Pattern Recognition 43 (2010) 3768–3786

Contents lists available at ScienceDirect

Pattern Recognition

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

The structural clustering and analysis of metric based on granular space

Xu-Qing Tang [a,*], Ping Zhu [a], Jia-Xing Cheng [b]

[a] School of Science, Jiangnan University, Wuxi 214122, Jiangsu Province, PR China
[b] Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei 230039, Anhui Province, PR China

ARTICLE INFO

Article history:
Received 21 March 2009
Received in revised form
4 August 2009
Accepted 14 October 2009

Keywords:
Granular computing
Granular space
Normalized metric space
Clustering structural analysis
Consistent cluster
Optimal cluster
Clustering fusion

ABSTRACT

In this paper, the research on granular space theory and structural clusters of metric based on granular space is introduced, and a comprehensive of theoretical and analyzing methodologies is developed. The granular space theory is established based on normalized equicrural metric, and the consistent cluster characteristics of ordered granular spaces are derived. Details are shared for some related subject, such as the granular representation of structural cluster from normalized metric, the optimal cluster determination, the fusion of structural cluster, and the structural clustering analysis of metric space based on granular space, etc. Theories and methodologies are established for structural clustering based on metric space, and developed a series of mathematical models and formal description tools for the research on its potential applications. Direct and geometric interpretations for structural clustering analysis are provided to assist deeper understanding the essence of clustering (or classifying) procedure.

1. Introduction

Granular computing (GrC) [1–3] is an emerging conceptual and computing paradigm of information processing. It expands to all aspects of theory, methodology, technology and tools that make use of granules, i.e. groups, classes, or clusters of a universe. It plays an important role in information processing for fuzzy, uncertainty, partial truth and huge volume of data, and has been one of the hot topics in artificial intelligence [4]. Currently, there are the following three main GrC theories:

(1) The theory of fuzzy granulation [1–3]. Zadeh proposed three basic concepts of human perceptions, i.e., *granulation* groups of which the total is divided into, *organization* the total integrated from the groups, and *causation* the correlation between cause and result, and later proposed the terminology for granular computing. Yao et al. conducted a series of research works [5–10], and applied it into data mining and other fields. The key of Yao's works is to establish IF–THEN relationship between concepts and inclusion based on granular sets. Yao solved the consistent classification problems by using the lattice consisted of all classifications. Those works provided new methods and solutions for knowledge mining.

(2) Rough sets theory. It is introduced by Pawlak [11–13], based on a hypothesis that human intelligence (or knowledge) is essentially the ability of classification. Given an equivalence relation R on X, he presented a knowledge base (X, R) and discussed how to represent any concept x by using knowledge base, and established the rough sets theory. So far, this theory has been broadly applied to many fields, especially in the field of data mining [14,15].

(3) The theory of quotient space. It was developed by Zhang and Zhang [16,17], when researching problem solving. They pointed out "one of the basic characteristics of human problem solving is the ability to conceptualize the world at different granularities and translate from one abstraction level to the others quite easily, i.e., deal with them hierarchically. It is the hierarchy that the human power in problem solving lies in". On the basis of that observation, they described the problems by using the three-index symbols (X, f, T), where X denotes a conceptual set, $f(\cdot)$ denotes the attribute on X, X denotes the structure of X and states the relationship of any two elements on X. By introducing the equivalent relation R on X, they introduced the classifications of X about equivalent classes, i.e. the quotient space (granulation or knowledge base) $[X]$, where quotient spaces are differential with the different of its equivalent classes. They also studied the relationship among various quotient spaces (or granulation) when T is a topological structure or a half-order structure. Therefore, the granular computing is transformed into studying the relationship among different granulations (i.e. $([X],[f],[T])$) and transformation between them. For the same problem, we may obtain the solution based on analyzing and reasoning using its different granulations. The granular world model (GRWM) of quotient space theory includes a series of

* Corresponding author. Tel.: +86 510 85910595; fax: +86 510 85910227.
E-mail addresses: txq5139@yahoo.com.cn, txq5139@jiangnan.edu.cn (X.-Q. Tang), zhuping@jiangnan.edu.cn (P. Zhu), cjx@ahu.edu.cn (J.-X. Cheng).

0031-3203/$ - see front matter
doi:10.1016/j.patcog.2009.10.009

theories and algorithms, and has been successfully applied to heuristic search (SA) and path programming [16,17]. They later introduced quotient space theory into fuzzy system and presented the theory of fuzzy quotient space [18,19], which provided mathematical models and tools for fuzzy granular computing.

At present, the theory of fuzzy granulation is mainly used in granular computing with words, with further applications in research fields of fuzzy logic and reasoning, and fuzzy control, etc. The rough sets theory is primarily used in data classification (or cluster) and data mining. The theory of quotient space is a perfect theoretical system, and compatible with the two earlier theories-supporting functions of both. More importantly, it introduced structural analysis into system research with potential of broad applications and deep connotation. As an analogy, both graphite and diamond consist of carbon, but they have drastically different in physical and chemical characteristics due to their different constructive structures. Another analogy, genetic codes of all living species in the world consist of A, C, G and T (note: T is replaced by U in RNA), but they result in different colors and styles with different genetic sequences. Essentially, those analogies explain why strict mathematical model description of system structure analysis is always one of the hot topics in complex system analysis.

The data clustering (or classifying) analysis is both ancient and current research field [20–22]. It resulted in many important methods for studying discrete data structures. In fact, these clustering or classifying methods have been used extensively in solving actual problems, such as in random or fuzzy type discrete data analysis. The wide variety of methods being used today can be divided into two main categories—theoretical analysis and machine learning. The theoretical analysis category can be further divided into the following two types.

The way is the traditional clustering or classifying, which is based on an equivalence relation or a fuzzy equivalence relation [23–27]. Its basic idea is to cluster or classify by using the equivalent classes. For example, fuzzy clustering based on a equivalence relation gets its clustering classes by using its cut-relation (note: a crisp equivalence relation) [25,28,29]. A fuzzy equivalence relation is difficult to verify in actual problems, while similarity relation is easy to derive (e.g., from normalized metric space [29]). Therefore, the traditional clustering or classifying analysis is turned into analyzing the transitive closure of similarity relation [26,29,30]. But, there are two issues needing to be addressed for this analysis method: (1) given a transitive closure of fuzzy similarity relation, how is to select its suitable cluster or classification? or how to determine the threshold value λ_0? (2) The transitive closure of fuzzy similarity relation does not guarantee that the correlative degree between any two elements in a category is large. For example, let us study a equivalent class that the distance between any two elements is not more than the threshold d_0, when $d(x_1,x_2) \le d_0$, $d(x_2,x_3) \le d_0, \ldots, d(x_{n-1},x_n) \le d_0$, then $x_1,x_2,\ldots,x_n$ are in a equivalent class according to the transitivity. However, it does not guarantee that $d(x_1,x_n) \le d_0$.

The other way is C-means (or FCM) and their various derivative versions on large quantity data clustering or classifying problems [31–37]. The key of this method is to classify n data into k categories according to metric correlation, where $k < n$, such that ith data category is neighbor to ith centroid, and to obtain the optimal solution under the given goal function with a metric.

Recently, there have been publications of new research in clustering (or classifying) methods and analysis. He [25] reported a perturbation fuzzy clustering method based on fuzzy similarity relations. Guoyao proposed the optimization method for fuzzy clustering [37], where computational complexities are quite high. Tsekouras et al. [35] presented a hierarchical fuzzy clustering approach based on fuzzy similarity relations, Devillez et al. [32] proposed a hybrid hierarchical clustering approach. Gibson et al. [37–40] presented a clustering approach, which is to minimize the sum of radii, and Schaeffer [41] presented also the graph clustering. The basic idea of all those efforts is to analyze structural clustering based on fuzzy reasoning rules. These works provide a good foundation for future research on structural clustering (or classifying), e.g., fusion problem and optimal clustering number determination problem, etc. The advantage of these methods is that they provide direct geometric interpretation for data clustering (or classifying) analysis. However, they also have a obvious disadvantage that they are based on several (fuzzy) rules [28,30,32,37]. In fact, given an ordered equivalence relations set or fuzzy equivalence relation on the universe X, its corresponding hierarchical structure is defined (note: it is seen in Section 3.2), and the basis of clustering (or classifying) analysis on X is determined as well. As to selecting the suitable classification (or cluster) number, it essentially comes down to the understanding level (i.e., to select the threshold value), in other words, the selecting suitable clustering (or classifying) is determined solely on the basis of constructing a proper objective function, but on any additional (fuzzy) rules.

Previously, Chien et al. [42] proposed the research on fuzzy concept hierarchy and measurement, which is a kind of general form of knowledge representation based on knowledge structure. We also presented cluster analysis based on fuzzy quotient space [43], reported a direct method for solving clustering problems based on normalized metric, and gave the research on fuzzy granular space [44]. In fact, for given a metric space (X,d), then $1-\exp(-d)$ is a normalized metric on X (details can be seen from Lemma 2.2). So the problem of structural clustering based on metric can be transformed into the one based on normalized metric. On the basis of two essential characteristics of human intelligence that "human intelligence (or knowledge) is a classification ability" [12] and "one of the basic characteristics in human problem solving is the ability to conceptualize the world at different granularities and translate from one abstraction level to the others easily, i.e., deal with hierarchically" [17], we have conducted further research on structural clusters analysis of metric space based on granular spaces. The aim is to develop a comprehensive of descriptive tools for granular computing based on metric space. This paper is a summary of our recent studies, and organized as follows. In Section 2, some necessary preliminaries are given. In Section 3, the granular space theory based on normalized equicrural metric is established. In Section 4, a basic concept that consist cluster characteristic according to metric is described, and the structural clustering characteristics of granular space is discussed. In Section 5, the granular representation of structural cluster on normalized metric is presented. In Section 6, the optimal cluster determination problem based on granular space is reported. In Section 7, the fusion problem of structural clusters based on granular space is given. In Section 8, the clustering structure analysis based on metrics is discussed. In Section 9, we give the conclusions of this paper.

2. Preliminaries

We introduce some basic concepts, notions and results as follows:

Definition 2.1. (Zhang and Zhang [19]). Let (X,d) be a metric space, if d satisfies:

(1) $\forall x,\ y \in X$, $0 \le d(x,y) \le 1$;
(2) $\forall x,y,z \in X$, there is not a number within the array $\{d(x,y),d(y,z),d(z,x)\}$ such that it is greater than the maximum of other two numbers.

Then d is called a normalized and equicrural metric (short form: NEC metric) on X. The above mentioned condition (1) is also called the normalized condition, and the above mentioned condition (2) is called the equicrural condition. If the metric on X satisfies only normalized condition, then d is called normalized metric on X.

Definition 2.2. (Li and Wang [45]). Let X be a universe, $d: X \times X \to R^+$, where R^+ stand for a nonnegative real set, and satisfies:

(1) $\forall x \in X,\ d(x,x)=0$;
(2) $\forall x,y \in X,\ d(x,y)=d(y,x)$;
(3) $\forall x,y,z \in X,\ d(x,y) \le d(x,z)+d(z,y)$.

Then d is called a pseudo-metric on X. Furthermore, if d is a pseudo-metric satisfying the normalized condition on X, then d is also called normalized pseudo-metric on X; if d is a normalized pseudo-metric satisfying the equicrural condition on X, then d is also called normalized and equicrural pseudo-metric on X.

Definition 2.3. Assume R is a relation on X, and satisfies:

(1) $\forall x \in X,\ (x,y) \in R$;
(2) $\forall x,y \in X,\ (x,y) \in R \to (y,x) \in R$;
(3) $\forall x,y,z \in X,\ (x,y) \in R, (y,z) \in R \to (x,z) \in R$.

Then, R is called a crisp equivalence relation on X. If R only satisfies conditions (1) and (2), then R is called a similarity relation on X.

Definition 2.4. (Wu [23] and Tang et al. [44]). Assume $R \in F[X \times X]$, where $F[X \times X]$ stands for all fuzzy sets on X. If R satisfies:

(1) $\forall x \in X,\ R(x,x)=1$;
(2) $\forall x,y \in X,\ R(x,y)=R(y,x)$;
(3) $\forall x,y,z \in X,\ R(x,y) \ge \sup_{z \in X}\{\min\{R(x,z),R(z,y)\}\}$.

Then, R is called a fuzzy equivalence relation on X. If R satisfies only conditions (1) and (2), then R is called fuzzy similarity relation on X.

If R is a fuzzy equivalence relation on X and satisfies

$$\forall x,y \in X, R(x,y)=1 \leftrightarrow x=y \tag{2.1}$$

then R is called a fuzzy equivalence relation satisfying the finest condition on X, the condition (2.1) is called the finest condition. If R is a fuzzy similarity relation and satisfies the finest condition on X, then R is a fuzzy similarity relation satisfying the finest condition on X.

Let $D(X)$, $WD(X)$, $ND(X)$ and $WND(X)$ stands for the set of NEC metrics, the one of NEC pseudo-metrics, the one of normalized metrics, and the one of normalized pseudo-metrics on X, respectively. $GEF(X)$ and $EF(X)$ stands for the set of fuzzy equivalence relations satisfying the finest condition and fuzzy equivalence relations on X, respectively.

Lemma 2.1. (*Tang et al.* [44]). *The following three statements are equivalent.*

(1) *A NEC metric on X is given*;
(2) *A fuzzy equivalence relation satisfying the finest condition on X is given*;
(3) *An ordered granular space including the finest granulation on X is given.*

Lemma 2.2. *Let X be an universe, $D=1-\exp(-d)$. If d is a metric on X, then D* is a normalized metric on X.

Proof. Because d is a metric on X, it is obvious that D satisfies the symmetry. $\forall x,y \in X,\ 0 \le d(x,y) \le 1,\ D(x,y)=0 \leftrightarrow d(x,y)=0 \leftrightarrow x=y$, and $\forall x,y,z \in X,\ d(x,y) \le d(x,z) + d(z,y) \to 1-\exp(-d(x,y)) \le 1-\exp(-d(x,z)-d(z,y)) \le [1-\exp(-d(x,y))]+[1-\exp(-d(z,y))]$, then $D(x,y) \le D(x,z)+D(z,y)$, i.e., D satisfies the triangle inequality on X. Therefore, D is a normalized metric on X by Definition 2.1.

3. The granular space theory based on NEC metric

In this section, we establish the granular space theory based on NEC (pseudo-)metric, and organize it as follows. In Section 3.1, the granular space based on NEC (pseudo-)metric is introduced. In Section 3.2, the relationship between ordered equivalence relations set and fuzzy equivalence relations is given. In Section 3.3, the metric on ordered granular space is studied. In Section 3.4, the relationship between NEC (pseudo-)metric and its granular space on order is researched.

3.1. The granular space on NEC metric

Theorem 3.1. *Let $d \in D(X)$ (or $d \in WD(X)$) and $\forall \lambda \in [0,1]$, define R_λ: $\forall x,y \in X, (x,y) \in R_\lambda \leftrightarrow d(x,y) \le \lambda$, then R_λ is an equivalence relation on X corresponding to λ.*

Proof. The proof is directly obtained from Definition 2.2, details are omitted here. □

Theorem 3.1 states clearly that NEC metric (or NEC pseudo-metric) may derive equivalence relations, and the following corollary should hold.

Corollary 3.1. *Let $d \in ND(X)$ (or $d \in WND(X)$) and $\forall \lambda \in [0,1]$, define $R_\lambda: \forall x,y \in X,\ (x,y) \in R_\lambda \leftrightarrow d(x,y) \le \lambda$, then R_λ is a similarity relation on X corresponding to λ.*

Definition 3.1. In Theorem 3.1, the equivalent classes of R_λ is denoted by $[x]_\lambda = \{y | d(x,y) \le \lambda\}$, marked $X(\lambda)=\{[x]_\lambda | x \in X\}$, then $X(\lambda)$ is called the granulation (or quotient space) corresponding to λ deriving by d. The set $\{X(\lambda)|\lambda \in [0,1]\}$ is called a granular space on X deriving by d, marked $\aleph_d(X)$; and R_λ is also called an equivalent relation corresponding to λ on X deriving by d. The set of all deriving equivalence relations is denoted by $\Re_d(X)$, i.e. $\Re_d(X)=\{R_\lambda | \lambda \in [0,1]\}$.

In fact, a granulation (or quotient space) just stands for a partition of X here. In Definition 3.1, the equivalent class $[x]_\lambda$ stands for a set consisting of elements whose distances to x be no greater than λ, i.e., it is a closed neighborhood in which x is center and λ is radius. With different $\lambda \in [0,1]$, the granulation (or quotient space) $X(\lambda)$ are not perfectly identical, nor is R_λ. Then, how is the structure of granular space $\aleph_d(X)$? How is the structure of $\Re_d(X)$? In this section, we answer these questions.

Definition 3.2. Suppose $X(\lambda_1)$ and $X(\lambda_2)$ are two granulations on X.

(1) If $\forall x \in X,\ [x]_{\lambda_1} \subseteq [x]_{\lambda_2}$, then it is called that granulation $X(\lambda_2)$ is not finer than $X(\lambda_1)$, and noted as $X(\lambda_2) \le X(\lambda_1)$.

(2) If $X(\lambda_2) \le X(\lambda_1)$ and there exists $x_0 \in X$ such that $[x_0]_{\lambda_1} \subset [x_0]_{\lambda_2}$, then it is called that $X(\lambda_1)$ is finer than $X(\lambda_2)$, and marked as $X(\lambda_2) < X(\lambda_1)$.

Remark 1. If $d \in D(X)$, then $\forall x,y \in X,\ d(x,y)=0 \leftrightarrow x=y$, i.e., $\forall x \in X$, $[x]_0=\{x\}$. Therefore, we have $X(0)=\{\{x\}|x \in X\} \triangleq X$, satisfying $\forall \lambda \in [0,1],\ X(\lambda) \le X(0)$, that is $X(0)$ (or X) is the finest granulation of X.

Definition 3.3. If $\aleph(X)$ is a granular space of X and satisfies $\{\{x\}|x \in X\} \in \aleph(X)$, then $\aleph(X)$ is a granular space including the finest granulation on X.

The granular space including the finest granulation is easy to obtain by pre-processing (i.e., the universe X is just the finest granulation) for actual complex systems. In general, granular spaces including the finest granulation on X are not unique.

Proposition 3.1. *Give $d \in D(X)$ (or $d \in WD(X)$), then the deriving granular space $\aleph_d(X)$ is an ordered set, and satisfies $\forall \lambda_1, \lambda_2 \in [0,1]$, $\lambda_1 \le \lambda_2 \to X(\lambda_2) \le X(\lambda_1)$. Particularly, $\forall \lambda_1, \lambda_2 \in D$, $\lambda_1 < \lambda_2 \to X(\lambda_2) < X(\lambda_1)$, where $D = \{d(x,y)|x,y \in X\}$. Futhermore, $\aleph_d(X)$ includes the finest granulation of X when $d \in D(X)$.*

Proof. By Definitions 3.2 and 3.3 and Remark 1, the proof is easy to be obtained, hence omitted here. □

Proposition 3.2. *Give $d \in D(X)$ (or $d \in WD(X)$), $\aleph_d(X) = \{X(\lambda)|\lambda \in [0,1]\}$ stands for the granular space deriving from d, and $X(\lambda) = \{[x]_\lambda | x \in X\}$. Then, $\forall x,y \in X$, $d(x,y) = \inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_\lambda\}$.*

Proof. $\forall x,y \in X$, marked $d(x,y) = \lambda_0$. On the one hand, if $\lambda \in [0,1]$ and $y \in [x]_\lambda$, $d(x,y) \le \lambda$. Then $d(x,y) \le \inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_\lambda\}$; On the other hand, if $\lambda > \lambda_0$, we have $d(x,y) < \lambda$ from $d(x,y) = \lambda_0$, i.e., $y \notin [x]_\lambda$, so $\inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_\lambda\} \le \lambda_0 = d(x,y)$. Therefore, $d(x,y) = \inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_\lambda\}$. □

Theorem 3.2. *A NEC pseudo-metric on X is given $\Leftrightarrow$ a fuzzy equivalence relation on X is given.*

Proof. "$\Rightarrow$". Let $d \in WD(X)$, define R: $\forall x,y \in X, R(x,y) = 1 - d(x,y)$. It is obvious that R satisfy the reflexive and symmetry property, $0 \le R(x,y) \le 1$. And $\forall x,y,z \in X, d(x,y) \le \max\{d(x,z), d(z,y)\} \to 1 - d(x,y) \ge \min\{1 - d(x,z), 1 - d(z,y)\}$, i.e., $R(x,y) \ge \min\{R(x,z), R(z,y)\}$, therefore, $R(x,y) \ge \sup_{z \in X}\{\min\{1 - d(x,z), 1 - d(z,y)\}\}$, i.e., $R \in EF(X)$.

"$\Leftarrow$". Let $R \in EF(X)$, define d: $\forall x,y \in X$, $d(x,y) = 1 - R(x,y)$. It is obvious that d satisfy the symmetry property and $0 \le d(x,y) \le 1$, and

(1) $\forall x \in X, d(x,x) = 1 - R(x,x) = 0$;
(2) $\forall x,y,z \in X, R(x,y) \ge \sup_{z \in X}\{\min\{R(x,y), R(x,y)\}\} \ge \min\{R(x,y), R(x,y)\} \to 1 - R(x,y) \le \max\{1 - R(x,z), 1 - R(z,y)\}$, i.e., $d(x,y) \le \max\{d(x,z), d(z,y)\}$.

