Artificial Intelligence Review (2024) 57:46
https://doi.org/10.1007/s10462-023-10642-w

Semi-supervised attribute reduction for hybrid data

Zhaowen Li[1] · Jiali He[2] · Pei Wang[2] · Ching-Feng Wen[3,4]

Published online: 12 February 2024

Abstract
Due to the high cost of labelling data, a lot of partially hybrid data are existed in many practical applications. Uncertainty measure (UM) can supply new viewpoints for analyzing data. They can help us in disclosing the substantive characteristics of data. Although there are some UMs to evaluate the uncertainty of hybrid data, they cannot be trivially transplanted into partially hybrid data. The existing studies often replace missing labels with pseudo-labels, but pseudo-labels are not real labels. When encountering high label error rates, work will be difficult to sustain. In view of the above situation, this paper studies four UMs for partially hybrid data and proposed semi-supervised attribute reduction algorithms. A decision information system with partially labeled hybrid data (p-HIS) is first divided into two decision information systems: one is the decision information system with labeled hybrid data (l-HIS) and the other is the decision information system with unlabeled hybrid data (u-HIS). Then, four degrees of importance on a attribute subset in a p-HIS are defined based on indistinguishable relation, distinguishable relation, dependence function, information entropy and information amount. We discuss the difference and contact among these UMs. They are the weighted sum of l-HIS and u-HIS determined by the missing rate and can be considered as UMs of a p-HIS. Next, numerical experiments and statistical tests on 12 datasets verify the effectiveness of these UMs. Moreover, an adaptive semi-supervised attribute reduction algorithm of a p-HIS is proposed based on the selected important degrees, which can automatically adapt to various missing rates. Finally, the results of experiments and statistical tests on 12 datasets show the proposed algorithm is statistically better than some stat-of-the-art algorithms according to classification accuracy.

Keywords Partially labeled hybrid data · p-HIS · Semi-supervised attribute reduction · Indiscernibility relation · Dependence function.

1 Introduction

1.1 Research background

With the development of science and technology, the amount of information increases in geometric progression. The increase of information brings many uncertainties to information processing. To deal with the uncertain information, many scholars have proposed

---

Extended author information available on the last page of the article

many effective methods. such as rough set theory (R-theory), fuzzy set theory and uncertainty measurement (UM). These methods have been successfully used in the following fields: pattern recognition (Cament et al. 2014; Swiniarski and Skowron 2003), image processing (Navarrete et al. 2016), medical diagnosis (Hempelmann et al. 2016; Wang et al. 2019), data mining (Dai et al. 2012) and expert systems (Pawlak 1991).

An information system (IS) based on R-theory was defined by Pawlak (1982). It is the main application field of R-theory. From the above, many scholars have researched UM for an IS. For instance, Wierman (1999) put forward information granulation in an IS; Dai and Tian (2013) researched granularity measurement in a set-valued IS; Liang and Qian (2008) studied information granules in an IS; Qian et al. (2011) proposed fuzzy information granularity by constructing fuzzy information granules; Dai et al. (2013) investigated UM for an incomplete decision IS.

In the view of the rapid growth of data, data generates a large number of attributes. However, some features have little effect on describing data. Therefore, attribute reduction or feature selection becomes very important in information processing. It is particularly important to form a new attribute set by selecting features with important influence, which can keep the data information unchanged. From this, the dimensions of the data are reduced. Hu et al. (2008) studied attribute reduction by using neighborhood rough set model.Singh et al. (2020) obtained a novel attribute reduction method in a set-valued IS. Wang et al. (2019) researched attribute reduction based on local conditional entropy. Dai et al. (2016) discussed attribute reduction in an interval-valued IS. Wang et al. (2020) investigated attribute reduction of self-information based on neighborhood. Hu et al. (2022) presented fast and robust attribute reduction in a fuzzy decision IS based on the separability. Yuan et al. (2022) explored interactive attribute reduction via fuzzy complementary entropy for unlabeled mixed data. Houssein et al. (2022) studied centroid mutation-based search, and rescued optimization algorithm for feature selection and classification. Ershadi and Seifi (2022) gave applications of dynamic feature selection and clustering methods to medical diagnosis. Pashaei and Pashaei (2022) proposed an efficient binary chimp optimization algorithm for feature selection in biomedical data classification. Tiwari and Chaturvedi (2022) introduced a hybrid feature selection approach based on information theory and dynamic butterfly optimization algorithm for data classification.

Sang et al. (2021) considered incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets. Yuan et al. (2021) put forward fuzzy complementary entropy using hybrid-kernel function and its unsupervised attribute reduction.

The deficiency of the existing work mentioned above is that there is less attention paid to attribute reduction for partially hybrid data.

1.2 Motivation and contributions

There are many high-dimensional complex data in many fields. Because of hardware failures, human sabotage, operational errors, virus infections and software failures, the labels of some samples in the data are lost. If only labeled data is used for attribute reduction, the reduction results can't effectively reflect the distribution of data, and the classification performance could be weak. Finding new labeling methods has become extremely important.

Semi-supervised attribute reduction aims to effectively utilize unlabeled data to enhance the effectiveness of attribute reduction, in order to improve the classification performance of learning model. In recent years, semi-supervised attribute reduction has attracted the attention of many scholars. For instance, Dai et al. (2017) constructed a heuristic semi-supervised attribute reduction algorithm on the basis of distinguishing pairs. Wang et al. (2018) designed a semi-supervised attribute reduction algorithm based on information entropy. Zhang et al. (2016) combined *R*-theory with ensemble learning framework, and constructed an ensemble base classifier. Xu et al. (2010) proposed a semi-supervised attribute reduction algorithm based on manifold regularization, which measures the importance of features by maximizing the spacing between different categories. Han et al. (2015) presented a semi-supervised attribute reduction algorithm, which effectively uses the information in a large number of unmarked video data to distinguish target categories by combining semi-supervised scatter points. Wu et al. (2021) used minimal redundancy to research semi-supervised feature selection. Wan et al. (2021) proposed a semi-supervised attribute reduction method based on neighborhood rough set.

The comparative of this paper with the research results of UM or attribute reduction about some above literatures is shown in Table 1.

However, the existing studies face new challenges in attribute reduction. The discretization of continuous attributes can result in the loss of structural information in the data. Additionally, when the amount of data increases, the quality of reduction becomes low. Attribute reduction based on an identification matrix and identification pair significantly increases the operation cost. While the existing studies can evaluate the uncertainty of hybrid data, they lack an effective mechanism for converting partially hybrid data into hybrid data. Furthermore, the existing studies often replace missing labels with pseudo-labels, which are not real labels. This approach becomes difficult to sustain when encountering high label error rates.

Based on the above research motivation, this paper cleverly divides a p-HIS into the l-HIS and u-HIS, and views the sum of their importance degree as the importance degree of a p-HIS where the incomplete rate of labels is used as the weight. The major contributions are summarized as follows.

(1) The facts that a discernibility pair set for hybrid data is actually a distinguishable relation and a p-HIS can induce the l-HIS and u-HIS are showed.
(2) Four kinds of important degrees on each attribute subset in a p-HIS are defined. They use the weighted importance degree sum of the induced l-HIS and u-HIS to deal with UM for each subsystem. The l-HIS and u-HIS can more deeply reflect the importance or classification ability of an attribute subset.
(3) The performance of four kinds of important degrees in a p-HIS is tested. Numerical experiments and statistical tests verify the effectiveness of these important degrees.
(4) Based on the selected important degrees, heuristic algorithms of semi-supervised attribute reduction in a p-HIS are constructed. The experiment results show the constructed algorithm is statistically better than some stat-of-the-art algorithms according to classification accuracy.

Table 1 The comparison of this paper with some literatures

Literature	Data type	Method	Research tool
Dai et al. (2016)	Interval-valued data	Information entropy	R-theory Information theory
Dai et al. (2017)	Partially labeled categorical data	The importance degree sum	R-theory
Pashaei and Pashaei (2022)	Biomedical data	Binary chimp optimization algorithm	Optimization theory
Sang et al. (2021)	Dynamic real-valued data	Conditional information entropy	Information theory
Singh et al. (2020)	Set-valued data	Dependency degree	R-theory
Tiwari and Chaturvedi (2022)	Hybrid data	Dynamic butterfly optimization algorithm	Optimization theory Information theory
Wu et al. (2021)	Real-valued data	Minimal redundancy	Information theory
Wang et al. (2020)	Real-valued data	Neighborhood self-information	Information theory
Wang et al. (2018)	Partially labeled categorical data	Information entropy	R-theory Information theory
Wang et al. (2019)	Real-valued data	Local conditional information entropy	Information theory
Xu et al. (2010)	Partially labeled hybrid data	Manifold regularization Maximizing the spacing between different categories	Information theory
Yuan et al. (2022)	Unlabeled mixed data	Fuzzy complementary entropy	Information theory
Our paper	Partially labeled hybrid data	The weighted importance degree sum	R-theory Information theory

1.3 Organization

In Sect. 2, the indiscernibility relation and discernibility relation in a p-HIS are defined. In Sect. 3, four degrees of importance in a p-HIS are introduced. In Sect. 4, UM for a p-HIS is investigated. A new algorithm is proposed which is related to semi-supervised attribute reduction in a p-HIS. In Sect. 5, the experiments on algorithm performance and effectiveness analysis of the proposed degrees of importance are conducted. In Sect. 6, the statistical test of algorithm performance is conducted. In Sect. 7, this paper is summarized. The logical structure of this paper is shown in Fig. 1.

2 Preliminaries

In this paper, $U=\{u_1,u_2,\cdots,u_n\}$, 2^U is the set which contains all subsets of U and $|X|$ represents the cardinality of X. Let

$$\delta = U\times U,\ \triangle=\{(e,e)\,:\,e\in U\}, R\subseteq\delta.$$

If $R=\delta$, R is called universal relation; if $R=\triangle$, R is called identity relation.

Given that R is an equivalence relation and any $u\in U$, $[u]_R$ is an equivalence class of u which is denoted as: $[u]_R=\{v\in U\,:\,uRv\}$. $U/R=\{[u]_R\,:\,u\in U\}$ denotes all equivalence classes of R.

