Measures of uncertainty for a four-hybrid information system and their applications

Bin Qin1 · Fanping Zeng1 · Kesong Yan1

Accepted: 23 January 2022 / Published online: 15 February 2022 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
A four-hybrid information system (4HIS) is an information system where the dataset of object descriptions consists of categorical, boolean, real-valued and missing data or attributes. This paper studies measures of uncertainty for a 4HIS and its application in attribute reduction. The distance function for each type of attribute in a 4HIS is first provided. Then, this distance is used to produce the tolerance relation induced by a given subsystem in a 4HIS. Next, information structure of this subsystem is proposed in terms of a set vector and dependence between information structures is introduced. Moreover, granulation and entropy measures in a 4HIS are investigated on the basis of information structures. In order to verify the feasibility of the proposed measures, effectiveness analysis is performed from a statistical perspective. Finally, an application of the proposed measures for attribute reduction in a 4HIS is given.

Keywords 4HIS · UM · Tolerance relation · Information structure · Effectiveness analysis · Attribute reduction

1 Introduction

1.1 Research background
Due to the complex diversity of the objective world, uncertainty exists in real life. Randomness, vagueness and imprecision are the most important concepts for uncertainty which can be appeared everywhere. Uncertainty plays a vital role in practical problems. Measuring uncertainty (UM) is helpful for understanding the nature of various kinds of information and then offer new visual angle for data analysis. UM a significant issue in many research fields, such as machine learning (Xie and Wang 2014), pattern recognition (Gu et al. 2015), medical diagnosis (Hempelmann et al. 2016), data mining (Delgado and Romero 2016) and so on.

Granular computing (GrC), proposed by Zadeh (1997, 1998), is a mode of thinking or method for solving practical problems based on granularity structure. Because GrC reflects the global view and approximate solution ability of human beings when dealing with multilevel and multiperspective problems, GrC has gradually become an important theory for solving uncertain. Information granulation is the basic content of GrC. An object is divided into a series of different information granules under given granulation criteria which is called the process of information granulation. Under dissimilar granulation criteria, the different granularity layers can be obtained, and then multi-granularity grid structure. Granular structure is the collection of information granules. Lin (1998) and Yao (2003) talked about the importance of GrC, it caught people’s attention. GrC is a superset that integrates many theoretical methods in artificial intelligence fields such as rough set theory (RST) Pawlak (1982), fuzzy set theory (Zadeh 1965), concept lattice (Ma et al. 2007; Wu et al. 2009) and quotient space theory (Zhang and Zhang 2007).

RST is a considerable mathematical tool. Not only does it offers new scientific logic and research methods for information science and cognitive science, but also provides a tool dealing with uncertainty. Its essential idea is to construct a partition of the universe by means of indistinguishable relations, obtain equivalence classes, and then establish an
approximate space. Information system (IS) based on RST is also called knowledge representation system (Pawlak 1982). An IS can be represented by a data table. Furthermore, the data table contains rows labeled by objects of interest, columns labeled by attributes, and entries of the table indicating attribute values. There are many applications in RST, for instance, uncertainty modeling (Dubois and Prade 1990), reasoning with uncertainty (Greco et al. 2006), rule extraction (Blaszczynski et al. 2011; Mi et al. 2005), classification and feature selection (Dai et al. 2018; Hu et al. 2010) are associated with IS.

In order to systematically assess uncertainty, the notion of entropy to communicate theory Shannon (Shannon 1948) was introduced to deal with UM. Beaubouef et al. (1998) investigated other methods to rough sets’ uncertainty. Miao and Hou (2004) proposed some more effective and significance measure tools, including information, combination and rough entropy. Liang and Qu (2002) introduced a rough metric method for knowledge in an incomplete information system (IIS). Mi et al. (2005) gave some properties of fuzzy approximation operators and a method of uncertainty measurement for generalized fuzzy rough sets. Li et al. (2019) studied uncertainty measurement for a fuzzy relation IS. Li et al. (2020) measured uncertainty of a fully fuzzy IS by using Gaussian kernel. Dai and Tian (2013) studied entropy measures and granularity measures for a set-valued IS. Li et al. (2019) investigated UM for a covering IS.

For GrC in an IS, the information structure is a significant research topic. An equivalence relation is a special kind of similarity between objects from a dataset. Given an IS, each attribute subset determines an equivalence relation. The object set of this IS is divided into disjoint classes by this equivalence relation, and these classes are said to be equivalence classes. If two objects belong to the same equivalence class, then we may say that they cannot be distinguished under this equivalence relation. Thus, each equivalence class is seen as an information granule consisting of indistinguishable objects. The family of all these information granules constitutes a vector; this vector is said to be an information structure in the IS induced by this attribute subset. Equally, information structures in an IS are also granular structures in the meaning of GrC. Yu (2018) proposed information structures in an IIS. Zhang et al. (2018) investigated information structures and uncertainty measures in a fully fuzzy IS. Li et al. (2020) investigated information structures in a covering IS.

1.2 Motivation and inspiration

If an IS has many kinds of attributes or data, such as boolean attributes, categorical attributes, real-valued attributes, missing value and so on, then this IS can be called a multiple data IS. Zeng et al. (2015) called such an IS as a hybrid information system (HIS). How to process this kind of hybrid data? Zeng et al. (2015) investigated the measurement problem of mixed data and the incremental updating method when IS changed. Juhola and Laurikkala (2007) introduced two distance measures in the presence of missing values is very useful to study for medical data of mixed-type variables. Han and Kamber (2011) introduced a useful approach to process hybrid data that a database consisting of six data types. Yu (2019) considered information structures and UM of a hybrid information system with images (HISI).

In practical applications, hybrid data exist anywhere. It is very meaningful topic to discuss UM of an IS. The main purpose of this paper is to study UM of a 4HIS.

In recent years, some scholars have discussed topics related to information structures and uncertainty in an IS, such as Aggarwal (2017), Chen et al. (2017), Xie et al. (2019). However, their research lacks numerical experiments and big data analysis support. To make sure our work is more convincing and complete, this paper gives numerical experiments and data analysis.

1.3 Discussion and contribution

In this part, we discuss several references for hybrid data, so as to see the contribution or innovation of this paper more clearly.

(1) Zeng et al. (2015) defined a new distance based on the value difference metric and then constructed a novel fuzzy rough set by combining the distance and Gaussian kernel. Considering an IS often vary with time, they analyzed the updating mechanisms for feature selection with the variation of the attribute set. Moreover, they presented fuzzy rough set approaches for incremental feature selection on HIS and proposed two corresponding algorithms. Finally, extensive experiments on eight datasets show that the incremental approaches significantly outperform non-incremental approaches with feature selection in the computational time.

(2) Zeng et al. (2017) analyzed the changing mechanisms of the attribute values and fuzzy equivalence relations in fuzzy rough set and then presented fuzzy rough set approaches for incrementally updating approximations in an HIS. Moreover, they gave two corresponding incremental algorithms. Finally, extensive experiments on eight data sets show that incremental approaches can effectively improve the performance of updating approximations and not only significantly shorten the computational time, but also increase approximation classification accuracies.
(3) Yu (2019) considered a hybrid information system with images (HISI). First, he developed new hybrid distance in an HISI. Then, he obtained the fuzzy $T_{cos}$-equivalence relation by using Gaussian kernel. Next, he described information structures in an HISI by set vectors and studied dependence between them by using inclusion degree. Finally, he investigated UM for an HISI by means of its information structures.

(4) Yuan et al. [38] introduced fuzzy rough sets to deal with the problem of outlier detection in hybrid data (numerical, categorical). First, they defined the granule outlier degree to characterize the outlier degree of fuzzy rough granules by employing the fuzzy approximation accuracy. Then, they constructed the outlier factor based on fuzzy rough granules by integrating the outlier degree and the corresponding weights to characterize the outlier degree of objects. Furthermore, they designed the corresponding outlier detection algorithm. Finally, they evaluated the effectiveness of the algorithm through experiments on 16 real-world datasets. The experimental results show that the algorithm is more flexible for detecting outliers and is suitable for hybrid data.