Similar to the above proof, we have

$$d(x,z) \le \max\{d(x,y), d(z,y)\}, d(z,y) \le \max\{d(x,y), d(x,z)\}$$

Then, $d \in WD(X)$. □

By Proposition 3.1, Lemma 2.1 and Theorem 3.2, the following theorem can be derived.

Theorem 3.3. *The following three statements are equivalent.*

(1) *A NEC pseudo-metric on X is given*;
(2) *A fuzzy equivalence relation on X is given*;
(3) *An ordered granular space on X is given.*

Obviously, Theorem 3.3 has more common meanings than Lemma 2.1.

3.2. The ordered equivalence relations set and fuzzy equivalence relations

In Definition 3.1, we introduced the equivalence relations set $\Re_d(X)$ deriving from the NEC (pseudo-)metric d on X. We study the relationship between equivalence relations set and fuzzy equivalence relations as follows.

Definition 3.4. Let R_1 and R_2 be two equivalence relations on X.

(1) If $\forall x,y \in X, (x,y) \in R_1 \to (x,y) \in R_2$, then it is called that the relation R_2 is not finer than R_1, and marked as $R_2 \le R_1$.
(2) If $R_2 \le R_1$ and there exists $(x_0,y_0) \in R_2$ such that but $(x_0,y_0) \notin R_1$, then it is called that R_1 is finer than R_2, and marked as $R_2 < R_1$.

Remark 2. When $d \in D(X)$, $R_0 = \{(x,x)|x \in X\} \in \Re_d(X)$ and $\forall \lambda \in [0,1]$, $R_\lambda \le R_0$. Therefore, R_0 is the finest equivalence relation in $\Re_d(X)$.

Definition 3.5. If $\Re(X)$ is an equivalence relations set of X and satisfies $\{(x,x)|x \in X\} \in \Re(X)$, then $\Re(X)$ is called an equivalence relations set including the finest equivalence relation on X.

Proposition 3.3. *If $d \in D(X)$ (or $d \in WD(X)$), then the deriving equivalence relations set $\Re_d(X)$ is an ordered set, and satisfies $\forall \lambda_1, \lambda_2 \in [0,1]$, $\lambda_1 \le \lambda_2 \to R_{\lambda_2} \le R_{\lambda_1}$. Particularly, $\forall \lambda_1, \lambda_2 \in D$, $\lambda_1 < \lambda_2 \to R(\lambda_2) < R(\lambda_1)$, where $D = \{d(x,y)|x,y \in X\}$. Furthermore, $\Re_d(X)$ includes the finest equivalence relation of X when $d \in D(X)$.*

Proof. By Definitions 3.4 and 3.5, and Remark 2, the proof is easy to be obtained, hence omitted here. □

The following corollary is directly obtained from Propositions 3.1 and 3.3.

Corollary 3.2. $\forall \lambda_1, \lambda_2 \in [0,1]$, $R_{\lambda_1} \le R_{\lambda_2} \Leftrightarrow X(\lambda_1) \le X(\lambda_2)$.

Theorem 3.4. *Assume $\{R_\lambda | \lambda \in [0,1]\}$ is an ordered equivalent relations set including the finest equivalent relation on X. Define R^+: $\forall x,y \in X$, $R^+(x,y) = 1 - \inf_{\lambda \in [0,1]}\{\lambda | (x,y) \in R_\lambda\}$. Then R^+ is a fuzzy equivalence relation satisfying the finest condition on X, and $R^+_{1-\lambda} = R_\lambda$, where R^+_λ stands for the cut relation of R^+ corresponding to λ.*

Proof. $\forall \lambda \in [0,1]$, the mark $[x]_\lambda$ and $X(\lambda)$ denotes the equivalent class of x and granulation corresponding to R_λ on X respectively, i.e., $[x]_\lambda = \{y | (x,y) \in R_\lambda, y \in X\}$, $X(\lambda) = \{[x]_\lambda | x \in X\}$. Because $(x,y) \in R_\lambda \leftrightarrow [x]_\lambda = [y]_\lambda$, therefore $R^+(x,y)$ is transformed into the equivalent form of $R^+(x,y) = 1 - \inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_\lambda\}$.

(1) $\forall x \in X$, $R^+(x,x) = 1 - \inf_{\lambda \in [0,1]}\{\lambda | x \in [x]_\lambda\} = 1$. Because $\{R_\lambda | \lambda \in [0,1]\}$ includes the finest equivalent relation on X, we have $\forall x,y \in X$, $R^+(x,y) = 1 \leftrightarrow \forall \lambda \in [0,1]$, $(x,y) \in R_\lambda \leftrightarrow x = y$;

(2) $\forall x,y \in X, R^+(x,y) = 1 - \inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_\lambda\} = 1 - \inf_{\lambda \in [0,1]}\{\lambda | x \in [y]_\lambda\} = R^+(y,x)$;

(3) $\forall x,y,z \in X, \min\{R^+(x,z), R^+(z,y)\} = \min\{1 - \inf_{\lambda \in [0,1]}\{\lambda | x \in [z]_\lambda\}, 1 - \inf_{\lambda \in [0,1]}\{\lambda | y \in [z]_\lambda\}\} = 1 - \max\{\inf_{\lambda \in [0,1]}\{\lambda | x \in [z]_\lambda\}, \inf_{\lambda \in [0,1]}\{\lambda | y \in [z]_\lambda\}\}$. Let $\lambda_0 = \inf_{\lambda \in [0,1]}\{\lambda | x \in [z]_\lambda\} \ge \inf_{\lambda \in [0,1]}\{\lambda | y \in [z]_\lambda\}$.

On the one hand, from $\lambda_0 = \inf_{\lambda \in [0,1]}\{\lambda | x \in [z]_\lambda\}$ and the definition of inferior, $\forall \varepsilon > 0$, we know there exists $\forall \lambda_1 \in [0,1]$ such that

$$\lambda_0 + \varepsilon \ge \lambda_1 \ge \lambda_0, \quad x \in [z]_{\lambda_1} \tag{3.1}$$

On the other hand, by $\inf_{\lambda \in [0,1]}\{\lambda | y \in [z]_\lambda\} \le \lambda_0$, we know there exists $\lambda_2 \in [0,1]$ such that

$$\lambda_2 \le \lambda_0, \quad y \in [z]_{\lambda_2} \tag{3.2}$$

Based on (3.1), (3.2), Propositions 3.1 and 3.3, we get $y \in [z]_{\lambda_2} \subseteq [z]_{\lambda_1} \subseteq [z]_{\lambda_0+\varepsilon}$, that is, $\inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_\lambda\} \le \lambda_0 + \varepsilon$.

Therefore, $\min\{R^+(x,z), R^+(z,y)\} = 1 - \lambda_0 \le 1 - \inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_\lambda\} + \varepsilon = R^+(x,y) + \varepsilon$. Furthermore, letting $\varepsilon \to 0^+$, we get $\min\{R^+(x,z), R^+(z,y)\} \le R^+(x,y)$. Because z is arbitrary in X, then $R^+(x,y) \ge \sup_{z \in X}\{\min\{R^+(x,z), R^+(z,y)\}\}$.

Synthesizing (1), (2) and (3), R^+ is a fuzzy equivalence relation satisfying the finest condition on X.

Because $(x,y) \in R^+_{1-\lambda} \leftrightarrow R^+(x,y) \geq 1-\lambda \leftrightarrow \inf_{\lambda_1 \in [0,1]}\{\lambda_1 | (x,y) \in R_{\lambda_1}\} \leq \lambda \leftrightarrow (x,y) \in R_\lambda$ (from Proposition 3.3), therefore, $R^+_{1-\lambda} = R_\lambda$. □

In fact, the inverse of Theorem 3.4 also holds, i.e.,

Theorem 3.5. *If R^+ is a fuzzy equivalence relation satisfying the finest condition on X, then there exists an ordered equivalent relations set $\{R_\lambda | \lambda \in [0,1]\}$ including the finest condition on X such that $R^+_{1-\lambda} = R_\lambda$, where R^+_λ stands for the cut relation of R^+ corresponding to λ.*

Proof. The proof is similar to that for Theorem 3.4, and omitted here. □

The following corollary is directly obtained from Theorems 3.4 and 3.5.

Corollary 3.3. *An ordered equivalent relation set $\{R_\lambda | \lambda \in [0,1]\}$ on X is given $\Leftrightarrow$ A fuzzy equivalence relation R^+ on X is given, satisfying $R^+_{1-\lambda} = R_\lambda$, where R^+_λ stands for the cut relation of R^+ corresponding to λ.*

Theorems 3.4 and 3.5 show that an ordered equivalent relations set $\{R_\lambda | \lambda \in [0,1]\}$ including the finest equivalent relation is equivalent to a fuzzy equivalence relation R^+ satisfying the finest condition on X, and the $\{R_\lambda | \lambda \in [0,1]\}$ is the same as the cut relation set of R^+ but their orders are in opposite direction. By Lemma 2.1, Theorems 3.4 and 3.5, we directly obtain the theorem as follows.

Theorem 3.6. (Basic Theorem 1). *The following four statements are equivalent.*

(1) *A fuzzy equivalence relation satisfying the finest condition on X is given*;
(2) *A NEC metric on X is given*;
(3) *An ordered granular space including the finest granulation on X is given*;
(4) *An ordered equivalence relations set including the finest equivalence relation on X is given.*

Remark 3. Here, the ordered granular space is equivalent to the hierarchical structure in Ref. [19]. Compared with Basic Theorem in [19], Theorem 3.6 is extended into the simplifying complex system (i.e., the Proposition (2)), and added the equivalent Proposition (4). In fact, "a quotient space" of the basic theorem in Ref. [19] is a special meaning of the fuzzy quotient (or granulation) $X(1)$ which is equal to the granulation $X(0)$ deriving by a NEC metric. In Theorem 3.6, by adding the intermediate state that ordered equivalence relations set between NEC metrics and fuzzy equivalent relations, we can see that fuzzy equivalent relation have thinner than crisp equivalent relation on general conceptual description. Meanwhile, from the equivalence between Proposition (2) and Proposition (4) in Theorem 3.6, it is more important interpretation for the conversion relation between "ordered structure on X" and "metric on X", that is the ordered structure on X can produce the metric on X. This states that the ordered granular space not only has ordered structure, but also has the topological structure which it is derived by metric [46]. Therefore, ordered granular space has superior properties to the others which we can see in Section 4.

From Theorem 3.3 and Corollary 3.3, the following theorem holds.

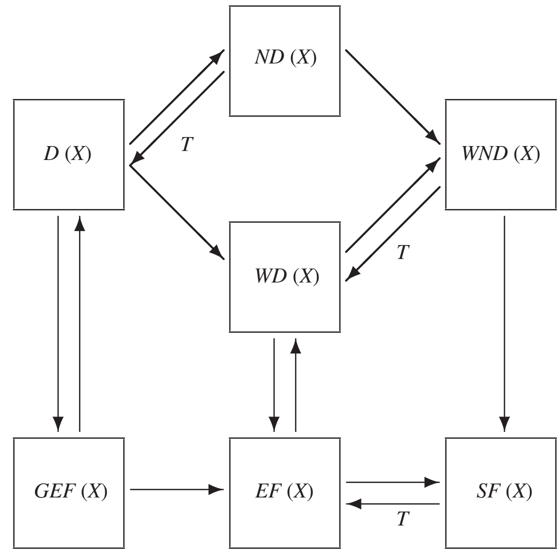

Fig. 1. The relationship graph showing $D(X)$, $WD(X)$, $GEF(X)$ and $EF(X)$.

Theorem 3.7. (Basic Theorem 2). *The following four statements are equivalent.*

(1) *A fuzzy equivalence relation on X is given*;
(2) *A NEC pseudo-metric on X is given*;
(3) *An ordered granular space on X is given*;
(4) *An ordered equivalence relations set on X is given.*

Remark 4. Basic Theorem 2 has more common meanings than Basic Theorem1 in granular space.

If "$A \rightarrow B$" denotes the satisfying condition of A is stronger than that of B, "$A \xrightarrow{T} B$" stands that the B can be get from the A under the transitive closure operation, and "$A \leftrightarrow B$" notes the satisfying condition of A is equivalent to that of B, we get Fig. 1 by Theorems 3.4, 3.5 and 3.2.

3.3. The metric on granular space deriving from NEC metric

By Theorems 3.6 and 3.7, given a NEC (pseudo-)metric d on X, then an ordered granular space $\aleph_d(X)$ is given. However, how to define the metric on granular space $\aleph_d(X)$? If there exists the metric on $\aleph_d(X)$, what is the relationship between it and d? Or, how to determine the metric on granular space $\aleph_d(X)$ by d? To answer these questions, we perform the research as follows.

Theorem 3.8. *Assume $d \in D(X)$ (or $d \in WD(X)$), $\aleph_d(X)$ denotes the deriving granular space by d. $\forall \lambda \in [0,1], a, b \in X(\lambda) \in \aleph_d(X)$, define*

$$d_\lambda: \quad d_\lambda(a,b) = \inf\{d(x,y) | x \in a, y \in b\} \tag{3.3}$$

then, d_λ is a NEC metric on $X(\lambda)$.

Proof. If $d \in D(X)$, then d_λ obviously satisfies the normalized condition and symmetry property from the definition of d_λ, and $\forall a \in X(\lambda), d_\lambda(a,a) = \inf\{d(x,y) | x,y \in a\} = 0$, and $\forall a,b \in X(\lambda)$, $0 = d_\lambda(a,b) = \inf\{d(x,y) | x \in a, y \in b\} \leftrightarrow a = b$ (Because a and b are closed sets).

We prove that d_λ satisfies the equicrural condition and triangle inequality on $X(\lambda)$ as follows.

First, we prove $\forall x_1, x_2 \in a \in X(\lambda), y \notin a$, $d(x_1,y) = d(x_2,y)$. From $x_1, x_2 \in a, y \notin a$, we have $d(x_1,x_2) \leq \lambda, d(x_1,y) > \lambda, d(x_2,y) > \lambda$. Because of $d \in D(X)$, we obtain

$$d(x_1,y) \leq \max\{d(x_1,x_2), d(x_2,y)\} = d(x_2,y),$$
$$d(x_2,y) \leq \max\{d(x_1,x_2), d(x_1,y)\} = d(x_1,y)$$

Therefore, $\forall x_1, x_2 \in a \in X(\lambda)$, $y \notin a, d(x_1, y) = d(x_2, y)$. Furthermore, we have $\forall a, b \in X(\lambda)$, $a \neq b, x_1, x_2 \in a, y_1, y_2 \in b, d(x_1, y_1) = d(x_2, y_2)$.

Next, $\forall a, b, c \in X(\lambda), x_1 \in a, y_1 \in b, z_1 \in c$,

$$d_\lambda(a,b) = \inf\{d(x,y)|x \in a, y \in b\} = d(x_1,y_1) \leq \max\{d(x_1,z_1), d(z_1,y_1)\} = \max\{d_\lambda(a,c), d_\lambda(c,b)\} \tag{3.4}$$

Similarly, we have $d_\lambda(a,c) \leq \max\{d_\lambda(a,b), d_\lambda(c,b)\}$, $d_\lambda(c,b) \leq \max\{d_\lambda(a,b), d_\lambda(a,c)\}$, i.e., d_λ satisfies the equicrural condition on $X(\lambda)$. And d_λ obviously satisfies the triangle inequality on $X(\lambda)$ from (3.4).

Therefore, d_λ is a NEC metric on $X(\lambda)$.

Similar to the above proof, we can prove that it also holds when $d \in WD(X)$). □

Definition 3.6. In Theorem 3.8, d_λ is called the projective metric by d on $X(\lambda)$.

Theorem 3.8 states the relationship between the metric d_λ on granulation $X(\lambda)$ and the NEC metric d on X, and gives the method to define a metric on every granulation of granular space $\aleph_d(X)$, i.e., the projective metric of d on granulation $X(\lambda)$ preserves the normalized equicrural characteristics. This definition of d_λ in Theorem 3.8 also follows the convention in mathematics, i.e. the distance between two sets is equal to the inferior of distances between any two elements taking from the two sets respectively. It is easy to get the following corollary in the proof of Theorem 3.8.

Corollary 3.4. *In Theorem* 3.8, $\forall a, b \in X(\lambda) \in \aleph_d(X)$, *then*

$$d(a,b) = \begin{cases} d(x,y), & \text{when } a \neq b,\ x \in a,\ y \in b \\ 0, & a = b \end{cases}$$

Definition 3.7. Assume d_λ is a NEC (pseudo-)metric on the granulation $X(\lambda)$ on the universe X. Define $d^*_\lambda : \forall x, y \in X, d^*_\lambda(x,y) = d_\lambda([x]_\lambda, [y]_\lambda)$, where $[x]_\lambda, [y]_\lambda \in X(\lambda)$. Then d^*_λ is called a extending metric on X by d_λ.

Proposition 3.4. *Assume $d \in D(X)$ (or $d \in WD(X)$), d_λ is the projective metric on granulation $X(\lambda)$ deriving by d, and d^*_λ is the extending metric on X by d_λ. Then, d^*_λ is a NEC pseudo- metric on X.*

Proof. When $d \in D(X)$, $\forall \lambda \in [0,1]$, d^*_λ obviously satisfies the normalized condition and the symmetry property from the definition of d_λ, and

(1) $\forall x \in X, d^*_\lambda(x,x) = d_\lambda([x]_\lambda, [x]_\lambda) = 0$;
(2) $\forall x, y, z \in X$

$$\begin{aligned} d^*_\lambda &= d_\lambda([x]_\lambda, [y]_\lambda) \leq \max\{d_\lambda([x]_\lambda, [z]_\lambda), d_\lambda([z]_\lambda, [y]_\lambda)\} \\ &= \max\{d^*_\lambda(x,z), d^*_\lambda(z,y)\} \end{aligned}$$

Similar to above proof, we have

$$d^*_\lambda(x,z) \leq \max\{d^*_\lambda(x,y), d^*_\lambda(z,y)\},$$
$$d^*_\lambda(z,y) \leq \max\{d^*_\lambda(x,y), d^*_\lambda(x,z)\}.$$

Therefore, d^*_λ is a NEC pseudo-metric on X according to Definition 2.2.

Similarly, we can prove it when $d \in WD(X)$, here omitted. □

Remark 5. From above results, derived from either NEC metrics or NEC pseudo-metric d on X, the order of their granular spaces $\aleph_d(X)$ are not effected. The metric on granulation $X(\lambda)$ is the projective metric on $X(\lambda)$ by d, it preserves the normalized and equicrural characteristics, and its extending metric is the NEC (pseudo-)metric. Simultaneously, from Theorem 3.8 and Proposition 3.4, the dynamic property of metric on ordered granular space is given, i.e., given a NEC (pseudo-)metric on a granulation $X(\lambda)$, then we can define the metric on any granulation that is more coarse than $X(\lambda)$ by its projective metric, and it is a NEC (pseudo-)metric; we can define the metric on any granulation which is finer than $X(\lambda)$ by its extending metric, and it is a NEC pseudo-metric.

All extending metrics set of projecting metrics derived from NEC (pseudo-)metric d on X is marked as $D_d(X)$, i.e., $D_d(X) = \{d^*_\lambda | 0 \leq \lambda \leq 1\}$.

Definition 3.8. Assume $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$, or $d_1, d_2 \in ND(X)$, or $d_1, d_2 \in WND(X)$).

(1) If $\forall x, y \in X$, $d_1(x,y) \leq d_2(x,y)$, then it is called that d_1 is not finer than d_2, and marked as $d_1 \leq d_2$;

(2) If $d_1 \leq d_2$, and there exists $x_0, y_0 \in X$ such that $d_1(x_0, y_0) < d_2(x_0, y_0)$, then it is called that d_2 is finer than d_1, and marked as $d_1 < d_2$.

Like an amplifying or reducing scale, the "coarse-fine" relationship among metrics shows a basic characteristic of metric, that is, for any two given metrics on a space, the bigger the distance value is, the finer the metric is.

Theorem 3.9. *Assume $d \in D(X)$ (or $d \in WD(X)$), then $D_d(X)$ is an ordered set, and satisfies $\forall \lambda_1, \lambda_2 \in [0,1]$, $\lambda_1 \leq \lambda_2 \to d^*_{\lambda_2} \leq d^*_{\lambda_1}$.*

Proof. Let $X(\lambda)$ be the granulation deriving from d on X about λ, i.e., $X(\lambda) = \{[x]_\lambda | x \in X\}$. For any $\lambda_1, \lambda_2 \in [0,1], \lambda_1 \leq \lambda_2$, we get $X(\lambda_2) \leq X(\lambda_1)$ by Proposition 3.1, then $\forall x, y \in X$, $[x]_{\lambda_1} \subseteq [x]_{\lambda_2}$, $[y]_{\lambda_1} \subseteq [y]_{\lambda_2}$. According to Corollary 3.4, $d^*_{\lambda_1}(x,y) = d_{\lambda_1}([x]_{\lambda_1}, [y]_{\lambda_1}) = \inf\{d(x_1,y_1) | x_1 \in [x]_{\lambda_1}, y_1 \in [y]_{\lambda_1}\} \geq \inf\{d(x_1,y_1) | x_1 \in [x]_{\lambda_2}, y_1 \in [y]_{\lambda_2}\} = d^*_{\lambda_2}(x,y)$.