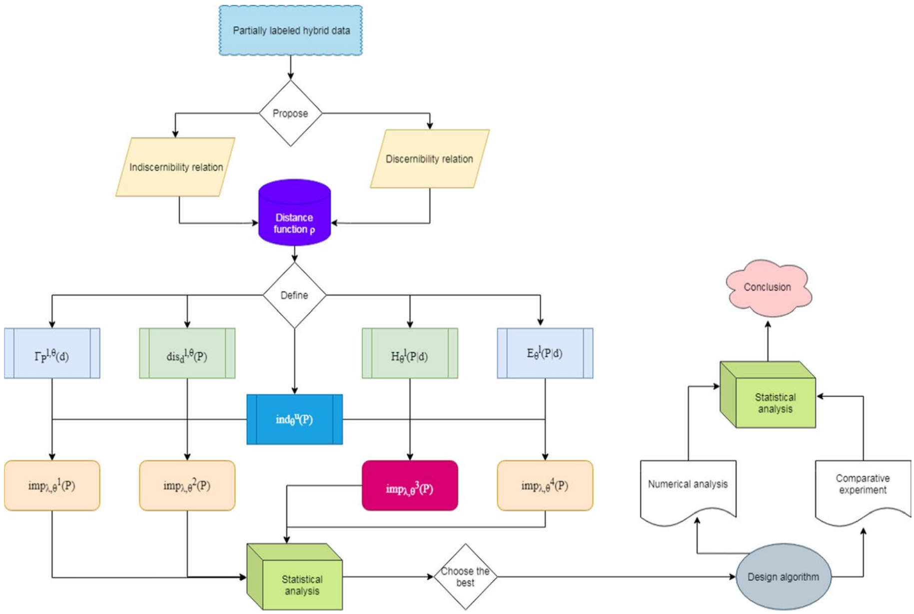


Fig. 1 The logical structure of the paper

Definition 2.1 [(Pawlak 1982)] Let U be a universe and A a finite attribute set. (U, A) is called an information system (IS). $a : U \to V_a$ is called an information function for any $a \in A$, where $V_a = \{a(e) : e \in U\}$.

Let (U, A) be an IS, for any $B \subseteq A$, $ind(B)$ can be defined as

$$ind(B) = \{(e, t) \in U \times U : \forall\, a \in B,\ a(e) = a(t)\}.$$

$ind(B)$ is called the indiscernibility relation of B.

Definition 2.2 [(Dai et al. 2017)] (U, A) is an IS, for any $B \subseteq A$. $dis_d(B)$ is called the discernibility relation of B if

$$dis(B) = \{(e, t) \in U \times U : \exists\, a \in B,\ a(e) \neq a(t)\}.$$

Let (U, A) be an IS and B a subset of A. B is called a coordinate subset of A if $ind(B) = ind(A)$. All coordinate subsets of A is denoted by $co(A)$. B is called a reducts of A if $B \in co(A)$ and for each $a \in B$, $B - \{a\} \notin co(A)$. All reducts of A is denoted by $red(A)$.

Proposition 2.3 *(U, A) is an IS and B a subset of A. Then $B \in co(A) \Leftrightarrow dis(B) = dis(A)$.*

Corollary 2.4 *(U, A) is an IS with $B \subseteq A$. Then*

$$B \in red(A) \Leftrightarrow dis(B) = dis(A) \text{ and } \forall\, a \in B,\ dis(B - \{a\}) \neq dis(A).$$

(U, C, d) is called a DIS, if $(U, C \cup \{d\})$ is an IS and d is a decision attribute.
Let (U, C, d) be a DIS with $B \subseteq C$ and $X \in 2^U$. Define

$$\begin{aligned}
R_B =& \{(e, t) \in U \times U : \forall\, a \in B,\ a(e) = a(t)\},\ [e]_B = \{t \in U : (e, t) \in R_B\};\\
R_d =& \{(e, t) \in U \times U : d(e) = d(t)\},\ [e]_d = \{t \in U : (e, t) \in R_d\};\\
U/R_d =& \{[e]_d : e \in U\};\\
\underline{R_B}(X) =& \{e \in U : [e]_B \subseteq X\}.
\end{aligned}$$

Definition 2.5 [(Dai et al. 2017)] Let (U, C, d) be a DIS with $B \subseteq C$. Define

$$dis_d(B) = \{(e, t) \in U \times U : \exists\, a \in B, a(e) \neq a(t) \bigwedge d(e) \neq d(t)\}.$$

Then $dis_d(B)$ is called the discernibility relation of B with d.

Let (U, C, d) be a DIS with $B \subseteq C$. Define $POS_B(d) = \bigcup_{D \in U/R_d} \underline{R_B}(D)$ and $\Gamma_B(d) = \frac{|POS_B(d)|}{|U|}$. B is called a coordinate subset of C if $R_B = R_C$, All coordinate subsets of C with d is denoted by $co_d(C)$. B is called a reduct of C with d if $B \in co_d(C)$ and for each $a \in B$, $B - \{a\} \notin co_d(C)$. All reducts of C with d is denoted by $red_d(C)$.
From Proposition 2.3 and Definition 2.5, we can get the following results.

Proposition 2.6 *Let (U, C, d) be a DIS with $B \subseteq C$. The following conditions are equivalent*:

(1) $B \in co_d(C)$;
(2) $POS_B(d) = POS_C(d)$;
(3) $\Gamma_B(d) = \Gamma_C(d)$;
(4) $dis_d(B) = dis_d(C)$.

Corollary 2.7 *(U, C, d) is a DIS with $B \subseteq C$. The following conditions are equivalent*:

(1) $B \in red_d(C)$;
(2) $POS_B(d) = POS_C(d)$, $POS_{B-\{a\}}(d) \neq POS_C(d)$ *for any* $a \in B$;
(3) $\Gamma_B(d) = \Gamma_C(d)$, $\Gamma_{B-\{a\}}(d) \neq \Gamma_C(d)$ *for any* $a \in B$;
(4) $dis_d(B) = dis_d(C)$, $dis_d(B - \{a\}) \neq dis_d(C)$ *for any* $a \in B$.

3 An information system for partially labeled hybrid data

The information system for partially labeled hybrid data plays an important role in real life. In order to better study it, we first define the following definitions.

3.1 The definition of a p-HIS

$\diamond$ represents an unknown information value, and $*$ stands for an unknown label.

Suppose that (U, C, d) is a DIS with $a \in C$. Put

$$\begin{aligned} V_a^\diamond &= \{a(u) : u \in U,\ a(u) \neq \diamond\}; \\ V_d^* &= \{d(u) : u \in U,\ d(u) \neq *\}. \end{aligned}$$

From the above, $V_a^\diamond$ and V_d^* are the sets which contain all known information values of a and all known labels of d respectively.

Definition 3.1 Let (U, C, d) be a DIS and $C = C^c \cup C^r$, where C^c and C^r are categorical and real-valued attributes, respectively. Put

$$U^l = \{u \in U : d(u) \neq *\}, \quad U^u = \{u \in U : d(u) = *\}.$$

Then $U^l \cup U^u = U, \quad U^l \cap U^u = \emptyset$.

(1) (U, C, d) is called a DIS with l-HIS, if $\exists\, a \in C$ and $u \in U$, $a(u) = \diamond$ and $U^l = U$.
(2) (U, C, d) is called a DIS with p-HIS, if $\exists\, a \in C$ and $u \in U$, $a(u) = \diamond$, $U^l \neq \emptyset$ and $U^u \neq \emptyset$.
(3) (U, C, d) is called a DIS with u-HIS, if $\exists\, a \in C$ and $u \in U$, $a(u) = \diamond$ and $U^u = U$.

Obviously,

$$V_d^* = \{d(u) : u \in U^l\}.$$

Since there are no labeled objects in u-HIS (U, C, d), we abbreviated (U, C, d) as (U, C).

Table 2 A p-HIS (U, C, d)

U	Headache (a_1)	Muscle pain (a_2)	Temperature (a_3)	Symptom (d)
u_1	Sick	Yes	39.5	Flu
u_2	Sick	Yes	40	Health
u_3	Middle	$\diamond$	39	*
u_4	No	Yes	38	Flu
u_5	$\diamond$	Yes	36.5	Rhinitis
u_6	Middle	No	$\diamond$	Rhinitis
u_7	No	No	36	Health
u_8	No	$\diamond$	$\diamond$	Health
u_9	$\diamond$	Yes	37	*

Definition 3.2 Let (U, C, d) be a p-HIS. Here, (U^l, C, d) and (U^u, C, d) are called the l-HIS and u-HIS induced by (U, C, d), respectively.

Example 3.3 Table 2 is a p-HIS (U, C, d), where $U = \{u_1, u_2, \cdots, u_9\}$, $C = C^c \cup C^r$, $C^c = \{a_1, a_2\}$ and $C^r = \{a_3\}$.

Example 3.4 (Continue with Example 3.3)

$$\begin{aligned} V^{\diamond}_{a_1} &= \{Sick, Middle, No\}, \quad V^{\diamond}_{a_2} = \{Yes, No\}, \\ V^{\diamond}_{a_3} &= \{39.5, 40, 39, 38, 36.5, 36, 37\}; \\ V^{*}_{d} &= \{Flu, Rhinitis, Health\} \neq V_d; \\ V^{l} &= \{u_1, u_2, u_4, u_5, u_6, u_7, u_8\}, \quad V^{u} = \{u_3, u_9\}. \end{aligned}$$

Definition 3.5 (U, C, d) is a p-HIS. The incomplete rate λ of labels is defined as

$$\lambda = \frac{|U^u|}{|U|}.$$

3.2 A novel distance function in a p-HIS

In order to effectively distinguish the difference between objects in a p-HIS, we will give a new concept.

Definition 3.6 [(Zhang et al. 2022)] (U, C, d) is a p-HIS and $|V^*_d| = s$. Given $a \in C^c$ and $u \in U^l$ with $a(u) \neq \diamond$.

Denote

$$V^*_d = \{d(v) : v \in U^l\} = \{d_1, d_2, \cdots, d_s\}.$$

Define

$$N(a,u) = |\{v \in U^l : a(v) = a(u)\}|,$$
$$N_i(a,u) = |\{v \in U^l : a(v) = a(u),\ d(v) = d_i\}|.$$

By Definition 3.6, we have

$$N(a,u) = \sum_{i=1}^{s} N_i(a,u).$$

Definition 3.7 [(Zhang et al. 2022)] Let (U, C, d) be a p-HIS with $|V_d^*| = s$. Given $a \in C^c$, $e, t \in U^l$ with $a(e) \neq \diamond$ and $a(t) \neq \diamond$. Define the distance as follows

$$\rho_c^l(a(e), a(t)) = \frac{1}{2} \sum_{i=1}^{s} |\frac{N_i(a,e)}{N(a,e)} - \frac{N_i(a,t)}{N(a,t)}|.$$

Obviously,

$$\rho_c^l(a(u), a(u)) = 0,\ 0 \leq \rho_c^l(a(u), a(v)) \leq 1.$$

Definition 3.8 [(Zhang et al. 2022)] Let (U, C, d) be a p-HIS. Suppose $a \in C^r$ and $e, t \in U^l$ with $a(e) \neq \diamond$, $a(t) \neq \diamond$. Define the distance as follows

$$\rho_r^l(a(e), a(t)) = \frac{|a(e) - a(t)|}{\hat{a}},$$

where $\hat{a} = max\{a(e) : e \in U^l\} - min\{a(e) : e \in U^l\}$.