(5) Zhang et al. (2016) proposed a fuzzy rough set based information entropy for feature selection for hybrid data (nominal, real-valued). They first proved that the newly defined entropy meets the common requirement of monotonicity and can equivalently characterize the existing feature selection in the fuzzy rough set theory. Then, they formulated a feature selection algorithm based on the proposed entropy and a filter-wrapper method is suggested to select the best feature subset in terms of classification accuracy. Finally, they carried out an extensive numerical experiment to assess the performance of the feature selection algorithm.

(6) This paper deals with hybrid data (categorical, boolean, real-valued and missing data). The main details are based on the following considerations: a) a 4HIS itself has uncertainty; b) how to define a tolerance relation in a 4HIS; c) information structure is very helpful for knowledge discovery from a 4HIS; d) the magnitude of the measured value in a 4HIS can be compared by dependence between information structures; e) which measure is chosen to measure the uncertainty of a 4HIS; f) it is very necessary to analyze the effectiveness of the proposed measurement; h) it is very necessary to give an application of the proposed measures for attribute reduction in a 4HIS.

This paper first provides the distance function for each type of attribute in a 4HIS. This distance is used to produce the tolerance relation induced by a given subsystem. Then, information granules of a 4HIS based on the tolerance relation are constructed. By the way, the information structure formed by information granules composed of toleration classes is presented. Next, the dependence between them is discussed. By means of the dependence, four kinds of measurement to estimate the uncertainty of a 4HIS are put forward. Moreover, the effectiveness analysis about the proposed measures is carried out from a statistical perspective. We find the influence of $\theta$ value on the UM for a 4HIS, which may have potential application value in data mining. Finally, an application of the proposed measures for attribute reduction in a 4HIS is given.

1.4 Structure and organization

The work process of the paper is given in Figure 1.

The remaining part of this paper is organized as follows. Section 2 recalls some notions about a 4HIS. Section 3 constructs the distance between the information values of two objects about each type of attribute in a 4HIS and proposes the tolerance relation induced by a given subsystem of a 4HIS. Section 4 describes information structures in a 4HIS and studies the dependence between them. Section 5 introduces some tools for measuring uncertainty of a 4HIS. Section 6 conducts effectiveness analysis for showing the feasibility of these tools. Section 7 gives an application of the proposed measures for attribute reduction in a 4HIS. Section 8 concludes this paper.

2 Preliminaries

In this section, some basic concepts about a 4HIS are introduced.

Definition 2.1 (Pawlak 1991) Suppose that $X$ is a finite set of objects. Assume that $AT$ expresses a finite set of attributes.

![Fig. 1 The work process of the paper](image-url)
Then the ordered triple \((X, AT)\) is referred to as an information system (IS), if every attribute \(a \in AT\) is able to decide a function \(a : X \rightarrow Y_a\), where \(Y_a = \{a(x) : x \in X\}\).

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

**Definition 2.2** Suppose that \((X, AT)\) is a IIS. Then \((X, AT)\) is referred to as a four-hybrid information system (4HIS), if \(A = A^{cat} \cup A^{boo} \cup A^{rea}\), where \(A^{cat}\), \(A^{boo}\) and \(A^{rea}\) are the categorical, boolean and real-valued attribute set, respectively.

**Example 2.3** In Table 1, categorical attribute “Headache”, boolean attribute “Muscle pain”, real-valued attribute “Temperature” and categorical attribute “Symptom” are denoted as \(a_1\), \(a_2\), \(a_3\) and \(a_4\), respectively, and "*" indicates the missing value. Then, Table 1 depicts a 4HIS \((X, AT)\), where \(X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}\) is an object set, \(A = \{a_1, a_2, a_3, a_4\}\) expresses a attribute set.

\(Y^*_a\) is denoted as the set of non-missing information values of the attribute \(a\). Then

\[
Y^*_{a_1} = \{\text{Sick, Middle, No}\}, \quad Y^*_{a_2} = \{\text{Yes, No}\}, \\
Y^*_{a_3} = \{36.6, 36.8, 37, 38, 39, 39.5, 40\}, \\
Y^*_{a_4} = \{\text{Flu, Rhinitis, Health}\}.
\]

### 3 Tolerance relations in a 4HIS

In this section, the distance between the information values of two objects about each type of attribute in a 4HIS is first constructed. Then, the tolerance relation induced by a given subsystem of a 4HIS is proposed.

#### 3.1 The distance function for each type of attribute in a 4HIS

\(\forall A, B \in 2^X\), denote

\[A \oplus B = A \cup B - A \cap B.\]

\(\forall a \in A^{rea}\), denote

\[\hat{a} = \max Y^*_a - \min Y^*_a.\]

\(\forall a \in A^{cat} \text{ and } x \in X \text{ with } a(x) \neq *, \text{ denote}\)

\[[x]_a = \{y \in X : a(x) = a(y), a(y) \in Y^*_a\}.\]

For missing data, we have the following thoughts.

1) Consider \(x \neq y, a(x) = *, a(y) \neq *, a \in A\), because “\(a(x)\)” is treated as “do not care” condition, thus \(a(x)\) has the probability of \(\frac{1}{|Y_a|}\) to equal to one certain value of \(Y^*_a\).

2) Consider \(x \neq y, a(x) \neq *, a(y) = *, a \in A\), because “\(a(y)\)” has the probability of \(\frac{1}{|Y_a|}\) to equal to one certain value of \(Y^*_a\).

3) Consider \(x \neq y, a(x) = *, a(y) = *, a \in A\) and \(a(y)\) both have the probability of \(\frac{1}{|Y_a|}\) to equal to one certain value of \(Y^*_a\), so the joint probability of \(a(x)\) and \(a(y)\) is \(\frac{1}{|Y_a|^2}\).

For \(x \neq y, a(x) \neq *, a(y) \neq *, a \neq a, a \in A^{rea}\), define

\[\text{dis}(a(x), a(y)) = \frac{|a(x) - a(y)|}{\hat{a}}.\]

For \(x \neq y, a(x) \neq *, a(y) \neq *, a \neq a, a \in A^{boo}\), according to the opinion of Zeng et al. (2015), define

\[\text{dis}(a(x), a(y)) = 1.\]

For \(x \neq y, a(x) \neq *, a(y) \neq *, a \neq a, a \in A^{cat}\), according to the opinion of Yu (2019), define

\[\text{dis}(a(x), a(y)) = \frac{|[x]_a \bigoplus [y]_a|}{|[x]_a | \cup | [y]_a |}.\]

In this way, the following definition is proposed.

**Definition 3.1** Suppose that \((X, AT)\) is a 4HIS with \(AT = A^{cat} \cup A^{boo} \cup A^{rea}\). Then \(\forall u, v \in X, \forall a \in A\), the distance between \(a(x)\) and \(a(y)\) is defined as
**Table 1** A 4HIS

| X       | Headache($a_1$) | Muscle pain($a_2$) | Temperature($a_3$) | Symptom($a_4$) |
|---------|-----------------|--------------------|--------------------|----------------|
| $x_1$   | Sick            | Yes                | 40                 | Flu            |
| $x_2$   | Sick            | Yes                | 39.5               | Flu            |
| $x_3$   | Middle          | *                  | 39                 | Flu            |
| $x_4$   | Middle          | Yes                | 36.8               | Rhinitis       |
| $x_5$   | Middle          | No                 | *                  | Rhinitis       |
| $x_6$   | No              | No                 | 36.6               | Health         |
| $x_7$   | No              | *                  |                     | Health         |
| $x_8$   | No              | Yes                | 38                 | Flu            |
| $x_9$   | *               | Yes                | 37                 | Health         |

\[
dis(a(x), a(y)) = \begin{cases} 
0, & x = y, \ a \in A; \\
1 - \frac{1}{|Y_a^2|}, & x \neq y, a(x) = *, a(y) = *, \ a \in A; \\
1 - \frac{1}{|Y_a^2|}, & x \neq y, a(x) = *, a(y) \neq *, \ a \in A; \\
1 - \frac{1}{|Y_a^2|}, & x \neq y, a(x) \neq *, a(y) = *, \ a \in A; \\
0, & x \neq y, a(x) \neq *, a(y) \neq *, a(x) = a(y), \ a \in A; \\
1, & x \neq y, a(x) \neq *, a(y) \neq *, a(x) \neq a(y), \ a \in A^{boo} \ [43]; \\
\frac{|x_1|_a \delta |y|_a}{|x|_a \delta |y|_a}, & x \neq y, a(x) \neq *, a(y) \neq *, a(x) \neq a(y), \ a \in A^{cat} \ [37]; \\
\frac{|x|_a \delta |y|_a}{|x|_a \delta |y|_a}, & x \neq y, a(x) \neq *, a(y) \neq *, a(x) \neq a(y), \ a \in A^{rea}.
\end{cases}
\]