Therefore, $d^*_{\lambda_2} \leq d^*_{\lambda_1}$ by Definition 3.8. □

Theorem 3.10. *Assume $d \in D(X)$ (or $d \in WD(X)$), $\forall \lambda \in [0,1]$, d_λ stands for the projective metric deriving from d on granulation $X(\lambda)$, and d^*_λ is the extending metric of d_λ on X. $\aleph_{d_\lambda}(X)$ and $\aleph_{d^*_\lambda}(X)$ is the granular space deriving by d_λ and d^*_λ respectively, their granulations is marked as $X_\lambda(\mu)$ and $X^*_\lambda(\mu)$ respectively ($\mu \in [0,1]$). Then, $\forall \mu \in [0,1]$, $X_\lambda(\mu) = X^*_\lambda(\mu)$, i.e., $\aleph_{d_\lambda}(X) = \aleph_{d^*_\lambda}(X)$.*

Proof. $\forall \lambda \in [0,1], \mu \in [0,1]$, $[x]_{\lambda,\mu}$ and $[x]^*_{\lambda,\mu}$ stands for the equivalent class of $X_\lambda(\mu)$ and one of $X^*_\lambda(\mu)$ respectively. By Theorem 3.8 and Definition 3.6, $\forall x \in a \in X_\lambda(\mu) \leftrightarrow a = [x]_{\lambda,\mu} = \{y | d_\lambda([x]_{\lambda,\mu}, [y]_{\lambda,\mu}) \leq \mu\} = \{y | d^*_\lambda(x,y) \leq \mu\} = [x]^*_{\lambda,\mu} \leftrightarrow a \in X^*_\lambda(\mu)$, therefore, $X_\lambda(\mu) = X^*_\lambda(\mu)$. □

Theorem 3.10 indicates that the ordered granular space deriving d_λ is the same as the one of its extending metric.

3.4. The relationship between NEC metric and its granular space on order

Based on the aforementioned conclusions, we know that a NEC (pseudo-)metric could derive an ordered granular space, and Definition 3.8 also gives the definition of order on metric spaces. Next, we establish the relationship between NEC (pseudo-)metric and its granular space on order.

Definition 3.9. Suppose $\aleph_1(X) = \{X_1(\lambda) | \lambda \in [0,1]\}$ and $\aleph_2(X) = \{X_2(\lambda) | \lambda \in [0,1]\}$ are two ordered granular space on X, their granulations is marked as $X_1(\lambda)$ and $X_2(\lambda)$ ($\lambda \in [0,1]$).

(1) If $\forall \lambda \in [0,1], X_1(\lambda) \leq X_2(\lambda)$, then it is called the granular space $\aleph_1(X)$ is not finer than $\aleph_2(X)$, and marked as $\aleph_1(X) \leq \aleph_2(X)$;

(2) If $\aleph_1(X) \leq \aleph_2(X)$ and there exists $\lambda_0 \in [0,1]$ such that $X_1(\lambda_0) < X_2(\lambda_0)$, then it is called that $\aleph_2(X)$ is finer than $\aleph_1(X)$, and marked as $\aleph_1(X) < \aleph_2(X)$.

Theorem 3.11. *Assume $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$), their deriving granular spaces is $\aleph_{d_1}(X)$ and $\aleph_{d_2}(X)$, respectively. Then, $d_1 \le d_2 \Leftrightarrow \aleph_{d_1}(X) \le \aleph_{d_2}(X)$.*

Proof. Letting $d_1, d_2 \in D(X)$, and marked $X_i(\lambda) = \{[x]_{i\lambda} | x \in X\}, i = 1, 2$.

"$\Rightarrow$". $\forall \lambda \in [0,1]$, $x, y \in X$, by the condition $d_1 \le d_2 \leftrightarrow d_1(x,y) \le d_2(x,y)$, i.e., $\forall x \in X$, $[x]_{2\lambda} = \{y | d_2(x,y) \le \lambda\} \subseteq \{y | d_1(x,y) \le \lambda\} = [x]_{1\lambda} \in X_1(\lambda)$. Therefore, $\aleph_{d_1}(X) \le \aleph_{d_2}(X)$ by Definition 3.9.

"$\Leftarrow$". By the condition $\forall \lambda \in [0,1]$, $X_1(\lambda) \le X_2(\lambda)$, i.e., $\forall x \in X$, $[x]_{2\lambda} \subseteq [x]_{1\lambda}$. By Proposition 3.2, $\forall x, y \in X, d_1(x,y) = \inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_{1\lambda}\} \le \inf_{\lambda \in [0,1]}\{\lambda | y \in [x]_{2\lambda}\} = d_2(x,y)$, that is $d_1 \le d_2$. □

Theorem 3.11 shows the order of NEC (pseudo-) metrics on universe X is the same as the one of their deriving granular spaces, i.e., they are order- preserving.

4. The structural clustering characteristic of granular space

Yao et al. presented their research on consistent classification problem [5,6], and they solved the problem by using the lattice, which all divisions are composite. Ziyan et al. [47] proposed also the research on consistency clustering, and they solved the problem using the algorithm based on a generalized expectation maximization (EM) framework. In fact, given $d \in D(X)$ (or $d \in WD(X)$), $\forall x, y, z \in X$, $d(x,y) \le \lambda$, $d(y,z) \le \lambda$, then we have $d(x,z) \le \max\{d(x,y), d(y,z)\} \le \lambda$, i.e., $x \in [y]_\lambda$, $y \in [z]_\lambda \rightarrow x \in [z]_\lambda$. If $X(\lambda)$ is finite set, the granulation $X(\lambda)$ stands for a finite covering of X, where the covering is a closed disk (in plan) or a closed sphere (in space) with the radius λ, and it is consisted of mutually disjoint family and satisfied the transitivity. Therefore, the ordered granulation space of complex systems based on NEC (pseudo-)metric just describes the structural clustering (or classifying) characteristic according to the metric, and the characteristic just accords with the demands of consistent classification problems presented in Refs. [5,6,47].

Definition 4.1. Let (X, d) be a (pseudo-)metric space, $A(X) = \{C_\alpha(X) | \alpha \in I\}$, $C_\alpha(X) = \{a_{\alpha\beta} | \beta \in I_\alpha\}$, and satisfies

(1) $\forall a_{\alpha\beta_1}, a_{\alpha\beta_2} \in C_\alpha(X), \beta_1 \ne \beta_2$, $a_{\alpha\beta_1} \cap a_{\alpha\beta_2} = \emptyset$;

(2) $\bigcup_{\beta \in I_\alpha} a_{\alpha\beta} = X$;

(3) $\forall a_{\alpha\beta_1}, a_{\alpha\beta_2} \in C_\alpha(X), \beta_1 \ne \beta_2$, there exists $d_0 > 0$ such that $x, y \in a_{\alpha\beta_i}$, $d(x,y) \le d_0$ (i=1,2), and $\forall x \in a_{\alpha\beta_1}$, $y \in a_{\alpha\beta_2}$, $d(x,y) > d_0$;

(4) $A(X)$ is an ordered set.

Then, $A(X)$ is called a consistent cluster (or classification) of X corresponding to d.

In Definition 4.1, the conditions (1) and (2) state that $C_\alpha(X)$ is a division of X, i.e., a cluster (or classification) of X, and the condition (3) states that it is obtained according to the (pseudo-)metric d on X. The condition (4) stands for the clustering (or classifying) is consistence, therefore, the condition (4) is also called the consistent condition. The consistent cluster (or classification) belongs to the category of structural cluster (or classification).

Proposition 4.1. *Suppose $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$), then the corresponding ordered granulation space $\aleph_d(X)$ is a consistent cluster (or classification) of X corresponding to d.*

Proof. By Proposition 3.2, the granulation space $\aleph_d(X) = \{X(\lambda) | \lambda \in [0,1]\}$ deriving by d is an ordered set, where $X(\lambda) = \{[x]_\lambda | x \in X\}$, and

(1) $\forall X(\lambda) \in \aleph_d(X), a, b \in X(\lambda), x \in a, y \in b$, that is, $a = [x]_\lambda$, $b = [y]_\lambda$. If $a \ne b$, then $a \cap b = \emptyset$. Otherwise, taken $z \in a \cap b \ne \emptyset$, then $a = [z]_\lambda = b$, that is contradictory to the condition $a \ne b$;

(2) $\forall \lambda \in [0,1]$, $\bigcup_{a \in X(\lambda)} a = X$;

(3) $\forall \lambda \in [0,1]$, $a = [x]_\lambda$, $b = [y]_\lambda \in X(\lambda)$, $a \ne b$, taken $d_0 = \lambda \ge 0$, we obtain $\forall x, y \in a$, $d(x,y) \le d_0$ and $\forall x \in a$, $y \in b$, $d(x,y) > d_0$ by the definition of a and b. Otherwise, if there exists $x \in a$, $y \in b$, $d(x,y) \le d_0$, then a=b, that is contradictory to the condition $a \ne b$.

Therefore, $\aleph_d(X)$ is a consistent cluster (or classification) of X corresponding to d by Definition 4.1. □

From Proposition 4.1, we directly obtain the following corollary.

Corollary 4.1. *Given a NEC (pseudo-)metric on the universe X, then the consistent cluster (or classification) on X is given.*

Definition 4.2. Suppose $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$). The graph, which elements of X denote vertex of graph and the distance between two elements stand for their arc length of the graph, is called metric graph of the NEC (pseudo-)metric space (X, d), marked $G(X, d)$.

By Definition 4.2, Proposition 4.1, Definition 3.6 and Theorem 3.8, we directly obtain the following proposition.

Proposition 4.2. *Suppose $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$), the corresponding ordered granular space is marked $\aleph_d(X)$. $\forall X(\lambda) \in \aleph_d(X)$, then the values of d_λ are just correspond with arcs in the metric graph of $X(\lambda)$, marked $G(X(\lambda), d_\lambda)$, where d_λ is the projective metric of d on $X(\lambda)$.*

Theorem 4.1. *Suppose $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$), the corresponding ordered granular space is marked as $\aleph_d(X)$. $\forall \lambda_1, \lambda_2 \in [0,1]$, $\lambda_1 < \lambda_2$, $a, b \in X(\lambda_1)$, $a \ne b$, if there exists $x \in a$, $y \in b$ such that $d(x,y) \le \lambda_2$, then $a \cup b \subseteq [x]_{\lambda_2} = [y]_{\lambda_2}$.*

Proof. When $d_1, d_2 \in D(X)$. By $a, b \in X(\lambda_1)$, $a \ne b$, we know that $\forall z_1 \in a$, $z_2 \in b$, $d(z_1, z_2) > \lambda_1$ and $d(z_1, x) \le \lambda_1$, $d(y, z_2) \le \lambda_1$. Then, $\forall z \in a \cup b$, we have

(1) When $z \in a$, $d(z,x) \le \lambda_1 < \lambda_2$, that is, $z \in [x]_{\lambda_2}$;

(2) When $z \in b$, $d(y,z) \le \lambda_1 < \lambda_2$. By $d(x,y) \le \lambda_2$ and $d \in D(X)$ (or $d \in WD(X)$), then $d(x,z) \le \max\{d(x,y), d(y,z)\} \le \lambda_2$, i.e., $z \in [x]_{\lambda_2}$ also holds.

Summarizing (1) and (2), $\forall z \in a \cup b$, $z \in [x]_{\lambda_2}$, that is, $a \cup b \subseteq [x]_{\lambda_2}$, and it is obvious that $[x]_{\lambda_2} = [y]_{\lambda_2}$.

Similar to the above proof, we can get that the theorem also holds when $d_1, d_2 \in WD(X)$. □

Theorem 4.1 shows that the granulation $X(\lambda)$ in space $\aleph_d(X)$ is changed with the increasing of λ, that is, it is a merging procedure. When λ vary from 0 to 1, Theorem 4.1 also shows that the granulation $X(\lambda)$ is changed only when $\lambda \in D = \{d(x,y) | x, y \in X\}$. Therefore, to obtain the granular space deriving from a NEC (pseudo-)metric d, we investigate the change of granulation $X(\lambda)$, where $\lambda \in D$. And Proposition 4.2 shows that the NEC metric on the granulation space $\aleph_d(X)$ deriving by the NEC (pseudo-)metric d just describes the procedure of its consistent cluster (or classification). So, we may obtain the sequence metric graphs and structural cluster (or classification) graphs of a NEC (pseudo-)metric d by using its projective metric d_λ on granulation $X(\lambda)$ $\lambda \in [0,1]$, the corresponding algorithm is similar to Algorithm A in Section 5, and omitted here. We shall give a example to illustrate the above theoretical results.

Example 1. Let $X = \{1,2,3,4,5\}$, $d \in D(X)$, and $d: d(i,i) = 0$, $i \in X$, $d(1,2) = 0.2$, $d(3,4) = 0.4$, $d(1,3) = d(1,4) = d(2,3) = d(2,4) = 0.5$, $d(1,5) = d(2,5) = d(3,5) = d(4,5) = 0.7$. The corresponding granular space is $\aleph_d(X) = \{X(0) = X, X(0.2) = \{\{1,2\},\{3\},\{4\},\{5\}\}, X(0.4) = \{\{1,2\},\{3,4\},\{5\}\}, X(0.5) = \{\{1,2,3,4\},\{5\}\}, X(0.7) = \{X\}\}$. We obtain the sequence metric graphs (Fig. 2) and structural clustering graph (Fig. 3).

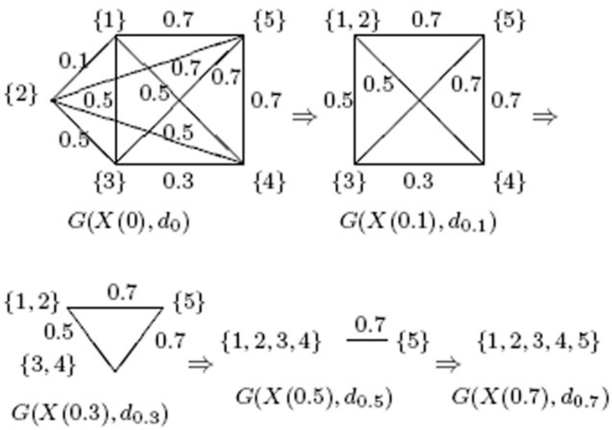

Fig. 2. The sequence metric graphs of d in Example 1.

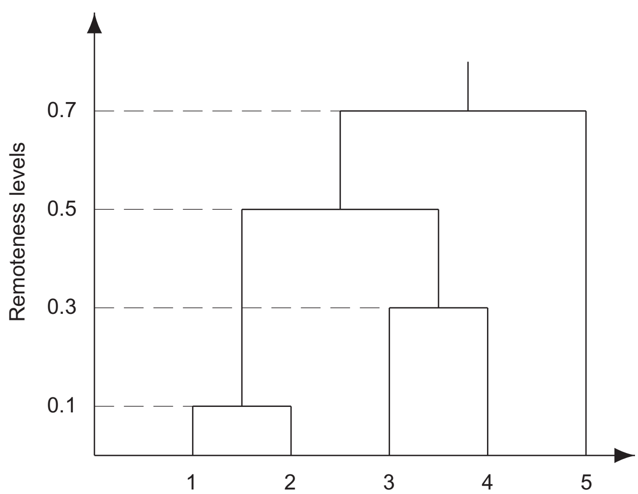

Fig. 3. The structural clustering graph of d in Example 1.

5. The granular representation of structural cluster on normalized metric

In Section 4, although the granular space deriving from NEC (pseudo-)metric is the consistent cluster (or classification), its equicrural condition is difficult to verify in actual applications. However, the normalized (pseudo-)metric is relatively easy to understand and derive. And Lemma 2.2 also shows that the structural clustering problem based on metric space can be transformed into the one based on normalized metric space. Therefore, in this section, we discuss how to get the consistent clustering (or classifying) analysis from normalized (pseudo-)metric space.

By Corollary 3.1, if $d \in ND(X)$ (or $d \in WND(X)$), $\forall \lambda \in [0,1]$, the cut-relation B_λ is a similarity relation on X (i.e., B_λ satisfies the reflexivity and symmetry), where $B_\lambda: (x,y) \in B_\lambda \leftrightarrow d(x,y) \le \lambda$, and B_λ is also called the cut similarity relation deriving from d corresponding to λ.

Proposition 5.1. *Assume B_1 and B_2 are two similarity relations on X, $t(B_1)$ and $t(B_2)$ stands for the transitive closure of B_1 and one of B_2, respectively. If $B_1 \subseteq B_2$, then $t(B_1) \subseteq t(B_2)$.*

Proof. $\forall (x.y) \in t(B_1)$, there exists $x = x_1, x_2, \ldots, x_m = y$ such that $(x_i, x_{i+1}) \in B_1$, $i = 1,2,\ldots,m-1$, where $x_i \in X$, $i = 1,2,\ldots,m$. By $B_1 \subseteq B_2$, there exists $x = x_1, x_2, \ldots, x_m = y$ such that $(x_i, x_{i+1}) \in B_2$, $i = 1,2,\ldots,m-1$, i.e., $(x,y) \in t(B_2)$. Therefore, $t(B_1) \subseteq t(B_2)$. □

Remark 6. Clearly, the inverse of Proposition 5.1 does not hold.

Definition 5.1. Assume $d \in ND(X)$ (or $d \in WND(X)$). $\forall \lambda \in [0,1]$, B_λ is the cut similarity relation deriving from d corresponding to λ. Define a relation D_λ on X:

$$D_\lambda = \{(x,y) | \exists x = x_1, x_2, \ldots, x_m = y, (x_i, x_{i+1}) \in B_\lambda, i = 1,2,\ldots,m-1\}$$

then D_λ is called the deriving relation based on the basis of B_λ.

Proposition 5.2. *D_λ in* Definition 5.2 *is an equivalence relation on X.*

Proof. It is obvious that this proposition holds according to Definition 5.1, here we omitted the details. □

In fact, D_λ is a transitive closure constructed from B_λ according to the transitivity, so it is marked as $D_\lambda = t(B_\lambda)$.

Definition 5.2. In Definition 5.1, $\forall x \in X$, the equivalent class of D_λ corresponding to x on X is marked $[x]_\lambda$, i.e., $[x]_\lambda = \{y | \exists x = x_1, x_2, \ldots, x_m = y, d(x_i, x_{i+1}) \le \lambda, i = 1,2,\ldots,m-1, y \in X\}$, the corresponding granulation is marked $X(\lambda) = \{[x]_\lambda | x \in X\}$, then the set $\{X(\lambda) | 0 \le \lambda \le 1\}$ is called the granular space deriving from d on X, and marked as $\aleph_{Td}(X)$.

Proposition 5.3. *Assume $d \in ND(X)$ (or $d \in WND(X)$), $\aleph_{Td}(X)$ is the deriving granular space from d according to Definition* 5.2. *Then, $\aleph_{Td}(X)$ is an ordered set, and $\forall \lambda_1, \lambda_2 \in [0,1], \lambda_1 \le \lambda_2 \to X(\lambda_2) \le X(\lambda_1)$.*

Proof. $\forall \lambda_1, \lambda_2 \in [0,1], \lambda_1 \le \lambda_2$, if $(x,y) \in B_1$, then $d(x.y) \le \lambda_1 \to d(x,y) \le \lambda_2$, i.e., $B_{\lambda_1} \subseteq B_{\lambda_2}$. By Proposition 5.1, $D_{\lambda_1} = t(B_{\lambda_1}) \subseteq t(B_{\lambda_2}) = D_{\lambda_2}$. Furthermore, we have $X(\lambda_2) \le X(\lambda_1)$ by Definitions 5.2 and 3.2. □

Theorem 5.1. *Assume $d \in ND(X)$ (or $d \in WND(X)$), $\aleph_{Td}(X)$ is its corresponding deriving granular space. Define $d^*: \forall x, y \in X$, $d^*(x,y) = \inf_{\lambda \in [0,1]} \{\lambda | \exists x = x_1, x_2, \ldots, x_m = y, d(x_i, x_{i+1}) \le \lambda, i = 1,2,\ldots,m-1$. Then $d^* \in D(X)$ (or $d^* \in WD(X)$).*

Proof. Note that the definition of d^* is equivalent to $d*(x,y) = \inf_{\lambda \in [0,1]} \{\lambda | (x,y) \in R_\lambda\} = \inf_{\lambda \in [0,1]} \{\lambda | y \in [x]_\lambda\}$. Some marks are seen in Definition 5.2, the proof is similar to the one of Proposition 3.2, hence omitted here. □

Definition 5.3. In Theorem 5.1, d^* is called the deriving NEC (pseudo-)metric from the normalized metric d on X according to the transitivity, and marked as $d^* = t(d)$.

Theorem 5.2. *Assume $d \in ND(X)$ (or $d \in WND(X)$), $d^* = t(d)$, their corresponding granular space is $\{X(\lambda) | 0 \le \lambda \le 1\}$ and $\{X^*(\lambda) | 0 \le \lambda \le 1\}$, respectively. If X is finite set, then $\forall \lambda \in [0,1]$, $X(\lambda) = X^*(\lambda)$.*

Proof. $[x]_\lambda$ and $[x]^*_\lambda$ stands for the deriving equivalent class from d and d^* corresponding to λ, respectively. $\forall a = [x]_\lambda \in X(\lambda)$, then $\forall y \in a$, there exists $x = x_1, x_2, \ldots, x_m = y$ such that $d(x_i, x_{i+1}) \le \lambda$, $i = 1,2,\ldots,m-1$, i.e., $y \in [x]^*_\lambda$, so $[x]_\lambda \subseteq [x]^*_\lambda$. Then

$$X^*(\lambda) \le X(\lambda) \tag{5.1}$$

Similarly, $\forall b = [x]^*_\lambda \in X^*(\lambda)$, then $\forall y \in b$, by Theorem 5.1, we have

$$\inf_{\lambda \in [0,1]} \{\lambda | \exists x = x_1, x_2, \ldots, x_m = y, d(x_i, x_{i+1}) \le \lambda, i = 1,2,\ldots,m-1\}$$
$$= d^*(x,y) \le \lambda$$

(1) When $d^*(x,y) < \lambda$, there exist λ_1 $(<\lambda)$ and $x = x_1, x_2, \ldots, x_m = y$ such that $d(x_i, x_{i+1}) \le \lambda_1 < \lambda$, $i = 1,2,\ldots,m-1$ according to the finite set X, that is, $y \in [x]_\lambda$;

(2) When $d^*(x,y) = \lambda$. By the definition of inferior, $\forall \varepsilon > 0$, there exists $\lambda_2 \in [0,1]$ such that $y \in [x]_{\lambda_2}$ and $\lambda - \varepsilon \le \lambda_2 < \lambda + \varepsilon$, i.e., there exists $x = x_1, x_2, \ldots, x_m = y$ such that $d(x_i, x_{i+1}) > \lambda - \varepsilon$, $i = 1,2,\ldots,m-1$. Because X is a finite set, let $\varepsilon \to 0^+$, there exists $x = x_1, x_2, \ldots, x_m = y$ such that $d(x_i, x_{i+1}) \le \lambda$, $i = 1,2,\ldots,m-1$. Therefore, $y \in [x]_\lambda$.

