ORIGINAL ARTICLE

Uncertainty measurement for incomplete set-valued data with application to attribute reduction

Yan Song[1] · Damei Luo[2] · Ningxin Xie[3] · Zhaowen Li[4]

Abstract

A set-valued information system (SVIS) is the generalization of a single-valued information system. A SVIS with missing information values is called an incomplete set-valued information system (ISVIS). This paper focuses on studying uncertainty measurement for an ISVIS with application to attribute reduction. First, the similarity degree between information values on each attribute is presented in an ISVIS. Then, the tolerance relation induced by each subsystem is given and rough approximations based on this relation is considered. Next, some tools to measure the uncertainty of an ISVIS are put forwarded. Moreover, the validity of the proposed measures is analyzed from the statistical point of view. Finally, information granulation and information entropy are applied to attribute reduction, the incomplete rate is adopted, and the effectiveness under different incomplete rates is analyzed and verified by k-means clustering algorithm and Mean Shift clustering algorithm.

Keywords RST · ISVIS · Similarity degree · Attribute reduction · Information granulation · Information entropy · Algorithm

✉ Zhaowen Li
  lizhaowen8846@126.com

  Yan Song
  songyan201805@126.com

  Damei Luo
  luodamei166@126.com

  Ningxin Xie
  ningxinxie100@126.com

1 School of Mathematics and Statistics, Yulin Normal University, Yulin, Guangxi 537000, People's Republic of China

2 School of Mathematics and Information Science, Guangxi University, Nanning, Guangxi 530004, People's Republic of China

3 School of Artificial Iintelligence, Guangxi Minzu University, Nanning, Guangxi 530006, People's Republic of China

4 Key Laboratory of Complex System Optimization and Big Data Processing, Department of Guangxi Education, Yulin Normal University, Yulin, Guangxi 537000, People's Republic of China

1 Introduction

Rough set theory (RST), an effective data analysis tool put forward by Pawlak [23, 24]. It is based on the idea that some objects in the universe have the corresponding information values. By generalizing equivalence classes or equivalence relations, this theory has been widely extended. On the one hand, equivalence relations are divided into tolerance relations, dominance relations and reflexive relations. On the other hand, the division of an equivalence class is extended to cover. RST, as an important method to manage uncertainty, has the merit of being directly based on the original data, rather than requiring preliminary or additional data information. Therefore, it is highly reliable. Many applications of RST are based on information systems (ISs) [6, 42, 43, 47]. In addition, some scholars studied multigranulation rough sets [5, 20, 41, 44].

An IS as a database that displays relationships between objects and attributes is also put forwarded by Pawlak so as to reveal large databases and knowledge discovery process mathematically. It is important to note that there may be missing information values in an IS. An IS with missing information values is called an incomplete IS (IIS). A set containing all possible information values can be utilized to represent the missing information values. By representing

all missing information values of a single valued IS as a set containing all possible information values, it can be evolved into a set-valued information system (SVIS). In this way, a SVIS can effectively deal with the noise caused by the missing information values.

As an important IS, scholars have paid great attention to SVISs. For instance, Yao [46] studied SVISs with upper and lower approximations; Leung et al. [16] found a method to select minimum feature set for SVISs; Couso et al. [7] looked into the rationality of SVISs from a statistical point of view; Huang et al. [15] obtained a probabilistic set-valued ISs by using probability distribution to describe set values; Qian et al. [25] introduced two kinds of set-valued ordered ISs and put forward an attribute reduction method that can simplify set-valued ordered ISs. Liu et al. [21] researched feature selection for a set-valued decision IS from the view of dominance relations. Xie et al. [38] gave uncertainty measures for interval-valued ISs. Chen et al. [4] investigated feature selection in a SVIS according to tolerance relations. A SVIS has been successfully employed to the data analysis of complete ISs, such as distinguishing the dependence between attributes, distinguishing the importance of attributes, and attribute reduction.

Uncertainty is mainly composed of four parts: randomness, fuzziness, incompleteness and inconsistency. It exists in every field of real life. Uncertainty measure has become a noticeable problem in many fields such as machine learning [40], image processing [22], medical diagnosis [14], and data mining [10]. With the development of research, some excellent results have been obtained. For example, information entropy put forward by Shannon [28] has been recognized as a very important research method to measure the uncertainty of ISs. Yao [45] considered granularity measure in terms of granularity. Until now, information granularity and information entropy have gradually become two important tools to consider the uncertainty of ISs. On the basis of these two tools, some outstanding scholars have promoted and applied them. Wierman [36] discussed granularity measure in RST. Düntsch et al. [11] explored the measurement of decision rules in RST based on information entropy. Dai et al. [8] brought up the two tools of entropy measure and granularity measure to measure the uncertainty of SVISs. Li et al. [21] researched entropy theory in fuzzy relation IS. Wang et al. [35] applied information entropy and information granularity to the measurement of interval and SVISs. Li et al. [18] analyzed the information structure of fuzzy set-valued ISs and the uncertainty measurement method of fuzzy set-valued ISs. Wu et al. [37] proposed a reliable approximation operator based on semi-monolayer covering for set-valued ISs.

Attribute reduction or feature selection, as an important technology of data processing in machine learning, can effectively reduce redundant attributes. It can also reduce the complexity of high-dimensional data for calculation and improve the accuracy of classification. To different data, many researchers study attribute reduction. For instance, Tang et al. [31] researched attribute reduction in set-valued decision ISs. Song et al. [30] applied attribute reduction in set-valued decision ISs. Cornelis et al. [6] studied a general definition for a fuzzy decision reduct. Wang et al. [34] presented an iterative reduction algorithm from the view of variable distance parameter. Giang et al. [13] obtained an algorithm with application to attribute reduction in a dynamic decision table. Qian et al. [26] explored an accelerator algorithm for attribute reduction based on RST. Chen et al. [2] brought up the concept of fuzzy kernel alignment and applied it to attribute reduction for heterogeneous data. Singh et al. [29] introduced a attribute selection method of rough set based on fuzzy similarity in SVISs. Wang et al. [32] proposed four uncertainty measures. On this basis, they designed a greedy algorithm for attribute reduction. Li et al. [19] constructed a new acceleration strategy for general attribute reduction algorithms. Li et al. [17] studied existing reduction methods to help researchers better understand and use these reduction methods to meet their own needs.

Set-valued data is an important data in practical applications. However, in some practical cases, set-valued may be described by missing information values, which can cause some critical information to be missing. An incomplete set-valued information system (ISVIS) is a SVIS with missing information values. Xie et al. [39] introduced the distance between the values of two information functions and applied it to obtain the information structures and uncertainty measure of incomplete probability set-valued ISs. Chen et al. [3] obtained some tools to measure the uncertainty of an ISVIS by means of Gaussian kernel.

This article focuses on studying uncertainty measurement of incomplete set-valued data and its attribute reduction. For an incomplete set-valued data, we treat it as an ISVIS. In an ISVIS, objects described by the same information are indiscernible. The indiscernibility relations produced in this mode constitute the mathematical foundation of RST. Therefore, the similarity degree between information values on each attribute is shown in an ISVIS base on RST. The tolerance relation induced by each subsystem is given and the tolerance relation is dealt with by introducing an approximate equality between fuzzy sets. Some tools are put passed on to measure the uncertainty of ISVISs. From the point of view of data's incomplete rate, some statistical methods are used to analyze the effectiveness of the proposed measures. Base on two measurement methods (i.e., information granulation and information entropy), two reduction algorithms are given, and their effectiveness under different incomplete rates is analyzed and verified by k-means clustering algorithm and Mean Shift clustering algorithm. The work process of the paper is given in Fig. 1.

Fig. 1 The work process of the paper
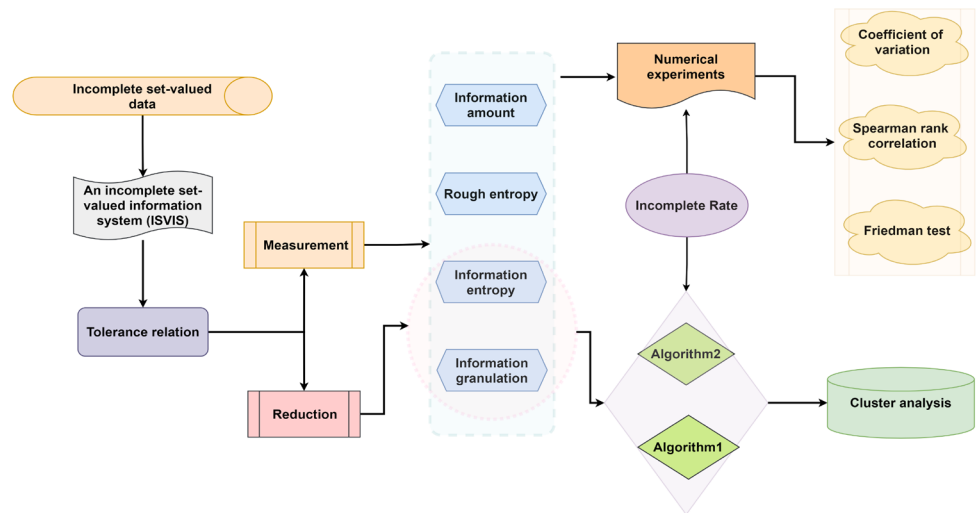
The rest of this paper is intended to be below. Section 2 retrospects the cardinal perceptions of fuzzy relations and ISVISs. Section 3 obtains similarity degree and equivalence relations in an ISVIS. Section 4 investigates uncertainty measure for an ISVIS. Section 5 gives experiments analysis for the proposed measures. Section 6 studies an application for attribute reduction in an ISVIS. Section 7 compares the proposed algorithms with the other two algorithms. Section 8 summaries this paper.

2 Preliminaries

In this section, we recall some basic notions about fuzzy relations and ISVISs.

Throughout this paper, U, A denote two non-empty finite sets, 2^U means the family of all subsets of U and $|X|$ expresses the cardinality of $X \in 2^U$.

In this paper, put

$$U = \{x_1, x_2, \ldots, x_n\}, \quad A = \{a_1, a_2, \ldots, a_m\}$$

2.1 Fuzzy relations

Recall that R is a binary relation on U whenever $R \subseteq U \times U$. If $(x, y) \in R$, then we denote it by xRy.

Let R be a binary relation on U. Then R is called

(1) reflexive, if xRx for any $x \in U$;

(2) symmetric, if xRy implies yRx for any $x, y \in U$;

(3) transitive, if xRy and yRz imply xRz for any $x, y, z \in U$.

Let R be a binary relation on U. Then R is called an equivalence relation on U, if R is reflexive, symmetric and transitive. Moreover, R is called a universal relation on U if $R = \delta$; R is said to be an identity relation on U if $R = \triangle$.

Recall that F is a fuzzy set whenever F is a function defined by $F : U \to I$.

In this article, I^U shows the collection of fuzzy sets on U.

If R is a fuzzy set in $U \times U$, then R is called a fuzzy relation on U, and R can be expressed by the following matrix

$$M(R) = (R(x_i, x_j))_{n \times n}.$$

In this article, $I^{U \times U}$ denotes the family of all fuzzy relations on U.

Definition 2.1 ([21]) Suppose $R \in I^{U \times U}$. For any $x \in U$, define

$$S_R(x)(y) = R(x, y), \forall \, y \in U.$$

Then $S_R(x)$ is called the fuzzy information granule of the point x with respect to R.

In [33], $S_R(x)$ is denote by $[x]_R$.

2.2 An ISVIS

Definition 2.2 ([24]) Let U be an object set and A an attribute set. Suppose that U and A are finite sets. Then the pair (U, A) is called an information system (IS), if each attribute $a \in A$ determines a information function $a : U \to V_a$, where $V_a = \{a(x) : x \in U\}$.

Let (U, A) be an IS. If there is $a \in A$ such that $* \in V_a$, here $*$ means a null or unknown value, then (U, A) is called an incomplete information system (IIS).

If (U, A) is an IIS, given $P \subseteq A$. Then a binary relation T_P on U can be defined as

$$(x, y) \in T_P \Leftrightarrow \forall a \in P, a(x) = a(y) \text{ or } a(x) = * \text{ or } a(y) = * .$$

Clearly, T_P is a tolerance relation on U. For each $x \in U$, denote

$$T_P(x) = \{y \in U : (x, y) \in T_P\}.$$

Then, $T_P(x)$ is called the tolerance class of x under the tolerance relation T_P.

For convenience, $T_{\{a\}}$ and $T_{\{a\}}(x)$ are denoted by T_a and $T_a(x)$, respectively.

Obviously,

$$T_P = \bigcap_{a \in P} T_a, \quad T_P(x) = \bigcap_{a \in P} T_a(x).$$

Let (U, A) be an IIS. For each $a \in A$, denote

$$V_a^* = V_a - \{a(x) : a(x) = *\}.$$

Then, V_a^* means the set of all non-missing information values of the attribute a.

Definition 2.3 ([45]) Suppose that (U, A) is an IIS. Then (U, A) is referred to as an incomplete set-valued information system (for short, an ISVIS), if for any $a \in A$ and $x \in U$, $a(x)$ is set.

If $P \subseteq A$, then (U, P) is referred to as the subsystem of (U, A).

Example 2.4 Table 1 depicts an ISVIS (U, A), where $U = \{x_1, x_2, \ldots, x_7\}$ and $A = \{a_1, a_2, \ldots, a_4\}$.

Example 2.5 (Continued from Example 2.4)

Table 1 An ISVIS (U, A)

	a_1	a_2	a_3	a_4
x_1	$\{1, 2\}$	$\{a, b, c\}$	$\{4, 5, 7\}$	$\{H, T\}$
x_2	$\{1, 2\}$	$\{a, c\}$	$\{2, 3\}$	$\{H, T\}$
x_3	$\{3, 4\}$	$\{a, c\}$	$\{2, 3\}$	$\{T, F\}$
x_4	$\{1, 2\}$	$\{a, b\}$	$\{1, 2, 5\}$	$\{T, F\}$
x_5	$\{1, 3\}$	$\{a, b\}$	$*$	$\{H, T, F\}$
x_6	$\{1, 2, 3\}$	$\{a, b, c\}$	$\{5, 6, 7\}$	$\{T, H\}$
x_7	$*$	$\{a, b, c\}$	$\{5, 6, 7\}$	$\{H, T\}$

$$V_{a_1}^* = \{\{1, 3\}, \{1, 2\}, \{3, 4\}, \{1, 2, 3\}\},$$

$$V_{a_2}^* = \{\{a, c\}, \{a, b\}, \{a, b, c\}\},$$

$$V_{a_3}^* = \{\{2, 3\}, \{5, 6, 7\}, \{4, 5, 7\}, \{1, 2, 5\}\},$$

$$V_{a_4}^* = \{\{T, F\}, \{T, H\}, \{H, T, F\}\}.$$

3 The equivalence relation induced by each subsystem of an ISVIS

In an ISVIS, objects described by the same information are indiscernible. The indiscernibility relation produced in this mode constitutes mathematical foundation of RST. Thus, this section constructs the similarity degree between information values on each attribute in an ISVIS and gives the equivalence relation induced by each subsystem.

Definition 3.1 Let (U, A) be an ISVIS. Then $\forall x, y \in U$, $a \in A$, the similarity degree between $a(x)$ and $a(y)$ is defined as follows:

$$s(a(x), a(y))$$

$$= \begin{cases} 1, & x = y; \\ \frac{1}{|V_a^*|^2}, & x \neq y, \ a(x) = *, \ a(y) = * ; \\ \frac{1}{|V_a^*|}, & x \neq y, \ a(x) \neq *, \ a(y) = * ; \\ \frac{1}{|V_a^*|}, & x \neq y, \ a(x) = *, \ a(y) \neq * ; \\ 1, & x \neq y, \ a(x) \neq *, \ a(y) \neq *, \ a(x) = a(y); \\ \frac{|a(x) \bigcap a(y)|}{|a(x) \bigcup a(y)|}, & x \neq y, \ a(x) \neq *, \ a(y) \neq *, \ a(x) \neq a(y). \end{cases}$$

For the convenience of expression, denote

$$s_{ij}^k = s(a_k(x_i), a_k(x_j)).$$

s_{ij}^k indicates the similarity degree between $a_k(x_i)$ and $a_k(x_j)$. This also expresses the similarity degree between two objects x_i and x_j with respect to the attribute a_k.

Example 3.2 (Continued from Example 2.4) By Definition 3.1, then s_{ij}^k $(i, j = 1, \ldots, 7, k = 1, \ldots, 4)$ is obtained as follows (see Tables 2, 3, 4, 5).

Let (U, A) be an ISVIS. For any $a \in A$, define

$$R_a(x, y) = s(a(x), a(y)).$$

Then R_a is a fuzzy relation on U.

Below, we attempt to deal with a fuzzy relation R_a by introducing the approximate equality between fuzzy sets.

Table 2 s_{ij}^1

s_{ij}^1	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1	1	0	1	0.33	0.67	0.25
x_2	1	1	0	1	0.33	0.67	0.25
x_3	0	0	1	0	0.33	0.25	0.25
x_4	1	1	0	1	0.33	0.67	0.25
x_5	0.33	0.33	0.33	0.33	1	0.67	0.25
x_6	0.67	0.67	0.25	0.67	0.67	1	0.25
x_7	0.25	0.25	0.25	0.25	0.25	0.25	1

Table 3 s_{ij}^2

s_{ij}^2	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1	0.67	0.67	0.67	0.67	1	1
x_2	0.67	1	1	0.33	0.33	0.67	0.67
x_3	0.67	1	1	0.33	0.33	0.67	0.67
x_4	0.67	0.33	0.33	1	1	0.67	0.67
x_5	0.67	0.33	0.33	1	1	0.67	0.67
x_6	1	0.67	0.67	0.67	0.67	1	1
x_7	1	0.67	0.67	0.67	0.67	1	1

Table 4 s_{ij}^3

s_{ij}^1	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1	0	0	0.2	0.25	0.5	0.5
x_2	0	1	1	0.25	0.25	0	0
x_3	0	1	1	0.25	0.25	0	0
x_4	0.2	0.25	0.25	1	0.25	0.2	0.2
x_5	0.25	0.25	0.25	0.25	1	0.25	0.25
x_6	0.5	0	0	0.2	0.25	1	1
x_7	0.5	0	0	0.2	0.25	1	1

Table 5 s_{ij}^4

s_{ij}^2	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1	1	0.33	0.33	0.67	1	1
x_2	1	1	0.33	0.33	0.67	1	1
x_3	0.33	0.33	1	1	0.67	0.33	0.33
x_4	0.33	0.33	1	1	0.67	0.33	0.33
x_5	0.67	0.67	0.67	0.67	1	0.67	0.67
x_6	1	1	0.33	0.33	0.67	1	1
x_7	1	1	0.33	0.33	0.67	1	1

Definition 3.3 ([49]) Suppose $k \in N$. Given $a, b \in [0, 1]$. If $a, b \in [0, \frac{1}{10^k})$ or $a, b \in [\frac{1}{10^k}, \frac{2}{10^k})$ or $\cdots\cdots$ or $a, b \in [\frac{10^k-1}{10^k}, 1)$ or $a = b = 1$, then a and b are said to be class-consistent, and k is said to be a threshold value. We denote it by $a \approx_k b$.

In this paper, we pick $k = 1$.

Definition 3.4 ([49]) Suppose $A, B \in I^U$. Then

$$A \approx_1 B \Leftrightarrow \forall x \in U, \ A(x) \approx_1 B(x).$$

Definition 3.5 Let (U, A) be an ISVIS. Given $P \subseteq A$. Define

$$R_a^* = \{(x, y) \in U \times U : S_{R_a}(x) \approx_1 S_{R_a}(y)\},$$
$$R_P^* = \bigcap_{a \in P} R_a^*.$$

It is easy to see that R_P^* is an equivalence relation on U. Then R_P^* is called the equivalence relation induced by the subsystem (U, P). And the partition on U induced by R_P^* is denoted by U/R_P^*.

For any $x \in U$, denote

$$R_P^*(x) = \{y \in U : (x, y) \in R_P^*\}.$$

Example 3.6 (Continued from Example 3.2) We can obtain $R_A^*(x_1) = \{x_1\}, R_A^*(x_2) = \{x_2\}, R_A^*(x_3) = \{x_3\}, R_A^*(x_4) = \{x_4\}, R_A^*(x_5) = \{x_5\}, R_A^*(x_6) = \{x_6\}, R_A^*(x_7) = \{x_7\}.$

4 Measuring uncertainty of an ISVIS

In this section, some tools for measuring uncertainty of an ISVIS are obtained.

4.1 Granulation measure for an ISVIS

Definition 4.1 Suppose that (U, A) is an ISVIS. Given $P \subseteq A$. Then information granulation of the subsystem (U, P) is specified as

$$G(P) = \frac{1}{n^2} \sum_{i=1}^{n} |R_P^*(x_i)|.$$

Proposition 4.2 *Let (U, A) be an ISVIS. Then for any $P \subseteq A$,*

$$\frac{1}{n} \leq G(P) \leq 1.$$

Furthermore, if R_P^* is an universal relation on U, G achieves the minimum value $\frac{1}{n}$; if R_P^* is a identity relation on U, G will achieve the maximum value 1.

Proof $\forall i, 1 \leq |R_P^*(x_i)| \leq n, n \leq \sum_{i=1}^{n} |R_P^*(x_i)| \leq n^2$. By Definition 4.1,

$$\frac{1}{n} \leq G(P) \leq 1.$$

If R_P^* is an identity relation on U, for any $i, |R_P^*(x_i)| = 1$, $G(P) = \frac{1}{n}$.

If R_P^* is a universal relation on U, for any $i, |R_P^*(x_i)| = n$, $G(P) = 1$. ☐

Proposition 4.3 *Let (U, A) be an ISVIS. If $Q \subseteq P \subseteq A$, then $G(P) \leq G(Q)$.*

Proof (1) Since $Q \subseteq P \subseteq A$, $\forall i$, we have $R_P^*(x_i) \subseteq R_Q^*(x_i)$. Then $|R_P^*(x_i)| \leq |R_Q^*(x_i)|$. By Definition 4.1,

$$G(P) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n} |R_P^*(x_i)| \leq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n} |R_Q^*(x_i)| = G(Q).$$

Thus $G(P) \leq G(Q)$. ☐

This Proposition 4.3 shows that information granulation increases with the coarsening of information and decreases with the refinement of information. This means that the uncertainty of ISVISs can be evaluated based on the information granulation introduced in Definition 4.1.

Example 4.4 (Continued from Example 3.6) By Definition 4.1, we can obtain

$$G(A) = \frac{1}{7^2} \sum_{i=1}^{7} |R_P^*(x_i)| = \frac{7}{49} \approx 0.14.$$

4.2 Entropy measure for an ISVIS

Entropy tends to measure the disorder degree of a system. The higher its value, the higher the disorder order of the system is. Shannon [28] applies this concept of entropy to information theory for calculating the measurement uncertainty of a system. Similarly, information entropy of a given ISVIS is defined as following.