**Example 3.2** (Continued from Examples 2.3)

1. Since $a_1$ is a categorical attribute, $a_1(x_1) = \text{Sick}$ and $(a_1(x_3)) = \text{Middle}$, we have

\[
[x_1]_{a_1} = \{x_1, x_2\}, \ [x_3]_{a_1} = \{x_3, x_4, x_5\}.
\]

By Definition 3.1,

\[
dis((a_1(x_1), a_1(x_3))) = \frac{|[x_1]_{a_1} \bigcup [x_3]_{a_1}|}{|[x_1]_{a_1} \bigcup [x_3]_{a_1}|} = \frac{5}{5} = 1;
\]

2. Since $a_2(x_1) = \text{Yes} \neq *$, $a_2(x_3) = *$, by Definition 3.1, we have

\[
dis((a_2(x_1), a_2(x_3))) = 1 - \frac{1}{|Y_{a_2}|} = 1 - \frac{1}{2} = 0.5;
\]

3. Since $a_3$ is a real-valued attribute, by Definition 3.1, we have

\[
dis((a_3(x_1), a_3(x_3))) = \frac{|a_3(x_1) - a_3(x_3)|}{a_3} = \frac{|40 - 39|}{40 - 36.6} \approx 0.2941;
\]

4. Since $a_4(x_1) = \text{Flu} = a_4(x_3)$, by Definition 3.1, we have

\[
dis((a_4(x_1), a_4(x_3))) = 0.
\]

Below, for convenience, the 4HIS $(X, C \cup D)$ with $C = A^{cat} \cup A^{boo} \cup A^{rea}$ is denoted as the 4HIS $(X, AT)$ where every element of $D$ is be viewed as a categorical attribute.

**Definition 3.3** Let $(X, AT)$ be a 4HIS. $\forall \ a \in A$, put

\[
M_a = (dis(a_i, a_j))_{n \times n}.
\]

Then $M_a$ is referred to as the distance matrix of the attribute $a$ in $(X, AT)$. 

Springer
Example 3.4 (Continued from Examples 2.3)

\[
M_{a_1} = \begin{pmatrix}
0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.6667 \\
0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.6667 \\
1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.6667 \\
1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.6667 \\
1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.6667 \\
1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.6667 \\
0.6667 & 0.6667 & 0.6667 & 0.6667 & 0.6667 & 0.6667 & 0.6667 & 0.6667 & 0.6667 & 0.0000
\end{pmatrix}
\]

\[
M_{a_2} = \begin{pmatrix}
0.0000 & 0.0000 & 0.5000 & 0.0000 & 1.0000 & 1.0000 & 0.5000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.5000 & 0.0000 & 1.0000 & 1.0000 & 0.5000 & 0.0000 & 0.0000 & 0.0000 \\
0.5000 & 0.5000 & 0.0000 & 0.5000 & 0.5000 & 0.5000 & 0.7500 & 0.5000 & 0.5000 & 0.5000 \\
0.0000 & 0.0000 & 0.5000 & 0.0000 & 1.0000 & 1.0000 & 0.5000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.5000 & 0.0000 & 1.0000 & 1.0000 & 0.5000 & 0.0000 & 0.0000 & 0.0000 \\
0.5000 & 0.5000 & 0.7500 & 0.5000 & 0.5000 & 0.5000 & 0.5000 & 0.0000 & 0.5000 & 0.5000 \\
0.0000 & 0.0000 & 0.5000 & 0.0000 & 1.0000 & 1.0000 & 0.5000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.5000 & 0.0000 & 1.0000 & 1.0000 & 0.5000 & 0.0000 & 0.0000 & 0.0000
\end{pmatrix}
\]

\[
M_{a_3} = \begin{pmatrix}
0.0000 & 0.1471 & 0.2941 & 0.9412 & 0.8571 & 1.0000 & 0.8571 & 0.5882 & 0.8824 & 0.1471 \\
0.1471 & 0.0000 & 0.1471 & 0.7941 & 0.8571 & 0.8529 & 0.8571 & 0.4412 & 0.7353 & 0.2941 \\
0.2941 & 0.1471 & 0.0000 & 0.6471 & 0.8571 & 0.7059 & 0.8571 & 0.2941 & 0.5882 & 0.9412 \\
0.9412 & 0.7941 & 0.6471 & 0.0000 & 0.8571 & 0.0588 & 0.8571 & 0.3529 & 0.0588 & 0.8571 \\
0.8571 & 0.8571 & 0.8571 & 0.0000 & 0.8571 & 0.9796 & 0.8571 & 0.8571 & 0.8571 & 0.8571 \\
0.8571 & 0.8571 & 0.8571 & 0.0000 & 0.8571 & 0.8571 & 0.8571 & 0.8571 & 0.8571 & 0.8571 \\
0.5882 & 0.4412 & 0.2941 & 0.3529 & 0.8571 & 0.4118 & 0.8571 & 0.0000 & 0.2941 & 0.8824 \\
0.8824 & 0.7353 & 0.5882 & 0.0588 & 0.8571 & 0.1176 & 0.8571 & 0.2941 & 0.0000 & 0.0000
\end{pmatrix}
\]

\[
M_{a_4} = \begin{pmatrix}
0.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\
0.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\
0.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\
1.0000 & 1.0000 & 1.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\
1.0000 & 1.0000 & 1.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\
1.0000 & 1.0000 & 1.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\
0.0000 & 0.0000 & 0.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\
1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000
\end{pmatrix}
\]

3.2 The tolerance relation induced by a given subsystem of a 4HIS

Below, the tolerance relation induced by a given subsystem of a 4HIS is established.

Definition 3.5 Suppose that \((X, AT)\) is a 4HIS and \(A \subseteq AT\). Pick \(\theta \in [0, 1]\). Put

\[
R^\theta_A = \{ (x, y) \in X \times X : \forall a \in A, \text{dist}(a(x), a(y)) \leq \theta \}.
\]

Then \(R^\theta_A\) is referred as to the relation induced by the subsystem \((X, A)\) with respect to \(\theta\).

Clearly, \(R^0_A\) is a tolerance relation on \(X\).

Denote

\[
R^\theta_A(x) = \{ y \in X : (x, y) \in R^\theta_A \}.
\]

Then \(R^\theta_A(x)\) is referred as to the tolerance class of the object \(x\) under the tolerance relation \(R^\theta_A\).

Proposition 3.6 Let \((X, AT)\) be a 4HIS. Then the following properties hold:
(1) If $A \subseteq B \subseteq AT$, then for any $\theta \in [0, 1]$ and $x \in X$,
\[ R^\theta_B(x) \subseteq R^\theta_A(x); \]

(2) If $0 \leq \theta_1 \leq \theta_2 \leq 1$, then for any $A \subseteq AT$ and $x \in X$,
\[ R^\theta_A(x) \subseteq R^{\bar{\theta}}_A(x). \]

**Proof** (1) Suppose $y \in R^\theta_B(x)$. Then $\forall a \in B, dis(a(x), a(y)) \leq \theta$.
Note that $A \subseteq B$. Then $\forall a \in A, dis(a(x), a(y)) \leq \theta$.
Thus $y \in R^\theta_A(x)$.
Hence $R^\theta_B(x) \subseteq R^\theta_A(x)$.