Summarizing (1) and (2), we have

$$X(\lambda) \le X^*(\lambda) \tag{5.2}$$

By the (5.1) and (5.2), we have completed the proof. □

From Theorems 5.1 and 5.2, we can obtain directly the following corollary.

Corollary 5.1. *Assume $d \in ND(X)$ (or $d \in WND(X)$), $d^* = t(d)$, B_λ and is the cut similarity relation deriving from d corresponding to λ, D^*_λ is the cut equivalence relation of d^*. If X is finite set, then $\forall \lambda \in [0,1]$, $D^*_\lambda = t(B_\lambda)$, i.e., $[t(d)]_\lambda = t(d_\lambda)$, where d_λ stands for the cut-relation of d.*

From Theorem 5.2, we know that the deriving ordered granular space can be perfectly determined from the transitive closure of its cut similarity relation when the universe X is a finite, where its deriving ordered granular space is just as the consistent classification (or cluster) mentioned in Section 4. Its deriving NEC (pseudo-)metric can also be obtained from Theorem 5.1, where metric on its deriving ordered granular space is defined according to Theorem 3.8. Once a normalized (pseudo-)metric on X is given, Corollary 5.1 ensures that its ordered granular space and NEC (pseudo-)metric obtained from the transitive closure are both unique, Corollary 5.1 also shows the fact that the cut operation and transitive closure are exchangeable for relations deriving by normalized (pseudo-)metric when the universe X is a finite.

Theorem 5.3. *Assume $d \in ND(X)$ (or $d \in WND(X)$). $\forall \lambda_1, \lambda_2 \in [0,1]$, $\lambda_1 < \lambda_2$, D_{λ_1} and D_{λ_2} are the deriving equivalence relations according to Definition 5.1, their corresponding granulations is marked as $X(\lambda_1)$ and $X(\lambda_2)$, respectively. $\forall [x]_{\lambda_1}, [y]_{\lambda_1} \in X(\lambda_1)$, $[x]_{\lambda_1} \ne [y]_{\lambda_1}$, if there exists $x_0 \in [x]_{\lambda_1}$, $y_0 \in [y]_{\lambda_1}$ such that $d(x_0, y_0) \le \lambda_2$, then $[x]_{\lambda_1} \cup [y]_{\lambda_1} \subseteq [x]_{\lambda_2}$, and $[x]_{\lambda_2} = [y]_{\lambda_2}$.*

Proof. From $[x]_{\lambda_1} \ne [y]_{\lambda_1}$, we get $\forall z_1 \in [x]_{\lambda_1}$, $z_2 \in [y]_{\lambda_1}$, $d(z_1, z_2) > \lambda_1$. $\forall z \in [x]_{\lambda_1} \cup [y]_{\lambda_1}$

(1) When $z \in [x]_{\lambda_1}$, $z \in [x]_{\lambda_1} \subseteq [x]_{\lambda_2}$;

(2) When $z \in [y]_{\lambda_1}$, there exists $z = y_1, y_2, \ldots, y_{m_1} = y$ such that

$$d(y_i, y_{i+1}) \le \lambda_1 < \lambda_2, \quad i = 1,2,\ldots,m_1 - 1 \tag{5.3}$$

From $x_0 \in [x]_{\lambda_1}$, there exists $x_0 = x_1, x_2, \ldots, x_{m_2} = x$ such that

$$d(x_j, x_{j+1}) < \lambda_2, \quad j = 1,2,\ldots,m_2 - 1 \tag{5.4}$$

By $y_0 \in [y]_{\lambda_1}$, there exists $y = z_1, z_2, \ldots, z_{m_3} = y_0$ such that

$$d(z_k, z_{k+1}) < \lambda_2, \quad k = 1,2,\ldots,m_3 - 1 \tag{5.5}$$

Summarizing (5.3)–(5.5) and the condition $d(x_0, y_0) \le \lambda_2$, we have $z \in [x]_{\lambda_2}$.

Therefore, $[x]_{\lambda_1} \cup [y]_{\lambda_1} \subseteq [x]_{\lambda_2}$. And it is obvious that $[x]_{\lambda_2} = [y]_{\lambda_2}$. □

Remark 7. Theorem 5.3 shows that granulation $X(\lambda)$ is changed from the fine to the coarse with the increasing of λ, that is, its equivalent classes is a merging procedure. It also shows that the granulation $X(\lambda)$ is changed only if B_λ is changed (Note: D_λ is also changed) when the threshold λ varies from 0 to 1. Therefore, to obtain the granular space deriving from a normalized (pseudo-) metric d, we need to investigate the change of granulation $X(\lambda)$, where $\lambda \in D = \{d(x,y) | x, y \in X\}$.

By Theorems 5.1, 5.2, 5.3 and Remark 7, we can get a reasonable algorithm to solve its deriving granular space on X from a normalized (pseudo-)metric d as follows.

Let d be a normalized (pseudo-)metric on finite set $X = \{x_1, x_2, \ldots, x_n\}$, $D = \{d(x,y) | x, y \in X\} = \{d_0, d_1, \ldots, d_m\}$, where $d_0 = 0 < d_1 < \cdots < d_m$. We can obtain the following algorithm to solve the deriving granular space $\aleph_{Td}(X)$ and NEC (pseudo-)metric d^*.

Algorithm A.

- *Step* 1. $i \Leftarrow 0$, $X(d_i) = C = \{a_1, a_2, \ldots, a_N\} (N \le n)$. For i=1 to N, if $x_k, x_j \in a_i$, $d^*(x_k, x_j) = 0$;
- *Step* 2. Output $X(d_i) = C$;
- *Step* 3. $A \Leftarrow C$, $i \Leftarrow i+1$, $C \Leftarrow \emptyset$;
- *Step* 4. $B \Leftarrow \emptyset$;
- *Step* 5. Taken $a_j \in A$, $B \Leftarrow B \cup a_j$, $A \Leftarrow A \backslash a_j$;
- *Step* 6. $\forall a_k \in A$, if there exists $x_j \in a_j$, $y_k \in a_k$ such that $d(x_j, y_k) \le d_i$, $B \Leftarrow B \cup a_k$, $A \Leftarrow A \backslash a_k$. $\forall x_j \in a_j$, $y_k \in a_k$, $d^*(x_j, x_k) \Leftarrow d_i$, otherwise goto Step 7;
- *Step* 7. $C \Leftarrow \{B\} \cup C$;
- *Step* 8. If $A \Leftarrow \emptyset$, then goto Step 4, otherwise, if $X(d_i) \ne X(d_{i-1})$, output $X(d_i) = C$;
- *Step* 9. If $i = m$ or $C = \{X\}$, then goto Step 10, otherwise goto Step 3;
- *Step* 10. Output $d^* = [d^*(x_i, x_j)]_{n \times n}$;
- *Step* 11. End.

By Algorithm A, we can rapidly get the deriving granular space $\aleph_{Td}(X) = \{X(\lambda) | \lambda \in D\}$ from the normalized (pseudo-)metric d, furthermore get all structural clusters (or categories) deriving from d. Here Steps 4–7 are used to compute the granulation corresponding to the threshold d_i, where the computational complexity is not larger than $n \times (n-1) \times m/2$. The terminating condition "i=m" of Algorithm A indicates that all probable granulations have been gone through all probable threshold values. And the condition "$C = \{X\}$" denotes that it is not necessary to continue operation when a granulation $X(\lambda) = \{X\}$ happens for any $\lambda > d_i$, $X(\lambda) = \{X\}$. The deriving NEC (pseudo-)metric d^* is given by the Step 10. The following example serves as an illustration of Algorithm A.

Example 2. Let d be a normalized metric $X = \{1,2,\ldots,12\}$, its metric representation is as follows:

$$d = [d_{ij}]_{n \times n} = \begin{bmatrix}
0.0000 \\
0.9856 & 0.0000 \\
0.9947 & 0.9795 & 0.0000 \\
0.9932 & 0.9840 & 0.8850 & 0.0000 \\
0.9658 & 0.9829 & 0.9957 & 0.9955 & 0.0000 \\
0.9800 & 0.9626 & 0.9829 & 0.9721 & 0.9901 & 0.0000 \\
0.9894 & 0.9596 & 0.9510 & 0.9165 & 0.9929 & 0.8550 & 0.0000 \\
0.9488 & 0.9718 & 0.9905 & 0.9864 & 0.9849 & 0.8440 & 0.9671 & 0.0000 \\
0.9864 & 0.9421 & 0.9915 & 0.9922 & 0.8920 & 0.9848 & 0.9875 & 0.9820 & 0.0000 \\
0.8750 & 0.9878 & 0.9960 & 0.9954 & 0.8490 & 0.9895 & 0.9931 & 0.9814 & 0.9703 & 0.0000 \\
0.9927 & 0.9060 & 0.8950 & 0.9691 & 0.9930 & 0.9760 & 0.9430 & 0.9863 & 0.9828 & 0.9942 & 0.0000 \\
0.9569 & 0.9107 & 0.9914 & 0.9909 & 0.9500 & 0.9658 & 0.9804 & 0.9513 & 0.9274 & 0.9632 & 0.9832 & 0.0000
\end{bmatrix}$$

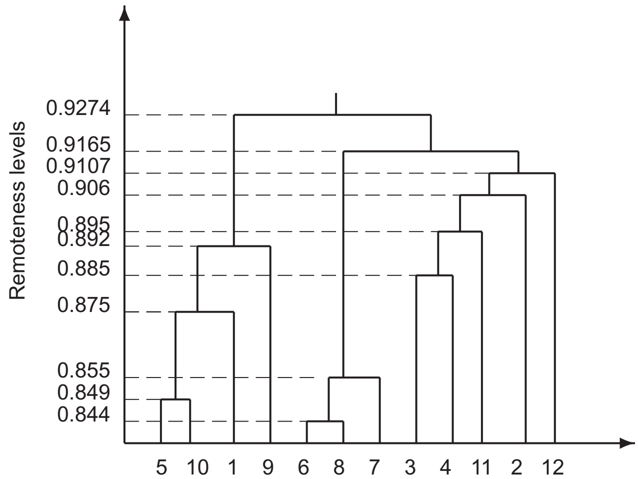

Fig. 4. The structural clustering graph of Example 2.

By Algorithm A, we get the deriving granular space $\aleph_{Td}(X)$ as follows: $X(0.0000)=X$, $X(0.8440)=\{\{1\},\ldots,\{5\},\{6,8\},\{7\},\{9\},\ldots,\{12\}\}$, $X(0.8490)=\{\{1\},\ldots,\{4\},\{5,10\},\{6,8\},\{7\},\{9\},\{11\},\{12\}\}$, $X(0.8550)=\{\{1\},\ldots,\{4\},\{5,10\},\{6,7,8\},\{9\},\{11\},\{12\}$, $X(0.8750)=\{\{1,5,10\},\{2\},\{3\},\{4\},\{6,7,8\},\{9\},\{11\},\{12\}\}$, $X(0.8850)=\{\{1,5,10\},\{2\},\{3,4\},\{6,7,8\},\{9\},\{11\},\{12\}\}$, $X(0.8920)=\{\{1,5,9,10\},\{2\},\{3,4\},\{6,7,8\},\{11\},\{12\}\}$, $X(0.8950)=\{\{1,5,9,10\},\{2\},\{3,4,11\},\{6,7,8\},\{12\}\}$, $X(0.9060)=\{\{1,5,9,10\},\{2,3,4,11\},\{6,7,8\},\{12\}\}$, $X(0.9107)=\{\{1,5,9,10\},\{2,3,4,11,12\},\{6,7,8\}\}$, $X(0.9165)=\{\{1,5,9,10\},\{2,3,4,6,7,8,11,12\}\}$, $X(0.9274)=\{X\}$.

We also get the deriving NEC metric d^* as follows. The corresponding structural clustering graph is shown in Fig. 4.

$$d^*=[d^*_{ij}]_{n\times n}=\begin{bmatrix}
0.0000 & & & & & & & & & & & \\
0.9274 & 0.0000 & & & & & & & & & & \\
0.9274 & 0.9060 & 0.0000 & & & & & & & & & \\
0.9274 & 0.9060 & 0.8850 & 0.0000 & & & & & & & & \\
0.8750 & 0.9274 & 0.9274 & 0.9274 & 0.0000 & & & & & & & \\
0.9274 & 0.9165 & 0.9165 & 0.9165 & 0.9274 & 0.0000 & & & & & & \\
0.9274 & 0.9165 & 0.9165 & 0.9165 & 0.9274 & 0.8550 & 0.0000 & & & & & \\
0.9274 & 0.9165 & 0.9165 & 0.9165 & 0.9274 & 0.8440 & 0.8550 & 0.0000 & & & & \\
0.8920 & 0.9274 & 0.9274 & 0.9274 & 0.8920 & 0.9274 & 0.9274 & 0.9274 & 0.0000 & & & \\
0.8750 & 0.9274 & 0.9274 & 0.9274 & 0.8490 & 0.9274 & 0.9274 & 0.9274 & 0.8920 & 0.0000 & & \\
0.9274 & 0.9060 & 0.8950 & 0.9274 & 0.9274 & 0.9165 & 0.9165 & 0.9165 & 0.9274 & 0.9274 & 0.0000 & \\
0.9274 & 0.9107 & 0.9107 & 0.9274 & 0.9274 & 0.9165 & 0.9165 & 0.9165 & 0.9274 & 0.9274 & 0.9107 & 0.0000
\end{bmatrix}$$

Remark 8. The data d in Example 2 stems from Ref. [32], where $d_{ij}=1-R_{ij}$ and element $1,2,\ldots,12$ stands for $a,b,\ldots,l$ in Ref. [32], respectively. The point set was divided into 12 subclasses before application corresponding algorithm. Note that the dendrogram in Ref. [32] is incorrect, the correct structural clustering graph should exchange the position of i and j each other (see our Fig. 4).

6. The optimal cluster determination based on granular space

In Section 5, we have investigated how to obtain the cluster (or classification) structure (or granular space) of X from a normalized (pseudo-)metric on X, and pointed out that this structure is the consistent cluster (or classification). Now the question is how to define the optimal cluster (or classification) number from the structure. It must be addressed when studying all cluster (or classification) issue by using the structural cluster method. It is essential about the optimal cluster determination based on granular space.

In the last ten years, many effectiveness indexes on structural cluster (or classification) have been proposed [37,48–51]. These indexes stemming from data sets in different problems are used to define the optimal classification number. However, it is difficult to solve them. Hardy [52] presented the comparative evaluation method under given cluster number. Kim et al. [37,51–55] proposed the optimal clustering determination method based on the clustering structure, and Bezdek et al. [28,30,48,56–58] proposed FCM based on fuzzy rules. The essence of these methods is to give optimum selecting principle or criterion in order to determine the best cluster (or classification). These methods have been widely used to define the optimal fuzzy classifying number [59–61]. At present, effectiveness index methods used to define the optimal clustering (or classifying) number are based on C-means according to fuzzy rules, and consists mainly of the partition coefficient index [61], fuzzy hyper-volume index [62], partition entropy index [63], Fukuyama and Sugeno index [64], compactness separability index [65], partition *CWB* index [56] and partition K_C index [32]. Descriptions and corresponding optimal cluster number models are listed for those indices in Table 1, respectively.

In Table 1, $\overline{v}$ denotes the center of all data, v_i is the center of the classification C_i, d_{ij} denotes the distance from point x_i to the center v_j, $\sum_i$ is the fuzzy covariance matrix of classification C_i, where $\sum_i=S_i/\sum_{k=1}^{c}u_{ik}^2$, $S_i=\sum_{k=1}^{c}u_{ik}^2\cdot(x_k-v_i)\cdot(x_k-v_i)^T$. In *CWB*, $\alpha=Dist(c_{\max})$, $Scat(c)=\sum_{k=1}^{c}[\sigma(v_k)^T\cdot\sigma(v_k)]^{1/2}\{c\cdot[\sigma(X)^T\cdot\sigma(X)]^{1/2}\}$, $Dist(c)=D_{\max}/D_{\min}\cdot\sum_{i=1}^{c}[\sum_{j=1}^{c}\|v_i-v_j\|]^{-1}$, where $D_{\max}$ and $D_{\min}$ are the maximum and minimum distances between the cluster centroids, $\sigma(X)=(\sum_{k=1}^{n}(x_k^1-\overline{x^1})^2/n,\ldots,\sum_{k=1}^{n}(x_k^p-\overline{x^p})^2/n)^T$, $\sigma(v_i)=(\sum_{k=1}^{n}(x_k^1-v_k^1)^2/n,\ldots,\sum_{k=1}^{n}(x_k^p-v_k^p)^2/n)^T$; $CS(U)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^2\cdot d_{ik}^2/[n\cdot\min_{v_i\in C_i,v_j\in C_j,i\neq j}\{\|v_i-v_j\|\}]$; $CO_{gl}=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^2\cdot d_{ik}^2/[c\cdot\sum_{j=1}^{n}u_{ij}]$, $CO_{av}=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^2\cdot d_{ik}^2/n$.

In this section, we give a study on the optimal cluster determination problem deriving from normalized (pseudo-)metric based on above section. We discuss the optimal selection criteria in the following.

Assume d is a normalized (pseudo-)metric on the finite set $X=\{x_1,x_2,\ldots,x_n\}$, its deriving granular space is $\aleph_{Td}(X)$. The mark $d_i=(d(x_i,x_1),d(x_i,x_2),\ldots,d(x_i,x_n))^T$ stands for the n-dimensional vector consisting of all distances between x_i to every element of X,

Table 1
Function description and corresponding optimal cluster models of various indexes.

Validity criteria	Function description	Optimal objection function
K_C index description	$K_C = 1 - CO_{av}/CO_{gl}$	$\min\{K_c, U, c\}$
Partition coefficient	$F = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^2/n$	$\max\{F, U, c\}$
Partition entropy	$H = -\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}\cdot\log u_{ik}/n$	$\max\{H, U, c\}$
Fuzzy hyper-volume	$F_{hv} = \sum_{i=1}^{c}[\det(\sum_i)^{1/2}]$	$\min\{F_{hv}, U, c\}$
Compactness separation	$CS(U)$	$\min\{CS, U, c\}$
Fukuyama and Sugeno	$FS(U) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^2\cdot(d_{ik}^2 - \lVert v_i - \overline{v}\rVert^2)$	$\min\{FS, U, c\}$
CWB index description	$F = \alpha\cdot Scat(c) + Dist(c)$	$\min\{CWB, U, c\}$

Table 2
The comparison table of $S(X(\lambda))$ with the others indexes in Example 3.

Number of clusters	2	3	4	5	6	7	8	9	10	11	12
$S(X(\lambda))$	1.570	2.120	2.140	2.120	2.030	2.000	1.670	1.570	1.470	1.170	0.870
K_C	0.050	0.003	0.065	0.095	0.122	0.086	0.110	0.065	0.032	0.013	0.010
F	0.887	0.782	0.765	0.746	0.725	0.702	0.688	0.659	0.623	0.601	0.572
H	0.207	0.417	0.485	0.551	0.614	0.680	0.726	0.799	0.894	0.954	1.030
F_{hv}	0.033	0.047	0.061	0.072	0.082	0.094	0.103	0.116	0.128	0.139	0.150
CS	0.087	0.092	0.088	0.084	0.080	0.077	0.083	0.078	0.110	0.105	0.107
FS	−78	−164	−163	−160	−153	−147	−140	−131	−128	−120	−119
CWB	4.390	4.260	4.100	5.410	6.070	5.860	6.190	6.520	7.370	7.210	7.480

$i = 1, 2, \ldots, n$. The mark $\overline{a} = \sum_{i=1}^{n} d_i/n$ denotes the center of n-dimension distance vectors.