Definition 4.5 Suppose that (U, A) is an ISVIS. Given $P \subseteq A$. Then information entropy of the subsystem (U, P) is defined as

$$H(P) = - \sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{|R_P^*(x_i)|}{n}.$$

Proposition 4.6 *Let (U, A) be an ISVIS. If $Q \subseteq P \subseteq A$, then $H(P) \leq H(Q)$.*

Proof Since $Q \subseteq P \subseteq A$, similar to the proof of Proposition 4.3, we again that $\forall i, 1 \leq |R_P^*(x_i)| \leq |R_Q^*(x_i)|$.

Then $\forall i, -\log_2 \frac{|R_P^*(x_i)|}{n} = \log_2 \frac{n}{|R_P^*(x_i)|} \geq \log_2 \frac{n}{|R_Q^*(x_i)|} = -\log_2 \frac{|R_Q^*(x_i)|}{n}$.

Consequently, $H(P) \leq H(Q)$. ☐

This statement clarifies that information entropy increases with the refinement of information and decreases with the coarsening of information. This means that the uncertainty of ISVISs can be evaluated based on the information entropy introduced in Definition 4.5.

Example 4.7 (Continued from Example 3.6) By Definition 4.5, we can obtain

$$H(A) = -\sum_{i=1}^{7} \frac{1}{7} \log_2 \frac{|R_P^*(x_i)|}{7} \approx 2.81.$$

Rough entropy is used to measure granularity of a given partition. It is also called co-entropy by some scholars.

Similarly, rough entropy of a given ISVIS is put forward in the following definition.

Definition 4.8 Let (U, A) be an ISVIS. Given $P \subseteq A$. Then rough entropy of the subsystem (U, P) is deemed as

$$E_r(P) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{|R_P^*(x_i)|}.$$

Proposition 4.9 Let (U, A) be an ISVIS. Given $P \subseteq A$. Then

$$0 \leq E_r(P) \leq \log_2 n.$$

What is more, if R_P^* is an identity relation on U, then E_r^* reaches the minimum value 0; if R_P^* is an universal relation on U, then E_r^* attains the maximum value $\log_2 n$.

Proof Note that R_P^* is a fuzzy equivalence relation on U. Then $\forall i, R_P^*(x_i)(x_i) = 1$.

Hence $\forall i$, $1 \leq |R_P^*(x_i)| \leq n$, then $0 \leq -\log_2 \frac{1}{|R_P^*(x_i)|} = \log_2 |R_P^*(x_i)| \leq \log_2 n$. Therefore, $0 \leq -\sum_{i=1}^{n} \log_2 \frac{1}{|R_P^*(x_i)|} \leq n \log_2 n$.

By Definition 4.8, we obtain that

$$0 \leq E_r(P) \leq \log_2 n.$$

If R_P^* is an identity relation on U, then $\forall i, |R_P^*(x_i)| = 1$. Thus $E_r(P) = 0$.

If R_P^* is an universal relation on U, then $\forall i, |R_P^*(x_i)| = n$. Thus $E_r(P) = \log_2 n$. □

Proposition 4.10 Let (U, A) be an ISVIS. If $P \subseteq Q \subseteq A$, then $E_r(Q) \leq E_r(P)$.

Proof Since $P \subseteq Q \subseteq A$, similar to the proof of Proposition 4.3, we obtain that $\forall i$,

$$1 \leq |R_Q^*(x_i)| \leq |R_P^*(x_i)|.$$

Then $\forall i$, $-\log_2 \frac{1}{|R_P^*(x_i)|} = \log_2 |R_P^*(x_i)| \geq \log_2 |R_Q^*(x_i)| = -\log_2 \frac{1}{|R_Q^*(x_i)|}$

As a result, $E_r(P) \geq E_r(Q)$. □

For Proposition 4.10, it can be found that the more uncertain the available information is, the bigger rough entropy value becomes. This means that rough entropy brought

forward in Definition 4.8 can be used to evaluate the uncertainty of an ISVIS.

Example 4.11 (Continued from Example 3.6) By Definition 4.8, we can obtain

$$E_r(P) = -\sum_{i=1}^{7} \frac{1}{7} \log_2 \frac{1}{|R_P^*(x_i)|} = 0.$$

Theorem 4.12 Let (U, A) be an ISVIS. Given $P \subseteq A$. Then

$$E_r(P) + H(P) = \log_2 n.$$

Proof

$$E_r(P) + H(P) = -\frac{1}{n} \sum_{i=1}^{n} \left(\log_2 \frac{1}{|R_P^*(x_i)|} + \log_2 \frac{|R_P^*(x_i)|}{n} \right)$$

$$= -\frac{1}{n} \sum_{i=1}^{n} \log_2 \frac{1}{n} = \log_2 n.$$

□

Corollary 4.13 Let (U, A) be an ISVIS. Given $P \subseteq A$. Then

$$0 \leq H(P) \leq \log_2 n.$$

Proof By Proposition 4.9, $0 \leq E_r(P) \leq \log_2 n$.

By Theorem 4.12, $H(P) = \log_2 n - E_r(P)$. Consequently, $0 \leq H(P) \leq \log_2 n$. □

4.3 Fuzzy information amount in an ISVIS

Similarly, information amount in a given ISVIS is stated in the following definition.

Definition 4.14 Let (U, A) be an ISVIS. Given $P \subseteq A$. Then information amount of the subsystem (U, P) is regarded as

$$E(P) = \sum_{i=1}^{n} \frac{1}{n} \left(1 - \frac{|R_P^*(x_i)|}{n} \right).$$

Proposition 4.15 Let (U, A) be an ISVIS. If $P \subseteq Q \subseteq A$, then $E(P) \leq E(Q)$.

Proof Since $P \subseteq Q \subseteq A$, similar to the proof of Proposition 4.3, we get that $\forall i, 1 \leq |R_Q^*(x_i)| \leq |R_P^*(x_i)|$. Then

$$E(P) = \sum_{i=1}^{n} \frac{1}{n} \left(1 - \frac{|R_P^*(x_i)|}{n} \right) \leq \sum_{i=1}^{n} \frac{1}{n} \left(1 - \frac{|R_Q^*(x_i)|}{n} \right) = E(Q).$$

Hence $E(P) \leq E(Q)$. □

It can be found that the more uncertain the available information is, the bigger information amount value becomes.

Theorem 4.16 *Let* (U, A) *be an ISVIS. Given* $P \subseteq A$. *Then* $G(P) + E(P) = 1$.

$$G(P) + E(P) = \frac{1}{n^2} \sum_{i=1}^{n} [|R_P^*(x_i)| + (n - |R_P^*(x_i)|)] = 1.$$

Proof

□

Corollary 4.17 *Let* (U, A) *be an ISVIS. Given* $P \subseteq A$. *Then* $0 \le E(P) \le 1 - \frac{1}{n}$.

Proof By Proposition 4.2, $\frac{1}{n} \le G(P) \le 1$.
By Theorem 4.16, $E(P) = 1 - G(P)$.
Thus $0 \le E(P) \le 1 - \frac{1}{n}$. □

From Proposition 4.15 and Corollary 4.17, we know that information amount introduced in Definition 4.14 can evaluate the uncertainty of an ISVIS.

Example 4.18 (Continued from Example 3.6) By Definition 4.14, we can obtain

$$E(P) = \sum_{i=1}^{7} \frac{1}{7} \left(1 - \frac{|R_P^*(x_i)|}{7}\right) \approx 0.86.$$

5 Experiments analysis

In this section, some numerical experiments are designed to evaluate the effectiveness of the proposed measures under different incomplete rates.

5.1 Incomplete rate

In this section, 8 incomplete data were selected in UCI to test the performance of the proposed measures, as shown in Table 6, where Obesity was randomly hollowed out with 20% of content and became an incomplete data.

For an ISVIS (U, A), the missing information values are randomly distributed on all attributes, and the incomplete rate of (U, A) (denoted by β) is defined as

$$\beta = \frac{Number\ of\ missing\ values}{mn}.$$

First, we transform the incomplete data into an ISVIS. Then, among all the information values, we delete 2%, 4%, 6%, 8%, 10%, 12%, 14% and 16% randomly, and we call the created data 'β-ISVIS' respectively, whose $\beta = 0.02k$ ($k = 1, 2, \ldots, 8$). If the incomplete rate of an incomplete data has exceeded the value of β we want to set, then the missing information value is randomly selected and set as the set composed of all possible values of this attribute. If the incomplete rate of an incomplete data does not reach the value of β we want to set, then the known information value is randomly selected as the missing information value to achieve the value of β to be set.

First of all, we explore the number of objects with missing information values at $\beta = 0.02k$ ($k = 1, 2, \ldots, 8$) based on the data in Table 6. The results are shown in Fig. 2, where the X-axis represents $\beta = 0.02k$ ($k = 1, 2, \ldots, 8$) and the Y-axis represents the percentage of objects with missing information values. The values in Fig. 2 are the average values of 10 training sets.

From Fig. 2, we can draw the following conclusions:

(1) As the incomplete rate increases, so does the percentage of objects that contain missing information values; (2) The more attributes a data has, the more decentralized distributed the missing information values are. As a result, the rate of missing objects is much higher; (3) When the incomplete rate is only 6%, the percentage of missing objects in all data except Sl is greater than 50%. When the incompleteness rate is only 10%, the percentage of missing objects in all data is greater than 70%; (4) Preprocessing methods such as deleting can seriously reduce the available information in the database.

Table 6 Eight data from UCI

Date sets	Abbr.	Objects	Features
Autism-adolescent	Aa	104	20
Autism-child	Ac	292	20
Dermatology	De	366	34
Hepatitis	He	155	19
Processed-cleveland	Pc	303	13
Soybean-large	Sl	307	35
Obesity	Ob	2111	16
Garments-worker	Gw	1197	14
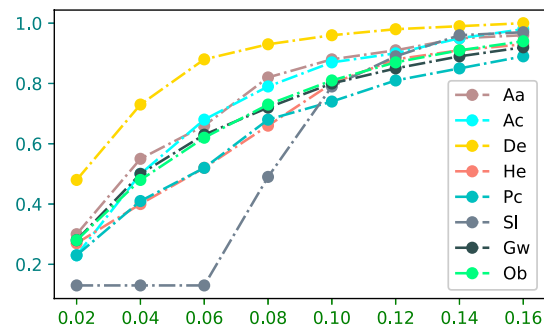
Fig. 2 Percentage of missing objects under different β

5.2 Numerical experiments

In Aa, pick $\beta = 0.02k$ $(k = 1, \ldots, 8)$, $A_i = \{a_1, \ldots, a_i\}$ $(i = 1, \ldots, 20)$. Denote

$$X_G^\beta(Aa) = \{G^\beta(A_1), \ldots, G^\beta(A_{20})\}, \quad X_E^\beta(Aa) = \{E^\beta(A_1), \ldots, E^\beta(A_{20})\},$$
$$X_{E_r}^\beta(Aa) = \{E_r^\beta(A_1), \ldots, E_r^\beta(A_{20})\}, \quad X_H^\beta(Aa) = \{H^\beta(A_1), \ldots, H^\beta(A_{20})\}.$$

In Ac, pick $\beta = 0.02k$ $(k = 1, \ldots, 8)$, $C_i = \{a_1, \ldots, a_i\}$, $(i = 1, \ldots, 20)$. Denote

$$X_G^\beta(Ac) = \{G^\beta(C_1), \ldots, G^\beta(C_{20})\}, \quad X_E^\beta(Ac) = \{E^\beta(C_1), \ldots, E^\beta(C_{20})\},$$
$$X_{E_r}^\beta(Ac) = \{E_r^\beta(C_1), \ldots, E_r^\beta(C_{20})\}, \quad X_H^\beta(Ac) = \{H^\beta(C_1), \ldots, H^\beta(C_{20})\}.$$

In De, pick $\beta = 0.02k$ $(k = 1, \ldots, 8)$, $D_i = \{a_1, \ldots, a_i\}$, $(i = 1, \ldots, 34)$. Denote

$$X_G^\beta(De) = \{G^\beta(D_1), \ldots, G^\beta(D_{34})\},$$
$$X_E^\beta(De) = \{E^\beta(D_1), \ldots, E^\beta(D_{34})\},$$
$$X_{E_r}^\beta(De) = \{E_r^\beta(D_1), \ldots, E_r^\beta(D_{34})\},$$
$$X_H^\beta(De) = \{H^\beta(D_1), \ldots, H^\beta(D_{34})\}.$$

In He, pick $\beta = 0.02k$ $(k = 1, \ldots, 8)$, $H_i = \{a_1, \ldots, a_i\}$, $(i = 1, \ldots, 19)$. Denote

$$X_G^\beta(He) = \{G^\beta(H_1), \ldots, G^\beta(H_{19})\},$$
$$X_E^\beta(He) = \{E^\beta(H_1), \ldots, E^\beta(H_{19})\},$$
$$X_{E_r}^\beta(He) = \{E_r^\beta(H_1), \ldots, E_r^\beta(H_{19})\},$$
$$X_H^\beta(He) = \{H^\beta(H_1), \ldots, H^\beta(H_{19})\}.$$

In Pc, pick $\beta = 0.02k$ $(k = 1, \ldots, 8)$, $P_i = \{a_1, \ldots, a_i\}$, $(i = 1, \ldots, 13)$. Denote

$$X_G^\beta(Pc) = \{G^\beta(P_1), \ldots, G^\beta(P_{13})\}, \quad X_E^\beta(Pc) = \{E^\beta(P_1), \ldots, E^\beta(P_{13})\},$$
$$X_{E_r}^\beta(Pc) = \{E_r^\beta(P_1), \ldots, E_r^\beta(P_{13})\}, \quad X_H^\beta(Pc) = \{H^\beta(P_1), \ldots, H^\beta(P_{13})\}.$$

In Sl, pick $\beta = 0.02k$ $(k = 1, \ldots, 8)$, $L_i = \{a_1, \ldots, a_i\}$, $(i = 1, \ldots, 35)$. Denote

$$X_G^\beta(Sl) = \{G^\beta(L_1), \ldots, G^\beta(L_{35})\}, \quad X_E^\beta(Sl) = \{E^\beta(L_1), \ldots, E^\beta(L_{35})\},$$
$$X_{E_r}^\beta(Sl) = \{E_r^\beta(L_1), \ldots, E_r^\beta(L_{35})\}, \quad X_H^\beta(Sl) = \{H^\beta(L_1), \ldots, H^\beta(L_{35})\}.$$

In Ob, pick $\beta = 0.02k$ $(k = 1, \ldots, 8)$, $B_i = \{a_1, \ldots, a_i\}$, $(i = 1, \ldots, 16)$. Denote

$$X_G^\beta(Ob) = \{G^\beta(B_1), \ldots, G^\beta(B_{16})\}, \quad X_E^\beta(Ob) = \{E^\beta(B_1), \ldots, E^\beta(B_{16})\},$$
$$X_{E_r}^\beta(Ob) = \{E_r^\beta(B_1), \ldots, E_r^\beta(B_{16})\}, \quad X_H^\beta(Ob) = \{H^\beta(B_1), \ldots, H^\beta(B_{16})\}.$$

In Gw, pick $\beta = 0.02k$ $(k = 1, \ldots, 8)$, $G_i = \{a_1, \ldots, a_i\}$, $(i = 1, \ldots, 14)$. Denote
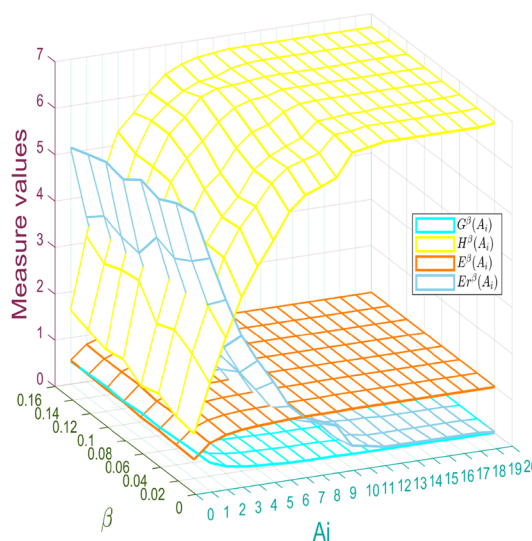
Fig. 3 Values of uncertainty measurement on Aa
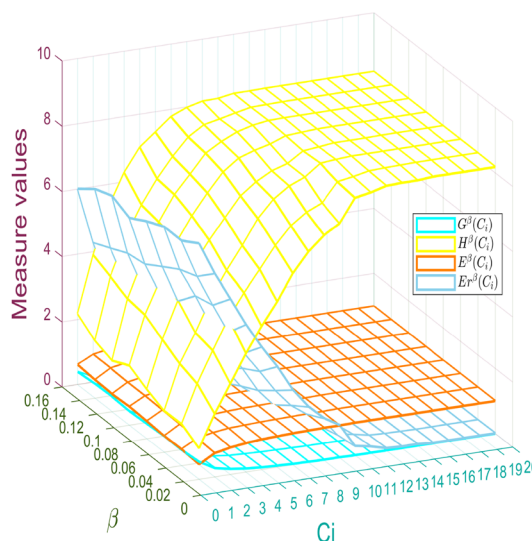
Fig. 4 Values of uncertainty measurement on Ac

$$X_G^\beta(Gw) = \{G^\beta(G_1), \ldots, G^\beta(G_{14})\},$$
$$X_E^\beta(Gw) = \{E^\beta(G_1), \ldots, E^\beta(G_{14})\},$$
$$X_{E_r}^\beta(Gw) = \{E_r^\beta(G_1), \ldots, E_r^\beta(G_{14})\},$$
$$X_H^\beta(Gw) = \{H^\beta(G_1), \ldots, H^\beta(G_{14})\}.$$

From Figs. 3,4, 5, 6, 7, 8, 9 10, the following conclusions are obtained:

Regardless of the incomplete rate, E_r and G decrease monotonously as the number of attributes in the attribute
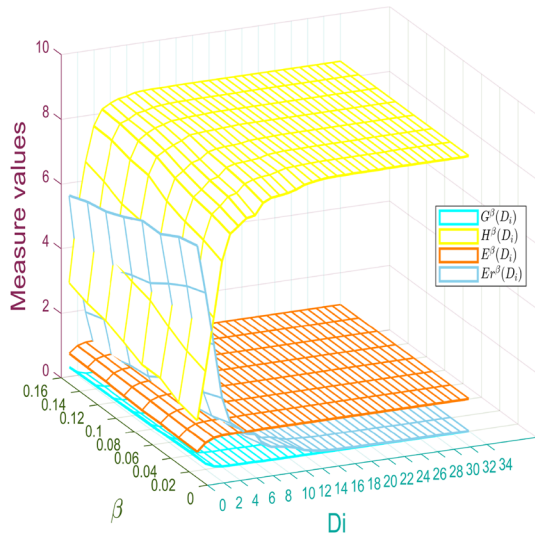
Fig. 5 Values of uncertainty measurement on De
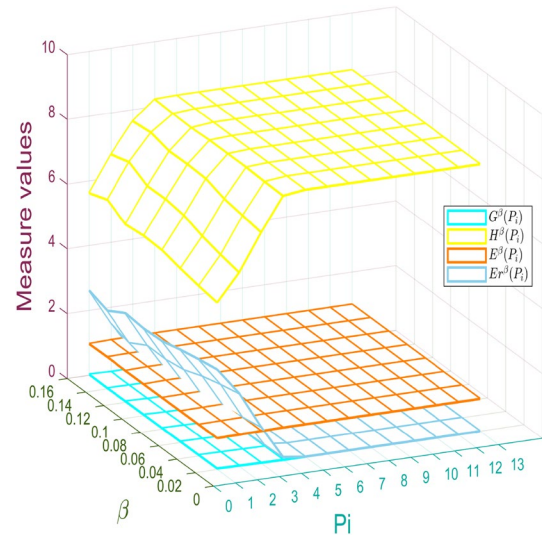
Fig. 7 Values of uncertainty measurement on Pc
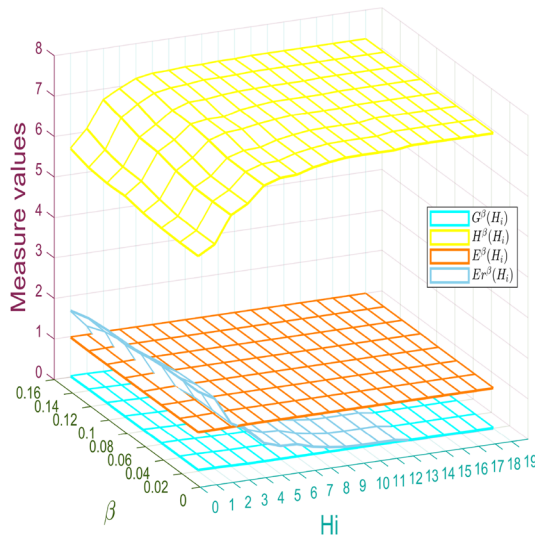
Fig. 6 Values of uncertainty measurement on He
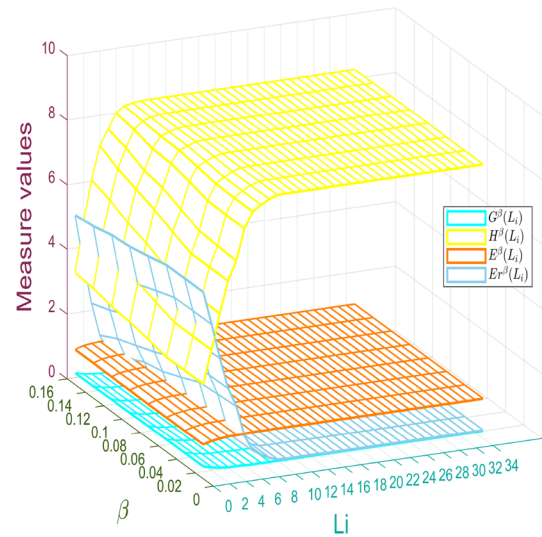
Fig. 8 Values of uncertainty measurement on Sl

subset increases. At the same time, both E and H increase monotonously with the increase of the number of attributes in the attribute subset. Therefore, the four measures proposed in this paper can be used to measure the uncertainty of ISVISs.

5.3 Dispersion analysis

Coefficient of variation CV, also known as discrete coefficient, is a statistic to measure the degree of variation of each observation in the data, which is obtained from the ratio of

standard deviation to mean value. Coefficient of variation can eliminate the influence of unit difference or average difference on the variation degree of two or more data.