(2) Suppose $y \in R^\theta_A(x)$. Then $\forall a \in A, dis(a(x), a(y)) \leq \theta_1$.
Note that $\theta_1 \leq \theta_2$. Then $\forall a \in A, dis(a(x), a(y)) \leq \theta_2$.
Thus $y \in R^\theta_B(x)$.
Hence $R^\theta_A(x) \subseteq R^{\bar{\theta}}_A(x)$.

\[ \square \]

**Example 3.7** (Continued from Example 2.3) Pick $\theta = 0.5$.
By Example 3.4, we have
\[ R^\theta_{AT}(x_1) = R^\theta_{AT}(x_2) = \{x_1, x_2\}, R^\theta_{AT}(x_3) = \{x_3\}, \]
\[ R^\theta_{AT}(x_4) = \{x_4\}, R^\theta_{AT}(x_5) = \{x_5\}, R^\theta_{AT}(x_6) = \{x_6\}, \]
\[ R^\theta_{AT}(x_7) = \{x_7\}, R^\theta_{AT}(x_8) = \{x_8\}, R^\theta_{AT}(x_9) = \{x_9\}. \]
In what follows, an algorithm of computing $R^\theta_A$ is designed as follows.

### 4 Information structures in a 4HIS

In this section, information structures in a 4HIS are studied.

#### 4.1 The concept of information structures in a 4HIS

**Definition 4.1** Suppose that $(X, AT)$ is a 4HIS and $A \subseteq AT$. Pick $\theta \in [0, 1]$. Put
\[ In^{\theta}S(A) = (R^\theta_A(x_1), R^\theta_A(x_2), \ldots, R^\theta_A(x_n)). \]
Then $In^{\theta}S(A)$ is referred to as $\theta$-information structure of the subsystem $(X, A)$.

**Example 4.2** (Continued from Example 3.7)
\[ In^{\theta}S(AT) = (\{x_1, x_2\}, \{x_1, x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6, x_8\}, \{x_7\}, \{x_6, x_8\}, \{x_9\}). \]

**Definition 4.3** Suppose that $(X, AT)$ is a 4HIS. Given $A, B \subseteq AT$ and $\theta_1, \theta_2 \in [0, 1]$. If for any $i$, $R^\theta_A(x_i) = R^\theta_B(x_i)$, then $In^{\theta_1}S(A)$ and $In^{\theta_2}S(B)$ are called to be the same. It’s written as $In^{\theta_1}S(A) = In^{\theta_2}S(B)$.

#### 4.2 Dependence between information structures in a 4HIS

**Definition 4.4** Let $(X, AT)$ be a 4HIS. Given $A, B \subseteq AT$ and $\theta_1, \theta_2 \in [0, 1]$.

(1) If for any $i$, $R^\theta_A(x_i) \subseteq R^\theta_B(x_i)$, then $In^{\theta_1}S(A)$ is referred to as depend on $In^{\theta_2}S(A)$. It is written as $In^{\theta_1}S(A) \leq In^{\theta_2}S(A)$.

(2) If $In^{\theta_1}S(A) \leq In^{\theta_2}S(B)$ and $In^{\theta_1}S(A) \neq In^{\theta_2}S(B)$, then $In^{\theta_2}S(B)$ is referred to as depend strictly on $In^{\theta_1}S(A)$. It is written as $In^{\theta_1}S(A) < In^{\theta_2}S(B)$.

Obviously,
\[ In^{\theta_1}S(A) = In^{\theta_2}S(B) \iff In^{\theta_1}S(A) \leq In^{\theta_2}S(B) \land In^{\theta_1}S(A) \neq In^{\theta_2}S(B). \]
\( < \text{InS}^{\theta_2}(B) \Rightarrow \text{InS}^{\theta_1}(A) \)
\( \leq \text{InS}^{\theta_2}(B) \).

**Theorem 4.5** Suppose that \((X, AT)\) is a 4HIS. Given \(A, B \subseteq AT\) and \(\theta_1, \theta_2 \in [0, 1]\). Then
\[
\text{InS}^{\theta_1}(A) = \text{InS}^{\theta_2}(B) \iff R^{\theta_1}_A = R^{\theta_2}_B.
\]

**Proof** Obviously. \(\square\)

**Theorem 4.6** Suppose that \((X, AT)\) is a 4HIS. Given \(A, B \subseteq AT\) and \(\theta_1, \theta_2 \in [0, 1]\). Then
\[
\text{InS}^{\theta_1}(A) \preceq \text{InS}^{\theta_2}(B) \iff R^{\theta_1}_A \subseteq R^{\theta_2}_B.
\]

**Proof** By Theorems 4.5 and 4.6, it is easy to prove. \(\square\)

**Corollary 4.7** Suppose that \((X, AT)\) is a 4HIS. Given \(A, B \subseteq AT\) and \(\theta_1, \theta_2 \in [0, 1]\). Then
\[
\text{InS}^{\theta_1}(A) \prec \text{InS}^{\theta_2}(B) \iff R^{\theta_1}_A \subset R^{\theta_2}_B.
\]

**Proof** By Theorems 4.5 and 4.6, it is easy to prove. \(\square\)

**Theorem 4.8** Suppose that \((X, AT)\) is a 4HIS. Then the following properties hold:

(1) If \(A \subseteq B \subseteq AT\), then for any \(\theta \in [0, 1]\), \(\text{InS}^{\theta}(B) \leq \text{InS}^{\theta}(A)\);
(2) If \(0 \leq \theta_1 \leq \theta_2 \leq 1\), then for any \(A \subseteq AT\), \(\text{InS}^{\theta_1}(A) \leq \text{InS}^{\theta_2}(A)\).

**Proof** (1) Since \(A \subseteq B\), by Proposition 3.6(1), \(\forall i\), we have
\[
R^{\theta}_B(x_i) \subseteq R^{\theta}_A(x_i).
\]
Thus \(\text{InS}^{\theta}(B) \leq \text{InS}^{\theta}(A)\).
(2) Since \(\theta_1 \leq \theta_2\), by Proposition 3.6(2), \(\forall P \subseteq A\), we have
\[
R^{\theta_1}_A(x) \subseteq R^{\theta_2}_A(x).
\]
Then \(\text{InS}^{\theta_1}(A) \leq \text{InS}^{\theta_2}(A)\). \(\square\)

## 5 UM of a 4HIS

In this section, we studies UM of a 4HIS.

### 5.1 Granulation measures of a 4HIS

**Definition 5.1** Let \((X, AT)\) be a 4HIS. Given \(\theta \in [0, 1]\). Suppose that \(M^\theta : 2^{AT} \to (-\infty, +\infty)\) is a mapping. Then \(M^\theta(A)\) is referred to as \(\theta\)-information granulation function of the subsystem \((X, A)\), if \(M^\theta\) satisfies the following conditions:

1. \(\forall A \subseteq AT, M^\theta(A) \geq 0\) (Non-negativity);
2. \(S^\theta(A) = S^\theta(B)\) implies \(M^\theta(A) = M^\theta(Q)\) (Invariability);
3. \(S^\theta(A) \prec S^\theta(B)\) implies \(M^\theta(A) < M^\theta(Q)\) (Monotonicity).

Similar to Definition 5 in Qian et al. (2011), \(\theta\)-information granulation in a 4HIS is given as follows.