Given $X(\lambda) \in \aleph_{Td}(X)$, we marked $X(\lambda) = \{a_1, a_2, \ldots, a_{C_\lambda}\}$, $a_k = \{x_{k_1}, x_{k_2}, \ldots, x_{k_{J_k}}\}$, where $k = 1, 2, \ldots, C_\lambda$ and $\sum_{k=1}^{C_\lambda} J_k = n$. $\overline{a}_k = \sum_{i=1}^{J_k} d_{k_i}/J_k$ stands for the center of classification a_k $(k = 1, 2, \ldots, C_\lambda)$. So, we can construct the deviation among classes and the deviation in classes of $X(\lambda)$ as follows, respectively: $S_{between} = \sum_{i=1}^{C_\lambda} J_i\cdot\lVert\overline{a}_i - \overline{a}\rVert_2^2/n$, $S_{in} = \sum_{i=1}^{C_\lambda}\sum_{j=1}^{J_i}\lVert d_{i_j} - \overline{a}_i\rVert_2^2/J_i$ where $\lVert\cdot\rVert_2$ stands for 2-norm. Then the total deviation of $X(\lambda)$ is:

$$S(X(\lambda)) = \sum_{i=1}^{C_\lambda} J_i\cdot\lVert\overline{a}_i - \overline{a}\rVert_2^2/n + \sum_{i=1}^{C_\lambda}\sum_{j=1}^{J_i}\lVert d_{i_j} - \overline{a}_i\rVert_2^2/J_i \tag{6.1}$$

From the perspective of classification, the finer $X(\lambda)$ is, the smaller S_{in} is and the greater $S_{between}$ is. Particularly, if X is seen as one classification, then $S_{between}$=0 and $S_{in} = \sum_{i=1}^{n}\lVert d_i - \overline{a}\rVert_2^2/n$; if X is divided into n classifications, then $S_{between} = \sum_{i=1}^{n}\lVert d_i - \overline{a}\rVert_2^2/n$ and S_{in}=0. Therefore, when X is divided into one or n classifications, their total deviation is the same, and the following theorem holds.

Theorem 6.1. *Assume* $d \in ND(X)$ (*or* $d \in WND(X)$), *corresponding granular space is marked as* $\aleph_{Td}(X)$. *Then,* $\forall X(\lambda) \in \aleph_{Td}(X)$, $S(X(\lambda)) \geq \sum_{i=1}^{n}\lVert d_i - \overline{a}\rVert_2^2/n$.

Proof.

$$\begin{aligned} S(X(\lambda)) &= \sum_{i=1}^{C_\lambda} J_i\cdot\lVert\overline{a}_i - \overline{a}\rVert_2^2/n + \sum_{i=1}^{C_\lambda}\sum_{j=1}^{J_i}\lVert d_{i_j} - \overline{a}_i\rVert_2^2/J_i \\ &\geq \sum_{i=1}^{C_\lambda} J_i\cdot\lVert\overline{a}_i - \overline{a}\rVert_2^2/n + \sum_{i=1}^{C_\lambda}\sum_{j=1}^{J_i}\lVert d_{i_j} - \overline{a}_i\rVert_2^2/n \\ &= \sum_{i=1}^{C_\lambda}\sum_{j=1}^{J_i}[\lVert\overline{a}_i - \overline{a}\rVert_2^2 + \lVert d_{i_j} - \overline{a}_i\rVert_2^2]/n \\ &= \sum_{i=1}^{C_\lambda}\sum_{j=1}^{J_i}\lVert d_{i_j} - \overline{a}\rVert_2^2/n = \sum_{i=1}^{n}\lVert d_i - \overline{a}\rVert_2^2/n. \quad\square \end{aligned}$$

From Theorem 6.1, we know that the total deviation is the smallest when X is divided into one or n classifications, and obviously this is not the results we had hoped for. From the perspective of a classification, any reasonable classification should reflect the greatest classifying capacity of its data. So, we establish the clustering optimization principle to achieve the cluster, which $S(X(\lambda))$ take the greatest value. Its mathematical model is as follows:

$$S(X(\lambda_0)) = \max_{X(\lambda)\in\aleph_{Td}(X)}\{S(X(\lambda))\} \tag{6.2}$$

The computational complexity of model (6.2) is computed as follows. For the given granulation $X(\lambda)$, the computational complexity of $S_{between}$ is $(C_\lambda+1)\times(n+1)+1 = n+2+(n+1)\times C_\lambda$ (i.e., the multiplication and division operation times of $S_{between}$), and the computational complexity of S_{in} is $\sum_{i=1}^{C_\lambda}[(J_i\times n+1)+n+1] = n^2+(n+2)\times C_\lambda$. So, the computational complexity of $S(X(\lambda))$ is $n^2+n+2+(2n+3)\times C_\lambda$ for the given granulation $X(\lambda)$. If there are m values in D, the computational complexity of model (6.2) is $[n^2+n+2+(2n+3)\times C_\lambda]\times(m-1)$. Because of $[n^2+n+2+(2n+3)\times C_\lambda]\times(m-1) \leq [n^2+n+2+(2n+3)\times n]\times(m-1) = O(n^2\times m)$, the computational complexity of model (6.2) is $O(n^2\times m)$. We shall give the following example to illustrate our optimizing mathematics model (6.2).

Example 3. In Example 2, we establish the optimizing mathematical model (6.2) to construct the total deviation by (6.1), and can compute the results in Table 2 by programming as follows.

From Table 2, we have obtained the optimal clustering number of 4. The other criteria attain the different optimal clustering numbers as follow: index F and F_{hv} correspond to 2, index K_c and FS correspond to 3, index CS corresponds to 7, index H corresponds to 12, and CWB attains its optimal value for 4 clusters, which is the same as our method. From Fig. 4, we can also know that the four classes is just as $X(.9060)$, that is, $C_1 = \{1,5,9,10\}$, $C_2 = \{6,7,8\}$, $C_3 = \{2,3,4,11\}$ and $C_4 = \{12\}$. Compared with Ref. [32], merged C_3 and C_4 into one subclass, we considered that C_3 and C_4 are each one subclass respectively. According to Ref. [35,54], it is also reasonable that the optimal clustering number of Example 2 is 4, and its classifying structure is reasonable when the threshold value $\lambda \in [0.9060, 0.9107)$. Note that the computational complexity of our objective function $S(X(\lambda))$ is the smallest of all, and it is easy to realize $S(X(\lambda))$.

Compared with previous Refs. [28,30,32,37], our example (Example 3) includes only an objective function (6.2), but without any additional (fuzzy) rules. However our method can select the optimal in all clustering structures (Note that it is the global optimization), it has great superiority, and very strong operability as demonstrated by Example 3.

7. The fusion problem of structural cluster on metrics

The fusion problem has recently become one of most hot research fields. Pedrycz [33] researched on collaborative fuzzy clustering, Devillez [32] worked on fusion clustering problem, and Miyamoto [66] also studied the information fusion based on fuzzy multi-sets. Esnaf et al. [67] proposed a fuzzy clustering-based hybrid method for a multi-facility location problem, Chuu [68] presented also the fusion problem based on fuzzy multi-attribute group decision-making with multi-granularity linguistic information. These fusion methods on structural clusters were developed to address a common question, that is, how to obtain a better cluster (or classification) from a variety of clusters (or classifications) on universe X. This question can be generalized. Just as we can derive a cluster (or classification) based on a piece of information on universe X, we can derive multiple clusters (or classifications) based on multiple pieces of information on universe X. Essentially, the question that we need to address is how to integrate/process these information towards deriving a cluster (or classification) or how to obtain the cluster (or classification) by integrating these information. The key is to obtain a thinner and precise cluster (or classification) from these given information. In this section, we study the fusion problem of structural clusters (or classifications) based on metrics, and organize it as follows. In Section 7.1, the fusion of structural clusters based on NEC (pseudo-)metrics is given. In Section 7.2, the fusion of structural clusters based on normalized (pseudo-)metrics is get.

7.1. The fusion of structural clusters based on NEC metrics

In this section, we analyze fusion problems of structural clusters based on granular spaces derived by NEC (pseudo-)metrics.

Theorem 7.1. *$D(X)$ (or $WD(X)$) composes a perfect semi-order lattice under the relation "$\leq$" defined by Definition 3.8.*

Proof. For any subset $\{d_\alpha, \alpha \in I\} \subseteq D(X)$, define $\overline{d}$ and $\underline{d}$: $\forall x, y \in X$,

$$\overline{d}(x,y) = \sup_{\alpha \in I}\{d_\alpha(x,y)\}, \underline{d}(x,y) = \inf_{\alpha \in I}\{d_\alpha(x,y)\}.$$

The following, we prove that $\overline{d}$ and $\underline{d}$ is the sup and inferior of $\{\{d_\alpha, \alpha \in I\}$ respectively.

(1) We prove that $\overline{d} \in D(X)$. Obviously, $\overline{d}$ satisfies the normalized condition, and is symmetric, and

(A1) $\forall x \in X$, $\overline{d}(x,x) = \sup_{\alpha \in I}\{d_\alpha(x,x)\} = 0$;

(A2) $\forall x,y,z \in X$, $\overline{d}(x,y) = \sup_{\alpha \in I}\{d_\alpha(x,y)\} \leq \sup_{\alpha \in I}\{\max\{d_\alpha(x,z), d_\alpha(z,y)\}\} = \max\{\sup_{\alpha \in I}\{d_\alpha(x,z)\}, \sup_{\alpha \in I}\{d_\alpha(z,y)\}\} \to \overline{d}(x,y) \leq \max\{\overline{d}(x,z), \overline{d}(z,y)\}$;

Similar to the above proof, we have

$$\overline{d}(x,z) \leq \max\{\overline{d}(x,y), \overline{d}(z,y)\}, \overline{d}(y,z) \leq \max\{\overline{d}(x,z), \overline{d}(x,y)\}.$$

Then, $\overline{d}$ satisfies the equicrural condition on X. Based on (A1), (A2) and Definition 2.1, we have $d_\alpha \in D(X)$.

(2) Now, we prove that $\overline{d}$ is the sup of $\{d_\alpha, \alpha \in I\}$.

(B1) From the definition of $\overline{d}$, $\forall x, y \in X$, $\overline{d}(x,y) = \sup_{\alpha \in I}\{d_\alpha(x,y)\} \geq d_\alpha(x,y) \to d_\alpha \leq \overline{d}$, i.e., $\overline{d}$ is an upper bound of $\{d_\alpha, \alpha \in I\}$.

(B2) For any upper bound d_1 of $\{d_\alpha, \alpha \in I\}$, i.e., $\forall \alpha \in I$, $d_\alpha \leq d_1$. Furthermore, $\forall x, y \in X$, $d_\alpha(x,y) \leq d_1(x,y) \to \overline{d}(x,y) = \sup_{\alpha \in I}\{d_\alpha(x,y)\} \leq d_1(x,y) \to \overline{d} \leq d_1$.

Based on (B1) and (B2), $\overline{d}$ is the sup of $\{d_\alpha, \alpha \in I\}$.

Next, we prove that $\underline{d}$ is the inferior of $\{d_\alpha, \alpha \in I\}$.

(3) We prove $\underline{d} \in D(X)$. Obviously, $\underline{d}$ satisfies the normalized condition and symmetry property, and

(C1) $\forall x \in X$, $\underline{d}(x,x) = \inf_{\alpha \in I}\{d_\alpha(x,x)\} = 0$;

(C2) $\forall x,y,z \in X$, $\underline{d}(x,y) = \inf_{\alpha \in I}\{d_\alpha(x,y)\} \leq \inf_{\alpha \in I}\{\max\{d_\alpha(x,z), d_\alpha(z,y)\}\} \leq \max\{\inf_{\alpha \in I}\{d_\alpha(x,z)\}, \inf_{\alpha \in I}\{d_\alpha(z,y)\}\} \to \underline{d}(x,y) \leq \max\{\underline{d}(x,z), \underline{d}(z,y)\}$;

Similar to the above proof, we have

$$\underline{d}(x,z) \leq \max\{\underline{d}(x,y), \underline{d}(z,y)\}, \underline{d}(y,z) \leq \max\{\underline{d}(x,z), \underline{d}(x,y)\}.$$

Then, $\underline{d}$ satisfies the equicrural condition on X. By (C1) and (C2), $\underline{d} \in D(X)$.

(4) We prove that $\underline{d}$ is the inferior of $\{d_\alpha, \alpha \in I\}$ in the following.

(D1) From the definition of $\underline{d}$, $\forall x, y \in X$, $\underline{d}(x,y) = \inf_{\alpha \in I}\{d_\alpha(x,y)\}\} \leq d_\alpha(x,y) \to \underline{d} \leq d_\alpha$, i.e., $\underline{d}$ is a lower bound of $\{d_\alpha, \alpha \in I\}$.

(D2) For any lower bound d_2 of $\{d_\alpha, \alpha \in I\}$, i.e., $\forall \alpha \in I$, $d_\alpha \geq d_2$. Furthermore, $\forall x, y \in X$, $d_\alpha(x,y) \geq d_2(x,y) \to \underline{d}(x,y) = \inf_{\alpha \in I}\{d_\alpha(x,y)\} \geq d_2(x,y) \to d_2 \leq \underline{d}$.

We proved that $\underline{d}$ is the inferior of $\{d_\alpha, \alpha \in I\}$ by (D1) and (D2).

Based on (1), (2), (3) and (4), we complete the proof.

Similar to the above proof, we may also give the proof for $WD(X)$. □

Definition 7.1. Assume $\{d_\alpha, \alpha \in I\} \subseteq D(X)$ (or $\{d_\alpha, \alpha \in I\} \subseteq WD(X)$). Define $\overline{d}$: $\forall x, y \in X$, $\overline{d}(x,y) = \sup_{\alpha \in I}\{d_\alpha(x,y)\}$, then $\overline{d}$ is also called the NEC (pseudo-)metric on X by intersection operation with $\{d_\alpha, \alpha \in I\}$, marked $\overline{d} = \cap_{\alpha \in I} d_\alpha$.

Theorem 7.2. *Suppose X_1 and X_2 are granulations on X, d_1 and d_2 is NEC (pseudo-)metric on X_1 and X_2, respectively. $X_1 \cap X_2 = \{a \cap b | a \in X_1, b \in X_2\}$. Define d : $\forall a, b \in X_1 \cap X_2$,*

$$d(a,b) = \max\{d_1(a_1,b_1), d_2(a_2,b_2)\} \tag{7.1}$$

where $a \subseteq a_i \in X_i$, $b \subseteq b_i \in X_i$, $i = 1, 2$. Then, d is a NEC (pseudo-)metric on $X_1 \cap X_2$.

Proof. When $d_1 \in D(X_1)$ and $d_2 \in D(X_2)$, it is obvious that d satisfies the normalized condition and symmetry property by definition of d.

(1) $\forall a \in X_1 \cap X_2$, $a \subseteq a_i \in X_i$, $i = 1, 2$, $d(a,a) = \max\{d_1(a_1,a_1), d_2(a_2,a_2)\} = 0$, and $\forall a, b \in X_1 \cap X_2, d(a,b) = \max\{d_1(a_1,b_1), d_2(a_2,b_2)\} = 0 \leftrightarrow a_1 = b_1$, $a_2 = b_2 \leftrightarrow a = b$, where $a \subseteq a_i \in X_i$, $b \subseteq b_i \in X_i$, $i = 1, 2$;

(2) $\forall a, b, c \in X_1 \cap X_2$, $a \subseteq a_i \in X_i$, $b \subseteq b_i \in X_i$, $c \subseteq c_i \in X_i$, $i = 1, 2$ we have $\max\{d(a,c), (c,b)\} = \max\{\max\{d_1(a_1,c_1), d_2(a_2,c_2)\}, \max\{d_1(c_1,b_1), d_2(c_2,b_2)\}\} = \max\{\max\{d_1(a_1,c_1), d_1(c_1,b_1)\}, \max\{d_2(a_2,c_2), d_2(c_2,b_2)\}\} \geq \max\{d_1(a_1,b_1), d_2(a_2,b_2)\} = d(a,b)$.

Similarly, we also have $\max\{d(a,b), d(c,b)\} \geq d(a,c)$, $\max\{d(a,b), d(a,c)\} \geq d(c,b)$.

So d satisfies the equicrural condition on $X_1 \cap X_2$. Therefore, d is a NEC metric on $X_1 \cap X_2$.

Similar to the above proof, it also holds when $d_1 \in WD(X_1)$ and $d_2 \in WD(X_2)$. □

Definition 7.2. In Theorem 7.2, d is also called the intersection operation with the NEC (pseudo-)metric d_1 and d_2, marked $d=d_1\cap d_2$.

Remark 9. The intersection operation in Definition 7.1 is different from the one in Definition 7.2. The former is defined on same granulation, and the latter is defined on different granulations of same universe.

Theorem 7.3. *Suppose X_1 and X_2 are two granulations on X, d_1 and d_2 is a NEC (pseudo-)metric on X_1 and X_2, respectively. d_1^* and d_2^* stands for the extending metric derived from d_1 and d_2 on $X_1\cap X_2$, respectively. $\overline{d}=d_1^*\cap d_2^*$, where the intersection operation of $d_1^*\cap d_2^*$ is defined by Definition 7.1. Then $\overline{d}=d_1\cap d_2$, where the intersection operation of $d_1\cap d_2$ is seen in Definition 7.2.*

Proof. From $\forall c_1,c_2\in X_1\cap X_2$,

$$\overline{d}(c_1,c_2)=\max\{d_1^*(c_1,c_2),d_2^*(c_1,c_2)\}=\max\{d_1(a_1,b_1),d_2(a_2,b_2)\}$$
$$=(d_1\cap d_2)(c_1,c_2)$$

therefore, $\overline{d}=d_1\cap d_2$, where $c_1\subseteq a_i\in X_i$, $c_2\subseteq b_i\in X_i$, $i=1,2$. □

Theorem 7.3 shows that the intersection operation in Definition 7.1 is equivalent to the one in Definition 7.2. From Theorems 7.2 and 7.3, and Definition 3.8, we directly obtain the following corollary.

Corollary 7.1. *In Theorem 7.3, $d_i^*\le d$, $i=1,2$, where $d=d_1\cap d_2$.*

Theorem 7.4. *Suppose X_1 and X_2 are two granulations on X, d_1 and d_2 is NEC (pseudo-)metric on X_1 and X_2, respectively, $d=d_1\cap d_2$. Then, $\aleph_{d_i}(X_i)\le\aleph_d(X_1\cap X_2)$, $i=1,2$, where $\aleph_{d_1}(X_1)$, $\aleph_{d_2}(X_2)$ and $\aleph_d(X_1\cap X_2)$ stands for the derived granular space by d_1 (on X_1), d_2 (on X_2) and d (on $X_1\cap X_2$), respectively.*

Proof. By Corollary 7.1, and Theorem 3.10 and 3.11, we may easily give the proof, and omitted it. □

Theorem 7.5. *Suppose X_1 and X_2 are two granulations on X, d_1 and d_2 is NEC (pseudo-)metric on X_1 and X_2, respectively. $d\in D(X_1\cap X_2)$ ($d\in WD(X_1\cap X_2)$). $\aleph_{d_1}(X_1),\aleph_{d_2}(X_2)$ and $\aleph_d(X_1\cap X_2)$ stands for the deriving granular space by d_1 (on X_1), d_2 (on X_2) and d (on $X_1\cap X_2$), respectively. If $\aleph_{d_i}(X_i)\le\aleph_d(X_1\cap X_2)$, $i=1,2$, then $d_1\cap d_2\le d$.*

Proof. The mark d_1^* and d_2^* stands for the extending metrics derived from d_1 and d_2 on $X_1\cap X_2$, respectively. $\aleph_{d_1^*}(X_1\cap X_2)$ and $\aleph_{d_2^*}(X_1\cap X_2)$ is the granular space derived from d_1^* and d_2^* on $X_1\cap X_2$, respectively. By Theorem 3.10, $\aleph_{d_i^*}(X_1\cap X_2)=\aleph_{d_i}(X_i)\le\aleph_d(X_1\cap X_2)$, $i=1,2$. Furthermore, we can derive $d_i^*\le d$, $i=1,2$, from Theorem 3.11. Therefore, $d_1\cap d_2=d_1^*\cap d_2^*\le d$ by Theorem 7.3. □

From Theorems 7.3 and 7.5, we directly get the following corollary.

Corollary 7.2. *In Theorem 7.2, $d=d_1\cap d_2$ is the sup of d_1 and d_2 on $X_1\cap X_2$.*

Theorem 7.5 shows that a new structural cluster (i.e., $\aleph_d(X_1\cap X_2)$) may be obtained according to Theorem 7.2 for given two structural clusters (i.e. $\aleph_{d_1}(X_1)$ and $\aleph_{d_2}(X_2)$). This new structural cluster is expected to satisfy $\aleph_{d_i}(X_i)\le\aleph_d(X_1\cap X_2)$, $i=1,2$, where it shows that the new structural cluster is finer than $\aleph_{d_1}(X_1)$ or $\aleph_{d_2}(X_2)$. By Corollary 7.2, the granular space $\aleph_d(X_1\cap X_2)$ derived from d is the finest obtained from d_1 (on X_1), d_2 (on X_2) on $X_1\cap X_2$, and satisfies $d_1\cap d_2=d_1^*\cap d_2^*$ by Theorem 7.3. We also gave the construction of d by Theorems 7.2 and 7.3, where d is derived through the intersection operation of the extending metric d_1^* and d_2^* on $X_1\cap X_2$, and d_1^* and d_2^* is derived from d_1 and d_2, respectively.

In fact, these results obtained above can be generalized to the fusion problem on an arbitrary number of clusters (or classifications) of universe X by Theorem 7.1, that is, the following theorem also holds.

Theorem 7.6. *Assume $\{X_\alpha,\alpha\in I\}$ is a granulations set on X. d_α is NEC (pseudo-)metric on X_α, d_α^* is the extending metric of d_α on $\cap_{\alpha\in I}X_\alpha$. Then*

(1) $d=\cap_{\alpha\in I}d_\alpha$ is the sup of $\{X_\alpha,\alpha\in I\}$ on $\cap_{\alpha\in I}X_\alpha$, and $\cap_{\alpha\in I}d_\alpha=\cap_{\alpha\in I}d_\alpha*$;

(2) $\aleph_{d_\alpha}(X_\alpha)\le\aleph_d(\cap_{\alpha\in I}X_\alpha)$, $\alpha\in I$, where $\aleph_{d_\alpha}(X_\alpha)$ and $\aleph_d(\cap_{\alpha\in I}X_\alpha)$ stands for the deriving granular space by d_α (on X_α) and d (on $\cap_{\alpha\in I}X_\alpha$), respectively.