Let $A = \{a_1, a_2, \ldots, a_n\}$ be a data set. Then, the coefficient of variation of A is denoted as $CV(A)$, which is defined as follows

$$CV(A) = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (a_i - \frac{1}{n} \sum_{i=1}^{n} a_i)^2}}{\frac{1}{n} \sum_{i=1}^{n} a_i}.$$
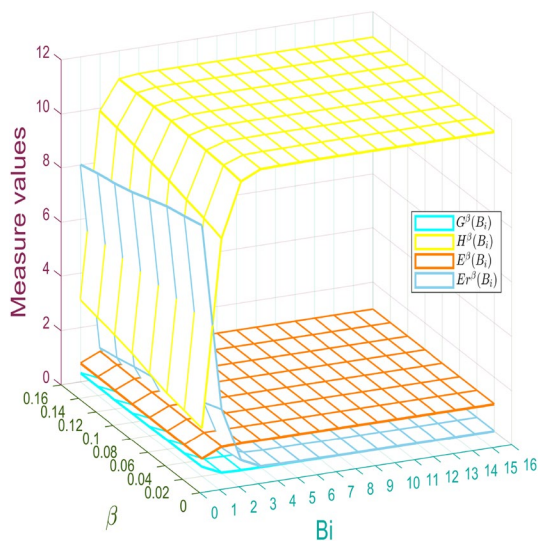
Fig. 9 Values of uncertainty measurement on Ob
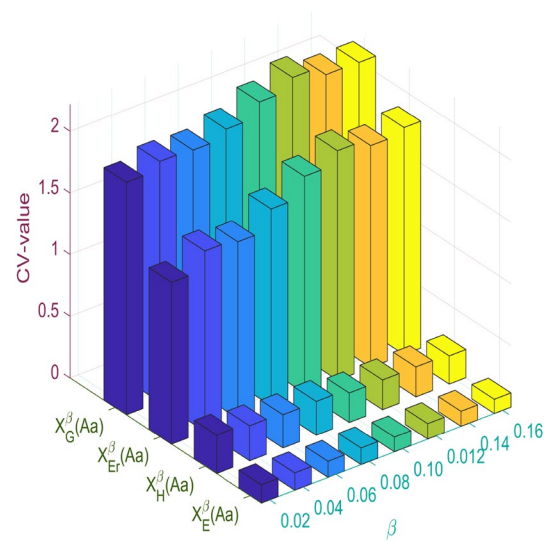
Fig. 11 *CV*-values of four measure sets of Aa
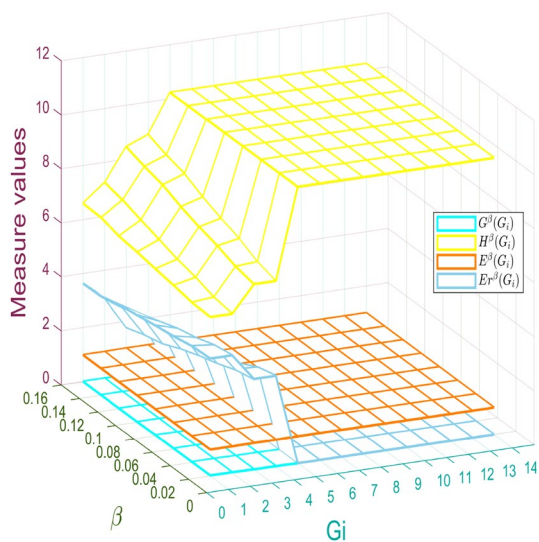
Fig. 10 Values of uncertainty measurement on Gw
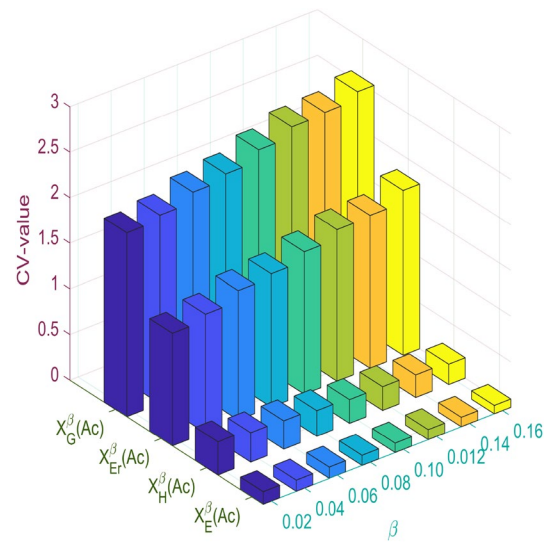
Fig. 12 *CV*-values of four measure sets of Ac

Continue the above experiment, *CV*-values of four measure sets are compared at different incomplete rates. The results are shown in Figs. 11, 12, 13, 14, 15, 16, 17, 18.

As can be seen from Figs. 11–18, some conclusions are obtained:

(1) When the data in Table 6 are at different incomplete rates, the *CV*-values of *Er* are all greater than 0.5, the *CV*-values of *E* and *H* are all less than 0.5, and the *CV*-values of *E* are all smaller than the *CV*-values of *H*.

(2) When the data in Table 6 are at different incomplete rates, *CV*-values of *G* are greater than 0.5 in all other gave data except He.

(3) For He with different incomplete rates, *CV*-values of *Er* are all greater than *CV*-values of *G*.

(4) For Pc with an incomplete rate 0.14 or 0.08, the *CV*-values of *Er* are greater than the *CV*-values of *G*, while there is little difference between the *CV*-values of *G* and the *CV*-values of *Er* at other incomplete rates.

Therefore, the coefficient of variation of *E* is the smallest in the given data with different incomplete rates, so *E* has the best measurement effect on the uncertainty of an ISVIS.
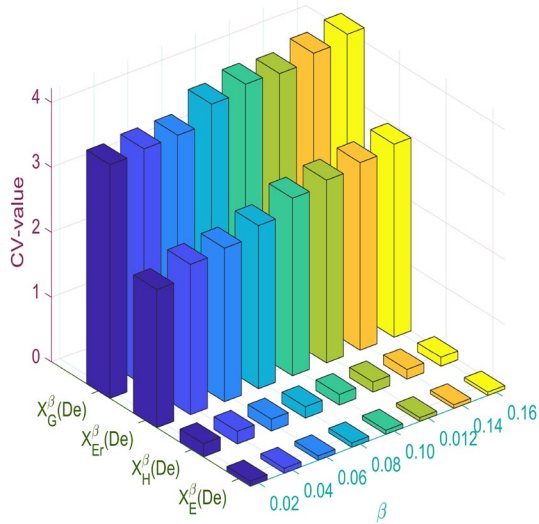
Fig. 13 *CV*-values of four measure sets of De
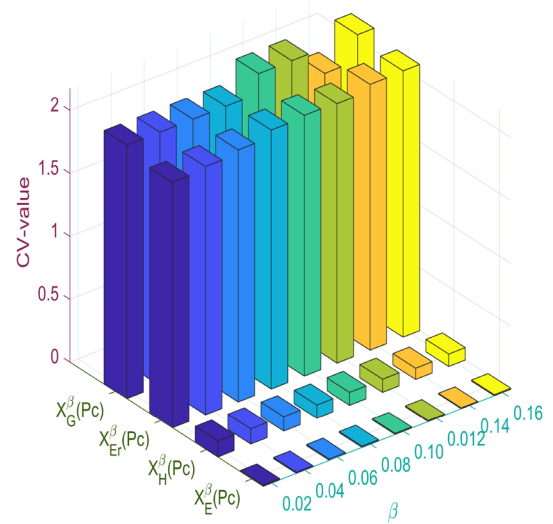
Fig. 15 *CV*-values of four measure sets of Pc
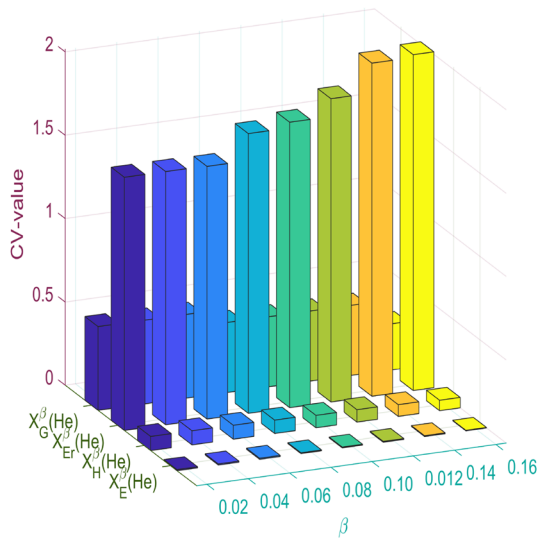
Fig. 14 *CV*-values of four measure sets of He
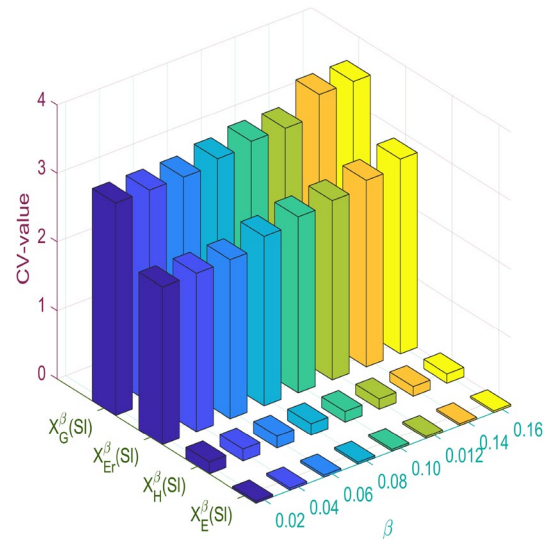
Fig. 16 *CV*-values of four measure sets of Sl

5.4 Correlation analysis

Spearman rank correlation [48], as an important statistical analysis method in statistics, is used to estimate the correlation between two statistical variables by using monotone equation.

Suppose that $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_n\}$ are two data sets. By sorting A and B (ascending or descending at the same time), two element ranking sets $R = \{R_1, R_2, \ldots, R_n\}$ and $Q = \{Q_1, Q_2, \ldots, Q_n\}$ are obtained, where R_i and Q_i $(i = 1, \ldots, n)$ are the ranking of a_i in A and b_i in B respectively. Spearman rank correlation coefficient

between A and B, denoted by $r_s(A, B)$, is defined as $r_s(A, B) = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}$, where $d_i = R_i - Q_i$. Obviously, $-1 \leq r_s(A, B) \leq 1$.

To test the significance of a correlation, we assume that there is no correlation between A and B. In the case of small samples, that is, the number of samples is less than 30, we can verify the hypothesis directly using the lookup method in Table 7 [48]. When $|r_s(A, B)|$ is greater than the threshold value of r_α, which indicates that the assumption is rejected, then correlation between A and B is significantly.

Continue the above experiment, r_s-values of four measure sets on Aa and Ac are compared. For Aa and Ac, the number
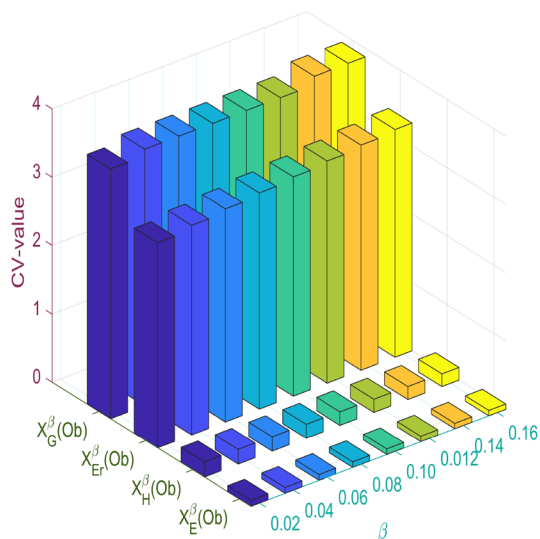
Fig. 17 *CV*-values of four measure sets of Ob
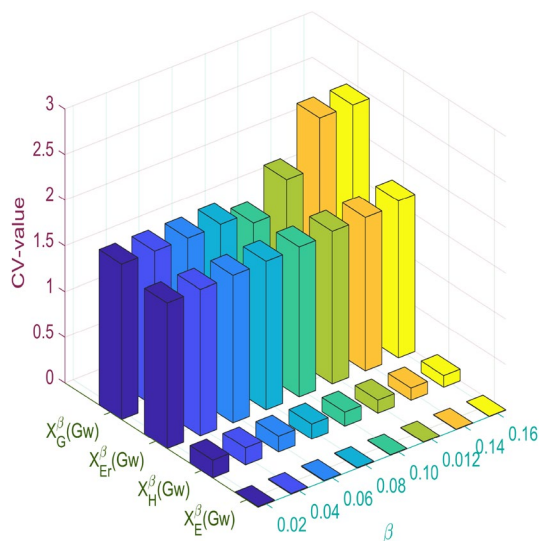
Fig. 18 *CV*-values of four measure sets of Gw

of four measure sets is 20. Then, from Table 7, we can obtain $r_{0.05} = 0.380$. Since $|r_s|$ in Aa and Ac exceeds 0.380, we can conclude that the pairwise correlations between these four measures are significant.

Example 5.1

If $0.7 \leq r_s(A, B) < 1$, $-1 \leq r_s(A, B) < -0.7$, $r_s(A, B) = 1$ and $r_s(A, B) = -1$ then the correlations between A and B are called height positive correlation (for short, HPC), height negative correlation (for short, HNC), completely positive correlation (for short, CNC) and completely positive negative correlation (for short, CPC), respectively. The following conclusions are obtained through calculation (see Tables 8, 9).

5.5 Friedman test and Nemenyi test

In this subsection, Friedman test [12] and Nemenyi test are used to further evaluate the performance of the proposed measures.

Friedman test, a nonparametric method to test whether there are significant differences among multiple algorithms by using rank, which is defined as

$$\chi_F^2 = \frac{12N}{k(k+1)} \left(\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4} \right),$$

where k, N and r_i are respectively the number of algorithms to be evaluated, the number of samples, and the average ranking of the i-th algorithm. This test is too conservative, which is why it is often replaced by the following statistic

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2}.$$

When F_F is greater than the threshold value of $F_\alpha(k-1, n-1)$, which indicates that the assumption that " all algorithms have the same performance" is rejected, then the performance of the algorithms is significantly different. Nemenyi test calculates the critical distance CD_α of average rank to judge which algorithm is better. If the corresponding average rank difference of the two algorithms reaches at least a critical distance, the performance of the two algorithms is significantly different. The critical distance CD_α is denoted as $CD_\alpha = q_\alpha \sqrt{\frac{k(k+1)}{6N}}$, where q_α and α are the critical tabulated value and significance level of Nemenyi test, respectively.

According to Figs. 11–18, we obtained the ranking of *CV*-values of the four measurement sets in the eight data (see Tables 10, 11, 12).

Table 7 Tables of critical values for Spearman rank correlation in which level of significance $\alpha = 0.05$-one-tailed test	n	4	5	6	7	8	9	10	11	12
	r_α	1	0.9	0.829	0.714	0.643	0.600	0.564	0.536	0.503
	n	13	14	15	16	17	18	19	20	
	r_α	0.484	0.464	0.443	0.429	0.414	0.401	0.391	0.380	

Table 8 r_s-values of four measure sets on Aa

r	$X_H^\beta(Aa)$	$X_{Er}^\beta(Aa)$	$X_G^\beta(Aa)$	$X_E^\beta(Aa)$
$X_H^\beta(Aa)$	CPC			
$X_{Er}^\beta(Aa)$	CNC	CPC		
$X_G^\beta(Aa)$	HNC	HPC	CPC	
$X_E^\beta(Aa)$	HPC	HNC	CNC	CPC

Table 9 r_s-values of four measure sets on Ac

r	$X_H^\beta(Ac)$	$X_{Er}^\beta(Ac)$	$X_G^\beta(Ac)$	$X_E^\beta(Ac)$
$X_H^\beta(Ac)$	CPC			
$X_{Er}^\beta(Ac)$	CNC	CPC		
$X_G^\beta(Ac)$	HNC	HPC	CNC	
$X_E^\beta(Ac)$	HPC	HNC	CPC	CPC

Table 10 The ranking of CV-values of the four measure sets on eight datasets with $\beta = 0.02k, k = 1, 2, 3, 6, 8$

Date	G	H	E_r	E
Aa	4	2	3	1
Ac	4	2	3	1
De	4	2	3	1
He	3	2	4	1
Pc	4	2	3	1
Sl	4	2	3	1
Ob	4	2	3	1
Gw	4	2	3	1
Average	3.875	2	3.125	1

Table 11 The ranking of CV-values of the four measure sets on eight datasets with $\beta = 0.02k, k = 4, 7$

Date	G	H	E_r	E
Aa	4	2	3	1
Ac	4	2	3	1
De	4	2	3	1
He	3	2	4	1
Pc	3	2	4	1
Sl	4	2	3	1
Ob	4	2	3	1
Gw	4	2	3	1
Average	3.75	2	3.25	1

Table 12 The ranking of CV-values of the four measure sets on eight datasets with $\beta = 0.02k, k = 5$

Date	G	H	E_r	E
Aa	4	2	3	1
Ac	4	2	3	1
De	4	2	3	1
He	3	2	4	1
Pc	4	2	3	1
Sl	4	2	3	1
Ob	4	2	3	1
Gw	3	2	4	1
Average	3.75	2	3.25	1
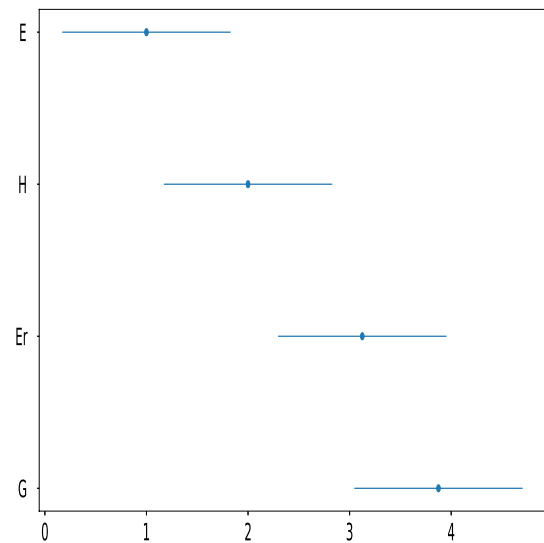
Fig. 19 Friedman test base on Table 10
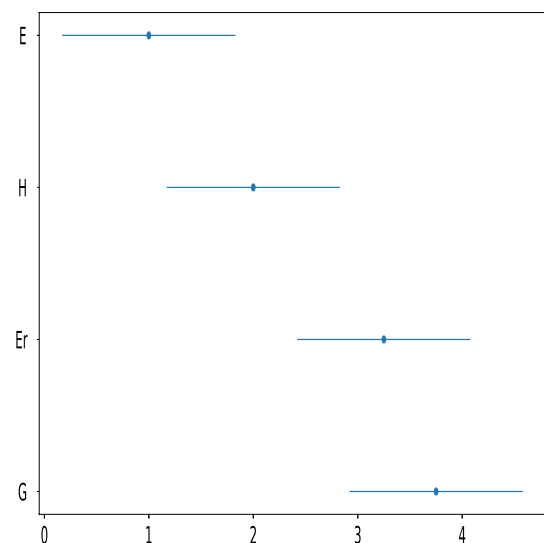
Fig. 20 Friedman test base on Tables 11 and 12

Friedman test was used to introduced whether there were significant differences in the four measures obtained in this paper. Since $k = 4, N = 8$. Then $k - 1 = 3$, $(k - 1)(N - 1) = 21$, $F_{0.05}(3, 21) = 3.072$. Therefore, for Table 10, $F_F \approx 109.29$ and $F_F > F_{0.05}(3, 21)$. For Table 11, $F_F \approx 61.67$ and $F_F > F_{0.05}(3, 21)$ and the results in Table 12 are the same as in Table 11. Therefore,

the assumption that " all algorithms have the same performance" is rejected at $\alpha = 0.05$, then the performance of the obtained algorithms is significantly different. Next, to further illustrate the significant differences between the four measures, Nemenyi test was investigate. Since $\alpha = 0.05$, then $q_\alpha = 2.569$ and $CD_\alpha \approx 1.658$. Based on these tests, we get Figs. 19, 20, where the dots represent the average ranking and the line segments represent the range of CD_α. If the two line segments do not overlap on the X-axis, there is a significant difference between the two uncertainty measures.

From Figs. 19, 20, in the case of data with an incomplete rate, the following conclusions can be drawn. (1) As far as performance is concerned, E is superior to G and Er. (2) In terms of performance, there are significant differences between E and Er, and between E and G.

6 An application in attribute reduction

In this section, an application of the proposed measures in attribute reduction is presented.

Definition 6.1 Suppose that (U, A) is an ISVIS. Then $P \subseteq A$ is referred to as consistent, if $R_P^* = R_A^*$.

Definition 6.2 Suppose that (U, A) is an ISVIS. Then $P \subseteq A$ is referred to as a reduct of A, if P is consistent and $\forall a \in P$, $P - \{a\}$ is not consistent.

In this paper, the family of all coordination subsets (resp., all reducts) of A is denoted by $co(A)$ (resp., $red(A)$).

Theorem 6.3 *Suppose that (U, A) is an ISVIS. Given $P \subseteq A$. Then the following conditions are equivalent*:

(1) $P \in co(A)$; (2) $G(P) = G(A)$; (3) $H(P) = H(A)$; (4) $E_r(P) = E_r(A)$; (5) $E(P) = E(A)$.

Proof $(1) \Rightarrow (2)$. Clearly. $(2) \Rightarrow (1)$. Suppose $G(P) = G(A)$. Then $\frac{1}{n^2} \sum_{i=1}^{n} |R_P^*(x_i)| = \frac{1}{n^2} \sum_{i=1}^{n} |R_A^*(x_i)|$.
So $\sum_{i=1}^{n} (|R_P^*(x_i)| - |R_A^*(x_i)|) = 0$.
Note that $R_A^* \subseteq R_P^*$. Then $\forall i$, $R_A^*(x_i) \subseteq R_P^*(x_i)$. This implies that

$$\forall i, \quad |R_P^*(x_i)| - |R_A^*(x_i)| \geq 0.$$

So $\forall i$, $|R_P^*(x_i)| - |R_A^*(x_i)| = 0$. It follows that $\forall i$, $R_P^*(x_i) = R_A^*(x_i)$.
Thus $R_P^* = R_A^*$. Hence

$P \in co(A)$.

$(2) \Leftrightarrow (5)$. It can be obtained by Theorem 4.16.
$(1) \Rightarrow (3)$. This is clear.
$(3) \Rightarrow (1)$. Suppose $H(P) = H(A)$. Then

$$-\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{|R_P^*(x_i)|}{n} = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{|R_A^*(x_i)|}{n}.$$

So

$$\sum_{i=1}^{n} \log_2 \frac{|R_P^*(x_i)|}{|R_A^*(x_i)|} = 0.$$

Note that $R_A^* \subseteq R_P^*$. Then $\forall i$, $R_A^*(x_i) \subseteq R_P^*(x_i)$. This implies that

$$\forall i, \quad \log_2 \frac{|R_P^*(x_i)|}{|R_A^*(x_i)|} \geq 0.$$

So $\forall i$, $\log_2 \frac{|R_P^*(x_i)|}{|R_A^*(x_i)|} = 0$. It follows that $\forall i$, $R_P^*(x_i) = R_A^*(x_i)$.
Thus $R_P^* = R_A^*$. Hence $P \in co(A)$.
$(3) \Leftrightarrow (4)$. It follows from Theorem 4.12. $\square$

Corollary 6.4 *Suppose that (U, A) is an ISVIS. Given $P \subseteq A$. Then the following conditions are equivalent*:

(1) $P \in red(A)$;

(2) $G(P) = G(A)$ and $\forall a \in P$, $G(P - \{a\}) \neq G(A)$;

(3) $H(P) = H(A)$ and $\forall a \in P$, $H(P - \{a\}) \neq H(A)$;

(4) $E_r(P) = E_r(A)$ and $\forall a \in P$, $E_r(P - \{a\}) \neq E_r(A)$;

(5) $E(P) = E(A)$ and $\forall a \in P$, $E(P - \{a\}) \neq E(A)$.