**Definition 5.2** Let \((X, AT)\) be a 4HIS and \(A \subseteq AT\). Pick \(\theta \in [0, 1]\). Then \(\theta\)-information granulation of \((X, A)\) is defined by
\[
G^\theta(A) = \frac{1}{n^2} \sum_{i=1}^{n} |R^{\theta}_A(x_i)|.
\]

**Proposition 5.3** Let \((X, AT)\) be a 4HIS and \(A \subseteq AT\). Pick \(\theta \in [0, 1]\). Then
\[
\frac{1}{n} \leq G^\theta(A) \leq 1.
\]

**Proof** \(\forall i, 1 \leq |R^{\theta}_A(x_i)| \leq n\). Then \(n \leq \sum_{i=1}^{n} |R^{\theta}_A(x_i)| \leq n^2\).

By Definition 5.2,
\[
\frac{1}{n} \leq G^\theta(A) \leq 1.
\]

\(\forall i, |R^{\theta}_A(x_i)| = 1\). Then \(G^\theta(A) = \frac{1}{n}\).
\(\forall i, |R^{\theta}_A(x_i)| = n\). Then \(G^\theta(A) = 1\). \(\square\)

**Theorem 5.4** Let \((X, AT)\) be a 4HIS and \(A \subseteq AT\). Pick \(\theta_1, \theta_2 \in [0, 1]\). Then the following properties hold:

(1) If \(\text{InS}^{\theta_1}(A) \leq \text{InS}^{\theta_2}(B)\), then \(G^{\theta_1}(A) \leq G^{\theta_2}(B)\);
(2) If \(\text{InS}^{\theta_1}(A) \prec \text{InS}^{\theta_2}(B)\), then \(G^{\theta_1}(A) < G^{\theta_2}(B)\).

**Proof** (1) This is obvious.
(2) By Definition 5.2,
\[
G^{\theta_1}(A) = \frac{1}{n^2} \sum_{i=1}^{n} |R^{\theta_1}_A(x_i)|,
\]
\[
G^{\theta_2}(B) = \frac{1}{n^2} \sum_{i=1}^{n} |R^{\theta_2}_B(x_i)|.
\]

Note that \(\text{InS}^{\theta_1}(A) \prec \text{InS}^{\theta_2}(B)\). Then \(\forall i, R^{\theta_1}_A(x_i) \subseteq R^{\theta_2}_B(x_i)\) and \(\exists j, R^{\theta_1}_A(x_j) \subset R^{\theta_2}_B(x_j)\). Thus \(\forall i, |R^{\theta_1}_A(x_i)| \leq |R^{\theta_2}_B(x_j)|\) and \(\exists j, |R^{\theta_1}_A(x_j)| < |R^{\theta_2}_B(x_j)|\).

Hence \(G^{\theta_1}(A) < G^{\theta_2}(B)\). \(\square\)
This theorem shows that when the available information becomes coarse, the \( \theta \)-information granularity increases, and when the available information becomes finer, the \( \theta \)-information granularity decreases. In other words, the greater the uncertainty of the existing information, the greater the value of the \( \theta \)-information granulation. Therefore, we can draw the conclusion that the \( \theta \)-information granulation introduced in definition 5.2 can be used to evaluate the degree of a 4HIS.

**Proposition 5.5** Let \( (X, AT) \) be a 4HIS. Then the following properties hold:

1. If \( A \subseteq B \subseteq AT \), then for any \( \theta \in [0,1] \), \( G^\theta(B) \leq G^\theta(A) \);
2. If \( 0 \leq \theta_1 \leq \theta_2 \leq 1 \), then for any \( A \subseteq AT \), \( G^{\theta_1}(A) \leq G^{\theta_2}(A) \).

**Proof** These follow from Theorems 4.8 and 5.4(1).

**Definition 5.6** Let \( (X, AT) \) be a 4HIS and \( A \subseteq AT \). Pick \( \theta_1, \theta_2 \in [0, 1] \). Then \( \theta \)-information amount of \( (X, A) \) is defined by

\[
E^{\theta_i}(A) = \sum_{i=1}^{n} \frac{1}{n} \left( 1 - \frac{|R_{A_i}^{\theta_i}(x_i)|}{n} \right).
\]

**Theorem 5.7** Let \( (X, AT) \) be a 4HIS and \( A \subseteq AT \). Pick \( \theta_1, \theta_2 \in [0, 1] \). Then the following properties hold:

1. If \( InS^{\theta_1}(A) \leq InS^{\theta_2}(B) \), then \( E^{\theta_1}(B) \leq E^{\theta_1}(A) \);
2. If \( InS^{\theta_1}(A) \prec InS^{\theta_2}(B) \), then \( E^{\theta_2}(B) \prec E^{\theta_2}(A) \).

**Proof**

1. This is clear.

2. By Definition 5.6,

\[
E^{\theta_1}(A) = \sum_{i=1}^{n} \frac{1}{n} \left( 1 - \frac{|R_{A_i}^{\theta_1}(x_i)|}{n} \right),
\]

\[
E^{\theta_2}(B) = \sum_{i=1}^{n} \frac{1}{n} \left( 1 - \frac{|R_{B_i}^{\theta_2}(x_i)|}{n} \right).
\]

Note that \( InS^{\theta_1}(A) \prec InS^{\theta_2}(B) \). Then \( \forall i, R_{A_i}^{\theta_1}(x_i) \leq R_{B_i}^{\theta_2}(x_i) \) and \( \exists j, R_{A_j}^{\theta_1}(x_j) \leq R_{B_j}^{\theta_2}(x_j) \). Thus \( \forall i, |R_{A_i}^{\theta_1}(x_i)| \leq |R_{B_i}^{\theta_2}(x_i)| \) and \( \exists j, |R_{A_j}^{\theta_1}(x_j)| < |R_{B_j}^{\theta_2}(x_j)| \). Hence \( E^{\theta_2}(B) \prec E^{\theta_1}(A) \).

This theorem shows that when the structure of hybrid information becomes finer, the \( \theta \)-information amount increases, and when the hybrid information structure becomes rough, the \( \theta \)-information amount decreases.

**Proposition 5.8** Let \( (X, AT) \) be a 4HIS. Then the following properties hold:

1. If \( A \subseteq B \subseteq AT \), then for any \( \theta \in [0,1] \), \( E^\theta(B) \leq E^\theta(A) \);
2. If \( 0 \leq \theta_1 \leq \theta_2 \leq 1 \), then for any \( A \subseteq AT \), \( E^{\theta_2}(A) \leq E^{\theta_1}(A) \).

**Proof** These follow from Theorems 4.8 and 5.7(1).

**Theorem 5.9** Let \( (X, AT) \) be a 4HIS. Given \( A \subseteq AT \) and \( \theta \in [0,1] \). Then

\[
G^\theta(A) + E^\theta(A) = 1.
\]

**Proof**

\[
G^\theta(A) + E^\theta(A) = \frac{1}{n} \sum_{i=1}^{n} |R_{A_i}^{\theta}(x_i)| + \frac{1}{n} \sum_{i=1}^{n} \left( 1 - \frac{|R_{A_i}^{\theta}(x_i)|}{n} \right)
= \frac{1}{n} \sum_{i=1}^{n} (|R_{A_i}^{\theta}(x_i)| + n - |R_{A_i}^{\theta}(x_i)|)
= \frac{1}{n} \sum_{i=1}^{n} n = 1.
\]

**Corollary 5.10** Let \( (X, AT) \) be a 4HIS. Given \( A \subseteq AT \) and \( \theta \in [0,1] \). Then

\[
0 \leq E^\theta(A) \leq 1 - \frac{1}{n}.
\]

**Proof** By Proposition 5.3, \( \frac{1}{n} \leq G^\theta(A) \leq 1 \). We have

\[
-1 \leq -G^\theta(A) \leq -\frac{1}{n}.
\]

By Theorem 5.9, \( G^\theta(A) + E^\theta(A) = 1 \). Hence

\[
0 \leq E^\theta(A) \leq 1 - \frac{1}{n}.
\]

**5.2 Entropy measures of a 4HIS**

**Definition 5.11** Let \( (X, AT) \) be a 4HIS. Given \( A \subseteq AT \) and \( \theta \in [0,1] \). Then \( \theta \)-rough entropy of \( (X, A) \) is defined by

\[
E^\theta_r(A) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{|R_{A_i}^{\theta}(x_i)|}.
\]
Proposition 5.12 Let \((X, AT)\) be a 4HIS. Given \(A \subseteq AT\) and \(\theta \in [0, 1]\). Then

\[0 \leq E^\theta_r(A) \leq \log_2 n.\]

**Proof** \(\forall i, 1 \leq |R^\theta_A(x_i)| \leq n\). Then

\[0 \leq -\log_2 \frac{1}{|R^\theta_A(x_i)|} = \log_2 |R^\theta_A(x_i)| \leq \log_2 n.\]

By Definition 5.11,

\[0 \leq E^\theta_r(A) \leq \log_2 n.\]

\(\forall i, |R^\theta_A(x_i)| = 1\). Then \(E^\theta_r(A) = 0\).