Remark 10. For given granulations set $\{X_\alpha,\alpha\in I\}$ on X and their corresponding NEC (pseudo-) metrics, we may get the finest granulation through intersection operation of extending metrics on intersection universe $\cap_{\alpha\in I}X_\alpha$.

The following example serves as an illustration of the above theoretical results.

Example 4. Let $X_1=\{a_1=\{1,2\},a_2=\{3,4\},a_3=\{5\}\}$ and $X_2=\{b_1=\{1,2\},b_2=\{3\},b_3=\{4,5\}\}$ are two granulations on $X=\{1,2,3,4,5\}$; d_1 and d_2 is NEC metric on X_1 and X_2 respectively, where $d_1: d_1(a_i,a_i)=0$, $i=1,2,3$, $d_1(a_1,a_2)=0.3$, $d_1(a_1,a_3)=d_1(a_2,a_3)=0.4$; $d_2: d_2(b_i,b_i)=0$, $i=1,2,3$, $d_2(b_1,b_2)=d_2(b_1,b_3)=0.5$, $d_2(b_2,b_3)=0.2$. The granular spaces of d_1 and d_2 as follows respectively: $\aleph_{d_1}(X_1)=\{X_1(0)=X_1,X_1(0.3)=\{\{a_1,a_2\},\{a_3\}\},X_1(0.4)=\{X_1\}\}$, $\aleph_{d_2}(X_2)=\{X_2(0)=X_2,X_2(0.2)=\{\{b_1\},\{b_2,b_3\}\},X_2(0.5)=\{X_2\}\}$.

By Theorem 7.2, we get $X_1\cap X_2=\{c_1=\{1,2\},c_2=\{3\},\ c_3=\{4\},c_4=\{5\}\}$. Furthermore, we get a NEC metric $d=d_1\cap d_2$ on $X_1\cap X_2$ by the formula (7.1), i.e., $d: d(c_i,c_i)=0$, $i=1,2,3,4$, $d(c_2,c_3)=0.2, d(c_2,c_4)=d(c_3,c_4)=0.4, d(c_1,c_2)=d(c_1,c_3)=\ d(c_1,c_4)=0.5$. Its deriving granular space is $\aleph_d(X_1\cap X_2)=\{X(0)=\ X_1\cap X_2,X(0.2)=\{\{c_1\},\{c_2,c_3\},\{c_4\}\},X(0.4)=\ \{\{c_1\},\{c_2,c_3,c_4\}\},\ X(0.5)=\ \{X_1\cap X_2\}\}$, and the comparison graph of corresponding structural clustering of d_1, d_2 and d in Example 4 is shown in Fig. 5, where a, b and c stand for their structural clustering. From Fig. 5, we also have $\aleph_{d_i}(X_i)\le\aleph_d(X_1\cap X_2)$, $i=1,2$.

7.2. The fusion of structural clusters based on normalized metrics

From Section 5, we known that $\aleph_{Td}(X)$ is an ordered granular space if d is a normalized (pseudo-)metric on X, which is a consistent clustering (or classifying) of X in Section 4. Because the structural clustering problem based on metric space can be transformed into the one based on normalized metric space, we address the fusion problems of structural clusters based on

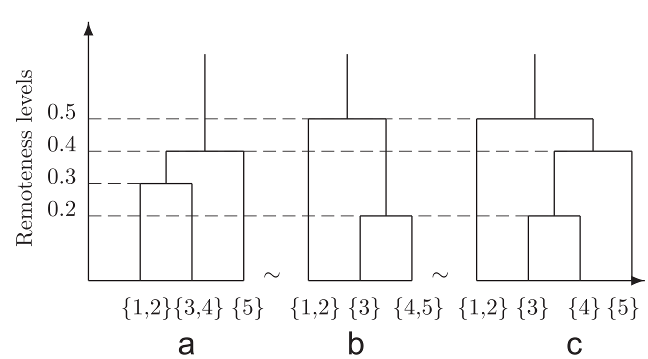

Fig. 5. The comparison graph of corresponding structural clustering of d_1, d_2 and d in Example 4.

granular spaces derived by normalized (pseudo-)metrics in this section.

Theorem 7.7. *Suppose X_1 and X_2 are two granulations on X, d_1 and d_2 is normalized (pseudo-) metric on X_1 and X_2, respectively, and $X_1 \cap X_2 = \{a \cap b | a \in X_1, b \in X_2\}$. Define d:*

$\forall a, b \in X_1 \cap X_2, d(a,b) = \max\{d_1(a_1,b_1), d_2(a_2,b_2)\}$

where $a \subseteq a_i \in X_i$, $b \subseteq b_i \in X_i$, $i=1,2$. Then, d is a normalized (pseudo-)metric on $X_1 \cap X_2$.

If the result is hold, then d is also called that the normalized (pseudo-)metric obtained from d_1 and d_2 by intersection operation, noted as $d = d_1 \cap d_2$.

Proof. When $d_1 \in ND(X_1)$ and $d_2 \in ND(X_2)$, it is obvious that d satisfies the normalized condition and symmetry property by definition of d, and

(1) $\forall a \in X_1 \cap X_2$, $a \subseteq a_i \in X_i$, $i=1,2, d(a,a)=0$, and $\forall a, b \in X_1 \cap X_2$, $d(a,b) = \max\{d_1(a_1,b_1), d_2(a_2,b_2)\} = 0 \leftrightarrow a_1 = b_1, a_2 = b_2 \leftrightarrow a = b$, where $a \subseteq a_i \in X_i$, $b \subseteq b_i \in X_i$, $i=1,2$. The latter derivation is based on $a = a_1 \cap a_2 = b_1 \cap b_2 = b$;

(2) $\forall a,b,c \in X_1 \cap X_2$, $a \subseteq a_i \in X_i$, $b \subseteq b_i \in X_i$, $c \subseteq c_i \in X_i$, $i=1,2$, we have

$$\begin{aligned} d(a,b) &= \max\{d_1(a_1,b_1), d_2(a_2,b_2)\} \\ &\le \max\{d_1(a_1,c_1)+d_1(c_1,b_1), d_2(a_2,c_2)+d_2(c_2,b_2)\} \\ &\le \max\{d_1(a_1,c_1), d_2(a_2,c_2)\} + \max\{d_1(c_1,b_1), d_2(c_2,b_2)\} \\ &= d(a,c)+d(c,b) \end{aligned}$$

where d satisfies the triangle inequality. Therefore, $d \in ND(X_1 \cap X_2)$.

Similar to the above proof, it also holds when $d_1 \in WND(X_1)$ and $d_2 \in WND(X_2)$. □

Theorem 7.8. *Suppose X_1 and X_2 are two granulations on X, d_1 and d_2 is normalized (pseudo-) metric on X_1 and X_2, respectively. $d = d_1 \cap d_2$. $d_1^* = t(d_1)$, $d_2^* = t(d_2)$ and $d^* = t(d)$. $X_1^*(\lambda)$, $X_2^*(\lambda)$ and $X^*(\lambda)$ stands for the granulation derived from d_1 (on X_1), d_2 (on X_2) and d (on $X_1 \cap X_2$), respectively ($\lambda \in [0,1]$). Then*

(1) $\forall \lambda \in [0,1]$, $X_i^*(\lambda) \le X^*(\lambda)$, $i=1,2$;
(2) $d_1^* \cap d_2^* \le d^*$ *or* $t(d_1) \cap t(d_2) \le t(d_1 \cap d_2)$.

Proof. (1) $\forall \lambda \in [0,1]$, $[x]_{1\lambda}$, $[x]_{2\lambda}$ and $[x]_\lambda$ stands for the equivalent class of x corresponding to λ derived from d_1^*, d_2^* and d^*, respectively. Based on Definition 5.1, $D_{1\lambda}$, $D_{2\lambda}$ and D_λ stands for the cut equivalence relation derived from d_1, d_2 and d corresponding to λ, respectively. $\forall x \in X$, marked as $a = [x]_\lambda \in X(\lambda)$. $\forall y \in a$, there exists $x = x_1, x_2, \ldots, x_m = y$ and $x_i \in c_i \in X_1 \cap X_2$, $i = 1,2,\ldots,m$ such that $d(x_i, x_{i+1}) \le d(c_i, c_{i+1}) \le \lambda$, $i=1,2,\ldots,m-1$. By $d = d_1 \cap d_2$, marked $x_i \in c_i \subseteq a_i \in X_1$, $i=1,2,\ldots,m$, then $d_1(x_i,x_{i+1}) \le d_1(a_i,a_{i+1}) \le \lambda$, $i=1,2,\ldots,m-1$ for given $x = x_1, x_2, \ldots, x_m = y$, i.e., $y \in [x]_{1\lambda}$, so $[x]_\lambda \subseteq [x]_{1\lambda}$. Therefore, we have $\forall \lambda \in [0,1]$, $X_1^*(\lambda) \le X^*(\lambda)$. Similarly, we can prove $\forall \lambda \in [0,1]$, $X_2^*(\lambda) \le X^*(\lambda)$.

(2) We can directly get the proof of (2) from (1) and Theorem 7.5. Details are omitted here. □

For given two granulations and corresponding normalized (pseudo-)metrics d_1 and d_2 on them, we can construct a normalized (pseudo-)metric $d = d_1 \cap d_2$ according to Theorem 7.7, using the intersection operation, and give its construction methods. Furthermore, we may obtain the corresponding structural cluster of d using the results reported in Section 5, which is finer than ones of d_1 and d_2 according to Theorem 7.8. We can also define the optimal cluster of d using the research method described in Section 6. Our proposed method of studying structural cluster fusion by intersection operation is superior over the ones in Refs. [32,33,67,68].

Example 5. Assume $X_1 = \{a_1 = \{1,2\}, a_2 = \{3,4\}, a_3 = \{5\}\}$ and $X_2 = \{b_1 = \{1,2\}, b_2 = \{3\}, b_3 = \{4,5\}\}$ are two granulations on $X = \{1,2,3,4,5\}$, d_1 and d_2 is normalized metric on X_1 and X_2 respectively, where $d_1 : d_1(a_i,a_i) = 0$, $i=1,2,3$, $d_1(a_1,a_2) = 0.3$, $d_1(a_1,a_3) = 0.4$, $d_1(a_2,a_3) = 0.5$; $d_2 : d_2(b_i,b_i) = 0$, $i=1,2,3$, $d_2(b_1,b_2) = 0.4$, $d_2(b_1,b_3) = 0.5$, $d_2(b_2,b_3) = 0.7$.

The granular spaces and NEC metrics deriving from d_1 and d_2 respectively are as follow: $\aleph_{Td_1}(X_1) = \{X_1(0) = X_1, X_1(0.3) = \{\{a_1,a_2\},\{a_3\}\}, X_1(0.4) = \{X_1\}\}$, $d_1^* : d_1^*(a_i,a_i) = 0$, $i=1,2,3$, $d_1^*(a_1,a_2) = 0.3$, $d_1^*(a_1,a_3) = d_1^*(a_2,a_3) = 0.4$; $\aleph_{Td_2}(X_2) = \{X_2(0) = X_2, X_2(0.4) = \{\{b_1,b_2\},\{b_3\}, X_2(0.5) = \{X_2\}\}$, $d_2^* : d_2^*(b_i,b_i) = 0$, $i=1,2,3$, $d_2^*(b_1,b_2) = 0.4$, $d_2^*(b_1,b_3) = d_2^*(b_2,b_3) = 0.5$.

By $X_1 \cap X_2 = \{c_1 = \{1,2\}, c_2 = \{3\}, c_3 = \{4\}, c_4 = \{5\}\}$. Furthermore, we get a normalized metric $d = d_1 \cap d_2$ on $X_1 \cap X_2$ according to Theorem 7.7, that is, $d : d(c_i,c_i) = 0$, $i=1,\ldots,4$, $d(c_1,c_2) = 0.4$, $d(c_1,c_3) = d(c_1,c_4) = d(c_3,c_4) = 0.5$, $d(c_2,c_3) = d(c_2,c_4) = 0.7$. Its derived granular space and NEC metric are $\aleph_{Td}(X_1 \cap X_2) = \{X(0) = X_1 \cap X_2, X(0.4) = \{\{c_1,c_2\}\{c_3\},\{c_4\}\}, X(0.5) = \{X_1 \cap X_2\}$, and $d^* : d^*(c_i,c_i) = 0$, $i=1,\ldots,4$, $d^*(c_1,c_2) = 0.4$, $d^*(c_1,c_3) = d^*(c_1,c_4) = d^*(c_2,c_3) = d^*(c_2,c_4) = d^*(c_3,c_4) = 0.5$. Note that $\overline{d} = d_1^* \cap d_2^* : \overline{d}(c_i,c_i) = 0$, $i=1,\ldots,4$, $\overline{d}(c_1,c_2) = \overline{d}(c_3,c_4) = 0.4$, $\overline{d}(c_1,c_3) = \overline{d}(c_1,c_4) = \overline{d}(c_2,c_3) = \overline{d}(c_2,c_4) = 0.5$, it is quite straightforward to conclude that $\forall \lambda \in [0,1]$, $X_i(\lambda) \le X(\lambda)$ $(i=1,2)$ from the above results, i.e., $\aleph_{Td_i}(X_i) \le \aleph_{Td}(X_1 \cap X_2)$, $i=1,2$, and $t(d_1) \cap t(d_2) = d_1^* \cap d_2^* < t(d_1 \cap d_2)$. □

Example 5 suggests that the result $t(d_1) \cap t(d_2) = t(d_1 \cap d_2)$ in (2) of Theorem 7.8 does not hold in general, but the result $t(d_1) \cap t(d_2) \le t(d_1 \cap d_2)$ in (2) of Theorem 7.8 is sufficient to indicate the regularity of structural cluster fusion based on normalized (pseudo-)metric under the transitive closure operation.

8. The clustering structure analysis based on metric

In Refs. [17,19], Zhang and Zhang proposed the structural analysis of fuzzy sets based on fuzzy quotient space, and gave the isomorphism principle and similarity principle of fuzzy sets to interpret that people can obtain the same conclusion from different membership functions on a fuzzy concept. In Ref. [69], DÖing et al. presented the data analysis with fuzzy clustering. In Ref. [70], we also presented the research on clustering analysis based on fuzzy granular space, and gave clustering structural analysis of fuzzy equivalence relation based on fuzzy granular space. We have pointed out "the structural clustering problem based on metric space can be transformed into the one based on normalized metric space" earlier in Section 5. In this section, we give the clustering structure analysis theory based on metric by leveraging corresponding basic concepts (i.e., isomorphism and ε-similarity of two fuzzy equivalence relations) and findings reported in Ref. [19] into (pseudo-)metric space, and organize it as follows. In Section 8.1, the clustering structure analysis based on NEC (pseudo-)metric is given. In Section 8.2, the clustering structural analysis theory based on normalized (pseudo-)metric is get.

8.1. The clustering structure analysis based on NEC metric

Definition 8.1. Assume $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$), $\aleph_{d_1}(X)$ and $\aleph_{d_2}(X)$ stands for the derived granular space from d_1 and d_2, respectively. If $\aleph_{d_1}(X) = \aleph_{d_2}(X)$, or $\forall \lambda \in [0,1]$, there exists $\mu \in [0,1]$

such that $X_2(\mu)=X_1(\lambda)$, and *vice versa*, then d_1 and d_2 are called isomorphic on clustering (or classifying) structures. d_1 and d_2 are also simply called as isomorphic, marked $d_1 \cong d_2$.

Theorem 8.1. (*Isomorphism Discrimination Theorem I*). *Suppose* $d_1,d_2 \in D(X)$ (*or* $d_1,d_2 \in WD(X)$), *then* $d_1 \cong d_2 \Leftrightarrow \forall x,y,u,v \in X$, $d_1(x,y) \le d_1(u,v) \leftrightarrow d_2(x,y) \le d_2(u,v)$.

Proof. Marked the granular space derived from d_i as $\aleph_{d_i}(X)=\{X_i(\lambda)|\lambda \in [0,1]\}=\{X_i(\lambda)|\lambda \in D_i\}$, where $X_i(\lambda)=\{[x]_{i\lambda}|x \in X\}$, $D_i=\{d_i(x,y)|x,y \in X\}$, $i=1,2$. We will only give the proof for $d_1,d_2 \in D(X)$ as follows, but it also applicable to $d_1,d_2 \in WD(X)$.

"$\Rightarrow$". $\forall x,y,u,v \in X$, $d_1(x,y) \le d_1(u,v)$.

(1) When $d_1(x,y)=\lambda < d_1(u,v)$, then $v \notin [u]_{1\lambda}$, $y \in [x]_{1\lambda}$. From $\aleph_{d_1}(X)=\aleph_{d_2}(X)$, there exists $\mu \in [0,1]$ such that $X_2(\mu)=X_1(\lambda)$, hence, $v \notin [u]_{2\mu}$, $y \in [x]_{2\mu} \to d_2(u,v) > \mu \ge d_2(x,y)$. So $d_1(x,y) < d_1(u,v) \to d_2(x,y) < d_2(u,v)$.

Similarly, we can prove $d_2(x,y) < d_2(u,v) \to d_1(x,y) < d_1(u,v)$.

(2) When $d_1(x,y)=d_1(u,v))$, then $d_2(x,y)=d_2(u,v)$. Otherwise, letting $d_2(x,y) < d_2(u,v)$, we get $d_1(x,y) < d_1(u,v)$ by the above proof, which contradicts with $d_1(x,y)=d_1(u,v)$. Therefore, $d_1(x,y)=d_1(u,v) \to d_2(x,y)=d_2(u,v)$.

Similarly, we can also prove $d_2(x,y)=d_2(u,v) \to d_1(x,y)=d_1(u,v)$.

From (1) and (2), $\forall x,y,u,v \in X$, $d_1(x,y) \le d_1(u,v) \leftrightarrow d_2(x,y) \le d_2(u,v)$.

"$\Leftarrow$" For any $\forall \lambda \in D_1$, there exists u, $v \in X$ such that $d_1(u,v)=\lambda$. Letting $\mu=d_2(u,v) \in D_2$. $\forall a \in X_1(\lambda)$, $x \in a=[x]_{1\lambda}$, then $\forall y \in a \leftrightarrow d_1(x,y) \le \lambda = d_1(u,v) \leftrightarrow d_2(x,y) \le d_2(u,v)=\mu \leftrightarrow y \in [x]_{2\mu}$, i.e., $a \in X_2(\mu)$, so $\aleph_{d_1}(X)=\aleph_{d_2}(X)$. Therefore $d_1 \cong d_2$. □

In Theorem 8.1, the isomorphism discrimination condition of two NEC (pseudo-)metrics is difficult to verify, particularly when X is a continuous field. In order to simplify it, we give the isomorphism discrimination theorem as follows.

Theorem 8.2. (*Isomorphism Discrimination Theorem II*). *Suppose* $d_1,d_2 \in D(X)$ (*or* $d_1,d_2 \in WD(X)$), *mark* $D_i=\{d_i(x,y)|x,y \in X\}$, $i=1,2$. *Then* $d_1 \cong d_2 \Leftrightarrow$ *there exists an one-to-one and strictly increasing mapping* $f: D_1 \to D_2$ *such that* $\forall x,y \in X$, $d_2(x,y)=f(d_1(x,y))$, *where* $f(0)=0$.

Proof. Mark the granular space deriving from d_i as $\aleph_{d_i}(X)=\{X_i(\lambda)|\lambda \in D_i\}$, where $X_i(\lambda)=\{[x]_{i\lambda}|x \in X\}$, $i=1,2$. We only give the proof for $d_1,d_2 \in D(X)$ as follows, but it also holds when $d_1,d_2 \in WD(X)$.

"$\Rightarrow$". $\forall \lambda \in D_1$, there exists $x,y \in X$ such that $d_1(x,y)=\lambda$, i.e., $X_1(\lambda) \in \aleph_{d_1}(X)$ and $y \in [x]_{1\lambda}$. By $\aleph_{d_1}(X)=\aleph_{d_2}(X)$, there exists $\mu \in D_2$ such that $X_2(\mu)=X_1(\lambda)$, then $y \in [x]_{2\mu}$, i.e., $d_2(x,y) \le \mu$. We prove $d_2(x,y)=\mu$ next. Otherwise, letting $d_2(x,y) < \mu$, there certainly exists u, $v \in X$ such that $d_2(u,v)=\mu$ from $\mu \in D_2$, i.e., $d_2(x,y) < d_2(u,v)$. By Theorem 8.1, we get $d_1(x,y)=\lambda < d_1(u,v) \to v \notin [u]_{1\lambda}$. However, $d_2(u,v)=\mu \to v \in [u]_{2\mu}$, which contradicts with $X_2(\mu)=X_1(\lambda)$.

Therefore, $\forall \lambda \in D_1$, there exists unique $\mu \in D_2$ that corresponds with λ. Similarly, we can prove: $\forall \mu \in D_2$, there exists unique $\lambda \in D_1$ that corresponds with μ, i.e., there exists an one-to-one mapping $f: D_1 \to D_2$ such that $\forall x,y \in X$, $d_2(x,y)=f(d_1(x,y))$ satisfying $f(0)=0$.