If $\hat{a} = 0$, let $\rho_r^l(a(e), a(t)) = 0$. Obviously,

$$\rho_r^l(a(e), a(e)) = 0,\ \ \rho_r^l(a(e), a(t)) \leq 1.$$

Definition 3.9 Let (U, C, d) be a p-HIS. Given $a \in C$ and $e, t \in U^l$. Define the distance as follows

$\rho(a(e), a(t)) =$

$$\begin{cases} 0, & e = t; \\ 0, & e \neq t,\ a \in C,\ a(e) = *\ \ or\ a(t) = *,\ d(e) = d(t); \\ 1 - \frac{1}{|V_a^*|^2}, & e \neq t,\ a \in C,\ a(e) = *,\ a(t) = *,\ d(e) \neq d(t); \\ 1 - \frac{1}{|V_a^*|}, & e \neq t,\ a \in C,\ a(e) \neq *,\ a(t) = *,\ d(e) \neq d(t); \\ 1 - \frac{1}{|V_a^*|}, & e \neq t,\ a \in C,\ a(e) = *,\ a(t) \neq *,\ d(e) \neq d(t); \\ 0, & e \neq t,\ a \in C,\ a(e) \neq *,\ a(t) \neq *,\ a(e) = a(t),\ d(e) = d(t); \\ 0, & e \neq t,\ a \in C,\ a(e) \neq *,\ a(t) \neq *,\ a(e) = a(t),\ d(e) \neq d(t); \\ \rho_c^l(a(e), a(t)), & e \neq t,\ a \in C^c,\ a(e) \neq *, a(t) \neq *,\ a(e) \neq a(t),\ d(e) = d(t); \\ \rho_c^l(a(e), a(t)), & e \neq t,\ a \in C^c,\ a(e) \neq *, a(t) \neq *,\ a(e) \neq a(t),\ d(e) \neq d(t); \\ \rho_r^l(a(e), a(t)), & e \neq t,\ a \in C^r,\ a(e) \neq *, a(t) \neq *,\ a(e) \neq a(t),\ d(e) = d(t) \\ \rho_r^l(a(e), a(t)), & e \neq t,\ a \in C^r,\ a(e) \neq *, a(t) \neq *,\ a(e) \neq a(t),\ d(e) \neq d(t). \end{cases}$$

3.3 Some concepts of a p-HIS and related results

Definition 3.10 Let P be a subset of C and (U, C, d) a p-HIS. (U^l, C, d) is the l-HIS induced by (U, C, d). Denote

$$\begin{aligned}
|U^l| =& n_l;\\
R_P^{l,\theta} =& \{(e,t) \in U^l \times U^l : \forall\, a \in P,\ \rho(a(e), a(t)) \leq \theta\},\\
[e]_P^{l,\theta} =& \{t \in U^l : (e,t) \in R_P^{l,\theta}\};\\
R_d^{l,\theta} =& \{(e,t) \in U^l \times U^l : d(e) = d(t)\},\\
[e]_d^{l,\theta} =& \{t \in U^l : (e,t) \in R_d^{l,\theta}\};\\
\underline{R_P^{l,\theta}}(X) =& \{e \in U^l : [e]_P^{l,\theta} \subseteq X\},\ X \subseteq U^l;\\
U^l/d =& \{[e]_d^{l,\theta} : e \in U^l\} = \{D_1, D_2, \cdots, D_r\};\\
POS_P^{l,\theta}(d) =& \bigcup_{i=1}^{r} \underline{R_P^{l,\theta}}(D_i).
\end{aligned}$$

In the above definition, θ is a parameter to control the distance between $a(e)$ and $a(t)$.

Definition 3.11 Let P be a subset of C and (U, C, d) a p-HIS. (U^l, C, d) is the l-HIS induced by (U, C, d). Denote

$$dis_d^{l,\theta}(P) = \{(e,t) \in U^l \times U^l : \exists\, a \in P, \rho(a(e), a(t)) > \theta \bigwedge d(e) \neq d(t)\}.$$

$dis_d^{l,\theta}(P)$ is called the relative discernibility relation of P to d on U^l.

Definition 3.12 Let P be a subset of C and (U, C, d) a p-HIS. (U^u, C, d) is the u-HIS induced by (U^u, C, d). Then (U^u, C, d) can be seen as (U^u, C), and

$$ind_\theta^u(P) = \{(e,t) \in U^u \times U^u : \forall\, a \in P, \rho(a(e), a(t)) \leq \theta\}$$

is called the discernibility relation of P on U^u.

P is a subset of C and (U, C, d) is a p-HIS, According to Kryszkiewicz's ideal Kryszkiewicz (1999), $\partial_P^{l,\theta} : U^l \to 2^{V_d^*}$ is defined as follows:

$$\partial_P^{l,\theta}(u) = d([u]_P^{l,\theta}),$$

Then $\partial_P^{l,\theta}(u)$ is called generalized decision of u in (U^l, P, d), and $\partial_P^{l,\theta} = \{\partial_P^{l,\theta}(u) : u \in U^l\}$.

Definition 3.13 (U, C, d) is a p-HIS. $\forall\, u \in U^l$, $|\partial_C^{l,\theta}(u)| = 1$, (U, C, d) is a θ-consistent; otherwise, (U, C, d) is called θ-inconsistent.

Proposition 3.14 *(U, C, d) is a p-HIS and P is a subset of C. $\forall\, u \in U^l$.*

$$[u]_P^{l,\theta} \subseteq [u]_d^{l,\theta} \quad \Leftrightarrow \quad |\partial_P^{l,\theta}(u)| = 1.$$

Proof }} $\Rightarrow$ ε. Let $[u]_P^{l,\theta} \subseteq [u]_d^{l,\theta}$. Suppose $w \in \partial_P^{l,\theta}(u)$. Then $\exists\, v \in [u]_P^{l,\theta}$, $w = d(v)$. $v \in [u]_P^{l,\theta}$ implies that $v \in [u]_d^{l,\theta}$. So $w = d(v) = d(u)$. Thus $|\partial_P^{l,\theta}(u)| = 1$.

}} $\Leftarrow$ ε. Let $|\partial_P^{l,\theta}(u)| = 1$. $\forall\ v \in [u]_P^{l,\theta}$. Then $d(v) \in \partial_P^{l,\theta}(u)$. Since $d(u) \in \partial_P^{l,\theta}(u)$ and $|\partial_P^{l,\theta}(u)| = 1$, therefore $d(u) = d(v)$. Then $v \in [u]_d^{l,\theta}$. Thus $[u]_P^{l,\theta} \subseteq [u]_d^{l,\theta}$. □

Corollary 3.15 *(U, C, d) is a p-HIS and P is a subset of A.*

$$R_P^{l,\theta} \subseteq R_d^{l,\theta} \quad \Leftrightarrow \quad \forall\, u \in U^l,\ |\partial_P^{l,\theta}(u)| = 1.$$

Proof Obviously. □

Proposition 3.16 *Let (U, C, d) be a p-HIS. (U, C, d) is θ-consistent $\Leftrightarrow$ $R_C^{l,\theta} \subseteq R_d^{l,\theta}$.*

Proof It can be proved by Corollary 3.15. □

4 Uncertainty measurement for a p-HIS

In this part, UM for a p-HIS is investigated by using four kinds of important degrees on the given attribute subset.

4.1 The type 1 importance of a subsystem in a p-HIS

Definition 4.1 Let (U, C, d) be a p-HIS. P is a subset of C and $|U^l| = n_l$. Put

$$\Gamma_P^{l,\theta}(d) = \frac{|POS_P^{l,\theta}(d)|}{n_l};$$

Then $\Gamma_P^{l,\theta}(d)$ is called the dependence of P to d in U^l.

Proposition 4.2 *Let (U, C, d) be a p-HIS with $|U^l| = n_l$. Denote*

$$U^l/R_d^{l,\theta} = \{D_1, D_2, \cdots, D_r\}.$$

(1) $\Gamma_P^{l,\theta}(d) = \dfrac{\sum\limits_{i=1}^{r} |\underline{R_P^{l,\theta}}(D_i)|}{n_l}$.

(2) $0 \leq \Gamma_P^{l,\theta}(d) \leq 1$.

(3) *If $P \subseteq Q \subseteq C$, then*

$$\Gamma_P^{l,\theta}(d) \leq \Gamma_Q^{l,\theta}(d).$$

Proof (1) Obviously, $\forall\, i$, $\underline{R_P^{l,\theta}}(D_i) \subseteq D_i$.

$\{D_1, D_2 \cdots, D_r\}$ is a partition of U^l. Then

$$|POS_P^{l,\theta}(d)| = |\bigcup_{i=1}^{r} \underline{R_P^{l,\theta}}(D_i)| = \sum_{i=1}^{r} |\underline{R_P^{l,\theta}}(D_i)|.$$

Thus

$$\Gamma_P^{l,\theta}(d) = \frac{\sum_{i=1}^{r} |\underline{R_P^{l,\theta}}(D_i)|}{n_l}.$$

(2) This holds by (1).

(3) Suppose $P \subseteq Q \subseteq C$. Then $\forall\, u \in U^l, [u]_Q^{l,\theta} \subseteq [u]_P^{l,\theta}$. So

$$\forall\, i,\ \underline{R_P^{l,\theta}}(D_i) \subseteq \underline{R_Q^{l,\theta}}(D_i).$$

It implies that

$$\forall\, i,\ |\underline{R_P}(D_i) \leq |\underline{R_Q}(D_i)|.$$

By (1),

$$\Gamma_P^{l,\theta}(d) \leq \Gamma_Q^{l,\theta}(d).$$

□

Definition 4.3 Let (U, C, d) be a p-HIS with $\lambda = \frac{|U^u|}{|U|}$ and $P \subseteq C$. Then the type 1 importance of the subsystem (U, P, d) is defined as

$$imp_{\lambda,\theta}^{(1)}(P) = (1-\lambda)\frac{\Gamma_P^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)} + \lambda\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}.$$

In the above definition, $\frac{\Gamma_P^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)}$ and $\frac{|ind^u(C)|}{|ind^u(P)|}$ can be viewed as the importance of (U^l, P, d) and (U^u, P, d), respectively. λ means the missing rate of labels, which is processed as a weight. We define the sum of the importance of (U^l, P, d) and (U^u, P, d) with the missing rate of labels as the type 1 importance of (U, P, d).