Proof It can be proved by Theorem 6.3. $\square$

Below, we study reduction algorithms in an ISVIS based on its uncertainty measurement. By Theorems 4.12 and 4.16, we have

$$E_r(P) + H(P) = \log_2 n, \quad G(P) + E(P) = 1,$$

where (U, A) is an ISVIS and $P \subseteq A$. Then, we only need to consider reduction algorithms based on information granulation and information entropy, respectively.

Reduction algorithms based on information granulation and information entropy are given as follows.

Algorithm 1: Reduction algorithm based on information granulation in an ISVIS

   Input: An ISVIS (U, A).
   Output: One reduct P.
1 Let $P \leftarrow A$ **for** *each feature $a_k \in A$* **do**
2   **for** *each $x_i, x_j \in U$* **do**
3   Compute s_{ij}^k
4   **end**
5 **end**
6 Compute Compute $G(P)$
7 **for** *each feature $a \in P$* **do**
8   Compute the equivalence relation $R_{P-\{a\}}^*$
9   Compute $G(P - \{a\})$
10   **if** $G(P - \{a\}) = G(P)$ **then**
11 $P \leftarrow P - \{a\}$
12   **else**
13   break;
14   **end**
15 **end**

Algorithm 2: Reduction algorithm based on information entropy in an ISVIS

   Input: An ISVIS (U, A).
   Output: One reduct P.
1 Let $P \leftarrow \emptyset$ **for** *each feature $a_k \in A$* **do**
2   **for** *each $x_i, x_j \in U$* **do**
3   Compute s_{ij}^k
4   **end**
5 **end**
6 Compute $H(A)$
7 **for** *each feature $a \in A - P$* **do**
8   Compute $H(P \cup \{a\})$
9   Pick $a*$, make $H(P \cup \{a*\}) = max\{H(P \cup \{a\}) : a \in A - P\}$
10   **if** $H(P \cup \{a*\}) \neq H(A)$ **then**
11 $P \leftarrow P \cup \{a*\}$
12   **else**
13   break;
14   **end**
15 **end**

For Algorithms 1–2, we assume that the number of attributes is m and the number of samples is n. First of all, we need to calculate each attribute a_k similarity matrix s_{ij}^k, its time complexity and space complexity are both $O(mn^2)$. Algorithm 1 randomly selects an attribute in each loop and judges whether the attribute can be discarded according to G. If not, the loop is terminated and the reduction set P is obtained. Therefore the worst search time for a reduct will need m evaluations. Algorithm 2 selects an attribute that meets the condition to add P in each loop according to H. If none of the attributes meet the condition, the loop is terminated and the reduction set P is obtained. Therefore the worst search time for a reduct will need $(m^2 + m)/2$ evaluations. Obviously, the overall time complexity of Algorithm 1 and Algorithm 2 is $O(mn^2 + m)$ and $O(mn^2 + m^2 + m)$ respectively. Since the space occupied above can be reused, the total space complexity of Algorithm 1 and Algorithm 2 is $O(mn^2)$.

6.1 Cluster analysis

In this subsection, in order to verify the effectiveness of the proposed algorithms, t-distributed stochastic neighbor embedding, k-means clustering algorithm ad Mean Shift clustering algorithm are used to cluster and analyze the reducts of the obtained algorithms.

In this paper, we give eight data from UCI described in Table 6. Each data can be regarded as an ISVIS. In order to verify that those algorithms are still effective in the case of different data missing, we also conducted experiments with Algorithm 1 and Algorithm 2 according to the incomplete rate from 0.02 to 0.16 with step size of 0.02. Then, for each algorithm, we obtained 8 reducts for each data and listed them in Tables 13, 14, respectively. As the reducts in Tables 13, 14 can be observed, Algorithms 1–2 can effectively reduce the dimension of incomplete data. Therefore, the results obtained from Algorithms 1–2 are employed to cluster analysis.

Fig. 21 The reduced clustering image of Algorithm 1 on Aa
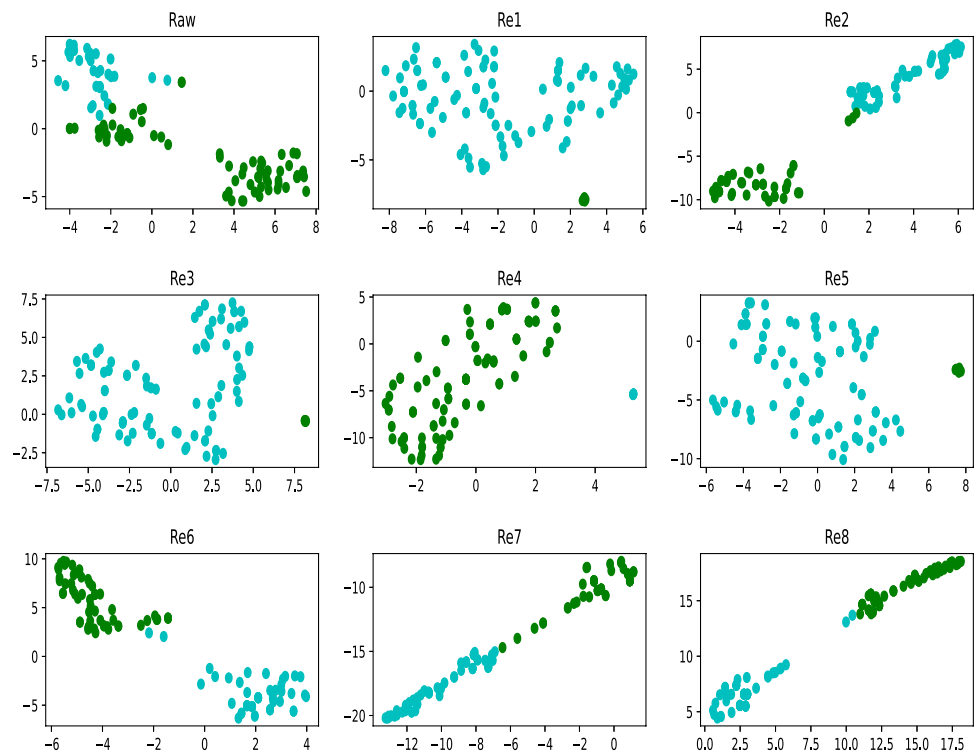
Fig. 22 The reduced clustering image of Algorithm 1 on Ac
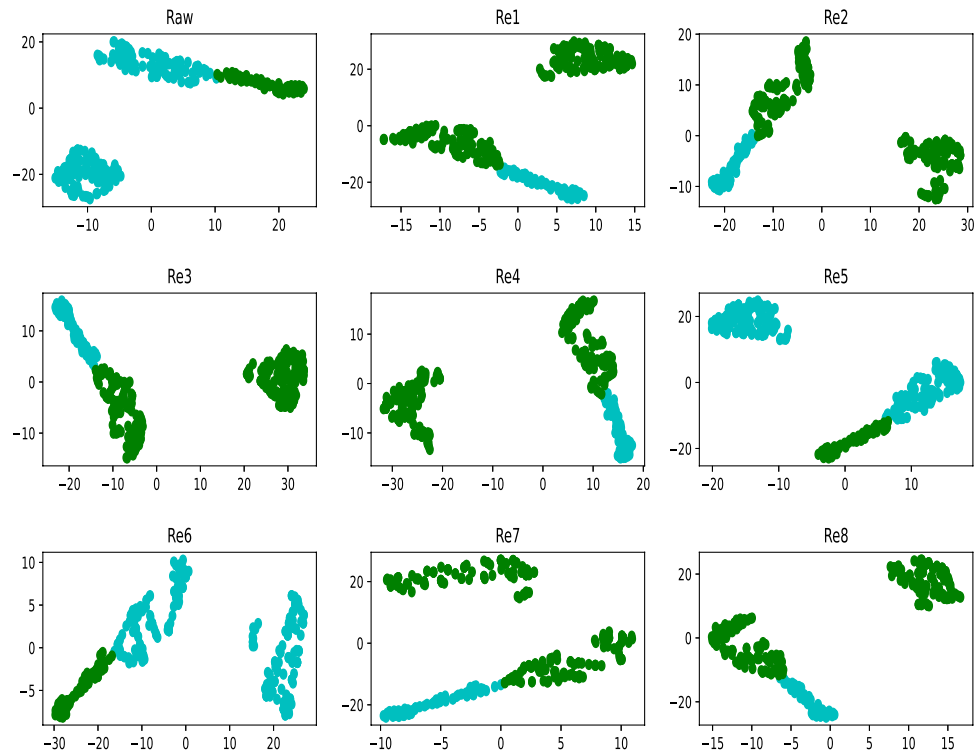
6.1.1 k-means cluster

we first use k-means clustering algorithm and the data before and after reduction to cluster. Then, we use three indexes, namely silhouette coefficient [27], calinski-harabasz index [1] and daviesbouldin index [9] to evaluate the clustering effect, so as to verify the effectiveness of the proposed algorithms. Among the three indicators, the larger the silhouette

Fig. 23 The reduced clustering image of Algorithm 1 on De
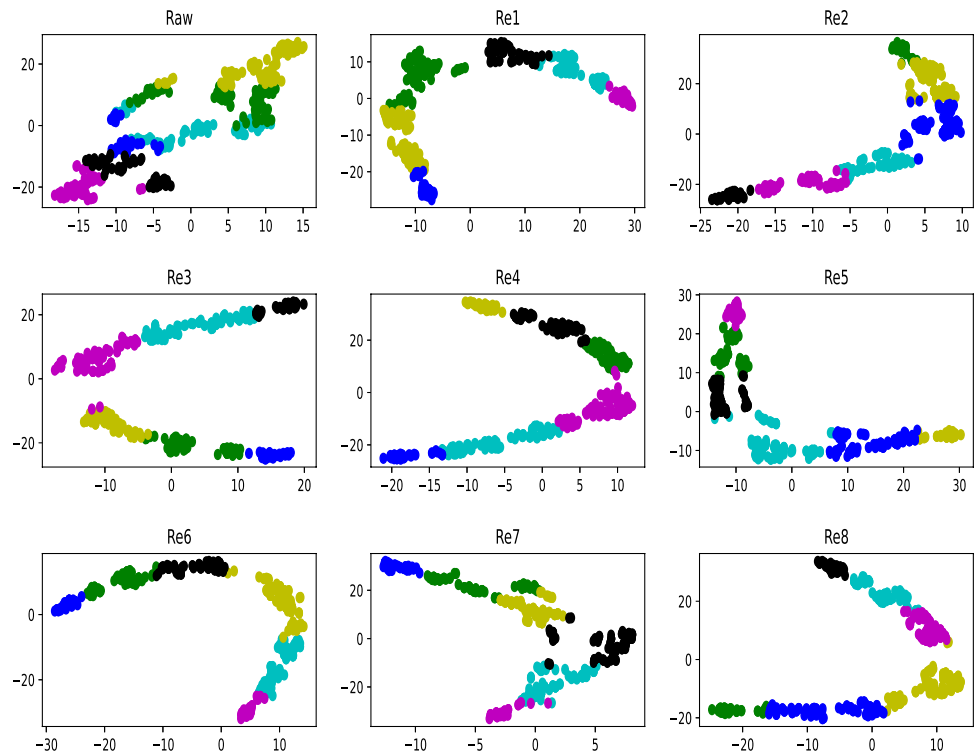
Fig. 24 The reduced clustering image of Algorithm 1 on He
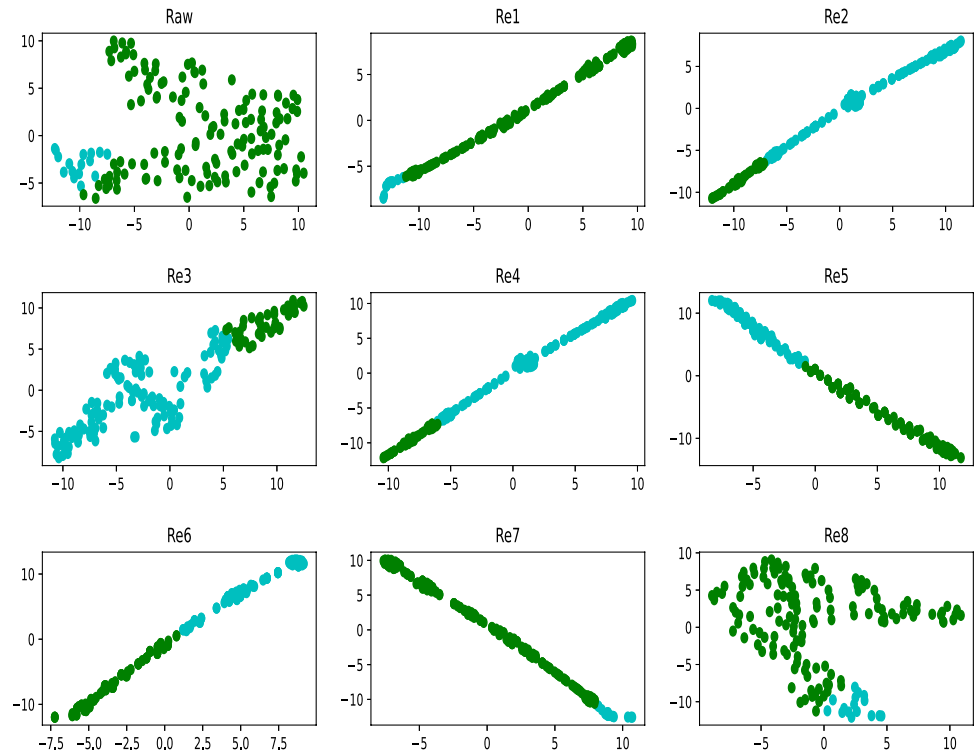
coefficient and calinski-harabasz index are, the better clustering effect is, while the daviesbouldin index is on the contrary. In order to make the experiment more reasonable, we set the number of clustering to the number of categories inherent in the data.

Clustering results are visualized by t-distributed stochastic neighbor embedding, as shown in Figs. 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36. The three indices values corresponding to the clustering results of Algorithms 1–2 are shown in Figs. 37, 38, 39, 40, 41, 42, 43, 44.

Fig. 25 The reduced clustering image of Algorithm 1 on Pc
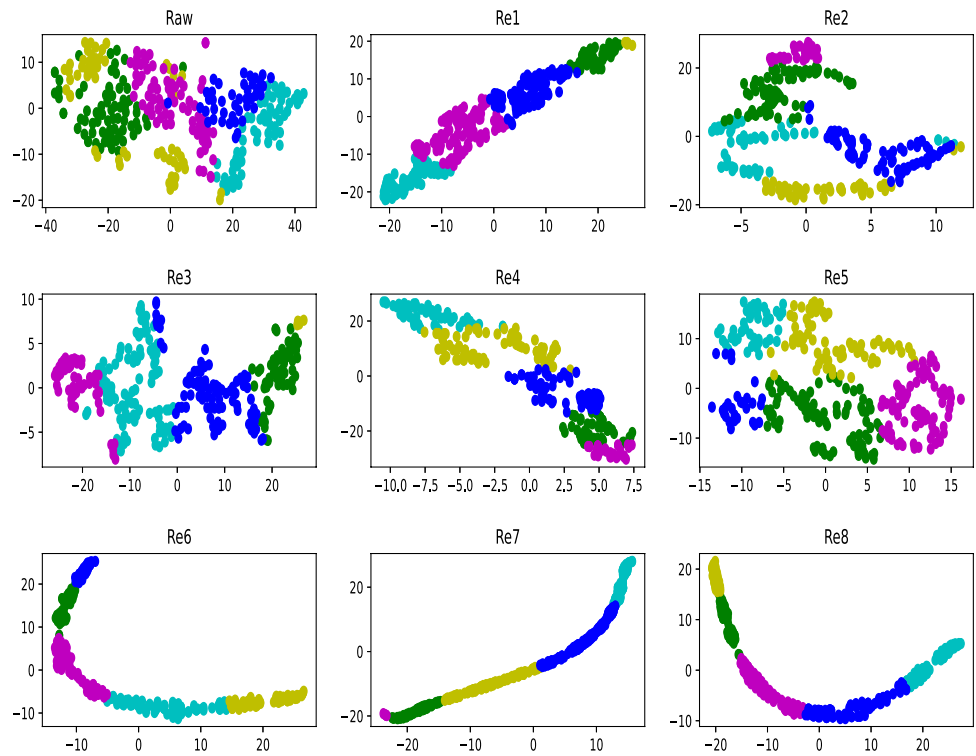
Fig. 26 The reduced clustering image of Algorithm 1 on Sl
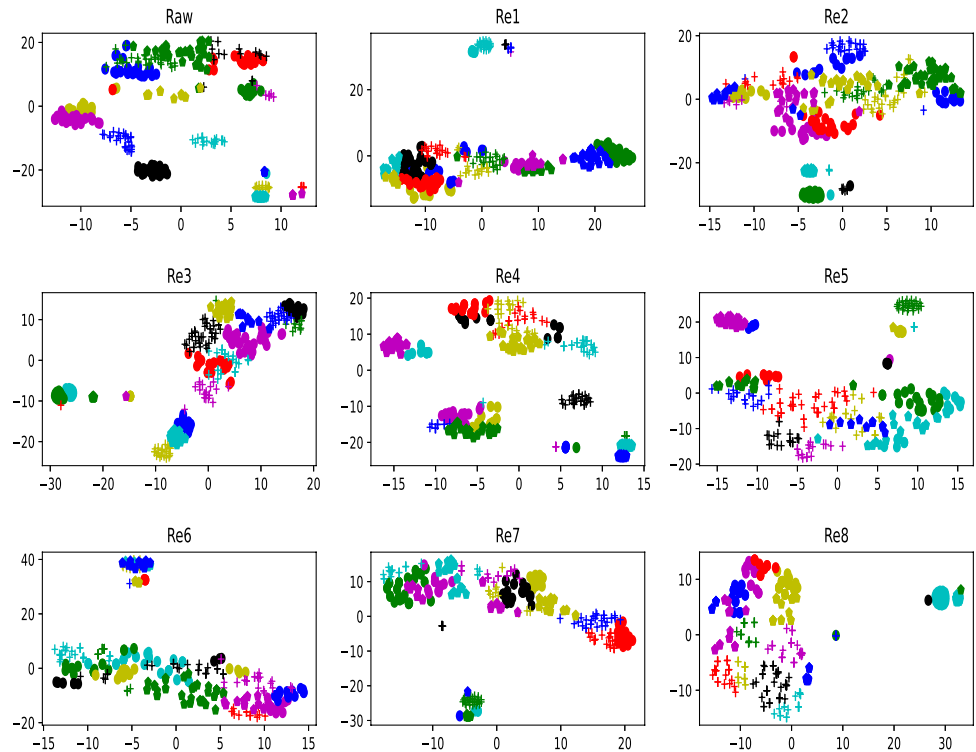
6.1.2 Mean Shift

we first use Mean Shift clustering algorithm and the data before and after reduction to cluster. Then, we also evaluate the clustering effect with these three indexes to verify the effectiveness of the proposed algorithms.

Clustering results are visualized by t-distributed stochastic neighbor embedding, as shown in Figs. 45, 46, 47, 48, 49,

Table 13 Reduction results of Algorithm 1 at the incomplete rates of $\beta = 0.02k$ $(k = 1, 2, \ldots, 8)$

Date sets	Reduction results at the incomplete rate of $0.02k, k = 1, 2, \ldots, 8$
Aa	Re1={1, 4, 11, 13, 18, 19}, Re2={8, 13, 16, 18}, Re3={2, 6, 9, 16, 18}
	Re4={1, 7, 10, 12, 17, 18, 19}, Re5={11, 18, 19, 20}, Re6={8, 12, 16, 18}
	Re7={2, 14, 16, 20}, Re8={2, 4, 8, 9, 16}
Ac	Re1={2, 11, 16, 18}, Re2={2, 9, 10, 13, 16, 18}, Re3={1, 7, 11, 16, 18, 20}
	Re4={8, 10, 13, 14, 16, 17, 18}, Re5={4, 5, 10, 11, 16, 18}, Re6={2, 11, 12, 14, 16}
	Re7={7, 11, 15, 16, 17}, Re8={5, 10, 12, 15, 16, 18, 20}
De	Re1={8, 10, 12, 14, 24, 31, 34}, Re2={14, 22, 25, 33, 34}, Re3={8, 9, 11, 17, 30, 34}
	Re4={2, 4, 11, 34}, Re5={6, 12, 23, 26, 29, 34}, Re6={1, 11, 27, 34}
	Re7={5, 12, 27, 29, 34}, Re8={4, 27, 34}
He	Re1={5, 7, 16, 17}, Re2={6, 11, 13, 15}, Re3={1, 2, 10, 15}, Re4={9, 15}
	Re5={1, 5, 12}, Re6={11, 18}, Re7={6, 16}, Re8={15, 16}
Pc	Re1={1, 3, 5}, Re2={1, 4, 6, 9}, Re3={2, 4, 5, 10}, Re4={1, 7, 10}
	Re5={4, 8, 12}, Re6={3, 6, 8, 12}, Re7={5, 7}, Re8={2, 7, 8, 9}
Sl	Re1={1, 4, 5, 6, 8, 9, 11, 21, 26, 32}, Re2={1, 2, 3, 4, 5, 6, 7, 9, 12, 13, 16}
	Re3={1, 2, 4, 5, 6, 9, 15, 29, 30, 33, 34}, Re4={1, 5, 6, 7, 16, 21, 22, 26, 28}
	Re5={1, 2, 3, 4, 9, 11, 19, 27, 28, 34}, Re6={1, 4, 6, 9, 20, 24, 26, 27}
	Re7={4, 5, 6, 7, 10, 21, 23}, Re8={1, 5, 7, 10, 25, 34}
Ob	Re1={2, 3, 10}, Re2={3, 8, 10, 16}, Re3={1, 5, 10, 12, 14, 15}
	Re4={3, 10, 15}, Re5={3, 15}, Re6={1, 2, 5}, Re7={1, 9, 11, 12}, Re8={2, 8, 10}
Gw	Re1={2, 3, 8, 11}, Re2={2, 3, 6, 8, 10, 11, 12, 13}, Re3={3, 8}
	Re4={6, 8}, Re5={6, 7, 8, 13}, Re6={8, 10}, Re7={4, 8}, Re8={4, 10, 14}

Table 14 Reduction results of Algorithm 2 at the incomplete rates of $\beta = 0.02k, k = 1, 2, \ldots, 8$