\(\forall i, |R^\theta_A(x_i)| = n\). Then \(E^\theta_r(A) = \log_2 n.\) \(\Box\)

Theorem 5.13 Let \((X, AT)\) be a 4HIS and \(A, B \subseteq AT\). Pick \(\theta_1, \theta_2 \in [0, 1]\). Then the following properties hold:

(1) If \(InS^\theta(A) \leq InS^\theta(B)\), then \(E^\theta_r(A) \leq E^\theta_r(B)\);
(2) If \(InS^\theta(A) < InS^\theta(B)\), then \(E^\theta_r(A) < E^\theta_r(B)\).

**Proof** (1) Obviously.

(2) By Definition 5.15,

\[H^\theta_r(A) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{|R^\theta_A(x_i)|}{n},\]

\[H^\theta_r(B) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{|R^\theta_B(x_i)|}{n}.\]

It should be noted that \(InS^\theta(A) < InS^\theta(B)\). Then \(\forall i, R^\theta_B(x_i) \subseteq R^\theta_B(x_i)\) and \(\exists j, R^\theta_A(x_j) \subseteq R^\theta_A(x_j)\). Thus \(\forall i, |R^\theta_A(x_i)| \leq |R^\theta_B(x_i)|\) and \(\exists j, |R^\theta_A(x_j)| < |R^\theta_B(x_j)|\).

Hence, \(E^\theta_r(A) < E^\theta_r(B)\). \(\Box\)

This theorem shows that the greater the uncertainty of the available information, the greater the \(\theta\)-rough entropy. Therefore, we can draw the conclusion that the \(\theta\)-rough entropy proposed in Definition 5.11 can be used to evaluate the degree of determination of a 4HIS.

Proposition 5.14 Let \((X, AT)\) be a 4HIS. Then the following properties hold:

(1) If \(A \subseteq B \subseteq AT\), then for any \(\theta \in [0, 1]\), \(E^\theta_r(B) \leq E^\theta_r(A)\);
(2) If \(0 \leq \theta_1 \leq \theta_2 \leq 1\), then for any \(A \subseteq AT\), \(E^\theta_r(A) \leq E^\theta_r(A)\).

**Proof** It can be proved by Theorems 4.8 and 5.13(1). \(\Box\)
Proof

\[ E^\theta_r(A) + H^\theta(A) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{|R^\theta_i(x_i)|} - \sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{|R^\theta_i(x_i)|} \]
\[ = -\sum_{i=1}^{n} \frac{1}{n} \left( \log_2 \frac{1}{|R^\theta_i(x_i)|} + \log_2 \frac{|R^\theta_i(x_i)|}{n} \right) \]
\[ = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{n} \]
\[ = \log_2 n. \]

\[ \Box \]

Corollary 5.19 Suppose that \((X, AT)\) is a 4HIS. Given \(A \subseteq AT\) and \(\theta \in [0, 1]\). Then

\[ 0 \leq H^\theta(A) \leq \log_2 n. \]

**Proof** By Proposition 5.12, \(0 \leq E^\theta_r(A) \leq \log_2 n\). Then

\[ -\log_2 n \leq -E^\theta_r(A) \leq 0. \]

By Theorem 5.18, \(E^\theta_r(A) + H^\theta(A) = \log_2 n\). Thus

\[ 0 \leq H^\theta(A) \leq \log_2 n. \]

\[ \Box \]

6 Experiments and analysis

In this section, we design a numerical experiment and do effectiveness analysis to evaluate the proposed measures.

6.1 A numerical experiment

In order to show the performance of the proposed measures for the uncertainty in a 4HIS, we select nine data sets that come from UCI (Repository of machine learning databases) which is described in Table 2, where each data set can be expressed as a 4HIS. We carry out a numerical experiment on the nine data sets.

In Annealing, pick \(L_i = \{a_1, \ldots, a_{33}\} \) \((i = 1, \ldots, 13)\) and \(\theta_j = 0.1 \times j \) \((j = 1, \ldots, 9)\). Measure sets on Annealing are defined as follows:

\[ X^\theta_H(An) = \{H^\theta(L_1), \ldots, H^\theta(L_{13})\} \]

\[ X^\theta_M(An) = \{M^\theta(L_1), \ldots, M^\theta(L_{13})\} \]

\[ X^\theta_A(An) = \{A^\theta(L_1), \ldots, A^\theta(L_{13})\} \]

\[ X^\theta_R(An) = \{R^\theta(L_1), \ldots, R^\theta(L_{13})\} \]

\[ \text{Table 2} \] Nine data sets from UCI

| Date sets          | Objects | Scale | Features |
|-------------------|---------|-------|----------|
| Annealing          | 798     | Ordinal | 39       |
| Automobile        | 205     | Ordinal | 26       |
| Contraceptive Method Choice | 1473   | Nominal | 10       |
| Credit Approval   | 690     | Ordinal | 16       |
| Dermatology       | 366     | Ordinal | 35       |
| Echocardiogram    | 132     | Ordinal | 13       |
| Hepatitis         | 155     | Ordinal | 20       |
| Meta-data         | 528     | Ordinal | 22       |
| Post-Operative Patient | 90     | Ordinal | 9        |

In Automobile, pick \(M_i = \{a_1, \ldots, a_{33}\} \) \((i = 1, \ldots, 13)\) and \(\theta_j = 0.1 \times j \) \((j = 1, \ldots, 9)\). Measure sets on Automobile are defined as follows:

\[ X^\theta_G(An) = \{G^\theta(N_1), \ldots, G^\theta(N_{10})\} \]

\[ X^\theta_M(An) = \{M^\theta(N_1), \ldots, M^\theta(N_{10})\} \]

\[ X^\theta_A(An) = \{A^\theta(N_1), \ldots, A^\theta(N_{10})\} \]

\[ X^\theta_R(An) = \{R^\theta(N_1), \ldots, R^\theta(N_{10})\} \]

In Contraceptive Method Choice, pick \(N_i = \{a_1, \ldots, a_{13}\} \) \((i = 1, \ldots, 10)\) and \(\theta_j = 0.1 \times j \) \((j = 1, \ldots, 9)\). Measure sets on Contraceptive Method Choice are defined as follows:

\[ X^\theta_G(An) = \{G^\theta(N_1), \ldots, G^\theta(N_{10})\} \]

\[ X^\theta_M(An) = \{M^\theta(N_1), \ldots, M^\theta(N_{10})\} \]

\[ X^\theta_A(An) = \{A^\theta(N_1), \ldots, A^\theta(N_{10})\} \]

\[ X^\theta_R(An) = \{R^\theta(N_1), \ldots, R^\theta(N_{10})\} \]