Example 4.4 (Continue with Example 3.3) Given $\theta = 0.5$ and $\lambda = \frac{2}{9} \approx 0.2222$. Then

$$\begin{aligned}
&\rho(a_1(u_1), a_1(u_1)) = 0,\ \rho(a_1(u_1), a_1(u_2)) = 0,\ \rho(a_1(u_1), a_1(u_4)) = 0.1667,\\
&\rho(a_1(u_1), a_1(u_5)) = 0.6667,\ \rho(a_1(u_1), a_1(u_6)) = 1,\\
&\rho(a_1(u_1), a_1(u_7)) = 0.1667,\\
&\rho(a_1(u_1), a_1(u_8)) = 0.1667,\ \rho(a_1(u_2), a_1(u_2)) = 0,\\
&\rho(a_1(u_2), a_1(u_4)) = 0.1667,\\
&\rho(a_1(u_2), a_1(u_5)) = 0.6667,\ \rho(a_1(u_2), a_1(u_6)) = 1,\\
&\rho(a_1(u_2), a_1(u_7)) = 0.1667,\\
&\rho(a_1(u_1), a_1(u_8)) = 0.1667,\ \rho(a_1(u_4), a_1(u_4)) = 0,\\
&\rho(a_1(u_4), a_1(u_5)) = 0.6667,\\
&\rho(a_1(u_4), a_1(u_6)) = 1,\ \rho(a_1(u_4), a_1(u_7)) = 0,\\
&\rho(a_1(u_4), a_1(u_8)) = 0,\\
&\rho(a_1(u_5), a_1(u_5)) = 0,\ \rho(a_1(u_5), a_1(u_6)) = 0,\\
&\rho(a_1(u_5), a_1(u_7)) = 0.6667,\\
&\rho(a_1(u_5), a_1(u_8)) = 0.6667,\ \rho(a_1(u_6), a_1(u_6)) = 0,\\
&\rho(a_1(u_6), a_1(u_7)) = 1,\\
&\rho(a_1(u_6), a_1(u_8)) = 1,\ \rho(a_1(u_7), a_1(u_7)) = 0,\\
&\rho(a_1(u_7), a_1(u_8)) = 0,\\
&\rho(a_1(u_8), a_1(u_8)) = 0.
\end{aligned}$$

Thus

$R^{l,\theta}_{\{a_1\}} = \{(u_1,u_1),(u_1,u_2),(u_1,u_4),(u_1,u_7),(u_1,u_8),(u_2,u_2),(u_2,u_4),(u_2,u_7),(u_2,u_8),(u_4,u_4),(u_4,u_7),(u_4,u_8),(u_5,u_5),(u_5,u_6),(u_6,u_6),(u_7,u_7),(u_7,u_8),(u_8,u_8)\}$.

$$\begin{aligned}
&[u_1]^{l,\theta}_{\{a_1\}} = \{u_1,u_2,u_4,u_7,u_8\},\ [u_2]^{l,\theta}_{\{a_1\}} = \{u_1,u_2,u_4,u_7,u_8\},\\
&[u_4]^{l,\theta}_{\{a_1\}} = \{u_1,u_2,u_4,u_7,u_8\},\ [u_5]^{l,\theta}_{\{a_1\}} = \{u_5,u_6\},\ [u_6]^{l,\theta}_{\{a_1\}} = \{u_5,u_6\},\\
&[u_7]^{l,\theta}_{\{a_1\}} = \{u_1,u_2,u_4,u_7,u_8\},\ [u_8]^{l,\theta}_{\{a_1\}} = \{u_1,u_2,u_4,u_7,u_8\}.
\end{aligned}$$

$$U^l/R^{l,\theta}_d = \{D_1, D_2, D_3\},$$

where $D_1 = \{u_1,u_4\}$, $D_2 = \{u_2,u_7,u_8\}$, $D_3 = \{u_5,u_6\}$.

$$\underline{R^{l,\theta}_{\{a_1\}}}(D_1) = \emptyset,\ \underline{R^{l,\theta}_{\{a_1\}}}(D_2) = \{u_5,u_6\},\ \underline{R^{l,\theta}_{\{a_1\}}}(D_3) = \emptyset.$$

$$\Gamma^{l,\theta}_{\{a_1\}}(d) = \frac{|\underline{R^{l,\theta}_{\{a_1\}}}(D_1)| + |\underline{R^{l,\theta}_{\{a_1\}}}(D_2)| + |\underline{R^{l,\theta}_{\{a_1\}}}(D_3)|}{n_l} = \frac{0+2+0}{7} \approx 0.2857.$$

Similarly, $\Gamma^{l,\theta}_{\{a_2\}}(d) = 0,\ \Gamma^{l,\theta}_{\{a_3\}}(d) \approx 0.2857,\ \Gamma^{l,\theta}_{C}(d) \approx 0.4286.$

Next,

$$\begin{aligned}
\rho(a_1(u_3), a_1(u_3)) &= 0,\ \rho(a_1(u_3), a_1(u_9)) = 0,\\
\rho(a_1(u_9), a_1(u_3)) &= 0,\ \rho(a_1(u_9), a_1(u_9)) = 0.
\end{aligned}$$

Then

$$ind_\theta^u(\{a_1\}) = \{(u_3, u_3), (u_3, u_4), (u_4, u_3), (u_4, u_4)\}.$$

Thus $|ind_\theta^u(\{a_1\})| = 4$.

Similarly, $|ind_\theta^u(\{a_2\})| = 4, |ind_\theta^u(\{a_3\})| = 4,\ |ind^u(C)| = 4$.

Finally, the $imp_{\lambda,\theta}^{(1)}(P)$ of each attribute is calculated as follows.

$$\begin{aligned} imp_{\lambda,\theta}^{(1)}(\{a_1\}) =&(1 - 0.2222) * \frac{0.2857}{0.4286} + 0.2222 * \frac{4}{4} \approx 0.7407, \\ imp_{\lambda,\theta}^{(1)}(\{a_2\}) =&(1 - 0.2222) * \frac{0}{0.4286} + 0.2222 * \frac{4}{4} = 0.2222, \\ imp_{\lambda,\theta}^{(1)}(\{a_3\}) =&(1 - 0.2222) * \frac{0.2857}{0.4286} + 0.2222 * \frac{4}{4} \approx 0.7407. \end{aligned}$$

Proposition 4.5 *Let (U, C, d) be a p-HIS with $\lambda = \frac{|U^u|}{|U|}$. Then the following properties hold.*

(1) $0 \leq imp_{\lambda,\theta}^{(1)}(P) \leq 1$;

(2) $imp_{\lambda,\theta}^{(1)}(C) = 1$;

(3) $P \subseteq Q \subseteq C$ *implies to* $imp_{\lambda,\theta}^{(1)}(P) \leq imp_{\lambda,\theta}^{(1)}(Q)$;

(4) $imp_{\lambda,\theta}^{(1)}(P) = 1 \Leftrightarrow \Gamma_P^{l,\theta}(d) = \Gamma_C^{l,\theta}(d), |ind_\theta^u(P)| = |ind_\theta^u(C)|$.

Proof "(1) and (2)" are obvious.

(3) Since $P \subseteq Q \subseteq C$, we have

$$\Gamma_P^{l,\theta}(d) \leq \Gamma_Q^{l,\theta}(d),\ \ |ind_\theta^u(Q)| \leq |ind_\theta^u(P)|.$$

Then

$$\frac{\Gamma_P^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)} \leq \frac{\Gamma_Q^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)},\ \ \frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|} \leq \frac{|ind_\theta^u(C)|}{|ind_\theta^u(Q)|}.$$

Thus

$$(1-\lambda)\frac{\Gamma_P^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)} \leq (1-\lambda)\frac{\Gamma_Q^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)},\ \ \lambda\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|} \leq \lambda\frac{|ind_\theta^u(C)|}{|ind_\theta^u(Q)|}.$$

Hence $imp_{\lambda,\theta}^{(1)}(P) \leq imp_{\lambda,\theta}^{(1)}(Q)$.

(4) }} $\Leftarrow$ ε is clear. Below, we prove }} $\Rightarrow$ ε.

Suppose $imp_{\lambda,\theta}^{(1)}(P) = 1$. Then

$$(1-\lambda)\frac{\Gamma_P^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)} + \lambda\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|} = 1 = (1-\lambda) + \lambda.$$

This implies that

$$(1-\lambda)(1-\frac{\Gamma_P^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)})+\lambda(1-\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|})=0.$$

Note that $1-\frac{\Gamma_P^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)}=\frac{\Gamma_C^{l,\theta}(d)-\Gamma_P^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)}\geq 0$, $1-\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}=\frac{|ind_\theta^u(P)|-|ind_\theta^u(C)|}{|ind_\theta^u(P)|}\geq 0$, Then $1-\frac{\Gamma_P^{l,\theta}(d)}{\Gamma_C^{l,\theta}(d)}=0, 1-\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}=0$. Thus

$$\Gamma_P^{l,\theta}(d)=\Gamma_C^{l,\theta}(d)\ and\ |ind_\theta^u(P)|=|ind_\theta^u(C)|.$$

□

4.2 The type 2 importance of a subsystem in a p-HIS

In the following definition, $\frac{|dis_d^{l,\theta}(P)|}{|dis_d^{l,\theta}(C)|}$ and $\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}$ can be viewed as the importance of (U^l,P,d) and (U^u,P,d), respectively. λ means the missing rate of labels, which is processed as a weight. We define the sum of the importance of (U^l,P,d) and (U^u,P,d) with the missing rate of labels as the type 2 importance of (U, P, d).