Date sets	Reduction results at the incomplete rate of $0.02k, k = 1, 2, \ldots, 8$
Aa	Re1={1, 11, 13, 16, 18, 20}, Re2={8, 11, 13, 16, 18, 20}
	Re3={1, 2, 11, 13, 16, 18}, Re4={7, 8, 11, 13, 16, 18}
	Re5={2, 3, 5, 8, 11, 16, 18}, Re6={5, 8, 11, 16, 18, 20}
	Re7={2, 8, 11, 16, 18}, Re8={2, 11, 16, 18}
Ac	Re1={4, 6, 7, 10, 11, 13, 14, 16, 18, 20}, Re2={2, 8, 11, 13, 14, 16, 17, 18}
	Re3={1, 2, 4, 11, 13, 14, 16, 18}, Re4={1, 3, 4, 8, 11, 12, 13, 16, 18}
	Re5={1, 2, 3, 9, 11, 13, 14, 16, 18}, Re6={3, 4, 6, 9, 11, 13, 16, 18}
	Re7={7, 11, 13, 14, 16, 18}, Re8={4, 8, 11, 13, 16, 18}
De	Re1={3, 4, 18, 19, 23, 32, 34}, Re2={3, 4, 16, 17, 18, 19, 34}
	Re3={1, 2, 3, 4, 19, 21, 34}, Re4={4, 5, 19, 26, 32, 34}
	Re5={3, 4, 9, 14, 16, 32, 34}, Re6={4, 14, 16, 19, 33, 34}
	Re7={2, 4, 5, 16, 19, 34}, Re8={2, 4, 6, 16, 24, 34}
He	Re1={1, 14, 16}, Re2={1, 15, 16},Re3={1, 16, 18}, Re4={1, 14, 18}
	Re5={14, 16, 18}, Re6={15, 16, 18}, Re7={10, 16, 18}, Re8={1, 18}
Pc	Re1={1, 5, 7}, Re2={1, 5, 13}, Re3={4, 5, 8}, Re4={1, 5}
	Re5={1, 5, 10}, Re6={1, 3, 5}, Re7={1, 5}, Re8={5, 8, 12}
Sl	Re1={1, 4, 6, 7, 8, 9, 10, 13, 16, 22}, Re2={1, 3, 4, 6, 7, 8, 9, 10, 15, 16}
	Re3={1, 3, 6, 7, 8, 9, 10, 14, 16, 18, 22}
	Re4={1, 2, 3, 6, 7, 8, 9, 10, 12, 13, 14, 17, 18, 21, 22, 28, 30, 31}
	Re5={1, 3, 4, 6, 7, 9, 10, 14}, Re6={1, 3, 6, 7, 8, 10, 15, 22, 28}
	Re7={1, 4, 6, 7, 9, 10, 21, 29}, Re8={1, 6, 7, 8, 10, 14, 29}
Ob	Re1={2, 3, 4, 14, 15}, Re2={2, 3, 4, 5, 15}, Re3={2, 3, 4, 5, 14}
	Re4={2, 3, 4, 5, 9}, Re5={2, 3, 4, 5, 8}, Re6={2, 3, 4, 14}, Re7={2, 3, 4, 14}, Re8={2, 3, 4, 15}
Gw	Re1={1, 5, 8}, Re2={1, 5, 8},Re3={4, 6, 8, 9, 11}
	Re4={1, 5, 8}, Re5={1, 5, 8}, Re6={1, 8, 10}, Re7={5, 8, 9}, Re8={1, 3, 8}

Fig. 27 The reduced clustering image of Algorithm 1 on Ob
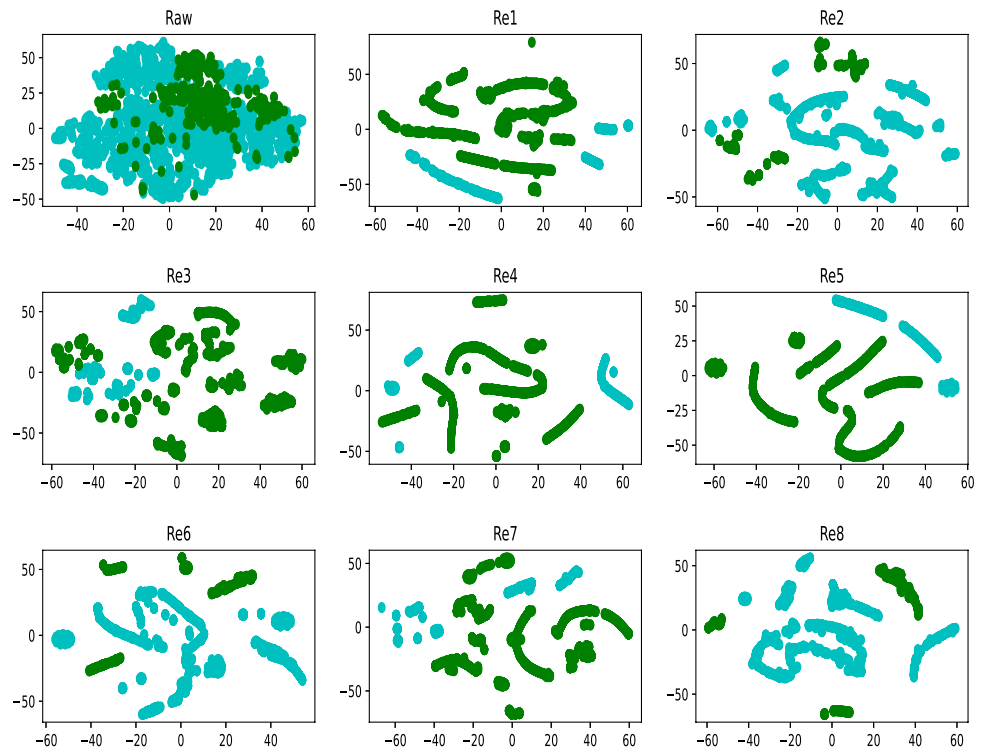
Fig. 28 The reduced clustering image of Algorithm 1 on Gw
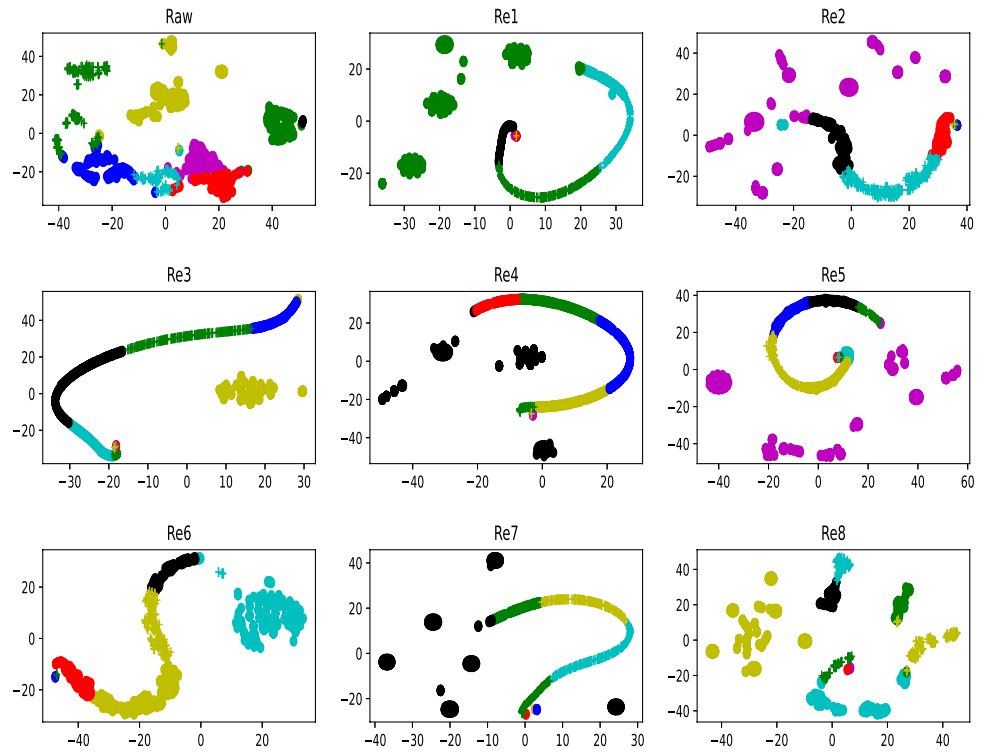
50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60. The three indices values corresponding to the clustering results of Algorithms 1–2 are shown in Figs. 61, 62, 63, 64, 65, 66, 67, 68.

6.1.3 Analysis of clustering results

According to 37-44 and 61 -68, the following findings can be made:

Fig. 29 The reduced clustering image of Algorithm 2 on Aa
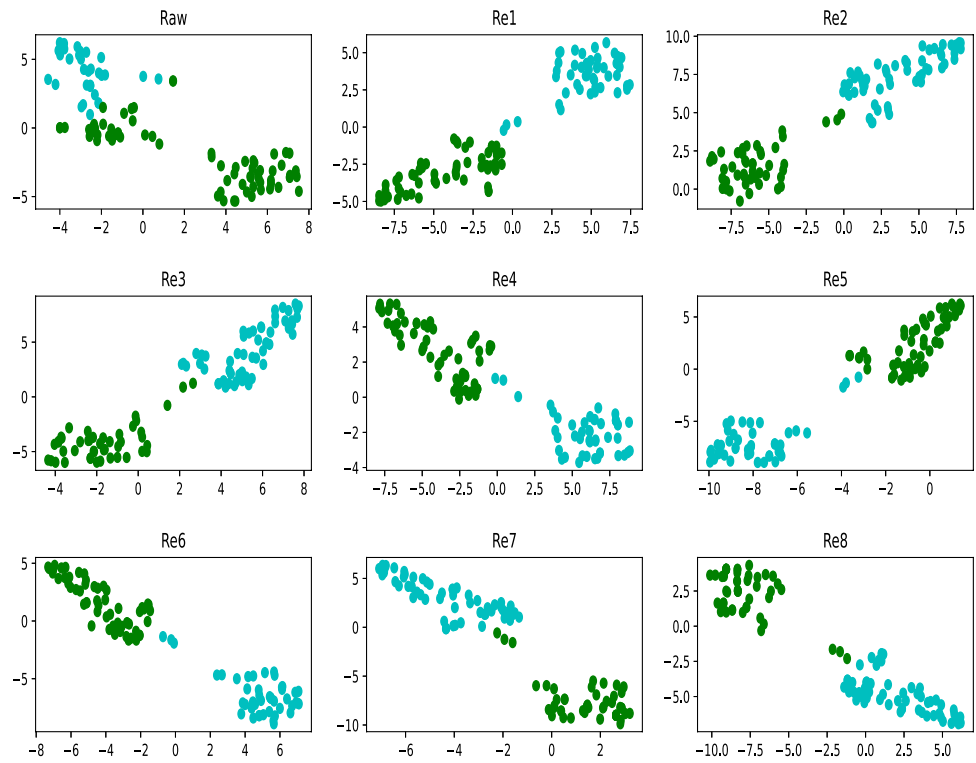
Fig. 30 The reduced clustering image of Algorithm 2 on Ac
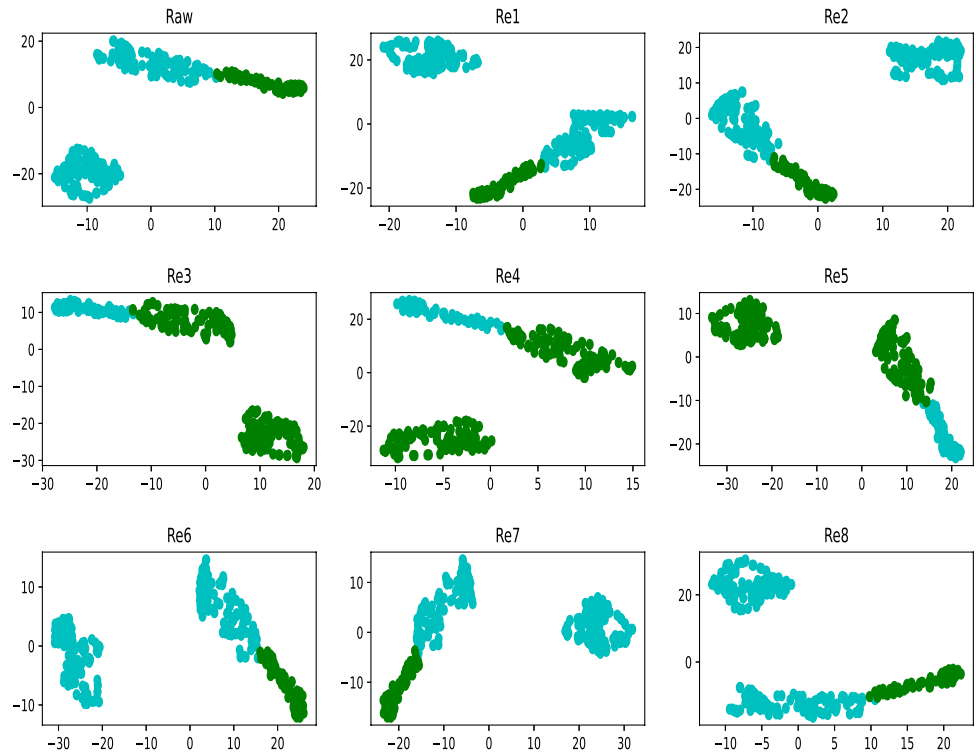
(1) For Aa and De, the results of the two clustering methods under the three indicators are superior to the original data.

(2) For Ac, only the calinski-Harabasz index under the Mean Shift clustering algorithm shows the reducts

of Algorithm 2 have little difference with the original data. Besides, other indices indicate all reducts are superior to the original data.

(3) For He, the three indices under the Mean Shift clustering algorithm show that all the reducts are better than

Fig. 31 The reduced clustering image of Algorithm 2 on De
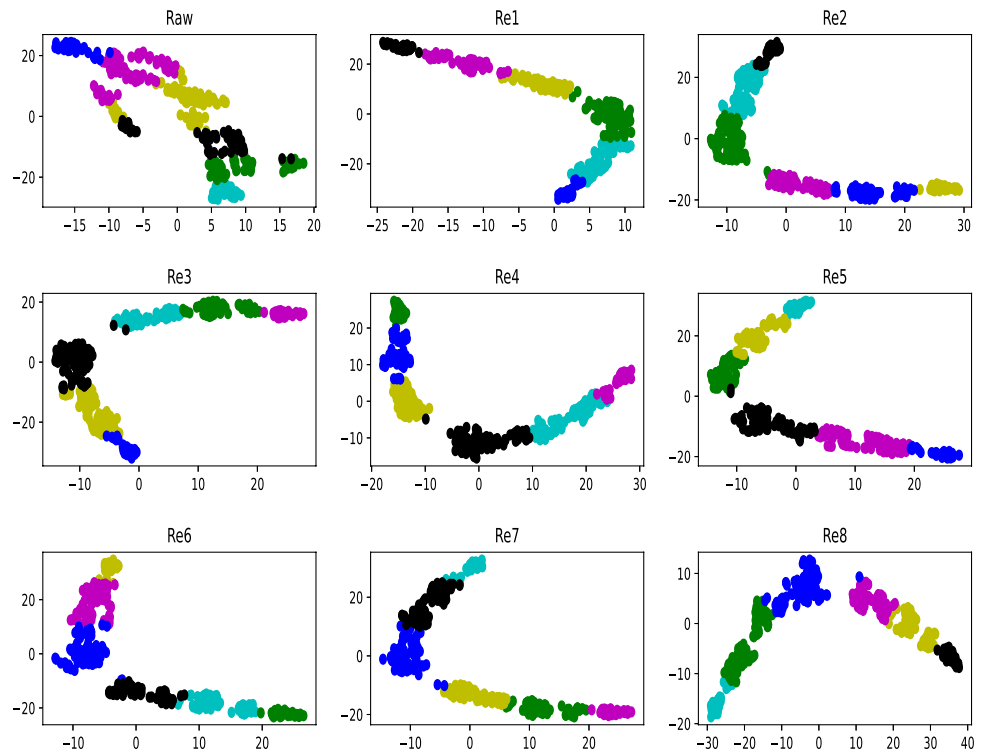
Fig. 32 The reduced clustering image of Algorithm 2 on He
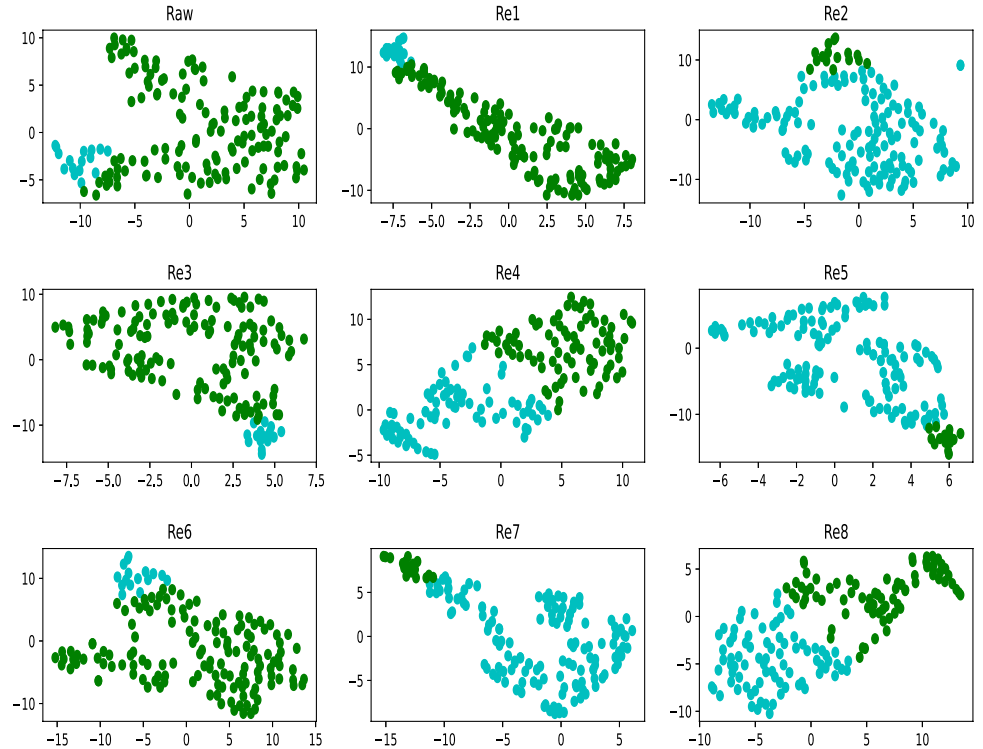
the original data. In addition, the silhouette coefficient index and daviesbouldin index of Algorithm 2's R4 and R8 under k-means clustering algorithm are worse than the original data, while other reducts are superior to the original data.

(4) For Pc, the three indices under the k-means clustering algorithm show that all the reducts are better than the original data. The calinski-Harabasz index under the Mean Shift clustering algorithm shows that the R2–R4, R6–R8 of Algorithm 1 and the R6–R8 of Algorithm 2

