, DOI: 10.1002/col.22059.
14. Pham Hong Phong, Le Hoang Son, Linguistic vector similarity measures and applications to linguistic information classification. International Journal of Intelligent Systems, 2017, 32, 67-81.
15. Li Shasha, Cui Tiejun, Ma Yundong, System reliability assessment method based on space fault tree. Journal of Safety Science and Technology, 2015, 11(6): 86-92.
16. Cui Tiejun, Ma Yundong, Research for connectivity reliability of directed acyclic network considering nodes and lines. Application Research of Computers, 2015(11). http://www.cnki.net/kcms/detail/51.1196.TP.20150507.1053.065.html.
17. Cui Tiejun, Ma Yundong, The definition and cognition of the factor importance distribution in Continuous Space Fault Tree. China Safety Science Journal, 2015, 25(3): 24-28.
18. Cui Tiejun, Ma Yundong, System safety classification decision rules considering the scope attribute. Journal of Safety Science and Technology, 2014, 10(11):6-9.
19. Cui Tiejun, Ma Yundong, Research on accident chain model based on CBM and human error. China Safety Science Journal, 2014, 24(8):37-42.
20. WANG Peizhuang. Factor spaces and Factor Data-bases, Journal of Liaoning Technical University(Natural Science), 2013, 32(10): 1-8.
21. Wang P Z, Liu Z L, Shi Y, et al. Factor space, the theoretical base of data science. Ann. Data Science, 2014, 1(2): 233-251.
22. Wang Peizhuang, Guo Sicong, Bao Yanke, et al. Factorial analysis in factor space. Journal of Liaoning Technical University: Natural Science, 2015, 34(2): 273-280.
23. WANG Peizhuang. Factor spaces and Factor Science, Journal of Liaoning Technical University(Natural Science), 2015, 34(2): 273-280.
24. WANG Peizhuang. Introduction to factor space and factor Library (special report). Intelligent science and mathematics forum, Huludao, 2014.6.
25. SHI Yong. Big data and new challenges. Science and technology development, 2014, (1): 25-30.
26. LI Hongxing. The space theory of factor space theory / factors and application. Intelligent science and mathematics forum, Huludao, 2014.
27. LI Deyi. Cognitive Physics (special report). Oriental thinking and fuzzy logic-international conference to commemorate the fifty anniversary of the birth of fuzzy set, Dalian, 2015.8.
28. Zeng W Y, Feng S. Approximate reasoning algorithm of interval-valued fuzzy sets based on least square method. Information Sciences, 2014, 272: 73-83.
29. Zeng W Y, Feng S. An improved comprehensive evaluation model and its application. International Journal of Computational Intelligence Systems, 2014, 7(4): 706-714.
30. Li D Q, Zeng W Y, Li J. Note on uncertain linguistic Bonferroni mean operators and their application to multiple attribute decision making. Applied Mathematical Modelling, 2015, 39(2): 894-900.
31. Li D Q, Zeng W Y, Zhao Y. Note on distance measure of hesitant fuzzy sets. Information Sciences, 2015, 32(1): 103-115.
32. Li D Q, Zeng W Y, Li J. New distance and similarity measures on hesitant fuzzy sets and their applications in multiple criteria decision making. Engineering Applications of Artificial Intelligence, 2015, 40(2): 11-16.
33. YU Gaofeng, LIU Wenqi, LI Dengfeng. A method of intuitionistic linguistic decision making based on eclectic variable weight vector. Control and decision, 2015, 30(12): 2233-2240.
34. YU Gaofeng, LIU Wenqi, SHI Mengting. Enterprise credit evaluation based on local variable weight model. Journal of management science. 2015,17(2): 85-94.
35. He Ping. Design of interactive learning system based on intuition concept space. Journal of computer, 2010, 21(5): 478-487.
36. OU Yanghe. Unified theory of uncertainty theory: mathematical foundations of factor spaces (special report), Oriental thinking and fuzzy logic-international conference to commemorate the fifty anniversary of the birth of fuzzy set, Dalian, 2015.8.
37. CUI Tie-jun, MA Yun-dong. System security classification decision rules considering the Scope attribute. Journal of Safety Science and Technology, 2014,10(11):6-9.
38. YANG Ju-wen, HE Feng, CUI Tie-jun, et al. Safety analysis of coal mine disaster based on factor analysis. Journal of Safety Science and Technology, 2015.11(4):84-89.
39. CUI Tie-jun, MA Yun-dong. Research on the classification method about coal mine safety situation based on the factor space. Systems Engineering-Theory & Practice, 2015,35(11):2891-2897.
40. CUI Tie-jun, MA Yun-dong. Research on the number of failures of repairable systems based on imperfect repair model. Systems Engineering-Theory & Practice, 2016,36(1): 184-188.
41. NIU Jun-lei,CHENG Long-sheng. Classification using improved Mahalanobis-Taguchi system based on omni-optimizer. Systems Engineering-Theory & Practice, 2012,32(6):1324-1336.