Definition 4.6 Let (U, C, d) be a p-HIS with $\lambda=\frac{|U^u|}{|U|}$ and $P\subseteq C$. Then the type 2 importance of the subsystem (U, P, d) is defined as

$$imp_{\lambda,\theta}^{(2)}(P)=(1-\lambda)\frac{|dis_d^{l,\theta}(P)|}{|dis_d^{l,\theta}(C)|}+\lambda\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}.$$

Example 4.7 (Continue with Example 4.4) Obviously,

$dis_d^{l,\theta}(\{a_1\})=\{(u_1,u_5),(u_1,u_6),(u_2,u_5),(u_2,u_6),(u_4,u_5),(u_4,u_6),(u_5,u_7),(u_5,u_8),(u_6,u_7),(u_6,u_8)\}$.

Since $dis_d^{l,\theta}(\{a_1\})$ is symmetric, we have $|dis_d^{l,\theta}(\{a_1\})|=20$.

Similarly, $|dis_d^{l,\theta}(\{a_2\})|=0, |dis_d^{l,\theta}(\{a_3\})|=22, |dis_d^{l,\theta}(C)|=26$.

This u-HIS is θ-consistent with the result in Example 4.4. Then

$$\begin{aligned}imp_{\lambda,\theta}^{(2)}(\{a_1\})=&(1-0.2222)*\frac{22}{26}+0.2222*\frac{4}{4}\approx 0.8205,\\imp_{\lambda,\theta}^{(2)}(\{a_2\})=&(1-0.2222)*\frac{0}{26}+0.2222*\frac{4}{4}=0.2222,\\imp_{\lambda,\theta}^{(2)}(\{a_3\})=&(1-0.2222)*\frac{22}{26}+0.2222*\frac{4}{4}\approx 0.8803.\end{aligned}$$

Proposition 4.8 *Let (U, C, d) be a p-HIS with $\lambda=\frac{|U^u|}{|U|}$. Then the following properties hold.*

(1) $0\leq imp_{\lambda,\theta}^{(2)}(P)\leq 1$;

(2) $imp_{\lambda,\theta}^{(2)}(C)=1$;

(3) $P\subseteq Q\subseteq C$ *implies to* $imp_{\lambda,\theta}^{(2)}(P)\leq imp_{\lambda,\theta}^{(2)}(Q)$;

(4) $imp_{\lambda,\theta}^{(2)}(P)=1\Leftrightarrow|dis_d^{l,\theta}(P)|=|dis_d^{l,\theta}(C)|, |ind_\theta^u(P)|=|ind_\theta^u(C)|$.

Proof "(1) and (2)" are obvious.

(3) Since $P \subseteq Q \subseteq C$, we have

$$|dis_d^{l,\theta}(P)| \le |dis_d^{l,\theta}(Q)|, \ \ |ind_\theta^u(Q)| \le |ind_\theta^u(P)|.$$

Then

$$\frac{|dis_d^{l,\theta}(P)|}{|dis_d^{l,\theta}(C)|} \le \frac{|dis_d^{l,\theta}(Q)|}{|dis_d^{l,\theta}(C)|}, \ \ \frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|} \le \frac{|ind_\theta^u(C)|}{|ind_\theta^u(Q)|}.$$

$$(1-\lambda)\frac{|dis_d^{l,\theta}(P)|}{|dis_d^{l,\theta}(C)|} \le (1-\lambda)\frac{|dis_d^{l,\theta}(Q)|}{|dis_d^{l,\theta}(C)|}, \ \ \lambda\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|} \le \lambda\frac{|ind_\theta^u(C)|}{|ind_\theta^u(Q)|}.$$

Hence $imp_{\lambda,\theta}^{(1)}(P) \le imp_{\lambda,\theta}^{(1)}(Q)$.

(4) }} $\Leftarrow$ ε is clear. Below, we prove }} $\Rightarrow$ ε.

Suppose $imp_{\lambda,\theta}^{(2)}(P) = 1$. Then

$$(1-\lambda)\frac{|dis_d^{l,\theta}(P)|}{|dis_d^{l,\theta}(C)|} + \lambda\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|} = 1 = (1-\lambda) + \lambda.$$

This implies that

$$(1-\lambda)(1-\frac{|dis_d^{l,\theta}(P)|}{|dis_d^{l,\theta}(C)|}) + \lambda(1-\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}) = 0.$$

Note that $1-\frac{|dis_d^{l,\theta}(P)|}{|dis_d^{l,\theta}(C)|} = \frac{|dis_d^{l,\theta}(C)|-|dis_d^{l,\theta}(P)|}{|dis_d^{l,\theta}(C)|} \ge 0$, $1-\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|} = \frac{|ind_\theta^u(P)|-|ind_\theta^u(C)|}{|ind_\theta^u(P)|} \ge 0$, Then $1-\frac{|dis_d^{l,\theta}(P)|}{|dis_d^{l,\theta}(C)|} = 0, 1-\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|} = 0$. Thus

$$|dis_d^{l,\theta}(P)| = |dis_d^{l,\theta}(C)| \ and \ |ind_\theta^u(P)| = |ind_\theta^u(C)|.$$

□

4.3 The type 3 importance of a subsystem in a p-HIS

Stipulate $0\log_2 0 = 0$.

Definition 4.9 (U, C, d) is a p-HIS with $|U^l| = n_l$. $\forall\ P \subseteq C$. $H_\theta^l(P)$ is called information entropy of P, if

$$H_\theta^l(P) = -\sum_{i=1}^{n_l} \frac{|[u_i]_P^{l,\theta}|}{n_l}\log_2\frac{|[u_i]_P^{l,\theta}|}{n_l}.$$

Proposition 4.10 *Let (U, C, d) be a p-HIS with $|U^l| = n_l$. Given $P \subseteq C$. Then*

$$0 \le H_\theta^l(P) \le n_l\log_2 n_l.$$

Moreover, if $R_P^{l,\theta} = \triangle$, *then* $H_\theta^l(P) = \log_2 n_l$; *if* $R_P^{l,\theta} = \delta$, *then* H *has a minimum.*

Proof $\forall\, i, 1 \leq |[u_i]_P^{l,\theta}| \leq n_l$, we have

$$\frac{1}{n_l} \leq \frac{|[u_i]_P^{l,\theta}|}{n_l} \leq 1,$$

$$0 \leq -\log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l} \leq \log_2 n_l.$$

Then

$$0 \leq -\frac{|[u_i]_P^{l,\theta}|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l} \leq \log_2 n_l.$$

By Definition 4.19,

$$0 \leq H_\theta^l(P) \leq n_l \log_2 n_l.$$

If $R_P^{l,\theta} = \triangle$, then $\forall\, i, |[u_i]_P^{l,\theta}| = 1$. So $H_\theta^l(P) = \log_2 n_l$.

If $R_P^{l,\theta} = \delta$, then $\forall\, i, |[u_i]_P^{l,\theta}| = n_l$. So $H_\theta^l(P) = 0$.

□

Definition 4.11 Let (U, C, d) be a p-HIS with $P \subseteq C$ and $|U^l| = n_l$. $H_\theta^l(P|d)$ is called conditional information entropy of P, if

$$H_\theta^l(P|d) = -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta} \cap D_j|}{|[u_i]_P^{l,\theta}|}.$$

Proposition 4.12 *(U, C, d) is a p-HIS with $|U^l| = n_l$. If $P \subseteq Q \subseteq C$, then*

$$H_\theta^l(Q|d) \leq H_\theta^l(P|d).$$

Proof Denote

$$U^l/R_d^{l,\theta} = \{D_1, D_2, \cdots, D_r\};$$
$$p_{ij}^{(1)} = |[u_i]_P^{l,\theta} \cap D_j|, \;\; p_{ij}^{(2)} = |[u_i]_P^{l,\theta} \cap (U^l - D_j)|;$$
$$q_{ij}^{(1)} = |R_Q^{l,\theta}(u_i) \cap D_j|, \;\; q_{ij}^{(2)} = |R_Q^{l,\theta}(u_i) \cap (U^l - D_j)|.$$

Then

$$\forall\, i,j, \;\; |[u_i]_P^{l,\theta}| = p_{ij}^{(1)} + p_{ij}^{(2)}, \;\; |R_Q^{l,\theta}(u_i)| = q_{ij}^{(1)} + q_{ij}^{(2)}.$$

Obviously, $\forall\, i$, $R_Q^{l,\theta}(u_i) \subseteq [u_i]_P^{l,\theta}$.

Then

$$\forall\, i,j, \;\; 0 \leq q_{ij}^{(1)} \leq p_{ij}^{(1)}, \;\; 0 \leq q_{ij}^{(2)} \leq p_{ij}^{(2)}.$$

$$H_\theta^l(P|d) = -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta} \cap D_j|}{|[u_i]_P^{l,\theta}|}$$

$$= -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{p_{ij}^{(1)}}{n_l} \log_2 \frac{p_{ij}^{(1)}}{p_{ij}^{(1)} + p_{ij}^{(2)}}$$

$$\triangleq -\sum_{i=1}^{n_l}\sum_{j=1}^{r} f(p_{ij}^{(1)}, p_{ij}^{(2)}).$$

$$H_\theta^l(Q|d) = -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|R_Q^{l,\theta}(u_i) \cap D_j|}{n_l} \log_2 \frac{|R_Q^{l,\theta}(u_i) \cap D_j|}{|R_Q^{l,\theta}(u_i)|}$$

$$= -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{s_{ij}^{(1)}}{n_l} \log_2 \frac{s_{ij}^{(1)}}{s_{ij}^{(1)} + s_{ij}^{(2)}}$$

$$\triangleq -\sum_{i=1}^{n_l}\sum_{j=1}^{r} f(s_{ij}^{(1)}, q_{ij}^{(2)}).$$

Put $f(x,y) = -x\log_2 \frac{x}{x+y} (x > 0, y \geq 0)$. Then $f(x, y)$ increases with respect to x and increases with respect to y, respectively.