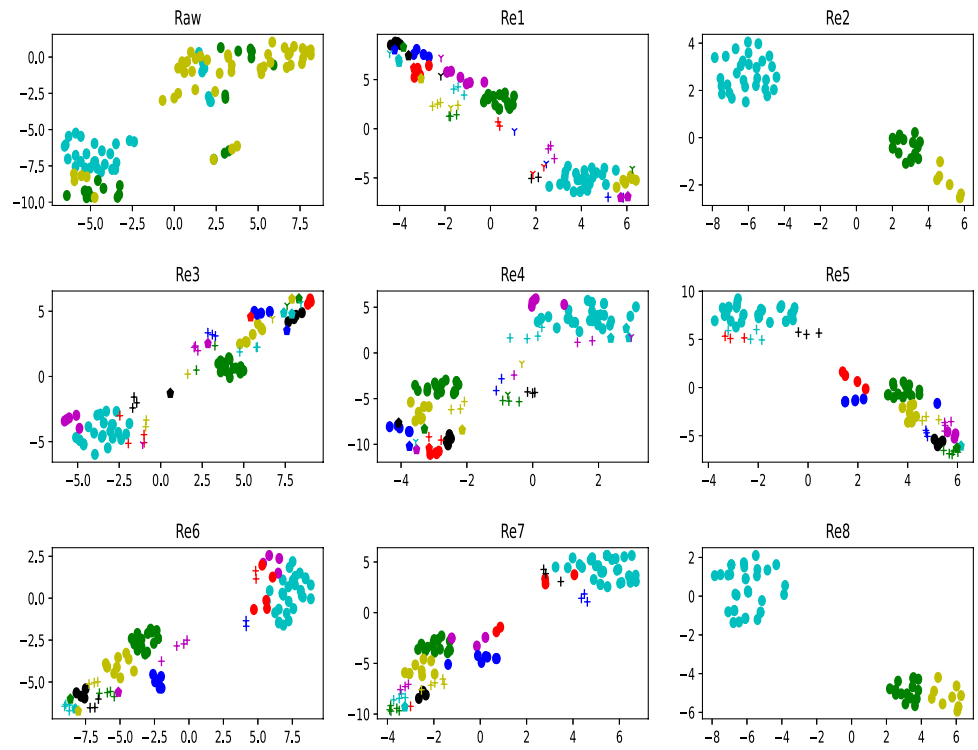**Fig. 33** The reduced clustering image of Algorithm 2 on Pc
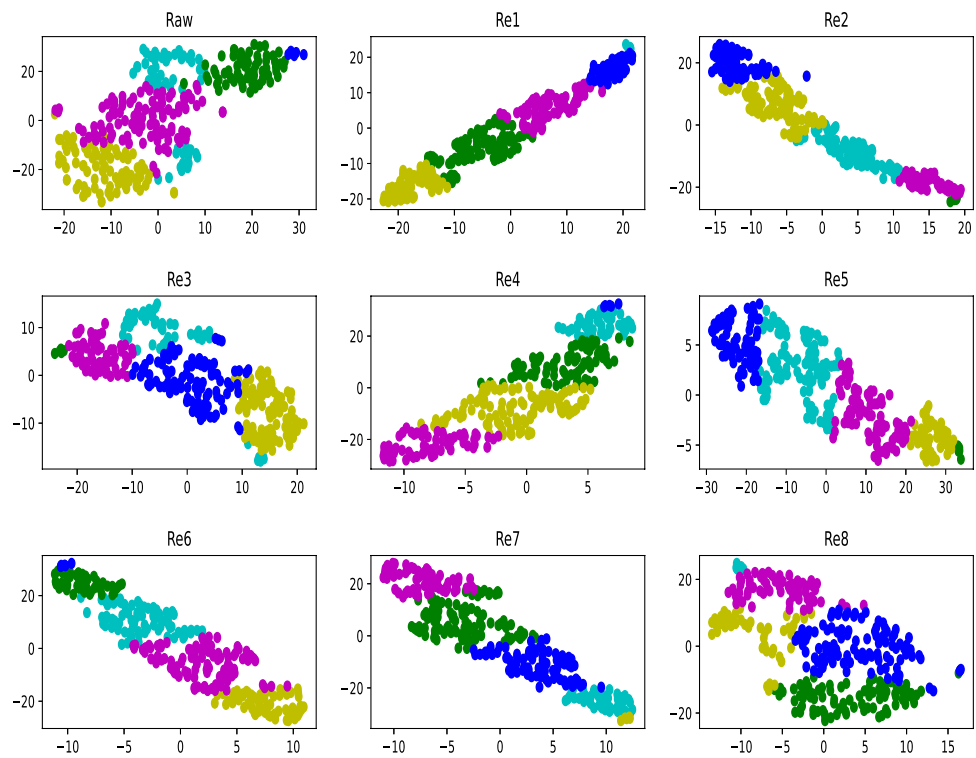
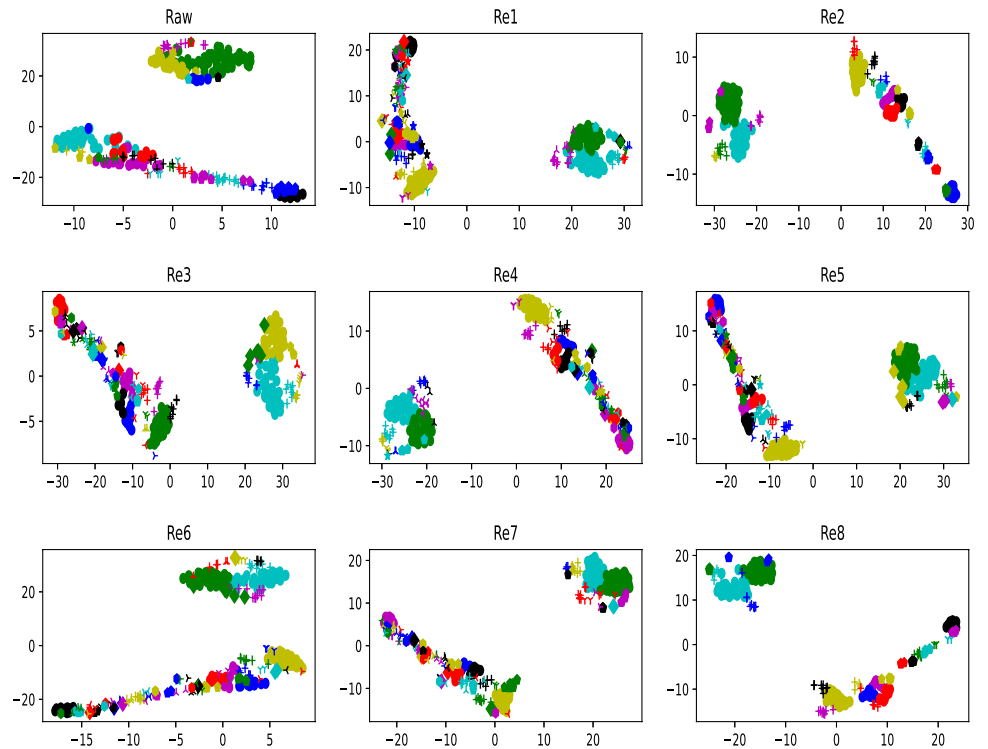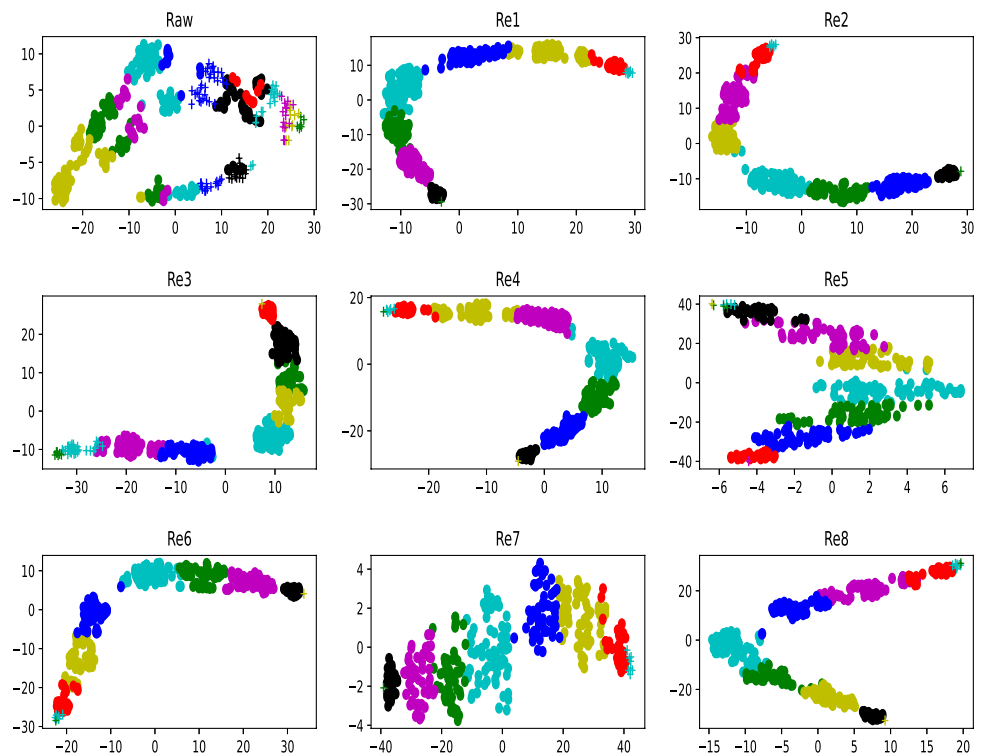**Fig. 34** The reduced clustering image of Algorithm 2 on Sl
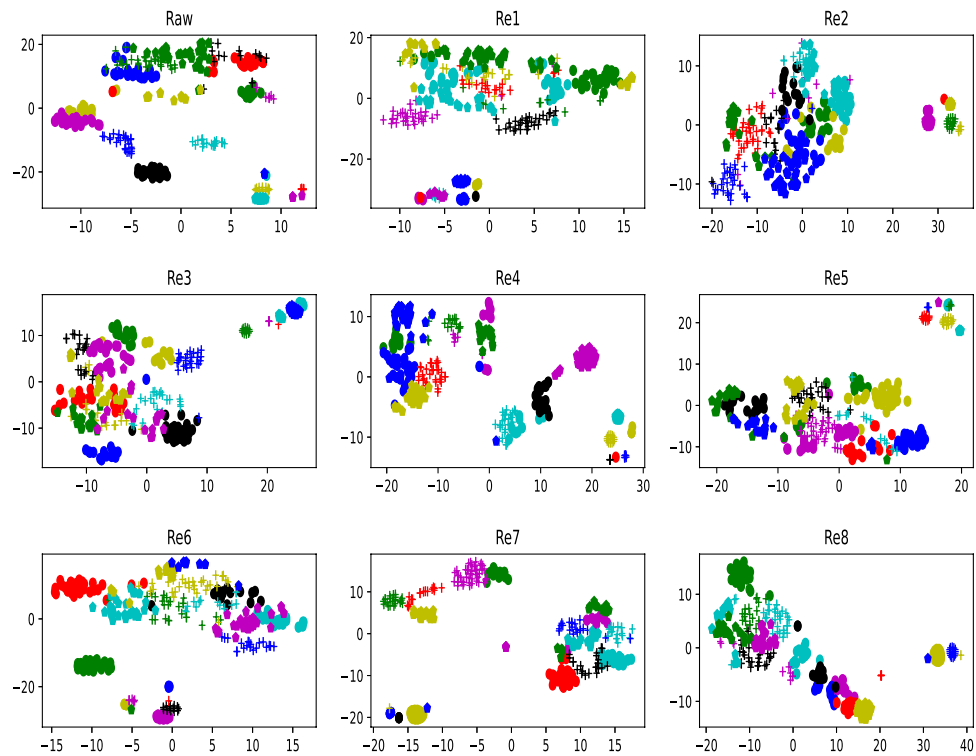
are better than the original data, while the other indicator has little difference with the original data. In addition, the other two indexes indicate that all the reducts are better than the original data.

(5) For Sl, the daviesbouldin index under the k-means clustering algorithm indicates that all the reducts are better than the original data. The coefficient index show that only the Re4 of Algorithm 2 has little difference with the original data, while other reducts are superior to

Fig. 35 The reduced clustering image of Algorithm 1 on Ob
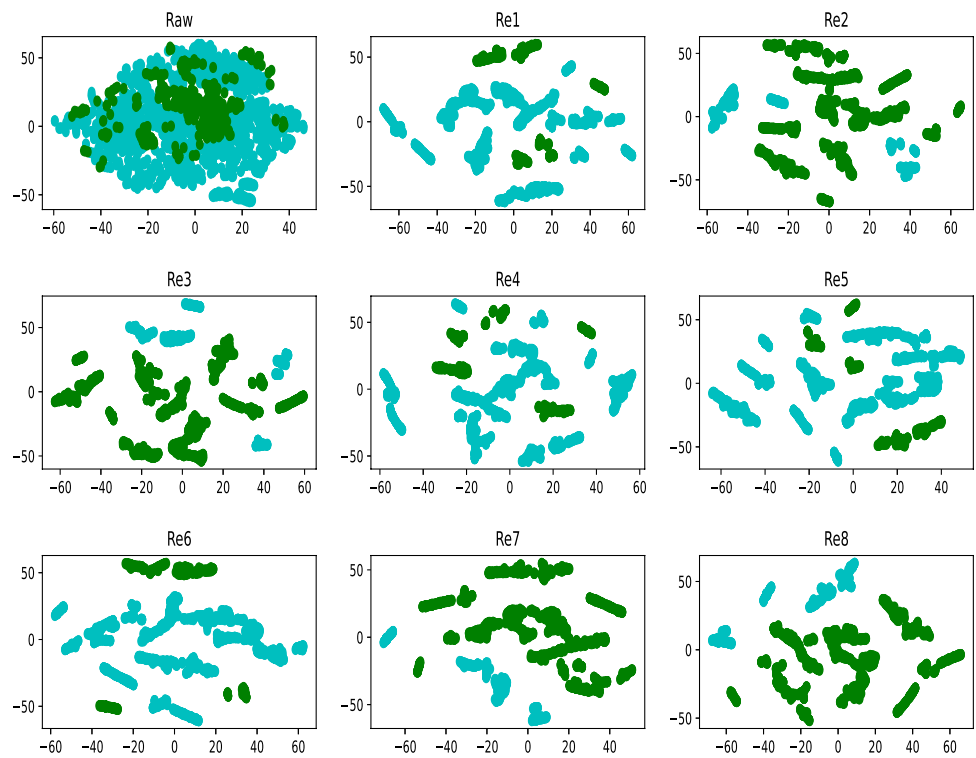
Fig. 36 The reduced clustering image of Algorithm 1 on Gw
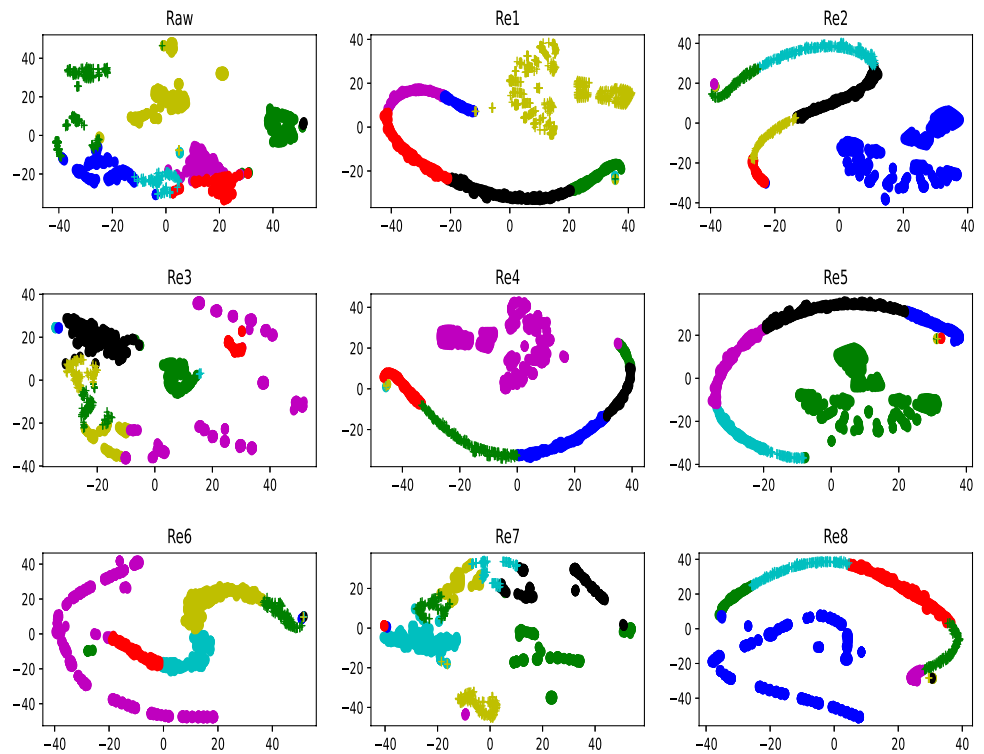
the original data. The silhouette coefficient index under the k-means clustering algorithm shows that the Re4, Re5, Re7 and Re8 of Algorithm 1 and Re7 and Re8 of

Algorithm 2 are better than the original data. The silhouette coefficient index and daviesbouldin index under the Mean Shift clustering algorithm show that all the
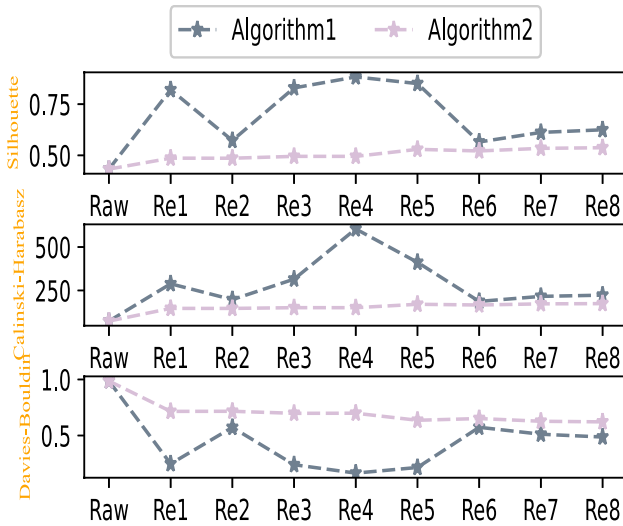
Fig. 37 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Aa
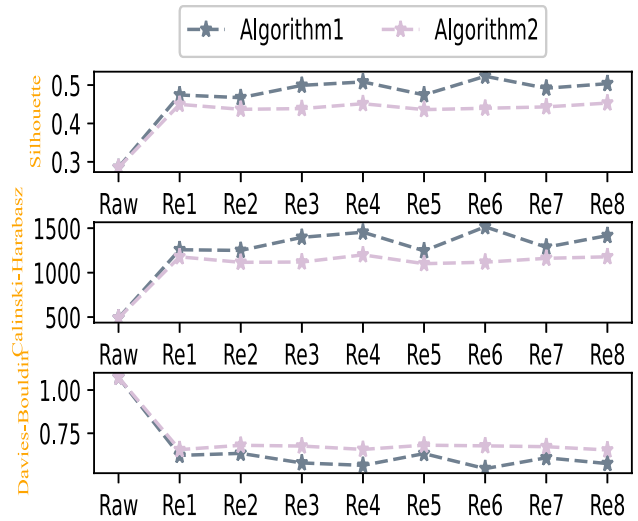
Fig. 39 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on De
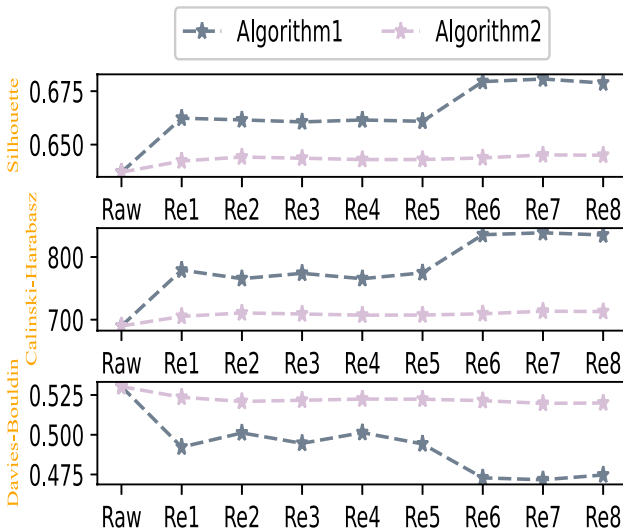
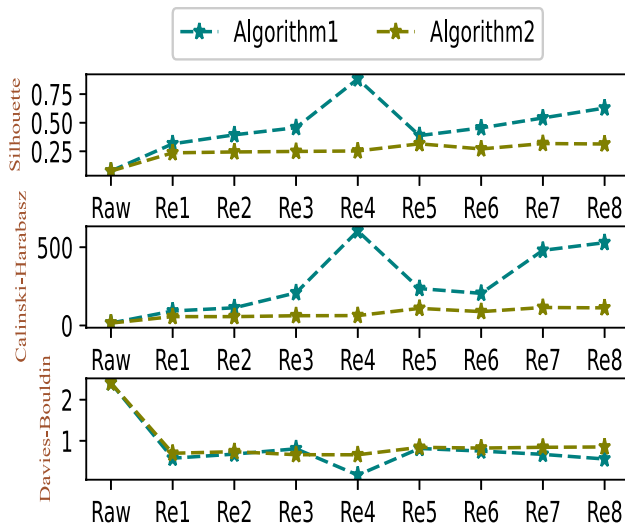**Fig. 38** The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Ac
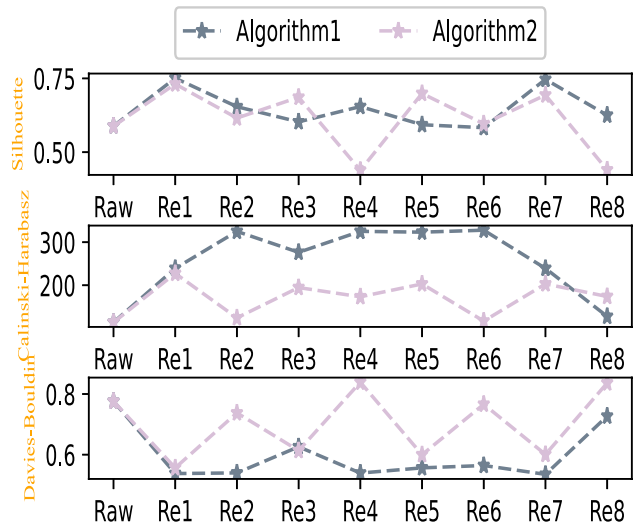
Fig. 40 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on He

reducts are better than the original data, while the other indicator has little difference with the original data.

(6) For Ob, the three indices under the k-means clustering algorithm show that all the reducts are better than the original data. The silhouette coefficient index and daviesbouldin index under the Mean Shift clustering algorithm show that all the reducts are better than the original data. The coefficient index show that only the Re4, Re5 and Re6 of Algorithm 1 are better than the

original data while other reductions are not much different from the original data.