Next, we prove that f is monotonic increasing. $\forall \lambda_1,\lambda_2 \in D_1, \lambda_1 < \lambda_2$, μ_1 and μ_2 corresponds with λ_1 and λ_2 under the mapping f, respectively, then $\lambda_1 < \lambda_2 \to X_1(\lambda_2) < X_1(\lambda_1) \to X_2(\mu_2)=X_1(\lambda_2) < X_1(\lambda_1)=X_2(\mu_1) \to f(\lambda_1)=\mu_1 < \mu_2 = f(\lambda_2)$, hence, f is a monotonic increasing mapping.

"$\Leftarrow$". $\forall a=[x]_{1\lambda} \in X_1(\lambda) \in \aleph_{d_1}(X)$, then $\forall y \in a$, $d_1(x,y) \le \lambda \leftrightarrow d_2(x,y)=f(d_1(x,y)) \le f(\lambda) \leftrightarrow y \in [x]_{2f(\lambda)}$, hence $a \in X_2(f(\lambda)) \in \aleph_{d_2}(X)$, i.e., $\aleph_{d_1}(X) \subseteq \aleph_{d_2}(X)$. Similarly, we can prove that $\aleph_{d_2}(X) \subseteq \aleph_{d_1}(X)$. Therefore, $\aleph_{d_1}(X)=\aleph_{d_2}(X)$, that is, $d_1 \cong d_2$. □

Theorems 8.1 and 8.2 show that structural clusters of two NEC (pseudo-)metrics are the same as long as their distances order relation between any two points of universe is kept. Furthermore, results inferred from these structures are also the same. For convenience of applications, we transform Theorem 8.2 into the following theorem.

Theorem 8.3. *Let* $d_1,d_2 \in D(X)$ (*or* $d_1,d_2 \in WD(X)$), *then* $d_1 \cong d_2 \Leftrightarrow$ *there exists an one-to-one and strictly increasing mapping* $f: [0,1] \to [0,1]$ *such that* $\forall x,y \in X$, $d_2(x,y)=f(d_1(x,y))$, *where* $f(0)=0$.

Essentially, Theorem 8.3 extends the mapping $f: D_1 \to D_2$ in Theorem 8.2 into $f:[0,1] \to [0,1]$, and makes it easier to handle the continuous universe. From Definition 8.1, two NEC (pseudo-)metrics are isomorphic on clustering (or classifying) structure if their ordered granular spaces are same. But, the isomorphic condition of two NEC (pseudo-)metrics is too strong. If their ordered granular spaces are different and results inferred from these structures are similar, then how shall we describe it when solving actual problems? Next, we will address this question.

Definition 8.2. Let $d_1,d_2 \in D(X)$ (or $d_1,d_2 \in WD(X)$) and $\varepsilon > 0$. If there exists $d_3 \in D(X)$ (or $d_3 \in WD(X)$) satisfying $d_3 \cong d_1$ and $\forall x,y \in X$, $|d_2(x,y)-d_3(x,y)| \le \varepsilon$ (or $d_3 \cong d_2$ and $\forall x,y \in X$, $|d_1(x,y)-d_3(x,y)| \le \varepsilon$), then d_1 and d_2 are called ε-similarity on clustering (or classifying) structures, short for d_1 and d_2 are ε-similarity, marked as $d_1 \sim d_2(\varepsilon)$.

Theorem 8.4. (ε-*Similarity Discrimination Theorem I*). *Suppose* $d_1,d_2 \in D(X)$ (*or* $d_1,d_2 \in WD(X)$), *the corresponding ordered granular space is* $\aleph_{d_i}(X)=\{X_i(\lambda)|\lambda \in [0,1]\}$, $i=1,2$. *Then* $d_1 \sim d_2(\varepsilon) \Leftrightarrow \forall \lambda \in [0,1]$, *there exists* $\mu \in [0,1]$ *such that* $X_2(\mu+\varepsilon) \le X_1(\lambda) \le X_2(\mu-\varepsilon)$, *or* $\forall \mu \in [0,1]$, *there exists* $\lambda \in [0,1]$ *such that* $X_1(\lambda+\varepsilon) \le X_2(\mu) \le X_1(\lambda-\varepsilon)$.

Proof. We only give the proof for $d_1,d_2 \in D(X)$ as follows, but it also applicable to $d_1,d_2 \in WD(X)$.

"$\Rightarrow$". If there exists $d_3 \in D(X)$ such that

$$d_3 \cong d_1 \quad \text{and} \quad \forall x,y \in X, |d_2(x,y)-d_3(x,y)| \le \varepsilon \tag{8.1}$$

Mark the ordered granular space derived from d_3 as $\aleph_{d_3}(X)=\{X_3(\lambda)|\lambda \in [0,1]\}$. $\forall \lambda \in [0,1]$, $R_{i,\lambda}=\{(x,y)|d_i(x,y) \le \lambda\}$, $i=1,2,3$, where $R_{i,\lambda}$ is a crisp equivalence relation on X by Theorem 3.1, and it is obvious that $X_i(\lambda)$ is the set of equivalent classes of $R_{i,\lambda}$ on X, $i=1,2,3$.

From $d_3 \cong d_1$, $\forall \lambda \in [0,1]$, there exists $\mu \in [0,1]$ such that $X_3(\mu)=X_1(\lambda)$. From the condition (8.1), $\forall x,y \in X$, $d_2(x,y)-\varepsilon \le d_3(x,y) \le d_2(x,y)+\varepsilon$, and $R_{2,\mu-\varepsilon} \subseteq R_{3,\mu} \subseteq R_{2,\mu+\varepsilon}$. So, $\forall \lambda \in [0,1]$, there exists $\mu \in [0,1]$ such that $X_2(\mu+\varepsilon) \le X_1(\lambda)=X_3(\mu) \le X_2(\mu-\varepsilon)$.

Similarly, we can prove the other half.

"$\Leftarrow$". If $\forall \lambda \in [0,1]$, there exists $\mu \in [0,1]$ such that $X_2(\mu+\varepsilon) \le X_1(\lambda) \le X_2(\mu-\varepsilon)$. Mark $X_3(\mu)=X_1(\lambda)$, so the granular space $\{X_3(\mu)|\mu \in [0,1]\}$ is an ordered set, the corresponding NEC metric is marked as d_3. According to Theorem 8.2, d_1 and d_3 are isomorphic obviously. And $\forall x,y \in X$, let $d_1(x,y)=\lambda \in [0,1]$, there exists $\mu \in [0,1]$ such that $X_2(\mu+\varepsilon) \le X_1(\lambda)=X_3(\mu) \le X_2(\mu-\varepsilon)$, i.e., $R_{2,\mu-\varepsilon} \subseteq R_{3,\mu} \subseteq R_{2,\mu+\varepsilon}$, then, $|d_2(x,y)-d_3(x,y)| \le \varepsilon$. By Definition 8.2, $d_1 \sim d_2(\varepsilon)$. □

From this theorem, ε-similarity of two NEC (pseudo-)metrics means that there is a restrictive relation among their granulations that any granulation of the one is always restricted by two granulations of the other. But the discrimination condition of ε-similarity in Theorem 8.4 is difficult to verify. In order to simplify it, we give the discrimination theorem as follows.

Theorem 8.5. (*ε-Similarity Discrimination Theorem II*). *Suppose $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$), then $d_1 \sim d_2(\varepsilon) \Leftrightarrow$ there exists an one-to-one and strictly increasing mapping $f:[0,1] \to [0,1]$ satisfying $f(0)=0$ such that $\forall x, y \in X$, $|f(d_1(x,y)) - d_2(x,y)| \le \varepsilon$; or there exists an one-to-one and strictly increasing mapping $g:[0,1]\to[0,1]$ satisfying $g(0)=0$ such that $\forall x, y \in X$, $|g(d_2(x,y)) - d_1(x,y)| \le \varepsilon$.*

Proof. The proof can be directly achieved from Theorem 8.3 and Definition 8.2, details omitted here. □

We shall give the following example to illustrate the effectiveness of Theorem 8.5.

Example 6. Let $X=\{1,2,3,4,5\}$. d_1 and d_2 are two NEC (pseudo-) metrics on X, where $d_1: d_1(i,i)=0$, $i=1,\dots,5$, $d_1(1,2)=0.2$, $d_1(3,4)=0.4$, $d_1(1,3)=d_1(1,4)=d_1(2,3)=d_1(2,4)=0.5$, $d_1(1,5)=$ $d_1(2,5)=d_1(3,5)=d_1(4,5)=0.7$; $d_2: d_2(i,i)=0$, $i=1,\dots,5$, $d_2(1,2)=0.3, d_2(1,3)=d_2(1,4)=$ $d_2(2,3)=d_2(2,4)=$ $d_2(3,4)=0.4, d_2(1,5)=d_2(2,5)=d_2(3,5)=$ $d_2(4,5)=0.5$. The granular space deriving from d_1 and d_2 are as follow: $\aleph_{d_1}(X)=\{X_1(0)=X, X_1(0.2)=\{\{1,2\},\{3\},\{4\},\{5\}\}, X_1(0.4)=\{\{1,2\},\{3,4\},\{5\}\}, X_1(0.5)=\{\{1,2,3,4\},\{5\}\}, X_1(0.7)=\{X\}\}$; $\aleph_{d_2}(X)=\{X_2(0)=X, X_2(0.3)=\{\{1,2\},\{3\},\{4\},\{5\}\}, X_2(0.4)=\{\{1,2,3,4\},\{5\}\}, X_1(0.5)=\{X\}\}$.

Because $D_1=\{0,0.2,0.4,0.5,0.7\}$ and $D_2=\{0,0.3,0.4,0.5\}$, there does not exist any one-to-one and strictly increasing mapping on $D_1 \to D_2$, that is, d_1 and d_2 are not isomorphic. But there exists an one-to-one and strictly increasing mapping $g:[0,1]\to[0,1]$ satisfying $g(0)=0$ and $0.3\to0.3$, $0.4\to0.5, 0.5\to0.7$ (Note: the former of mapping g belongs to D_2) such that $\forall x,y \in X$, $|g(d_2(x,y))-d_1(x,y)| \le 0.1$, i.e., $d_1 \sim d_2(0.1)$.

In Example 6, we can see that the mapping g is an one-to-one and strictly increasing mapping from [0,1] to [0,1] although there does not exist any one-to-one correspondence between D_1 and D_2. It is clear that Theorem 8.3 plays an important role in reasoning.

As we pointed out previously ε-similarity of two NEC (pseudo-)metrics means that there is a restrictive relation among their granulations that any granulation of the one is always restricted by two granulations of the other. However, this restrictive relationship is not mutual. Next, we introduce a ε-similarity of two NEC (pseudo-)metrics with mutual restrictive relation, so-called strong ε-similarity.

Definition 8.3. Let $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$) and $\varepsilon>0$, the granular space derived from d_i is marked as $\aleph_{d_i}(X)=\{X_i(\lambda)|\lambda\in[0,1]\}$, $i=1,2$, and satisfies

(1) $\forall \lambda \in [0,1]$, there exists $\mu \in [0,1]$ such that $X_2(\mu+\varepsilon) \le X_1(\lambda) \le X_2(\mu-\varepsilon)$;
(2) $\forall \mu \in [0,1]$, there exists $\lambda \in [0,1]$ such that $X_1(\lambda+\varepsilon) \le X_2(\mu) \le X_1(\lambda-\varepsilon)$.

Then d_1 and d_2 are called strong ε-similarity on clustering (or classifying) structures, marked $d_1 \approx d_2(\varepsilon)$.

From Theorems 8.4 and 8.5, and Definition 8.3, it is easy to obtain the following theorem.

Theorem 8.6. (*The Strong ε-Similarity Discrimination Theorem*). *Suppose $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$), then $d_1 \approx d_2(\varepsilon) \Leftrightarrow$ there exists one-to-one and strictly increasing mapping f and g on $[0,1]\to[0,1]$ satisfying $f(0)=g(0)=0$ such that $\forall x, y \in X$*

$|f(d_1(x,y)) - d_2(x,y)| \le \varepsilon$ and $|g(d_2(x,y)) - d_1(x,y)| \le \varepsilon$.

Theorem 8.7. *Suppose $d_1, d_2 \in D(X)$ (or $d_1, d_2 \in WD(X)$), then $d_1 \cong d_2 \Leftrightarrow \forall \varepsilon>0$, $d_1 \approx d_2(\varepsilon)$.*

Proof. "$\Rightarrow$". The proof is obtained directly from Theorems 8.3 and 8.6.

"$\Leftarrow$". $\forall \varepsilon_n > 0$, $\lim_{n\to\infty}\varepsilon_n = 0$. By $d_1 \approx d_2(\varepsilon_n)$ and Theorem 8.6, there exists an one-to-one and strictly increasing mapping f: $[0,1]\to[0,1]$ satisfying $f(0)=0$ such that

$$\forall x, y \in X, |f(d_1(x,y)) - d_2(x,y)| \le \varepsilon_n \quad (8.2)$$

In (8.2), letting $n\to\infty$, we have $\forall x,y\in X$, $d_2(x,y)=f(d_1(x,y))$. Therefore, $d_1 \cong d_2$ by Theorem 8.3. □

Theorem 8.7 shows the relationship between the isomorphism and strong ε-Similarity for two NEC (pseudo-)metrics on clustering (or classifying) structures.

8.2. The clustering structure analysis based on normalized metric

In Section 8.1, we gave the research on clustering structure analysis based on normalized (pseudo-) metric, introduced some basic concepts that isomorphism, ε-Similarity and strong ε-Similarity for two NEC (pseudo-)metrics on clustering (or classifying) structures, and gave their discrimination theorem. In Section 5, we researched the theory and algorithm of structural cluster based on normalized (pseudo-)metric. In this section, we establish the clustering structure analysis based on normalized (pseudo-)metric.

Definition 8.4. Assume $d_1, d_2 \in ND(X)$ (or $d_1, d_2 \in WND(X)$), $\aleph_{Td_1}(X)$ and $\aleph_{Td_2}(X)$ stands for the derived granular space from d_1 and d_2, respectively. If $\aleph_{Td_1}(X)=\aleph_{Td_2}(X)$, then, it is defined that d_1 and d_2 are isomorphic on clustering (or classifying) structures, or simply that d_1 and d_2 are isomorphic, marked as $d_1 \cong d_2(T)$.

The following example is used to illustrate the difference between normalized (pseudo-)metric and NEC (pseudo-)metric space on isomorphism.

Example 7. Let $X=\{1,2,3,4\}$. d_1 and d_2 are two normalized metrics on X, where $d_1: d_1(i,i)=0$, $i=1,\dots,4$, $d_1(1,2)=0.2$, $d_1(1,4)=d_1(2,3)=0.3$, $d_1(1,3)=0.5$, $d_1(2,4)=d_1(3,4)=0.4$; $d_2: d_2(i,i)=0$, $i=1,\dots,4$, $d_2(1,2)=0.2$, $d_2(1,3)=d_2(1,4)$ $=d_2(2,3)=d_2(3,4)=0.4$, $d_2(2,4)=0.5$. The granular space derived from d_1 and d_2 are as follow: $\aleph_{Td_1}(X)=\{X_1(0)=X, X_1(0.2)=\{\{1,2\},\{3\},\{4\}\}, X_1(0.3)=\{X\}\}$; $\aleph_{Td_2}(X)=\{X_2(0)=X, X_2(0.2)=\{\{1,2\},\{3\},\{4\}\}, X_2(0.4)=\{X\}\}$.

Because $\aleph_{Td_1}(X)=\aleph_{Td_2}(X)$, then $d_1 \cong d_2(T)$ by Definition 8.4. But there exist $d_1(2,4)=0.4<0.5=d_1(1,3)$ and $d_2(2,4)=0.5>0.4=d_2(1,3)$.

Example 7 shows that they do not have those discrimination theorems such as Theorems 8.1–8.3 if two normalized (pseudo-)metric are isomorphic, but the sufficient condition that two normalized (pseudo-)metric are isomorphic in the following is hold.

Theorem 8.8. (*Isomorphic Sufficient Theorem I*). *Suppose $d_1, d_2 \in ND(X)$ (or $d_1, d_2 \in WND(X)$). If $\forall x, y, u, v \in X$, $d_1(x,y) \le d_1(u,v) \leftrightarrow d_2(x,y) \le d_2(u,v)$, then $d_1 \cong d_2(T)$.*

Proof. We only give the proof for $d_1, d_2 \in ND(X)$ as follows, but it also holds when $d_1, d_2 \in WND(X)$. Mark the granular space derived from d_i as $\aleph_{Td_i}(X) = \{X_i(\lambda)|\lambda \in [0,1]\} = \{X_i(\lambda)|\lambda \in D_i\}$, and $\forall \lambda \in [0,1]$, $B_{i,\lambda} = \{(x,y)|d_i(x,y) \le \lambda\}$, $D_{i,\lambda} = t(B_{i,\lambda})$, where $X_i(\lambda)$ is the equivalent class of $D_{i,\lambda}$, and $D_i = \{d_i(x,y)|x,y \in X\}$ $(i=1,2)$.

According to Remark 7, in order to prove $d_1 \cong d_2(T)$, we only need to prove that the granulation derived from d_1 (on D_1) is identical to the one derived from d_2 (on D_2).

$\forall \lambda \in D_1$, there exists x_0, $y_0 \in X$ such that $d_1(x_0,y_0) = \lambda$, marked $d_2(x_0,y_0) = \mu \in D_2$. So, $\forall (x,y) \in B_{1,\lambda}$, $d_1(x,y) \le \lambda = d_1(x_0,y_0)$. We have $d_2(x,y) \le d_2(x_0,y_0) = \mu$ with the condition given in this theorem, i.e., $(x,y) \in B_{2,\mu}$. Then, $B_{1,\lambda} \subseteq B_{2,\mu}$.

Similar to the above proof, we have $B_{2,\mu} \subseteq B_{1,\lambda}$. So, $B_{1,\lambda} = B_{2,\mu}$. Furthermore, $D_{1,\lambda} = t(B_{1,\lambda}) = t(B_{2,\mu}) = D_{2,\mu}$, that is, $X_2(\mu) = X_1(\lambda)$. Therefore, $\forall \lambda \in D_1$, there exists $\mu \in D_2$ such that $X_2(\mu) = X_1(\lambda)$.

Similarly, we also can prove that $\forall \mu \in D_2$, there exists $\lambda \in D_1$ such that $X_1(\lambda) = X_2(\mu)$. Therefore, $\aleph_{Td_1}(X) = \aleph_{Td_2}(X)$, that is $d_1 \cong d_2(T)$. □

Remark 11. The inverse of Theorem 8.8 does not hold according to Remark 6.

Theorem 8.8 shows that two normalized (pseudo-)metrics are isomorphic as long as their distances order relationship between any two points of universe is kept, and Example 7 also shows that the inverse of Theorem 8.8 does not hold. But the isomorphic sufficient condition for two normalized (pseudo-)metrics in Theorem 8.8 is difficult to verify, particularly when X is a continuous field. In order to simplify it, we improve Theorem 8.8 as follows.

Theorem 8.9. (*Isomorphic Sufficient Theorem II*). *Suppose* $d_1, d_2 \in ND(X)$ (*or* $d_1, d_2 \in WND(X)$). *Let* $D_i = \{d_i(x,y)|x,y \in X\}$, $i=1,2$. *If there exists an one-to-one and strictly increasing mapping* f *satisfying* $f(0)=0$ *on* D_1 *to* D_2 (*or* [0,1] *to* [0,1]) *such that* $\forall x,y \in X$, $d_2(x,y) = f(d_1(x,y))$, *then* $d_1 \cong d_2(T)$.

Proof. We only give the proof for $d_1, d_2 \in ND(X)$ as follows, but it also applicable for $d_1, d_2 \in WND(X)$. Some marks are seen in the proof of Theorem 8.8.

$\forall \lambda \in D_1$, there exists x_0, $y_0 \in X$ such that $d_1(x_0,y_0) = \lambda$, marked $\mu = f(\lambda) \in D_2$. So, $\forall (x,y) \in B_{1\lambda}$, $d_1(x,y) \le \lambda = d_1(x_0,y_0)$. We have $d_2(x,y) = f(d_1(x,y)) \le f(d_1(x_0,y_0)) = f(\lambda) = \mu$ by the condition of satisfying f, i.e., $(x,y) \in B_{2,\mu}$. Then, $B_{1,\lambda} \subseteq B_{2,\mu}$. Similar to the above proof, we have $B_{2,\mu} \subseteq B_{1,\lambda}$. So, $B_{1,\lambda} = B_{2,\mu}$. Furthermore, $D_{1,\lambda} = t(B_{1,\lambda}) = t(B_{2,\mu}) = D_{2,\mu}$, i.e., $X_2(\mu) = X_1(\lambda)$. Therefore, $\forall \lambda \in D_1$, there exists $\mu \in D_2$ such that $X_2(\mu) = X_1(\lambda)$.

Similarly, we also can prove that there exists $\lambda \in D_1$ such that $X_1(\lambda) = X_2(\mu)$ for $\forall \mu \in D_2$. Therefore, $\aleph_{Td_1}(X) = \aleph_{Td_2}(X)$, that is $d_1 \cong d_2(T)$.

Similar to the above proof, if $f:[0,1] \to [0,1]$, we have the conclusion is also hold. □

We shall give the following example to illustrate the effectiveness of Theorem 8.9.