In Credit Approval, pick \(O_i = \{a_1, \ldots, a_{33}\} \) \((i = 1, \ldots, 8)\) and \(\theta_j = 0.1 \times j \) \((j = 1, \ldots, 9)\). Measure sets on Credit Approval are defined as follows:

\[ X^\theta_G(An) = \{G^\theta(O_1), \ldots, G^\theta(O_{10})\} \]

\[ X^\theta_M(An) = \{M^\theta(O_1), \ldots, M^\theta(O_{10})\} \]

\[ X^\theta_A(An) = \{A^\theta(O_1), \ldots, A^\theta(O_{10})\} \]

\[ X^\theta_R(An) = \{R^\theta(O_1), \ldots, R^\theta(O_{10})\} \]

In Dermatology, pick \(A_i = \{a_1, \ldots, a_{33}\} \) \((i = 1, \ldots, 7)\) and \(\theta_j = 0.1 \times j \) \((j = 1, \ldots, 9)\). Measure sets on Dermatology are defined as follows:

\[ X^\theta_G(An) = \{G^\theta(A_1), \ldots, G^\theta(A_{10})\} \]

\[ X^\theta_M(An) = \{M^\theta(A_1), \ldots, M^\theta(A_{10})\} \]

\[ X^\theta_A(An) = \{A^\theta(A_1), \ldots, A^\theta(A_{10})\} \]

\[ X^\theta_R(An) = \{R^\theta(A_1), \ldots, R^\theta(A_{10})\} \]
When the threshold $\theta$ defined as follows:

$$\theta_i = \text{CV}(Q_i)$$

$\text{CV}=\frac{\text{average value, standard deviation and standard deviation}}{\text{average value, standard deviation and standard deviation}}$

From Figs. 10, 11, 12, 13, 14, 15, 16, 17, and 18, the $\text{CV}$ values of measure sets on each data set are computed (see Figs. 10, 11, 12, 13, 14, 15, 16, 17, and 18).

From Figures 1, 2, 3, 4, 5, 6, 7, 8, and 9, the following conclusions are obtained:

1. When the threshold $\theta$ is a fixed value, $G^0$ and $E^\theta$ are both monotonically decreasing as the cardinality of attribute subset increases. Meanwhile, $E^\theta$ and $H^\theta$ are both monotonically increasing with the attribute subset growth. It shows that the uncertainty of a 4HIS decreases as the attribute subset growth.

2. When the attribute subset $A$ is given, $G^0$ and $E^\theta$ are both monotonically increasing as the threshold $\theta$ increases. Meanwhile, $E^\theta$ and $H^\theta$ are both monotonically decreasing with the threshold $\theta$ growth. It shows that the uncertainty of a 4HIS as the threshold $\theta$ increases.

Thus, $G^0$, $E^\theta$, $E^\theta$, and $H^\theta$ can be applied to measuring uncertainty of a 4HIS.

### 6.2 Dispersion analysis

In this part, the standard deviation is applied to do effectiveness analysis of the proposed measures.

Let $X = \{x_1, \cdots, x_n\}$ be a data set. Then arithmetic average value, standard deviation and standard deviation coefficient of $X$ are defined as follows:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \sigma(X) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2},$$

$$CV(X) = \frac{\sigma(X)}{\overline{x}}$$

Then, according to the above experiments, the $CV$ values of measure sets on each data set are computed (see Figs. 10, 11, 12, 13, 14, 15, 16, 17, and 18).

From Figs. 10, 11, 12, 13, 14, 15, 16, 17, and 18, the following conclusions are obtained:

1. When the threshold $\theta$ is changing, $X^\theta$ is minimum in four measure sets in each data set except for $X^\theta_{E}(Ec)$, $X^\theta_{E}(Ec)$, $X^\theta_{E}(Ec)$ and $X^\theta_{E}(Ec)$. It shows that the dispersion degree of $E^\theta$ is minimum in most cases.

2. When the attribute subset $A_j$ is changing, $X^A_{E}$ is minimum in four measure sets in each data set except for $X^A_{E}(Co)$, $X^A_{E}(Cr)$, $X^A_{E}(Cr)$, $X^A_{E}(Ec)$, $X^A_{E}(Ec)$ and $X^A_{E}(Ec)$. It shows that the dispersion degree of $E^\theta$ is minimum in most cases.
Therefore, $E^\theta$ has much better performance to measure 4HISs’ uncertainty.

By summarizing the above experiments, the following results are obtained:

1. If people need only monotonicity, then $G^\theta$, $E^\theta$, $E^\theta_r$ and $H^\theta$ can be used to measure the uncertainty of a 4HIS;

2. If people investigate only dispersion degree, then $E^\theta$ has better performance to measure the uncertainty of a 4HIS;

3. If people consider both monotonicity and dispersion degree, then $E^\theta$ has much better performance to measure the uncertainty of a 4HIS.
7 An application

In this part, an application of the proposed measures for attribute reduction in a 4HIS is given.

**Definition 7.1** Suppose that \((X, AT)\) is a 4HIS. Given \(A \subseteq AT\) and \(\theta \in [0, 1]\). Then \(A\) is called \(\theta\)-consistent, if \(R_A^\theta = R_{AT}^\theta\).

**Definition 7.2** Suppose that \((X, AT)\) is a 4HIS. Given \(a \in A \subseteq AT\). Then \(a\) is called \(\theta\)-independent in \(A\), if \(R_a^\theta \neq R_{A \setminus \{a\}}^\theta\).

**Definition 7.3** Suppose that \((X, AT)\) is a 4HIS. Given \(A \subseteq AT\) and \(\theta \in [0, 1]\). Then \(A\) is called \(\theta\)-independent, if for any \(a \in A\), \(a\) is \(\theta\)-independent in \(A\).
Definition 7.4 Suppose that \((X, AT)\) is a 4HIS. Given \(A \subseteq AT\) and \(\theta \in [0, 1]\). Then \(A\) is called a \(\theta\)-reduct of \(AT\), if \(A\) is both \(\theta\)-consistent and \(\theta\)-independent.

In this paper, the family of all \(\theta\)-coordination subsets (resp., all \(\theta\)-reducts) of \(AT\) is denoted by \(co^\theta(AT)\) (resp., \(red^\theta(AT)\)).

Obviously,

\[
A \in red^\theta(AT) \iff A \in co^\theta(AT) \quad \text{and} \quad \forall P' \subseteq P, \quad P' \notin co^\theta(AT).
\]
Fig. 11  $CV$-values of measure sets on Annealing

Fig. 12  $CV$-values of measure sets on Automobile

Fig. 13  $CV$-values of measure sets on Contraceptive Method Choice
Measures of uncertainty for a four-hybrid information system and their applications

**Theorem 7.5** Suppose that \((X, AT)\) is a 4HIS. Given \(A \subseteq AT\) and \(\theta \in [0, 1]\). Then the following conditions are equivalent:

1. \(A \in \text{co}^\theta (AT)\);
2. \(G^\theta (A) = G^\theta (AT)\);
3. \(H^\theta (A) = H^\theta (AT)\);
4. \(E^\theta_r (A) = E^\theta_r (AT)\);
5. \(E^\theta (A) = E^\theta (AT)\).

**Proof**

1. \(\Rightarrow\) (2). This is obvious.

2. \(\Rightarrow\) (1). Suppose \(G^\theta (A) = G^\theta (AT)\). Then

\[
\frac{1}{n^2} \sum_{i=1}^{n} |R^\theta_A(x_i)| = \frac{1}{n^2} \sum_{i=1}^{n} |R^\theta_{AT}(x_i)|.
\]

So

\[
\sum_{i=1}^{n} (|R^\theta_A(x_i)| - |R^\theta_{AT}(x_i)|) = 0.
\]

Note that \(R^\theta_{AT} \subseteq R^\theta_A\). Then \(\forall i, R^\theta_{AT}(x_i) \subseteq R^\theta_A(x_i)\). This implies that

\[
\forall i, \quad |R^\theta_A(x_i)| - |R^\theta_{AT}(x_i)| \geq 0.
\]

So \(\forall i, \quad |R^\theta_{AT}(x_i)| = 0\). It follows that

\[
\forall i, \quad R^\theta_{AT}(x_i) = R^\theta_A(x_i).
\]

Thus \(R^\theta_A = R^\theta_{AT}\). Hence

\[
A \in \text{co}^\theta (AT).
\]

2. \(\Leftrightarrow\) (5). It can be obtained by Theorem 5.9.