(7) For Gw, the three indices under the k-means clustering algorithm show that all the reducts are better than the original data. The silhouette coefficient index and daviesbouldin index under the Mean Shift clustering algorithm show that all the reducts are better than the original data. The coefficient index indicates that the R2 of Algorithm 1 and the R3, R6 and R7 of Algorithm 2 are worse than the original data.
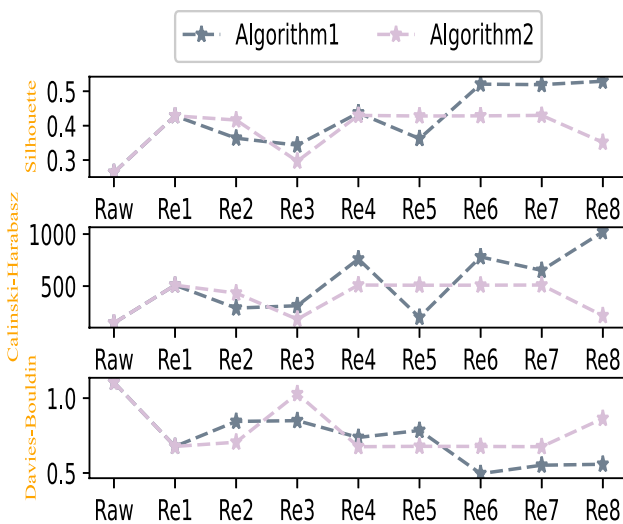
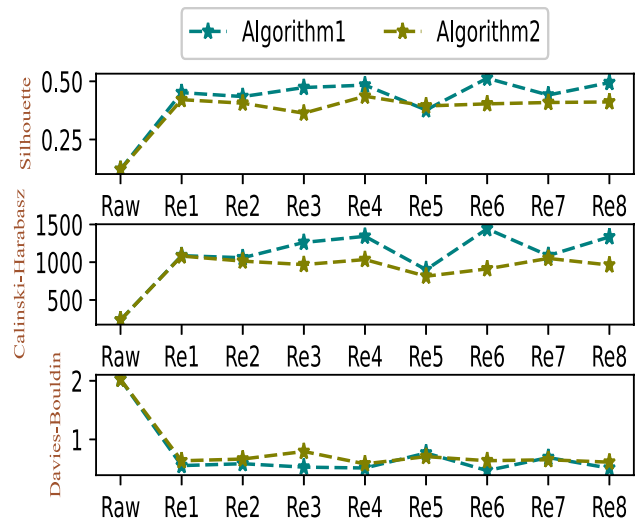**Fig. 41** The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Pc
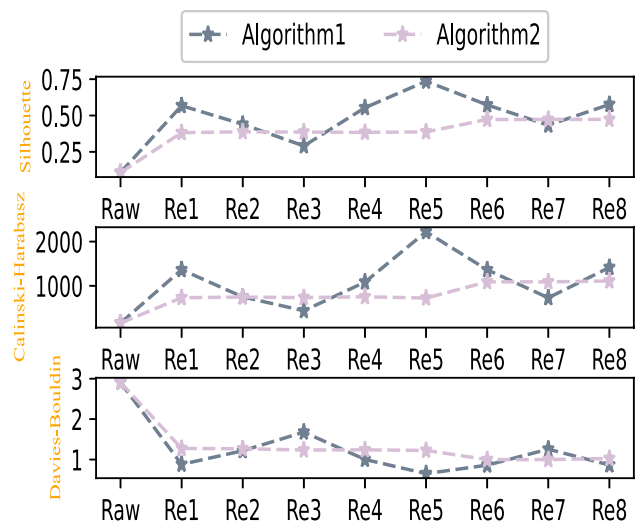
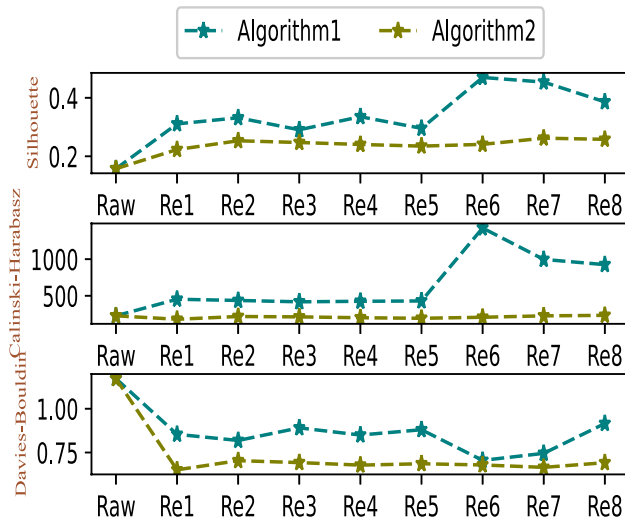**Fig. 43** The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Ob
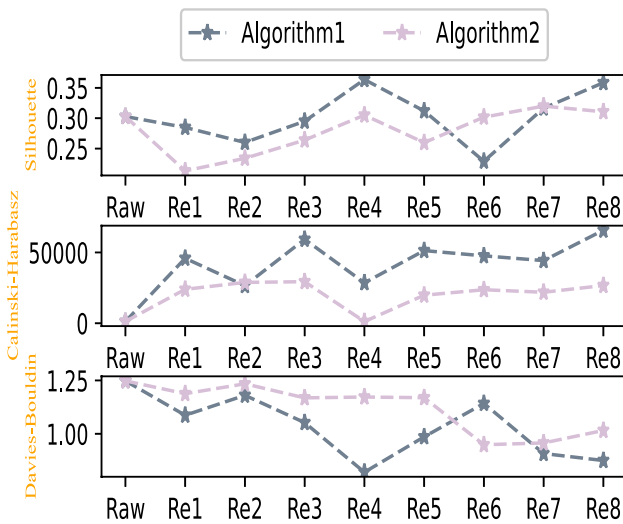
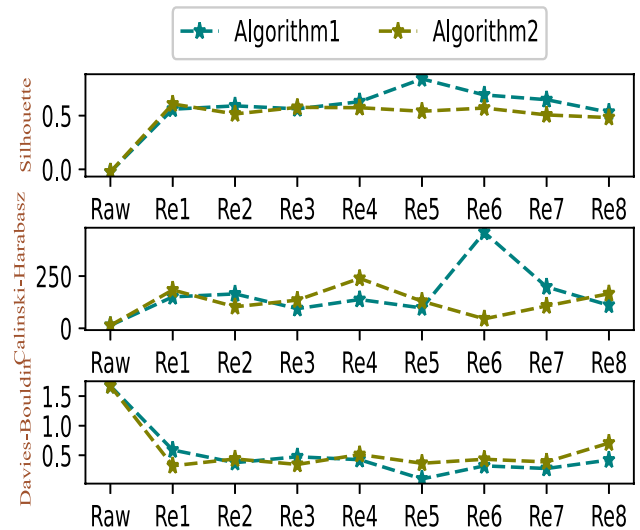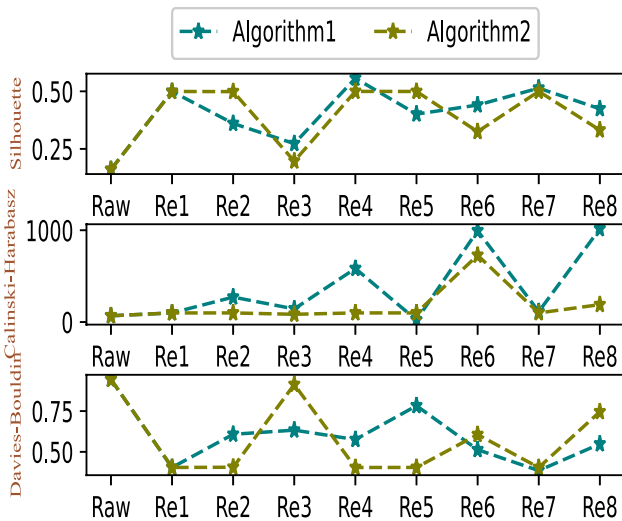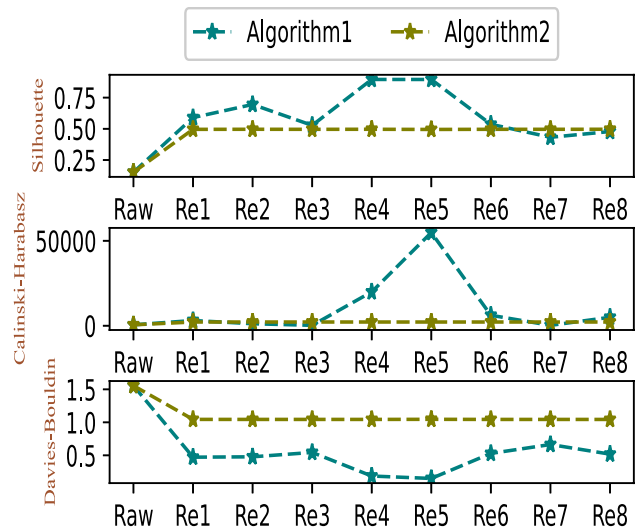**Fig. 42** The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Sl
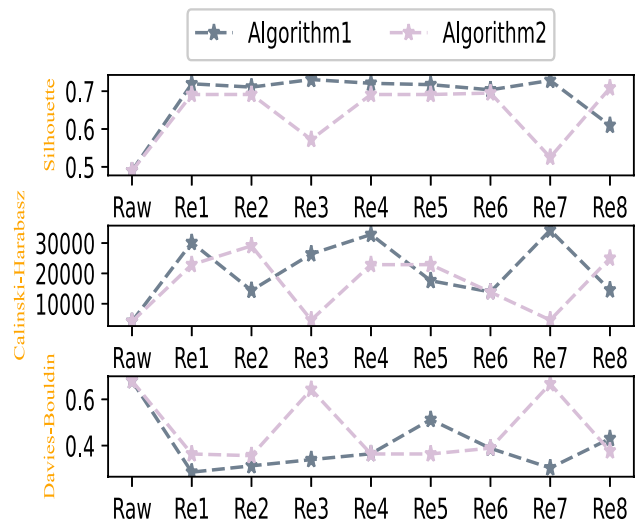
Fig. 44 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Gw

In a nutshell, the three indices values of those reducts are obviously better than those original data. Therefore, those reducts of Algorithms 1–2 are credible. This also means that the obtained algorithms can effectively perform attribute reduction at different miss rates, so it is of great significance to study the attribute reduction of these two algorithms at different miss rates. In addition, these three indices show that, under the same incomplete rate, the reduct of Algorithm 1 is better than Algorithm 2 in most cases. Therefore, Algorithm 1 can be preferred for attribute reduction to improve efficiency.

7 Comparison and discussion

In this subsection, we evaluate the performance of the proposed method and existing methods. We consider comparing our algorithm with two other algorithms. Dai et al. [8] brought up the two tools of entropy measure and granularity measure to measure the uncertainty of SVISs. They also explored the problem of attribute reduction for SVISs and proposed a representative attribute selection algorithm (FRSM) based on fuzzy rough sets. Liu et al. [21] researched feature selection for a set-valued decision IS from the view of dominance relations. They proposed a representative

Fig. 45 The reduced clustering image of Algorithm 1 on Aa
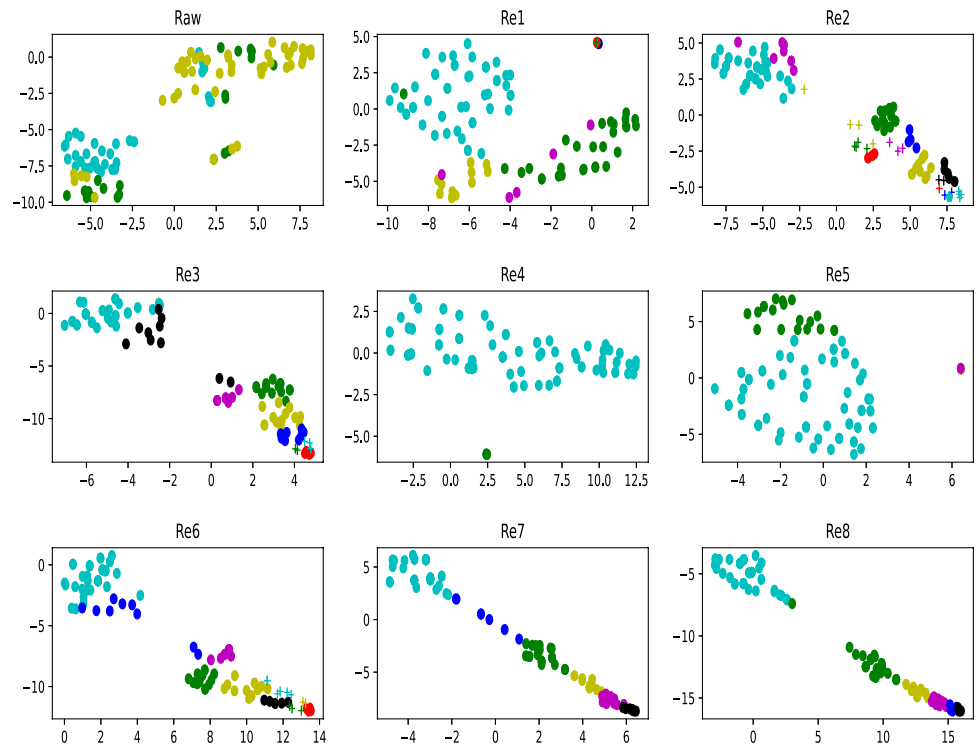
Fig. 46 The reduced clustering image of Algorithm 1 on Ac
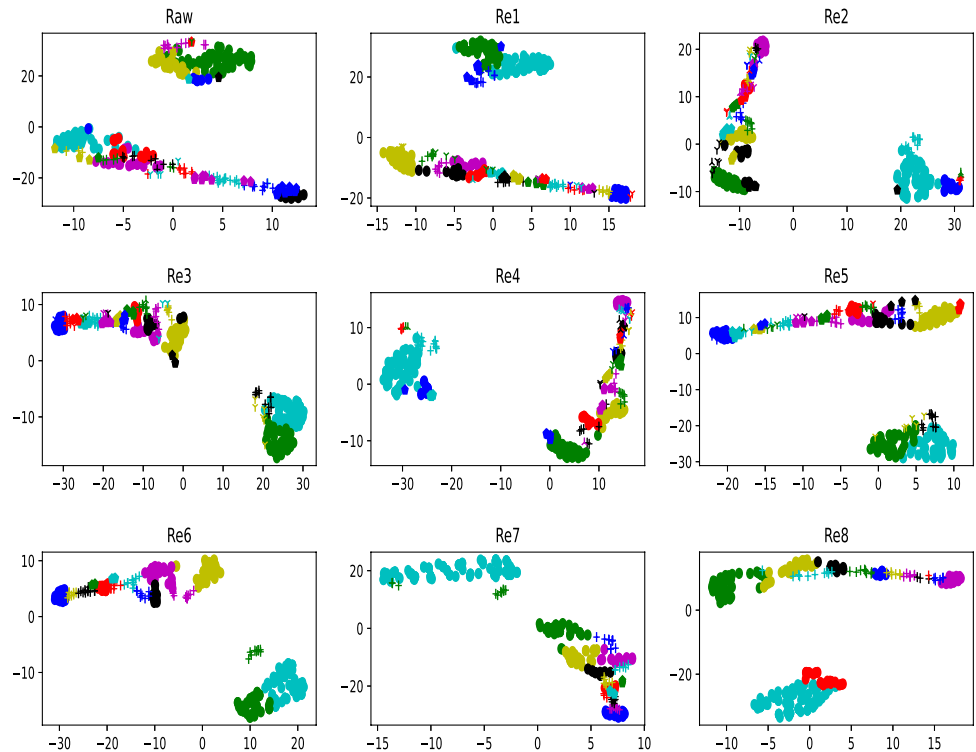
attribute selection algorithm (DRM) based on dominance relation.

Table 15 shows the reduction results of FRSM and DRM for the data in Table 6. By comparing the results in Tables 13, 14 and 15, we can see that the attributes of the reduction set obtained by Algorithms 1–2 are less than FRSM and DRM by in most case.

For Algorithms 1–2, we choose the reducts when the miss rate is 8%, and compare it with the reducts of FRSM and

Fig. 47 The reduced clustering image of Algorithm 1 on De
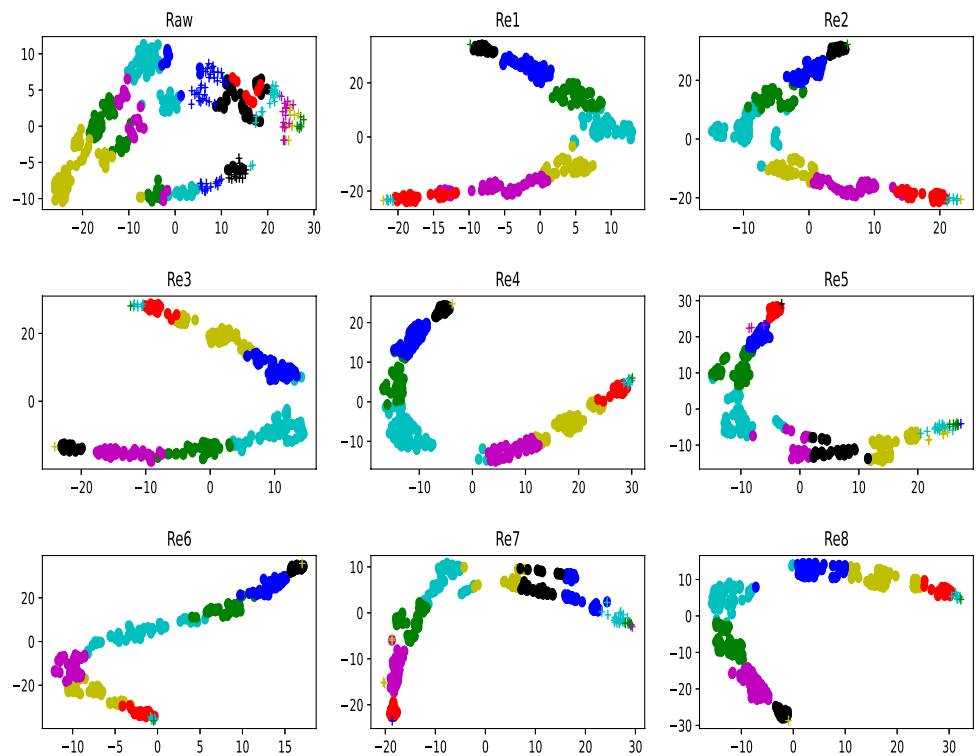
Fig. 48 The reduced clustering image of Algorithm 1 on He
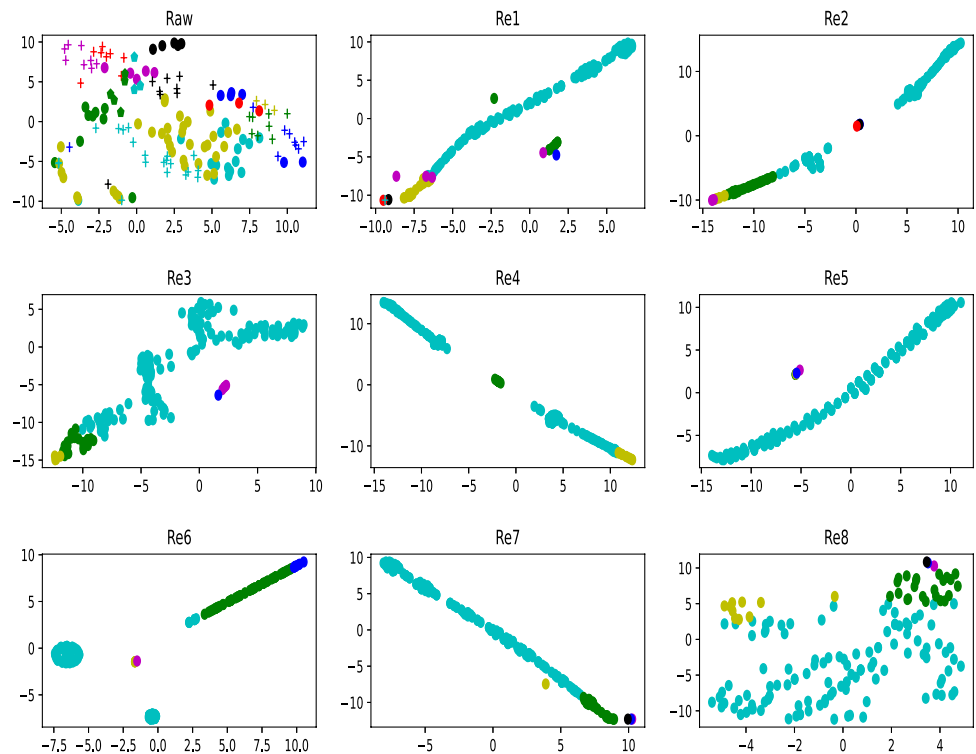
DRM. Through Mean Shift clustering algorithm and the three evaluation indexes, we get Fig. 69.

From Fig. 69, the following conclusions can be drawn. (1) The silhouette coefficient index shows that Algorithms 1–2 are superior to the other two algorithms in Aa, De, He

and Ob. In Ac and Gw, Algorithm 1 are superior to the other algorithms. Besides, in the other data, the results of the four algorithms are not very different. (2) The calinski-Harabasz index shows that Algorithms 1–2 are superior to the other two algorithms in Ob. In He, Algorithm 1 are superior to the

Fig. 49 The reduced clustering image of Algorithm 1 on Pc
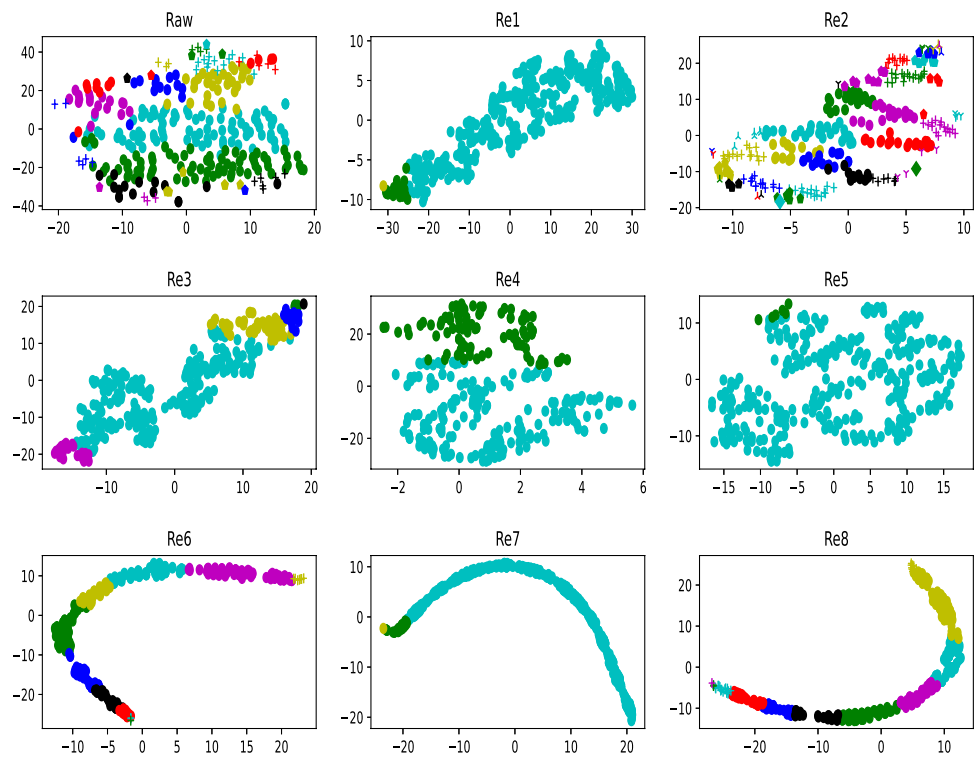
Fig. 50 The reduced clustering image of Algorithm 1 on Sl
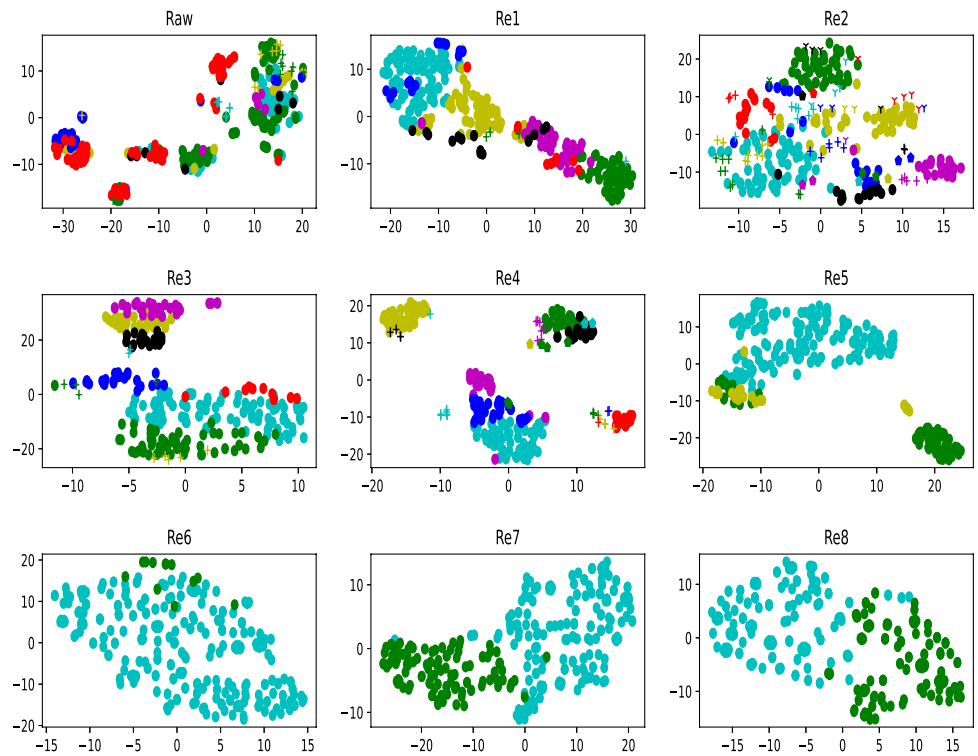
other algorithms. What's more, in the other data, the results of the four algorithms are not very different. (3) The davies-bouldin index shows that Algorithms 1–2 are superior to the other two algorithms in Aa, Ac, De and Ob. In the other data, the results of the four algorithms are not very different.

Therefore, Algorithms 1–2 are more effective in data reduction than FRSM and DRM algorithm. So it is meaningful to explore the application of ISVISs.