Since $q_{ij}^{(1)} \leq p_{ij}^{(1)}, q_{ij}^{(2)} \leq p_{ij}^{(2)}$, we have

$$f(q_{ij}^{(1)}, q_{ij}^{(2)}) \leq f(p_{ij}^{(1)}, q_{ij}^{(2)}) \leq f(p_{ij}^{(1)}, p_{ij}^{(2)}).$$

Thus

$$H_\theta^l(Q|d) \leq H_\theta^l(P|d).$$

□

Definition 4.13 (U, C, d) is a p-HIS with $|U^l| = n_l$. $\forall$ $P \subseteq C$. Denote $U^l/R_d^{l,\theta} = \{D_1, D_2, \cdots, D_r\}$. $H_\theta^l(P \cup d)$ is called joint information entropy of P with d, if

$$H_\theta^l(P \cup d) = -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l}.$$

Proposition 4.14 *(U, C, d) is a p-HIS with $|U^l| = n_l$. $\forall$ $P \subseteq C$. Then*

$$H_\theta^l(P|d) = H_\theta^l(P \cup d) - H_\theta^l(P).$$

Proof

$$U^l/R_d^{l,\theta} = \{D_1, D_2, \cdots, D_r\}.$$

Then $\{D_1, D_2 \cdots, D_r\}$ is a partition of U^l. $\forall$ i,

$$\sum_{j=1}^{r} |[u_i]_P^{l,\theta} \cap D_j| = |[u_i]_P^{l,\theta}|.$$

$$\begin{aligned}
H_\theta^l(P|d) =& -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta} \cap D_j|}{|[u_i]_P^{l,\theta}|} \\
=& -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} (\log_2 \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \\
& - \log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l}) \\
=& -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \\
& + \sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l} \\
=& H_\theta^l(P \cup d) + \sum_{i=1}^{n_l} \frac{|[u_i]_P^{l,\theta}|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l} \\
=& H_\theta^l(P \cup d) - H_\theta^l(P).
\end{aligned}$$

□

Proposition 4.15 *Let (U, C, d) be a p-HIS with $|U^l| = n_l$. $\forall\ P \subseteq C$. Then $H_\theta^l(P|d) \geq 0$.*

$$U^l/R_d^{l,\theta} = \{D_1, D_2, \cdots, D_r\}.$$

Proof

We have

$$H_\theta^l(P) = -\sum_{i=1}^{n_l} \frac{|[u_i]_P^{l,\theta}|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l}.$$

$\{D_1, D_2 \cdots, D_r\}$ is a partition of U^l. $\forall\ i$,

$$\sum_{j=1}^{r} |[u_i]_P^{l,\theta} \cap D_j| = |[u_i]_P^{l,\theta}|.$$

Then

$$\begin{aligned}
H_\theta^l(P) =& -\sum_{i=1}^{n_l} \frac{\sum_{j=1}^{r} |[u_i]_P^{l,\theta} \cap D_j|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l}. \\
=& -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l}.
\end{aligned}$$

By Definition 4.13,

$$H_\theta^l(P \cup d) = -\sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l}.$$

$\forall\, i, j$,

$$log_2 \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \leq \ log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l}.$$

Then

$$H_\theta^l(P) \leq H_\theta^l(P \cup d).$$

By Proposition 4.14,

$$H_\theta^l(P|d) = H_\theta^l(P \cup d) - H_\theta^l(P).$$

Hence $H_\theta^l(P|d) \geq 0$.

□

Definition 4.16 Let (U, C, d) be a p-HIS with $\lambda = \frac{|U^u|}{|U|}$ and $P \subseteq C$. Then the type 3 importance of the subsystem (U, P, d) is defined as

$$imp_{\lambda,\theta}^{(3)}(P) = (1-\lambda)\frac{H_\theta^l(C|d)}{H_\theta^l(P|d)} + \lambda\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}.$$

In the above definition, $\frac{H_\theta^l(C|d)}{H_\theta^l(P|d)}$ and $\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}$ can be viewed as the importance of (U^l, P, d) and (U^u, P, d), respectively. λ means the missing rate of labels, which is processed as a weight. We define the sum of the importance of (U^l, P, d) and (U^u, P, d) with the missing rate of labels as the type 3 importance of (U, P, d).

Example 4.17 (Continue with Example 4.4) We have

$$\begin{aligned} H_\theta^l(\{a_1\}|d) = -(&\tfrac{2}{7} * \log_2 \tfrac{2}{5} + \tfrac{2}{7} * \log_2 \tfrac{2}{5} + \tfrac{2}{7} * \log_2 \tfrac{2}{5} + \tfrac{0}{7} * \log_2 \tfrac{0}{2} + \tfrac{0}{7} * \log_2 \tfrac{0}{2} + \tfrac{2}{7} * \log_2 \tfrac{2}{5} + \tfrac{2}{7} * \log_2 \tfrac{2}{5} \\ &+ \tfrac{3}{7} * \log_2 \tfrac{3}{5} + \tfrac{3}{7} * \log_2 \tfrac{3}{5} + \tfrac{3}{7} * \log_2 \tfrac{3}{5} + \tfrac{0}{7} * \log_2 \tfrac{0}{2} + \tfrac{0}{7} * \log_2 \tfrac{0}{2} + \tfrac{3}{7} * \log_2 \tfrac{3}{5} + \tfrac{3}{7} * \log_2 \tfrac{3}{5} \\ &+ \tfrac{0}{7} * \log_2 \tfrac{0}{5} + \tfrac{0}{7} * \log_2 \tfrac{0}{5} + \tfrac{0}{7} * \log_2 \tfrac{0}{5} + \tfrac{2}{7} * \log_2 \tfrac{2}{2} + \tfrac{2}{7} * \log_2 \tfrac{2}{2} + \tfrac{0}{7} * \log_2 \tfrac{0}{5} + \tfrac{0}{7} * \log_2 \tfrac{0}{5}) \approx 3.4677 \end{aligned}.$$

Similarly, $H_\theta^l(\{a_2\}|d) \approx 10.8966$, $H_\theta^l(\{a_3\}|d) \approx 3.7664$, $H_\theta^l(C|d) \approx 1.93$.
Then

$$\begin{aligned} imp_{\lambda,\theta}^{(3)}(\{a_1\}) =& (1-0.2222) * \frac{1.93}{3.4677} + 0.2222 * \frac{4}{4} \approx 0.6551, \\ imp_{\lambda,\theta}^{(3)}(\{a_2\}) =& (1-0.2222) * \frac{1.93}{10.8966} + 0.2222 * \frac{4}{4} \approx 0.3600, \\ imp_{\lambda,\theta}^{(3)}(\{a_3\}) =& (1-0.2222) * \frac{1.93}{3.7664} + 0.2222 * \frac{4}{4} \approx 0.6208. \end{aligned}$$

Proposition 4.18 *Let (U, C, d) be a p-HIS with $|U^l| = n_l$. Then the following properties hold.*

(1) $0 \leq imp_{\lambda,\theta}^{(3)}(P) \leq 1$;

(2) $imp^{(3)}_{\lambda,\theta}(C) = 1$;

(3) $P \subseteq Q \subseteq C$ *implies to* $imp^{(3)}_{\lambda,\theta}(Q) \leq imp^{(3)}_{\lambda,\theta}(P)$;

(4) $imp^{(3)}_{\lambda,\theta}(P) = 1 \Leftrightarrow H^l_\theta(P|d) = H^l_\theta(C|d), |ind^u_\theta(P)| = |ind^u_\theta(C)|$.

Proof "(1) and (2)" are obvious.

(3) Since $P \subseteq Q \subseteq C$, we have

$$H^l_\theta(Q|d) \leq H^l_\theta(P|d), \;\; |ind^u_\theta(Q)| \leq |ind^u_\theta(P)|.$$

Then

$$\frac{H^l_\theta(Q|d)}{H^l_\theta(C|d)} \leq \frac{H^l_\theta(P|d)}{H^l_\theta(C|d)}, \;\; \frac{|ind^u_\theta(Q)|}{|ind^u_\theta(C)|} \leq \frac{|ind^u_\theta(P)|}{|ind^u_\theta(C)|}.$$

Thus

$$(1-\lambda)\frac{H^l_\theta(Q|d)}{H^l_\theta(C|d)} \leq (1-\lambda)\frac{H^l_\theta(P|d)}{H^l_\theta(C|d)}, \;\; \lambda\frac{|ind^u_\theta(Q)|}{|ind^u_\theta(C)|} \leq \lambda\frac{|ind^u_\theta(P)|}{|ind^u_\theta(C)|}.$$

Hence $imp^{(1)}_{\lambda,\theta}(Q) \leq imp^{(1)}_{\lambda,\theta}(P)$.

(4) }} $\Leftarrow$ ε is clear. Below, we prove }} $\Rightarrow$ ε.

Suppose $imp^{(3)}_{\lambda,\theta}(P) = 1$. Then

$$(1-\lambda)\frac{H^l_\theta(C|d)}{H^l_\theta(P|d)} + \lambda\frac{|ind^u_\theta(C)|}{|ind^u_\theta(P)|} = 1 = (1-\lambda) + \lambda.$$

This implies that

$$(1-\lambda)(1 - \frac{H^l_\theta(C|d)}{H^l_\theta(P|d)}) + \lambda(1 - \frac{|ind^u_\theta(C)|}{|ind^u_\theta(P)|}) = 0.$$

Note that $1 - \frac{H^l_\theta(C|d)}{H^l_\theta(P|d)} = \frac{H^l_\theta(P|d) - H^l_\theta(C|d)}{H^l_\theta(P|d)} \geq 0$, $\;1 - \frac{|ind^u_\theta(C)|}{|ind^u_\theta(P)|} = \frac{|ind^u_\theta(P)| - |ind^u_\theta(C)|}{|ind^u_\theta(P)|} \geq 0$. Then $1 - \frac{H^l_\theta(C|d)}{H^l_\theta(P|d)} = 0, 1 - \frac{|ind^u_\theta(C)|}{|ind^u_\theta(P)|} = 0$. Thus

$$H^l_\theta(C|d) = H^l_\theta(P|d), \;\; |ind^u_\theta(P)| = |ind^u_\theta(C)|.$$

□

4.4 The type 4 importance of a subsystem in a p-HIS

Definition 4.19 (U, C, d) is a p-HIS with $|U^l| = n_l$. $\forall\, P \subseteq C$. Then information amount of P is defined as

$$E_\theta^l(P) = \sum_{i=1}^{n_l} \frac{|[u_i]_P^{l,\theta}|}{n_l} \frac{|U^l - [u_i]_P^{l,\theta}|}{n_l}.$$

Obviously, $E_\theta^l(P) = \sum_{i=1}^{n_l} \frac{|[u_i]_P^{l,\theta}|}{n_l}(1 - \frac{|[u_i]_P^{l,\theta}|}{n_l})$.

Proposition 4.20 *Let (U, C, d) be a p-HIS with $|U^l| = n_l$. Given $P \subseteq C$. Then*

$$0 \leq E_\theta^l(P) \leq 1 - \frac{1}{n_l}.$$

Moreover, if $R_P^{l,\theta} = \triangle$, then E_θ^l achieves the maximum value $1 - \frac{1}{n_l}$; if $R_P^{l,\theta} = \delta$, then E_θ^l achieves the minimum value 0.