3. \(\Rightarrow\) (1). This is obvious.

(3) \(\Rightarrow\) (1). Suppose \(H^\theta (A) = H^\theta (AT)\). Then

\[
- \sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{|R^\theta_A(x_i)|}{n} = - \sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{|R^\theta_{AT}(x_i)|}{n}.
\]

So

\[
\sum_{i=1}^{n} \log_2 \frac{|R^\theta_A(x_i)|}{|R^\theta_{AT}(x_i)|} = 0.
\]
Fig. 16 CV-values of measure sets on Echocardiogram

Fig. 17 CV-values of measure sets on Hepatitis

Fig. 18 CV-values of measure sets on Meta-data

Note that $R_{AT}^\theta \subseteq R_A^\theta$. Then $\forall i, R_{AT}^\theta(x_i) \subseteq R_A^\theta(x_i)$. This implies that

$$\forall i, \log_2 \frac{|R_A^\theta(x_i)|}{|R_{AT}^\theta(x_i)|} \geq 0.$$ 

So $\forall i, \log_2 \frac{|R_A^\theta(x_i)|}{|R_{AT}^\theta(x_i)|} = 0$. It follows that $\forall i, R_A^\theta(x_i) = R_{AT}^\theta(x_i)$.

Thus $R_A^\theta = R_{AT}^\theta$. Hence

$A \in \text{co}^\theta(AT)$. 

 Springer
Corollary 7.6 Suppose that \((X, AT)\) is a 4HIS. Given \(A \subseteq AT\) and \(\theta \in [0, 1]\). Then the following conditions are equivalent:

1. \(A \in \text{red}^\theta(\text{AT})\);
2. \(G^\theta(A) = G^\theta(\text{AT})\) and \(\forall \ a \in A, G^\theta(A \setminus \{a\}) \neq G^\theta(\text{AT})\);
3. \(H^\theta(A) = H^\theta(\text{AT})\) and \(\forall \ a \in A, H^\theta(A \setminus \{a\}) \neq H^\theta(\text{AT})\);
4. \(E_r^\theta(A) = E_r^\theta(\text{AT})\) and \(\forall \ a \in A, E_r^\theta(A \setminus \{a\}) \neq E_r^\theta(\text{AT})\);
5. \(E^\theta(A) = E^\theta(\text{AT})\) and \(\forall \ a \in A, E^\theta(A \setminus \{a\}) \neq E^\theta(\text{AT})\).

Proof It can be proved by Theorem 7.5.

8 Conclusions

A HIS contains many types of attributes. It is more difficult to measure an HIS than an IS with a type of attribute. This paper has measured the uncertainty of a 4HIS that contains four types of data or attributes and given its application in attribute reduction. First, a novel distance function for each type of attribute in a 4HIS has been proposed. The proposed distance is more consistent with reality in measuring the difference between two information values on each type of attribute. And then, the tolerance relation has been produced by using the proposed distance. By the way, the information granules composed of the tolerance classes have been constructed, and the information structure formed by the information granules has been presented. Next, four UMs based on the structures have been investigated. Furthermore, the effectiveness of four measures has been verified by statistical analysis. As an application of the proposed UMs, attribute reduction has been studied. We have found the influence of \(\theta\) value on the UM for a 4HIS, which may have potential application value in data mining. This paper provides a new idea of UM for hybrid data. The disadvantage is that attribute reduction algorithms are not given. In the future, we will continue to explore attribute reduction algorithms in a 4HIS based on its UM.

Acknowledgements The authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions, which have helped immensely in improving the quality of the paper. This work is supported by National Natural Science Foundation of China (11761012, 11861010), National Natural Science Foundation of Guangxi Province (2018GXNSFFA281008) and Cultivation Plan of Thousands of Young Backbone Teachers in Higher Education Institutions of Guangxi Province.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Aggarwal M (2017) Representation of uncertainty with information and probabilistic information granules. Int J Fuzzy Syst 19:1617–1634
Bianucci D, Cattaneo G (2009) Information entropy and granulation co-entropy of partitions and coverings: a summary. Trans Rough Sets 10:15–66
Beaubouef T, Petry FE, Arora G (1998) Information-theoretic measures of uncertainty for rough sets and rough relational databases. Inf Sci 109:185–195
Blaszczyński J, Slowinski R, Szelag M (2011) Sequential covering rule induction algorithm for variable consistency rough set approaches. Inf Sci 181(5):987–1002
Chen YM, Xue Y, Ma Y, Xu FF (2017) Measures of uncertainty for neighborhood rough sets. Knowledge-Based Syst 120:226–235
Dai JH, Hu H, Wu WZ, Qian YH, Huang DB (2018) Maximal-discriminability-pair-based approach to attribute reduction in fuzzy rough sets. IEEE Trans Fuzzy Syst 26(4):2175–2187

Dubois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. Int J General Syst 17(2–3):191–209

Delgado A, Romero I (2016) Environmental conflict analysis using an integrated grey clustering and entropy-weight method: a case study of a mining project in Peru. Environ Modell Softw 77:108–121

Dai J, Tian H (2013) Entropy measures and granularity measures for set-valued information systems. Inf Sci 240:72–82

Greco S, Inuiuchi M, Slowinski R (2006) Fuzzy rough sets and multiple premise gradual decision rules. Int J Approx Reason 41:179–211

Gu B, Sheng VS, Wang ZJ, Ho D, Osman S (2015) Incremental learning for v-support vector regression. Neural Netw 67:140–150

Han JW, Kamber M (2011) Data mining: concepts and techniques. Elsevier, Amsterdam

Hu QH, Pedrycz W, Yu DR, Lang J (2010) Selecting discrete and continuous features based on neighborhood decision error minimization. IEEE Trans Syst Man Cybern Part B 40:137–150

Hempelmann CF, Sakoglu U, Gurupur VP, Jampana S (2016) An entropy-based evaluation method for knowledge bases of medical information systems. Exp Syst Appl 46:262–273

Juhola M, Laurikka J (2007) On metrics of two heterogeneous measures in the presence of missing values. Artif Intell Rev 28:163–178

Lin TY (1998) Granular computing on binary relations I: data mining and neighborhood systems. In: A. Skowron, L. Polkowski (eds.), Rough sets in knowledge discovery, Physica-Verlag, pp 107–121

Li ZW, Huang D, Liu XF, Xie NX, Zhang GQ (2020) Information structures in a covering information system. Inf Sci 507:449–471

Li ZW, Liu XF, Dai JH, Chen JL, Fujita H (2020) Measures of uncertainty based on Gaussian kernel for a fully fuzzy information system. Knowledge-Based Syst 196:105791

Liang JY, Qu KS (2010) Selecting discrete and continuous features based on neighborhood decision error minimization. IEEE Trans Syst Man Cybern Part B 40:137–150

Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90:111–127

Zadeh LA (1998) Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information & intelligent systems. Soft Comput 2:23–25

Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353

Zhang X, Mei CL, Chen DG, Li JH (2016) Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy. Pattern Recogn 56:1–15

Zeng AP, Li TR, Liu D, Zhang JB, Chen HM (2015) A fuzzy rough set approach for incremental feature selection on hybrid information systems. Fuzzy Sets Syst 258:39–60

Zeng AP, Li TR, Hu J, Chen HM, Luo C (2017) Dynamical updating fuzzy rough approximations for hybrid data under the variation of attribute values. Inf Sci 378:363–388

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

Zhang L, Zhang B (2007) Theory and application of problem solving-theory and application of granular computing in quotient spaces. Tsinghua University Publishers, Beijing

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.