Fig. 51 The reduced clustering image of Algorithm 1 on Ob
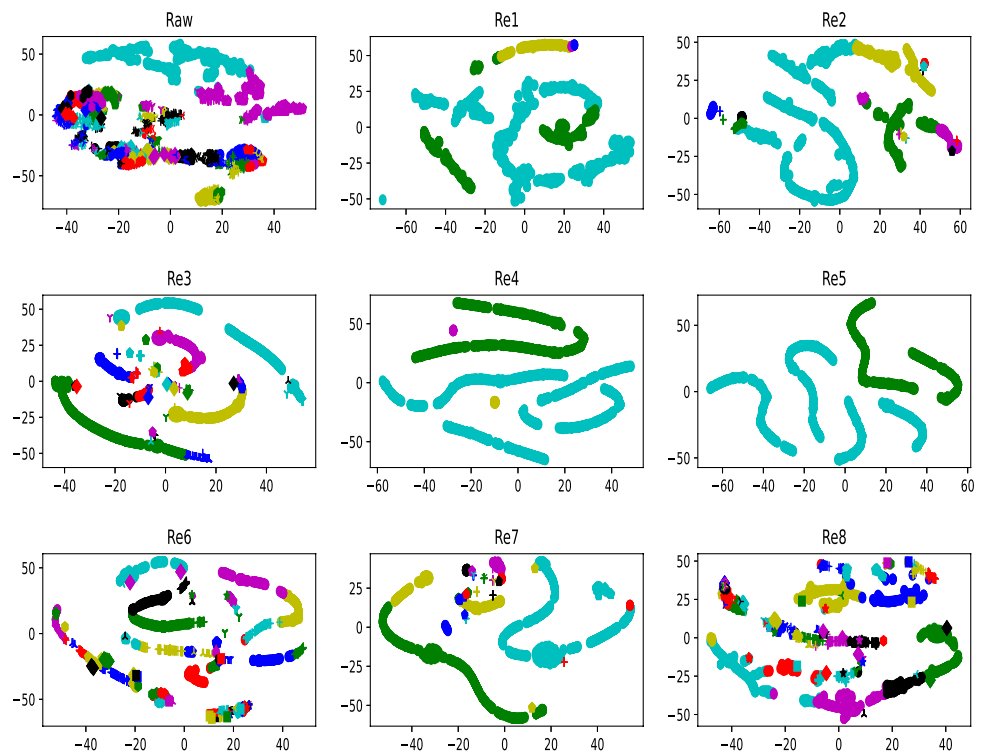
Fig. 52 The reduced clustering image of Algorithm 1 on Gw
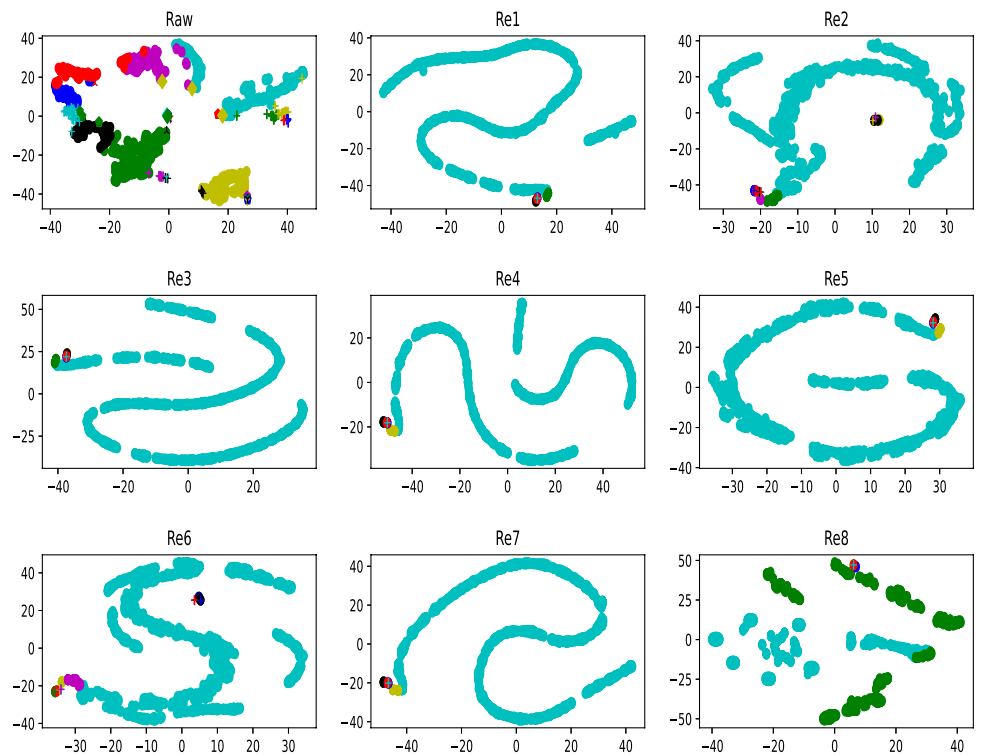
Fig. 53 The reduced clustering image of Algorithm 2 on Aa

Fig. 54 The reduced clustering image of Algorithm 2 on Ac

Fig. 55 The reduced clustering image of Algorithm 2 on De

Fig. 56 The reduced clustering image of Algorithm 2 on He
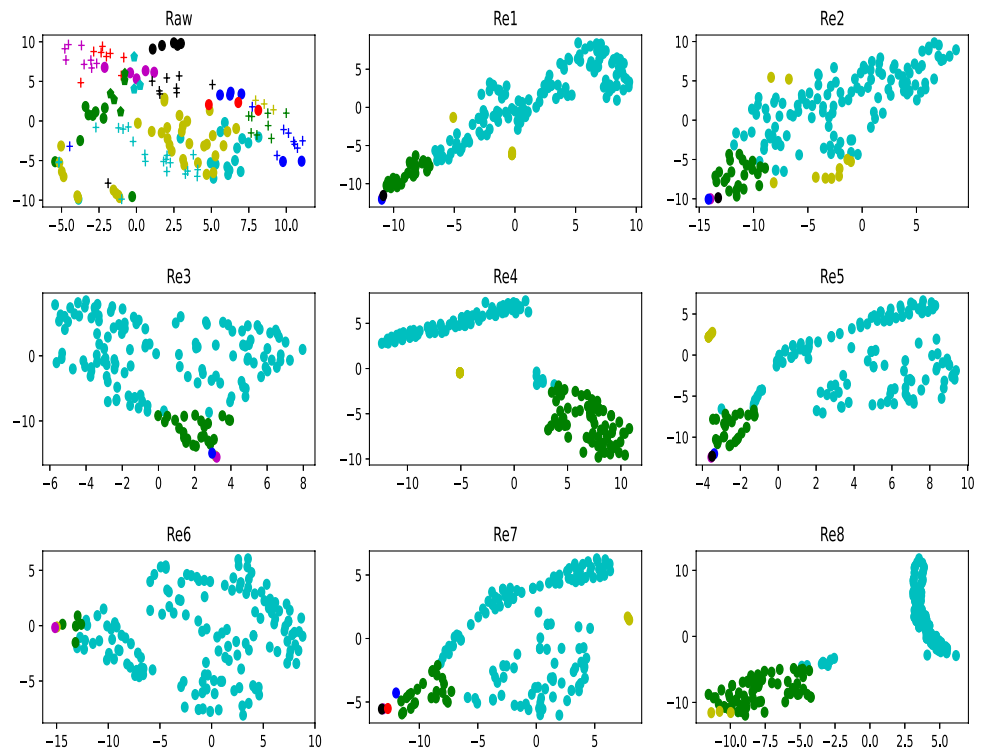
8 Conclusions

In this paper, incomplete set-valued data has been viewed as ISVISs. By this way, the similarity degree between information values on each attribute in incomplete set-valued data has been presented. The tolerance relation induced by each subsystem in an ISVIS has been given and rough approximations based on this relation have been investigated. To explore the potential information of this data, four

Fig. 57 The reduced clustering image of Algorithm 2 on Pc
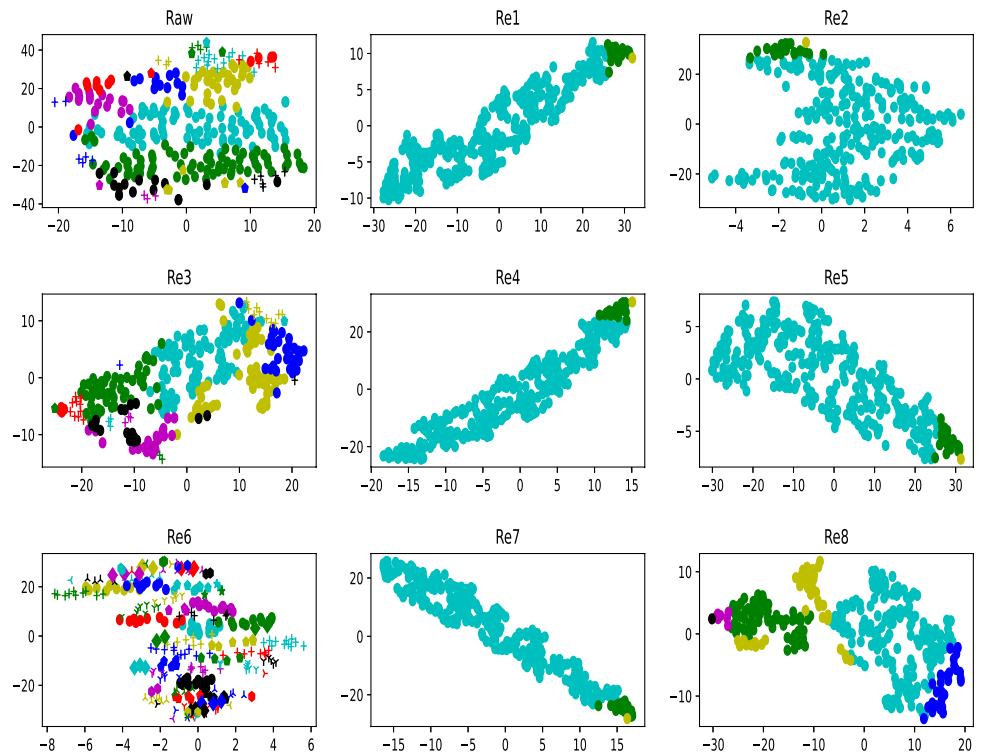
Fig. 58 The reduced clustering image of Algorithm 2 on Sl
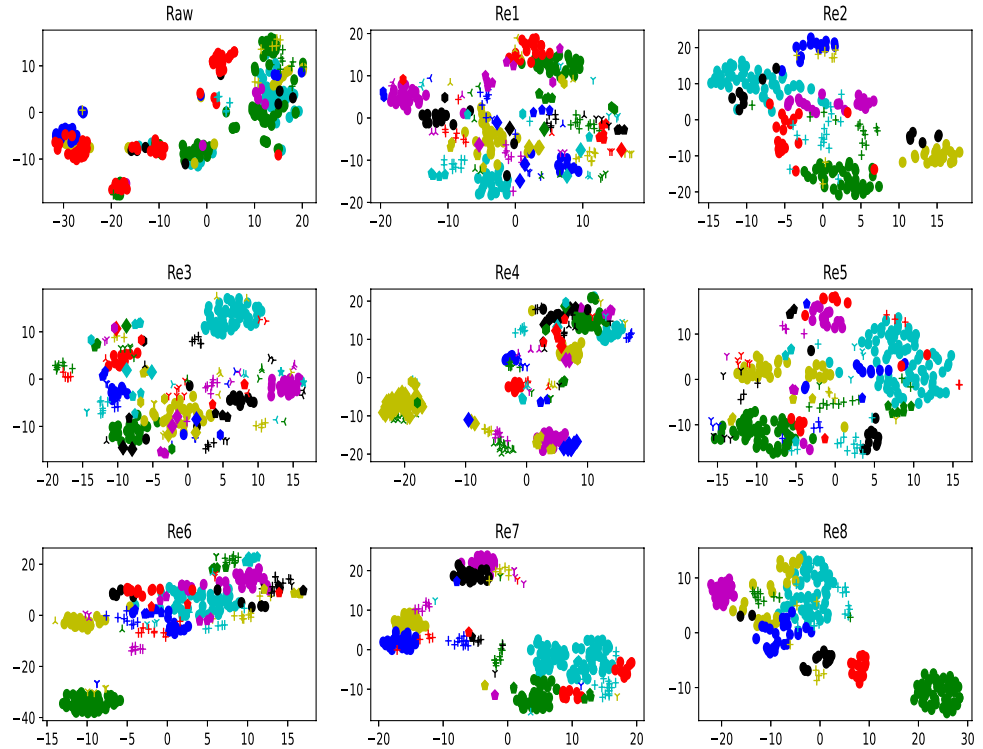
Fig. 59 The reduced clustering image of Algorithm 1 on Ob
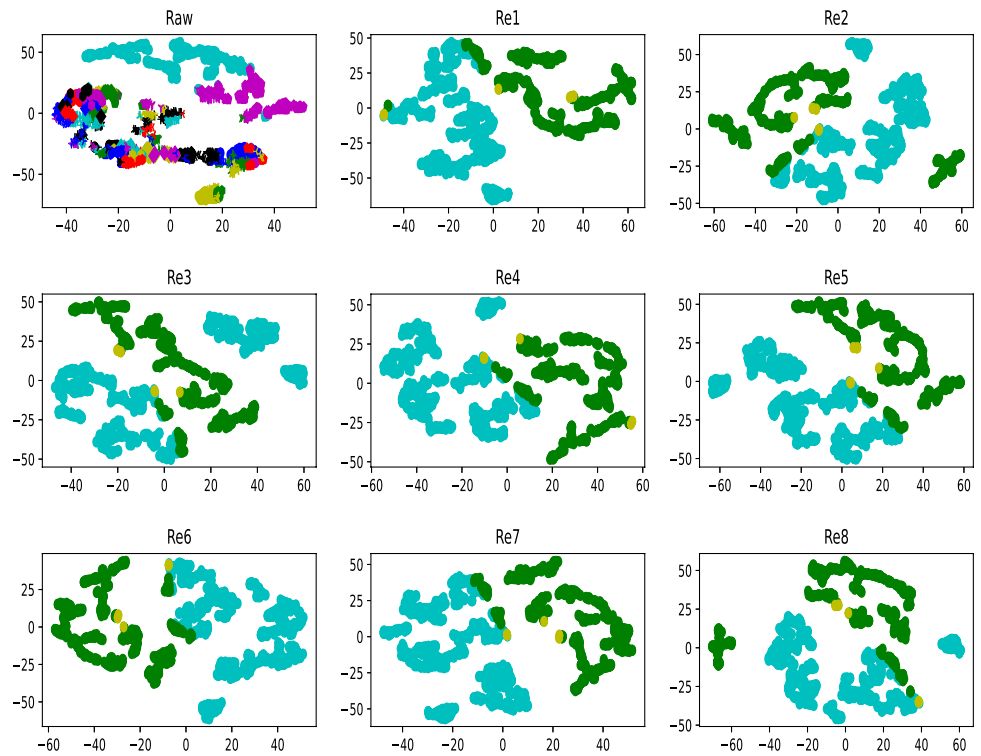
Fig. 60 The reduced clustering image of Algorithm 1 on Gw
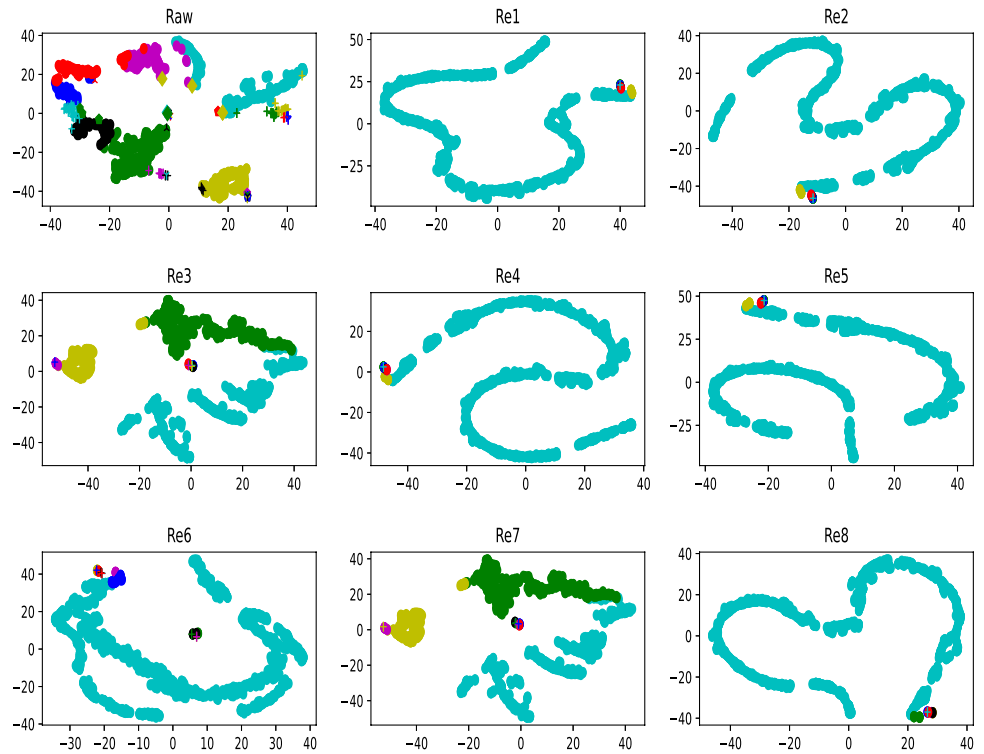
Fig. 61 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Aa

Fig. 63 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on De

Fig. 62 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Ac

Fig. 64 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on He

methods for measuring uncertainty in incomplete set-valued data have been studied, which is information granulation, information amount, rough entropy and information entropy respectively. The validity of the proposed measures has been statistically analyzed. It can be verified that the obtained measures can effectively measure the uncertainty of incomplete set-valued data with different incomplete rates. In order to study the attribute reduction of incomplete set-valued data, Algorithms 1 and 2 have been proposed based on information granulation and information entropy, respectively, and their validity has been analyzed and verified by k-means clustering algorithm and Mean Shift clustering algorithm. It is worth mentioning that the incomplete rate has been adopted. In the future, we will further explore more effective reduction methods for incomplete set-valued data.

Fig. 65 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Pc

Fig. 67 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Ob
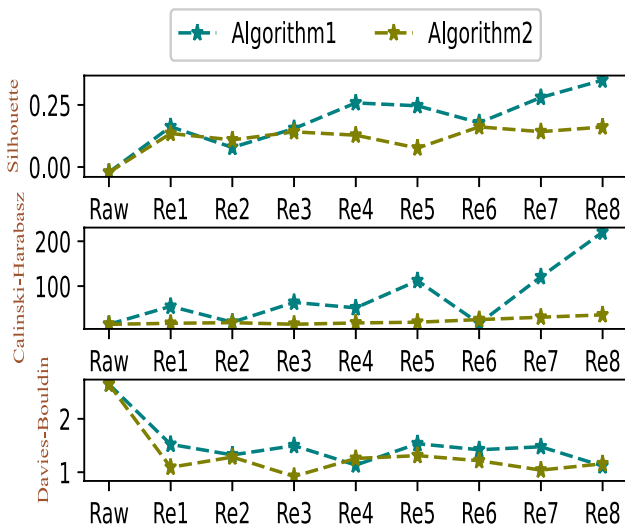
Fig. 66 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Sl
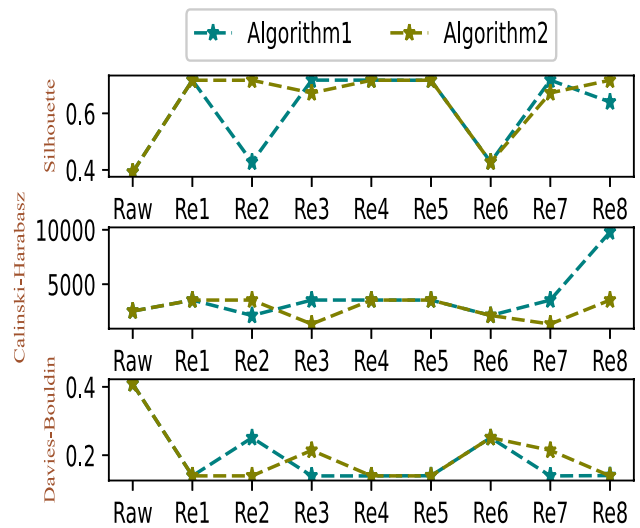
Fig. 68 The clustering evaluation coefficient of Algorithm 1 and Algorithm 2 on Gw

Table 15 Reduction results of FRSM and DRM

Date	FRSM	DRM
Aa	Re={1–6, 8–13, 18–20}	Re={1–3, 5, 7, 10–12, 16, 18, 20}
Ac	Re={1–5, 7–18, 20}	Re={1–5, 7–9, 11, 16–18, 20}
De	Re={1–2, 5–6, 9–10, 12–13, 15–16, 18–19, 23, 25–27, 30, 32, 34}	Re={2–3, 5, 9, 14–18, 21, 24, 26, 30–31, 34}
He	Re={1–3, 5–6, 14, 16, 17, 19}	Re={5, 8, 9, 11, 14–17}
Pc	Re={3, 5–7, 10}	Re={1, 3, 5–7, 9–11}
Sl	Re={1–10, 12, 15–17, 19–22, 24, 30, 35}	Re={1, 3–7, 10–13, 16, 18, 21–23, 27–29}
Ob	Re1={1–4, 7–8, 10–16}	Re2={1, 3–4, 6–9, 11, 13, 15}
Gw	Re1={1–3, 5, 7, 10, 11, 14}	Re2={1–3, 5, 7, 9, 10, 12}

36. Wierman MJ (1999) Measuring uncertainty in rough set theory. Int J Gen Syst 28:283–297

37. Wu ZJ, Wang H, Chen N, Luo JW (2021) Semi-monolayer covering rough set on set-valued information systems and its efficient computation. Int J Approx Reason 130:83–106

38. Xie NX, Liu M, Li ZW, Zhang GQ (2019) New measures of uncertainty for an interval-valued information system. Inf Sci 470:156–174

39. Xie XL, Li ZW, Zhang PF, Zhang GQ (2019) Information structures and uncertainty measures in an incomplete probabilistic set-valued information system. IEEE Access 7:27501–27514

40. Xie SD, Wang YX (2014) Construction of tree network with limited delivery latency in homogeneous wireless sensor networks. Wireless Pers Commun 78(1):231–246

41. Xu WH, Guo YT (2016) Generalized multigranulation double-quantitative decision-theoretic rough set. Knowl-Based Syst 105(1):190–205

42. Xu WH, Li WT (2016) Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets. IEEE Trans Cybern 46(2):366–379

43. Xu WH, Yu JH (2017) A novel approach to information fusion in multi-source datasets: a granular computing viewpoint. Inf Sci 378:410–423

44. Xu WH, Yuan KH, Li WT (2022) Dynamic updating approximations of local generalized multigranulation neighborhood rough set. Appl Intell. https://doi.org/10.1007/s10489-021-02861-x

45. Yao YY (2003) Probabilistic approaches to rough sets. Expert Syst 20:287–297

46. Yao YY, Li XN (1996) Comparison of rough-set and set-set models for uncertain reasoning. Fund Inf 27:289–298

47. Yuan KH, Xu WH, Li WT, Ding WP (2022) An incremental learning mechanism for object classification based on progressive fuzzy three-way concept. Inf Sci 584(1):127–147

48. Zar JH (1972) Significance testing of the Spearman rank correlation coefficient. J Am Stat Assoc 67(339):578–580

49. Zhang GQ, Li ZW, Wu WZ, Liu XF, Xie NX (2018) Information structures and uncertainty measures in a fully fuzzy information system. Int J Approx Reason 101:119–149