Proof Since $\forall\, i, 1 \leq |[u_i]_P^{l,\theta}| \leq n_l$, we have

$$\frac{1}{n_l} \leq \frac{|[u_i]_P^{l,\theta}|}{n_l} \leq 1,$$

$$0 \leq 1 - \frac{|[u_i]_P^{l,\theta}|}{n_l} \leq 1 - \frac{1}{n_l}.$$

Thus

$$0 \leq E_\theta^l(P) \leq 1 - \frac{1}{n_l}.$$

If $R_P^{l,\theta} = \triangle$, then $\forall\, i, |[u_i]_P^{l,\theta}| = 1$. So $E_\theta^l(P) = 1 - \frac{1}{n_l}$.

If $R_P^{l,\theta} = \delta$, then $\forall\, i, |[u_i]_P^{l,\theta}| = n_l$. So $E_\theta^l(P) = 0$.

□

Definition 4.21 Let (U, C, d) be a p-HIS with $|U^l| = n_l$. Given $P \subseteq C$. Denote

$$U^l/R_d^{l,\theta} = \{D_1, D_2, \cdots, D_r\}.$$

Put

$$E_\theta^l(P|d) = \sum_{i=1}^{n_l} \sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \frac{|[u_i]_P^{l,\theta} - D_j|}{n_l}.$$

Then $E_\theta^l(P|d)$ are called conditional information amount of P to d in U^l.

Proposition 4.22 *(U, C, d) is a p-HIS with $|U^l| = n_l$. If $P \subseteq Q \subseteq C$, then*

$$E_\theta^l(Q|d) \leq E_\theta^l(P|d).$$

Proof Denote

$$U^l/R_d^{l,\theta} = \{D_1, D_2, \cdots, D_r\}.$$

Suppose $P \subseteq Q \subseteq C$. Then $\forall\ i$, $R_Q^{l,\theta}(u_i) \subseteq [u_i]_P^{l,\theta}$. So

$$\forall\ i,j,\ \ R_Q^{l,\theta}(u_i) \cap D_j \subseteq [u_i]_P^{l,\theta} \cap D_j,\ R_Q^{l,\theta}(u_i) - D_j \subseteq [u_i]_P^{l,\theta} - D_j.$$

This implies that

$$\forall\ i,j,\ \ |R_Q^{l,\theta}(u_i) \cap D_j| \leq |[u_i]_P^{l,\theta} \cap D_j|,\ |R_Q^{l,\theta}(u_i) - D_j| \leq |[u_i]_P^{l,\theta} - D_j|.$$

Thus

$$E_\theta^l(Q|d) \leq E_\theta^l(P|d).$$

□

Definition 4.23 (U, C, d) is a p-HIS with $|U^l| = n_l$. $\forall\ P \subseteq C$. Denote

$$U^l/R_d^{l,\theta} = \{D_1, D_2, \cdots, D_r\}.$$

Then joint information amount of P and d is defined as

$$E_\theta^l(P \cup d) = \sum_{i=1}^{n_l} \sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l}\left(1 - \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l}\right).$$

Proposition 4.24 *(U, C, d) is a p-HIS with $|U^l| = n_l$. $\forall\ P \subseteq C$. Then*

$$E_\theta^l(P|d) = E_\theta^l(P \cup d) - E_\theta^l(P).$$

Proof $\{D_1, D_2 \cdots, D_r\}$ is a partition of U^l. $\forall\ i$,

$$\sum_{j=1}^{r} |[u_i]_P^{l,\theta} \cap D_j| = |[u_i]_P^{l,\theta}|.$$

$$\begin{aligned}
E_\theta^l(P|d) &= \sum_{i=1}^{n_l} \sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \frac{|[u_i]_P^{l,\theta} - [u_i]_P^{l,\theta} \cap D_j|}{n_l} \\
&= \sum_{i=1}^{n_l} \sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} \frac{|[u_i]_P^{l,\theta}| - |[u_i]_P^{l,\theta} \cap D_j|}{n_l} \\
&= \sum_{i=1}^{n_l} \sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} (\frac{|[u_i]_P^{l,\theta}|}{n_l} - \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l}) \\
&= \sum_{i=1}^{n_l} \sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} ((1 - \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l}) - (1 - \frac{|[u_i]_P^{l,\theta}|}{n_l})) \\
&= \sum_{i=1}^{n_l} \sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} (1 - \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l}) - \\
&\quad \sum_{i=1}^{n_l} \sum_{j=1}^{r} \frac{|[u_i]_P^{l,\theta} \cap D_j|}{n_l} (1 - \frac{|[u_i]_P^{l,\theta}|}{n_l}) \\
&= E_\theta^l(P \cup d) - \sum_{i=1}^{n_l} \frac{|[u_i]_P^{l,\theta}|}{n_l} (1 - \frac{|[u_i]_P^{l,\theta}|}{n_l}) \\
&= E_\theta^l(P \cup d) - E_\theta^l(P).
\end{aligned}$$

□

Proposition 4.25 *Let* (U, C, d) *be a p-HIS with* $|U^l| = n_l$. *Given* $P \subseteq C$. *Then* $E_\theta^l(P|d) \geq 0$.

$$U^l / R_d^{l,\theta} = \{D_1, D_2, \cdots, D_r\}.$$

Proof

By Definition 4.19, we have

$$E_\theta^l(P) = \sum_{i=1}^{n_l} \frac{|[u_i]_P^{l,\theta}|}{n_l} \log_2 \frac{|[u_i]_P^{l,\theta}|}{n_l}.$$

$\{D_1, D_2 \cdots, D_r\}$ is a partition of U^l. $\forall\, i$,

$$\sum_{j=1}^{r} |[u_i]_P^{l,\theta} \cap D_j| = |[u_i]_P^{l,\theta}|.$$

Then

$$E^l_\theta(P) = \sum_{i=1}^{n_l} \frac{\sum_{j=1}^{r} |[u_i]^{l,\theta}_P \cap D_j|}{n_l}(1 - \frac{|[u_i]^{l,\theta}_P|}{n_l}).$$
$$= \sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]^{l,\theta}_P \cap D_j|}{n_l}(1 - \frac{|[u_i]^{l,\theta}_P|}{n_l}).$$

By Definition 4.23, we have

$$E^l_\theta(P \cup d) = \sum_{i=1}^{n_l}\sum_{j=1}^{r} \frac{|[u_i]^{l,\theta}_P \cap D_j|}{n_l}(1 - \frac{|[u_i]^{l,\theta}_P \cap D_j|}{n_l}).$$

$\forall\ i, j$,

$$\frac{|[u_i]^{l,\theta}_P \cap D_j|}{n_l} \le \frac{|[u_i]^{l,\theta}_P|}{n_l}.$$

Then

$$E^l_\theta(P) \le E^l_\theta(P \cup d).$$

By Proposition 4.24,

$$E^l_\theta(P|d) = E^l_\theta(P \cup d) - E^l_\theta(P).$$

Hence $E^l_\theta(P|d) \ge 0$.

□

In the following definition, $\frac{E^l_\theta(C|d)}{E^l_\theta(P|d)}$ and $\frac{|ind^u_\theta(C)|}{|ind^u_\theta(P)|}$ can be viewed as the importance of (U^l, P, d) and (U^u, P, d), respectively. λ means the missing rate of labels, which is processed as a weight. We define the sum of the importance of (U^l, P, d) and (U^u, P, d) with the missing rate of labels as the type 4 importance of (U, P, d).

Definition 4.26 Let (U, C, d) be a p-HIS with $|U^l| = n_l$. $\forall\ P \subseteq C$. Then the type 4 importance of the subsystem (U, P, d) is defined as

$$imp^{(4)}_{\lambda,\theta}(P) = (1-\lambda)\frac{E^l_\theta(C|d)}{E^l_\theta(P|d)} + \lambda\frac{|ind^u_\theta(C)|}{|ind^u_\theta(P)|}.$$

Example 4.27 (Continue with Example 4.4) We have

$E^l_\theta(\{a_1\}|d) = \frac{2}{7} * \log_2\frac{3}{7} + \frac{2}{7} * \log_2\frac{3}{7} + \frac{2}{7} * \log_2\frac{3}{7} + \frac{0}{7} * \log_2\frac{2}{7} + \frac{0}{7} * \log_2\frac{2}{7} + \frac{2}{7} * \log_2\frac{3}{7} + \frac{2}{7} * \log_2\frac{3}{7} + \frac{3}{7} * \log_2\frac{2}{7} + \frac{3}{7} * \log_2\frac{2}{7} + \frac{3}{7} * \log_2\frac{2}{7} + \frac{0}{7} * \log_2\frac{2}{7} + \frac{0}{7} * \log_2\frac{2}{7} + \frac{3}{7} * \log_2\frac{2}{7} + \frac{3}{7} * \log_2\frac{2}{7} + \frac{0}{7} * \log_2\frac{5}{7} + \frac{0}{7} * \log_2\frac{5}{7} + \frac{0}{7} * \log_2\frac{5}{7} + \frac{2}{7} * \log_2\frac{0}{7} + \frac{2}{7} * \log_2\frac{0}{7} + \frac{0}{7} * \log_2\frac{5}{7} + \frac{0}{7} * \log_2\frac{5}{7} \approx 1.2245$

.

Similarly, $E^l_\theta(\{a_2\}|d) \approx 4.5714$, $E^l_\theta(\{a_3\}|d) \approx 0.9796$, $E^l_\theta(C|d) \approx 0.4898$.
Then

$$imp^{(4)}_{\lambda,\theta}(\{a_1\}) =(1 - 0.2222) * \frac{0.4898}{1.2245} + 0.2222 * \frac{4}{4} \approx 0.5333,$$
$$imp^{(4)}_{\lambda,\theta}(\{a_2\}) =(1 - 0.2222) * \frac{0.4898}{4.5714} + 0.2222 * \frac{4}{4} \approx 0.3056,$$
$$imp^{(4)}_{\lambda,\theta}(\{a_3\}) =(1 - 0.2222) * \frac{0.4898}{0.9796} + 0.2222 * \frac{4}{4} \approx 0.6111.$$

Proposition 4.28 *Let (U, C, d) be a p-HIS with $|U^l| = n_l$. Then the following properties hold.*

(1) $0 \leq imp^{(4)}_{\lambda,\theta}(P) \leq 1$;

(2) $imp^{(4)}_{\lambda,\theta}(C) = 1$;

(3) $P \subseteq Q \subseteq C$ *implies to* $imp^{(4)}_{\lambda,\theta}(Q) \leq imp^{(4)}_{\lambda,\theta}(P)$;

(4) $imp^{(4)}_{\lambda,\theta}(P) = 1 \Leftrightarrow E^l_\theta(P|d) = E^l_\theta(C|d), |ind^u_\theta(P)| = |ind^u_\theta(C)|$.