Example 8. Let $X = R^n$. $d_1, d_2 \in ND(X)$, and they are defined as follow: $\forall x,y \in X$,

$$d_1(x,y) = 1 - \exp(-\|x-y\|), d_2(x,y) = 1 - \exp(-\|x-y\|^2)$$

where $\|\cdot\|$ is a norm on X. Because d_1 and d_2 have the relationship as follows:

$$d_2 = f(d_1) = \begin{cases} 1 - \exp(-\ln^2(1-d_1)), & d_1 \in (0,1) \\ 0, & d_1 = 0 \end{cases}$$

It is obvious that f is an one-to-one and strictly increasing mapping f satisfying $f(0)=0$ on $D_1 = [0,1)$ to $D_2 = [0,1)$, then $d_1 \cong d_2(T)$ by Theorem 8.9.

Remark 12. The norm $\|\cdot\|$ in Example 8 is under general norm definition. Particularly, we may take matrix norms. Because matrix norms are equivalent [71], the data corresponding with $1 - d_2 = \exp(-\|x-y\|^2)$ is Gaussian [20]. According to the conclusions reached earlier in this paper, it is easy to know that all Gaussian types of data have consistent clustering structure under the transitive closure operation (note: the clustering structure of normalized (pseudo-)metric is not affected when it is multiplied by a constant). That may help explain why all data selected in analysis and experimental studies are almost all Gaussian types.

The isomorphic condition for two normalized (pseudo-)metrics (i.e., their granular spaces are same) is too strict. If their ordered granular spaces are different and results inferred from these structures are similar, then, how shall it be described? Similar to the discussion in Section 8.1, we will address this question as follows.

Definition 8.5. Let $d_1, d_2 \in ND(X)$ (or $d_1, d_2 \in WND(X)$) and $\varepsilon > 0$. If there exists $d_3 \in ND(X)$ (or $d_3 \in WND(X)$) satisfying $d_3 \cong d_1(T)$ and $\forall x,y \in X$, $|d_2(x,y) - d_3(x,y)| \le \varepsilon$ (or $d_3 \cong d_2(T)$ and $\forall x,y \in X$, $|d_1(x,y) - d_3(x,y)| \le \varepsilon$), then normalized (pseudo-)metric d_1 an d_2 are called ε-similarity on clustering (or classifying) structures, short for d_1 and d_2 are ε-similarity, marked as $d_1 \sim d_2(T,\varepsilon)$.

Theorem 8.10. (ε-*Similarity Necessary Theorem*). *Suppose* $d_1, d_2 \in ND(X)$ (*or* $d_1, d_2 \in WND(X)$), *the ordered granular space derived from* d_i *is marked as* $\aleph_{Td_i}(X) = \{X_i(\lambda)|\lambda \in [0,1]\}$, $i=1,2$. *If* $d_1 \sim d_2\ (T,\varepsilon)$, *then* $\forall \lambda \in [0,1]$, *there exists* $\mu \in [0,1]$ *such that* $X_2(\mu+\varepsilon) \le X_1(\lambda) \le X_2(\mu-\varepsilon)$, *or* $\forall \mu \in [0,1]$, *there exists* $\lambda \in [0,1]$ *such that* $X_1(\lambda+\varepsilon) \le X_2(\mu) \le X_1(\lambda-\varepsilon)$.

Proof. We only give the proof for $d_1, d_2 \in ND(X)$ as follows, but it also holds when $d_1, d_2 \in WND(X)$. If there exists $d_3 \in ND(X)$ such that

$$d_3 \cong d_1(T) \quad \text{and} \quad \forall x,y \in X,\ |d_2(x,y) - d_3(x,y)| \le \varepsilon \tag{8.3}$$

Marked the ordered granular space derived by d_3 as $\aleph_{Td_3}(X) = \{X_3(\lambda)|\lambda \in [0,1]\}$. $\forall \lambda \in [0,1]$, $B_{i,\lambda} = \{(x,y)|d_i(x,y) \le \lambda\}$, $D_{i,\lambda} = t(B_{i,\lambda})$, $i=1,2,3$, where $D_{i,\lambda}$ is a crisp equivalence relation on X by Proposition 5.2, and $X_i(\lambda)$ is the set of equivalent classes of $D_{i,\lambda}$ on X by Theorem 5.2, $i=1,2,3$.

From $d_3 \cong d_1(T)$, $\forall \lambda \in [0,1]$, there exists $\mu \in [0,1]$ such that $X_3(\mu) = X_1(\lambda)$. By the condition (8.3), we have $B_{2,\mu-\varepsilon} \subseteq B_{3,\mu} \subseteq B_{2,\mu+\varepsilon}$. Furthermore,

$$D_{2,\mu-\varepsilon} \subseteq D_{3,\mu} \subseteq D_{2,\mu+\varepsilon} \tag{8.4}$$

which is essentially same as $X_2(\mu+\varepsilon) \le X_3(\mu) = X_1(\lambda) \le X_2(\mu-\varepsilon)$. Therefore, $\forall \lambda \in [0,1]$, there exists $\mu \in [0,1]$ such that $X_2(\mu+\varepsilon) \le X_1(\lambda) \le X_2(\mu-\varepsilon)$.

Similarly, we can prove the other half. □

Based on this theorem, ε-similarity of two NEC (pseudo-)metrics means that there is a restrictive relation among their granulations. And any granulation of the one is always restricted by two granulations of the other, but the restrictive relation is not mutual. Because the inverse of formula (8.4) does not hold, neither does the inverse of Theorem 8.10. From Theorem 8.3 and Definition 8.5, we directly obtain the following sufficient condition of ε-Similarity for two normalized (pseudo-)metrics.

Theorem 8.11. (ε-*Similarity Sufficient Theorem*). *Suppose* $d_1, d_2 \in ND(X)$ (*or* $d_1, d_2 \in WND(X)$). *If there exists an one-to-one and strictly increasing mapping* $f:[0,1] \to [0,1]$ *satisfying* $f(0)=0$ *such that* $\forall x,y \in X$, $|f(d_1(x,y)) - d_2(x,y)| \le \varepsilon$, *or there exists an one-to-one and strictly increasing mapping* $g:[0,1] \to [0,1]$ *satisfying*

$g(0)=0$ *such that* $\forall x,y \in X$, $|g(d_2(x,y))-d_1(x,y)| \le \varepsilon$, *then* $d_1 \sim d_2(T,\varepsilon)$.

The following example is used to show the effectiveness of Theorem 8.11.

Example 9. In Example 7, taken $d_1: d_1(i,i)=0$, $i=1,\ldots,4$, $d_1(1,2)=0.2$, $d_1(1,4)=0.3$, $d_1(2,3)=0.4$, $d_1(1,3)=d_1(2,4)=d_1(3,4)=0.5$. Because $\forall x,y \in X$, $|d_1(x,y)-d_2(x,y)| \le 0.1$, by Theorem 8.11, then $d_1 \sim d_2(T,0.1)$.

By integrated Sections 8.1 and 8.2, we have achieved the clustering structure analysis theory based on metric.

9. Conclusion

In this paper, we researched structural clusters analysis of metric space based on granular space on the basis of quotient space theory, developed comprehensive theory and methods, and achieved six main accomplishments as follows.

(1) We established the granular space theory based on NEC (pseudo-)metric to study its ordered property, gave two basic conclusions (i.e., basic theorems 1 and 2), and defined the metric on granular space based on NEC (pseudo-)metric.
(2) We presented a basic concept that consistent cluster (or classification) characteristic according to metric, and researched that the ordered granular space has the consistent cluster (or classification) characteristic.
(3) We researched granular representation of structural cluster from normalized (pseudo-)metric, and obtained the algorithm to rapidly find its clustering structure that has the consistent cluster characteristic.
(4) We studied optimal cluster determination problem based on granular space, and developed a new method to obtain the optimal cluster number, which the new method has the global optimal.
(5) We researched fusion of structural cluster based on granular space, which it is a studying method according to the intersection operation of normalized (pseudo-)metrics or NEC (pseudo-)metrics.
(6) We established the clustering structures analysis theory based on granular space by introducing some basic concepts that the isomorphic and similar on clustering structures, and discussed their sufficient conditions and necessary conditions that two NEC metrics (or two normalized metrics) are isomorphic and similar.

Furthermore, we established a series of theories and methodologies for structural clustering (or classifying) analysis based on metric space, and developed powerful mathematical models and formal description tools for the research on its potential applications, particularly for large volume data processing, such as, structural bioinformatics processing, complex system analysis, etc. These results can help us pursue even deeper understanding of the essence of clustering (or classifying) procedure.

Acknowledgements

The research is supported in part by the National Natural Science Foundation of China (Grant no. 60475017), the National Grand Fundamental Research 973 Program of China (Grant no. 2004CB318108), the International Technology Collaboration Research Program of China (Grant no. 2007DFA70670), and Program for Innovative Research Team of Jiangnan University. We are especially indebted to Professor L. Zhang. He provided suggestions to improve this article after the careful reading throughout it. We are also thankful to the anonymous reviewers for their valuable comments which have helped in improving of presentation of the paper.

References

[1] L.A. Zadeh, Fuzzy logic=computing with words, IEEE Transactions on Fuzzy Systems 2 (1996) 103–111.
[2] L.A. Zadeh, Some reflections on information granulation and its contrility in granular computing, computing with words, the computational theory of perceptions and preciseated natural language, in: T.Y. Lin, Y.Y. Yao, L.A. Zadeh (Eds.), Data Mining, Rough Sets and Granular Computing, Physica-Verlag, Wurzburg, 2002.
[3] L.A. Zadeh, The roles of fuzzy logic and soft computing in the conception, design and deployment of intelligent systems, in: H.S. Nwana, N. Azarmi (Eds.), Software Agents and Soft Computing: Towards Enhancing Machine Intelligence, Concepts and Applications, Springer, Berlin, 1997, pp. 183–190.
[4] D.-G. Li, D.-Q. Miao, D.-X. Zhang, H.-Y. Zhang, An overview of granular computing, Computer Science (China) 32 (9) (2005) 1–12 (in Chinese).
[5] Y.Y. Yao, Granular computing: basic issues and possible solutions, in: P.P. Wang (Ed.), Proceedings of the 5th Joint Conference on Information Sciences, vol. I, Association for Intelligent Machinery, 2000, pp. 186–189.
[6] Y.Y. Yao, J.T. Yao, Granular computing as a basis for consistent classification problems, in: PAKDD 2002 Workshop entitled towards foundation of Data Mining, vol. 5(2), 2002, pp. 101–106.
[7] Y.Y. Yao, Modeling data mining with granular computing, in: Proceedings of the 25th Annual International Computer Software and Applications Conference CCMPSAC, 2001, pp. 638–643.
[8] Y.Y. Yao, N. Zhong, Potential applications of granular computing in knowledge discovery and data mining, in: M. Torres, B. Sanchez, J. Aguilar (Eds.), Proceedings of World Multiconference on Systemics, Cybernetics and Informatics, vol. 5, Computer Science and Engineering, 1999, pp. 573–580.
[9] Y.Y. Yao, Granular computing using neighborhood systems, in: R. Roy, T. Furuhashi, R.K. Chawdhry (Eds.), Advances in Soft Computing: Engineering Design and Facturing, Springer, London, 1999, pp. 111–116.
[10] Y.Y. Yao, Rough sets, neighborhood Systems and granular computing, in: M. Meng (Ed.), Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering, IEEE Press, 1999, pp. 1553–1558.
[11] Z. Pawlak, Rough sets, International Journal of Computer and Information Science 11 (1982) 341–356.
[12] Z. Pawlak, Rough Sets Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers, Dordrecht, 1991.
[13] Z. Pawlak, Granularity of knowledge, indiscernibility and rough sets, in: Proceedings of 1998 IEEE International Conference on Fuzzy Systems, 1998, pp. 106–110.
[14] Y. Ye, J.K. Tsotsos, Knowledge granularity and action selection, in: F. Giunchiglia (Ed.), AIMSA'98, LNAI 1480, 1998, pp. 475–488.
[15] F. Huang, S. Zhang, Clustering Web documents based on knowledge granularity, in: X. Zhou et al. (Eds.), APWeb2006, LNCS3841, 2006, pp. 85–96.
[16] B. Zhang, L. Zhang, The Theory and Application of Problem Solving, Elsevier Science Publishers, North-Holland, 1992.
[17] B. Zhang, L. Zhang, Theory of Problem Solving and Its Applications—The Theory and Methods of Quotient Granular Computing, second ed, Tsinghua University Press, Beijing, 2007 (in Chinese).
[18] L. Zhang, B. Zhang, Fuzzy quotient space theory (fuzzy granule computing methods), Chinese Journal of Software 14 (4) (2003) 770–776 (in Chinese).
[19] L. Zhang, B. Zhang, The structure analysis of fuzzy sets, International Journal of Approximate Reasoning 40 (2005) 92–108.
[20] S. Theodoridis, K. Koutroumbas, Pattern Recognition, second ed, Elsevier Science, USA, 2003.
[21] S.-Y. Yang, Pattern Recognition and Intelligent Computing—Technical Realization of Matlab, Publishing House of Electronics Industry, Beijing, 2008 (in Chinese).
[22] Z.-B. Xu, J.-S. Zhang, Y.-L. Zheng, Bionics in Intelligent Computing: Theory and Algorithms, first ed. (second printing), Science Press (China), Beijing, 2004 (in Chinese).
[23] C.-Z. Wu, Introduction of Fuzzy Set(I), Beijing Normal University Press, Beijing, 1984 (in Chinese).
[24] E.H. Ruspini, A new approach to clustering, Information and Control 15 (1) (1969) 22–32.
[25] Q. He, H.-X. Li, Z.-Z. Shi, E.S. Lee, Fuzzy clustering method based on Perturbation, Computers and Mathematics with Applications 46 (2003) 929–946.
[26] H.-S. Lee, An optimal algorithm for computing the Max–Min transitive closure of a fuzzy similarity maxtrix, Fuzzy Sets and Systems 123 (2001) 129–136.
[27] G.-S. Liang, T.-Y. Chou, T.-C. Han, Cluster analysis based on fuzzy equivalence relation, European Journal of Operational Research 166 (2005) 160–171.
[28] A.O. Esogbue, B.D. Liu, Cluster validity for fuzzy criterion clustering, Computers and Mathematics with Applications 37 (1999) 95–100.
[29] H. Kamimura, M. Kurano, Clustering by a fuzzy metric, Fuzzy Sets and Systems 120 (2001) 249–254.

[30] D.-W. Kim, K.H. Lee, D. Lee, Fuzzy clustering of categorical data using fuzzy centroids, Pattern Recognition Letters 25 (2004) 1263–1271.
[31] P. Angelov, An approach for fuzzy rule-base adaptation using on-line clustering, International Journal of Approximate Reasoning 35 (2004) 275–289.
[32] A. Devillez, P. Billaudel, G.V. Lecolier, A fuzzy hybrid hierarchical clustering method with a new criterion able to find the optimal partition, Fuzzy Sets and Systems 128 (2002) 323–338.
[33] W. Pedrycz, Collaborative fuzzy clustering, Pattern Recognition Letters 23 (2002) 1675–1686.
[34] O.S. Piangkh, Analytically traceable case of fuzzy C-means clustering, Pattern Recognition 39 (2006) 35–46.
[35] G. Tsekouras, H. Sarimveis, E. Kavakli, G. Bafas, A hierarchical fuzzy clustering approach to fuzzy modeling, Fuzzy Sets and Systems 150 (2005) 246–266.
[36] K.-L. Wu, J. Yu, M.-S. Yang, A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality tests, Pattern Recognition Letters 26 (2005) 639–652.
[37] F. Guoyao, Optimization methods for fuzzy clustering, Fuzzy Sets and Systems 93 (1998) 301–309.
[38] M. Gibson, G. Kanade, E. Krohn, I.A. Pirwani, K. Varadarajan, On metric clustering to minimize the sum of radii, Algorithm, published online: 12 February 2009, Springer Science+Business, Media, LLC2009.
[39] M. Gibson, G. Kanade, E. Krohn, I.A. Pirwani, K. Varadarajan, On clustering to minimize the sum of radii, in: SODA, SIAM, Philadelphia, 2008, pp. 819–825.
[40] M. Charikar, R. Panigrahy, Clustering to minimize the sum of cluster diameters, Journal of Computer and System Science 68 (2) (2004) 417–441.
[41] S.E. Schaeffer, Graphy clustering, Computer Science Review 1 (1) (2007) 27–64.
[42] B.-C. Chien, C.-H. Hu, M.-Y. Ju, Learning fuzzy concept hierarchy and measurement with node labeling, Inf. Syst. Front, Springer Science+Business, Media, LLC2008.
[43] X.-Q. Tang, P. Zhu, J.-X. Cheng, Cluster analysis based on fuzzy quotient space, Chinese Journal of Software 19 (4) (2008) 861–868.
[44] X.-Q. Tang, P. Zhu, J.-X. Cheng, Study on fuzzy granular space based on normalized equicrural metric, Computer Science 35 (4) (2008) 142–145 (in Chinese).
[45] J. Li, G.-J. Wang, Theory of α-truth degrees in n-valued Godel propositional logic, Chinese Journal of Software 18 (1) (2007) 33–39 (in Chinese).
[46] Z.-Z. Guan, Lecture on Functional Analysis, Higher Education Press (China), Beijing, 1958 (in Chinese).
[47] U. Ziyan, M.R. Sabuncu, W.E.L. Grimson, C.-F. Westin, Consistency clustering: a robust algorithm for group-wise registration, segmentation and automatic atlas construction in diffusion MRI, Int. J. Comput. Vir., published online: 26 March 2009, Springer Science+Business, Media, LLC2009.
[48] J.C. Bezdek, N.R. Pal, Some new indexes of cluster validity, IEEE Transactions on Systems Man Cybernet 28 (1998) 301–315.
[49] R. Gunderson, Application of fuzzy ISODATA algorithms to star-trackerpointing systems, in: Proceedings of 7th Triannual World IFAC Congress, Helsinki, 1978, pp. 1319–1323.
[50] R. Kothari R, D. Pitts, On finding the number of classes, Pattern Recognition Letters 20 (1999) 405–416.
[51] E. Trauwert, L. Kaufman, P. Rousseeuw, Fuzzy clustering algorithms based on the maximum likehood principle, Fuzzy Sets and Systems 42 (1991) 213–227.
[52] A. Hardy, On the number of classes, Computational Statistics & Data Analysis 23 (1996) 83–96.
[53] H. Sun, S. Wang, Q. Jiang, FCM-based model selection algorithms for determining the number of clusters, Pattern Recognition 37 (2004) 2027–2037.
[54] D.W. Kim, K.H. Lee, D. Lee, On cluster validity index for estimation of optimal number of fuzzy clusters, Pattern Recognition 37 (2004) 2009–2024.
[55] J. Yu, Q.-S. Cheng, The search scope of optimal cluster number in fuzzy clustering methods, Science in China (Series E) 32 (4) (2002) 274–280 (in Chinese).
[56] M.R. Rezaee, B.P.F. Lelieveldt, J.H.C. Reiber, A new cluster validity index for fuzzy C-mean, Pattern Recognition Letters 19 (1998) 237–246.
[57] N.R. Pal, J.C. Bezdek, On cluster validity for the fuzzy C-mean model, IEEE Transactions on Fuzzy System 3 (3) (1995) 370–379.
[58] J. Sheu, A hybrid fuzzy optimization approach to customer grouping-based logistics distribution to operations, Applied Mathematical Modeling 31 (2007) 1048–1066.
[59] E. Backer, A.K. Jain, A clustering performance measure based on fuzzy set decomposition, IEEE Transactions on Pattern Analysis and Machine Intelligence 3 (1) (1981) 66–75.
[60] J.C. Bezdek, Cluster validity with fuzzy sets, Journal of Cybernetics 3 (3) (1974) 58–73.
[61] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[62] M.P. Windham, Cluster validity for fuzzy clustering algorithms, Fuzzy Sets and Systems 5 (1981) 177–185.
[63] J.C. Bezdek, Mathematical models for systematics and taxonomy, in: Proceedings of 8th International Conference on Numberical Taxonomy, San Francisco, 1975, pp. 143–166.
[64] Y. Fukuyama, M. Sugeno, A new method of choosing the number of clusters for the fuzzy C means method, in: Proceedings of 5th Fuzzy Systems Symposium, 1989, pp. 247–250.
[65] L.X. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (8) (1991) 841–847.
[66] S. Miyamoto, Information clustering based on fuzzy multi-sets. Information and Management 39 (2003) 195–213.
[67] S. Esnaf, T. Kücükdeniz, A fuzzy Clustering based hybrid method for a multi-facility location problem, Journal of Intelligent Manufacturing 20 (2009) 259–265.
[68] S.-J. Shuu, Evaluating the flexibility in a manufacturing system using fuzzy multi-attribute group decision-making with multi-granularity linguistic information, International Journal of Advanced Manufacturing Technology 32 (2007) 409–421.
[69] C. DÖring, M. Lesot, R. Kruse, Data analysis with fuzzy clustering methods, Computational Statistics and Data Analysis 51 (2006) 192–214.
[70] X.-Q. Tang, X.-S. Fang, P. Zhu, Clustering characteristic analysis based on fuzzy granular space, in: Proceedings of 1st International Conference on Modelling and Simulation, Nanjing, 2008, pp. 371–377.
[71] J.J. Leader, Numerical Analysis and Scientific Computation, Tsinghua University Press, Beijing, 2008.

About the Author—XU-QING TANG, born in 1963 year, is an associate Professor. His current research interests include intelligent computing, granular computing, mathematical modeling and bio-informatics.

About the Author—PING ZHU, born in 1962, is a Professor, and the vice dean of the school of science. Her current research interests include algebra, bio-informatics, computer science.

About the Author—JIA-XING CHENG, born in 1946 year, is a professor and doctoral supervisor of Anhui University. His current research interests include intelligent computing, and optimizing theory and methodology.