Proof “(1) and (2)" are obvious.

(3) Since $P \subseteq Q \subseteq C$, we have

$$E^l_\theta(Q|d) \leq E^l_\theta(P|d), \;\; |ind^u_\theta(Q)| \leq |ind^u_\theta(P)|.$$

Then

$$\frac{E^l_\theta(Q|d)}{E^l_\theta(C|d)} \leq \frac{E^l_\theta(P|d)}{E^l_\theta(C|d)}, \;\; \frac{|ind^u_\theta(Q)|}{|ind^u_\theta(C)|} \leq \frac{|ind^u_\theta(P)|}{|ind^u_\theta(C)|}.$$

Thus

$$(1-\lambda)\frac{E^l_\theta(Q|d)}{E^l_\theta(C|d)} \leq (1-\lambda)\frac{E^l_\theta(P|d)}{E^l_\theta(C|d)}, \;\; \lambda\frac{|ind^u_\theta(Q)|}{|ind^u_\theta(C)|} \leq \lambda\frac{|ind^u_\theta(P)|}{|ind^u_\theta(C)|}.$$

Hence $imp^{(1)}_{\lambda,\theta}(Q) \leq imp^{(1)}_{\lambda,\theta}(P)$.

(4) }} $\Leftarrow$ ε is clear. Below, we prove }} $\Rightarrow$ ε.

Suppose $imp^{(4)}_{\lambda,\theta}(P) = 1$. Then

$$(1-\lambda)\frac{E^l_\theta(C|d)}{E^l_\theta(P|d)} + \lambda\frac{|ind^u_\theta(C)|}{|ind^u_\theta(P)|} = 1 = (1-\lambda) + \lambda.$$

This implies that

$$(1-\lambda)(1 - \frac{E^l_\theta(C|d)}{E^l_\theta(P|d)}) + \lambda(1 - \frac{|ind^u_\theta(C)|}{|ind^u_\theta(P)|}) = 0.$$

Note that $1-\frac{E_\theta^l(C|d)}{E_\theta^l(P|d)}=\frac{E_\theta^l(P|d)-E_\theta^l(C|d)}{E_\theta^l(P|d)}\geq 0$, $1-\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}=\frac{|ind_\theta^u(P)|-|ind_\theta^u(C)|}{|ind_\theta^u(P)|}\geq 0$. Then $1-\frac{E_\theta^l(C|d)}{E_\theta^l(P|d)}=0, 1-\frac{|ind_\theta^u(C)|}{|ind_\theta^u(P)|}=0$. Thus

$$E_\theta^l(C|d)=E_\theta^l(P|d),\ |ind_\theta^u(P)|=|ind_\theta^u(C)|.$$

□

5 Statistical analysis

In this section, we make statistical analysis on the four UMs. Experimental analysis is carried out to test the effect of measurements.

5.1 Measurement analysis

Twelve hybrid datasets (see Table 3) are selected from UCI (machine learning repository) database for experiments. Since the labels of these datasets are not missing, for the convenience of the experiment, we make the labels randomly missing by 50%. As a result, the following experiments were conducted with $\lambda = 0.5$. The experimental setup utilized a Lenovo computer equipped with an Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz and 16GB of memory. The code is programmed with MATLAB 2019 software.

In order to test the performance of these four measures, all datasets need to be preprocessed. For any dataset, let $P_i=\{p_1,\cdots,p_i\}(i=1,\cdots,n), n=|C|$, four measure sets are as follows:

$$X_j(dataset)=\{imp_{\lambda,\theta}^{(j)}(P_1),\cdots,imp_{\lambda,\theta}^{(j)}(P_n))\}(j=1,2,3,4).$$

Table 3 The datasets excerpted from the UCI

ID	Data set	Abbreviation	Instance	Attribute	Class	Type
1	Abalone	Aba	4177	8	29	Mix
2	Anneal	Ann	798	38	6	Mix
3	Arrhythmia	Arr	452	279	16	Mix
4	AustralianCreditApproval	ACA	690	14	2	Mix
5	Bands	Ban	539	39	2	Mix
6	SteelPlatesFaults	SPF	1940	27	7	Mix
7	Ionosphere	Ion	350	32	2	Mix
8	DiabeticRetinopathyDebrecen	DRD	1151	19	2	Mix
9	QSARbiodegradation	QSA	1055	41	2	Mix
11	Spambase	Spa	4601	57	2	Mix
10	Thyroid_sick	Thy	2800	29	2	Mix
12	EEGEyeState	EES	14980	14	2	Mix

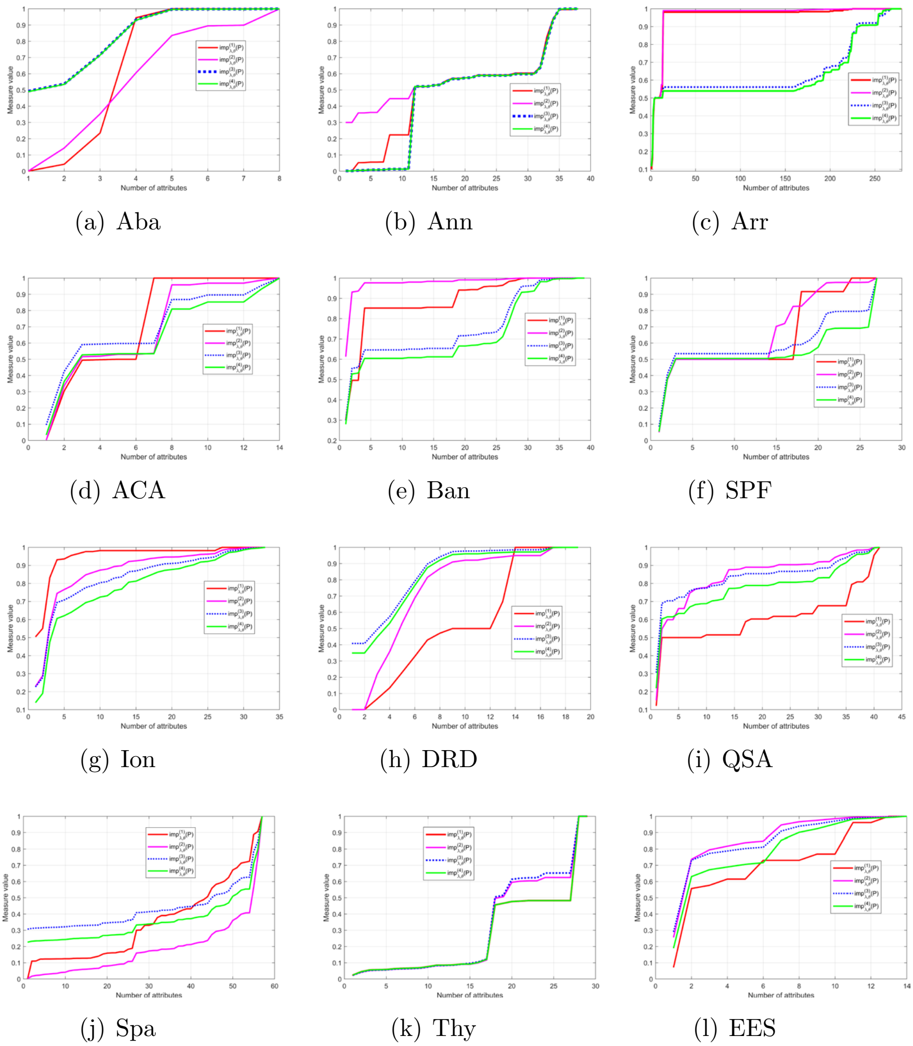


Fig. 2 Four values of measures on each of twelve datasets

This formula means that we need to observe the change of these measurements as the number of subsets increases. Using this formula, the measurement change curves for 12 data sets are plotted as shown in Fig. 2. The x-axis indicates the cardinality of subset, and the y-axis represents the value of $X_j(dataset)$. It can be seen that when the number of attributes gradually increases, the four UMs gradually rise to 1. It can be concluded that the certainty of A p-HIS increase with the growth of the attribute subset. We also find that the area enclosed by the blue curve with the x-axis on datasets Ann, Ban and Ion is slightly smaller than that of other curves. However, in other datasets, the blue line is above other color curves. This indicates that $imp^{(3)}_{\lambda,\theta}(P)$ sometimes have the better effect. Consequently, four UMs all can be used to measure the uncertainty of partially labeled hybrid data.

5.2 Dispersion analysis

Many literatures employ the standard deviation and discrete coefficient along with the mean to report statistical analysis results, and discrete analysis of hybrid data can use them. Next, we analyze the discretization of 12 datasets by using four UMs.

Suppose that $X_j(dataset) = \{x_1, x_2 \cdots x_n\}$ is a dataset, let $\bar{x}, \sigma(X), CV(X)$ be the arithmetic average value, standard deviation and coefficient of variation for X respectively. Their formulas are as follows: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\sigma(X) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})}$, $CV(X) = \frac{\sigma(X)}{\bar{x}}$. For the sake of simplicity, the coefficient of variation is called CV-value.

The smaller the coefficient of variation is, the more reliable the information system for measuring uncertainty is and the smaller the risk is. The CV-value of each dataset is calculated and displayed in Table 4. Table 4 reports four UMS measure the discretization of different datasets. As we can see that the minimum average value is 0.3517, and the maximum value is 0.4608. In order to more intuitively compare the advantages and disadvantages of these four UMS, we rank them according to the results in Table 4. Since the smaller the CV-value, the smaller the risk, the data in Table 4 are sorted from small to large, and the results are shown in Table 4. It is easy to find that the lowest and highest average ranking are $imp^{(3)}_{\lambda,\theta}(P)$ (1.6667) and $imp^{(1)}_{\lambda,\theta}(P)$ (3.0833) respectively. Consequently, it can be concluded that $imp^{(3)}_{\lambda,\theta}(P)$ is the most stable and the other UMS take more risk.

5.3 Statistical analysis of four UMs

Next, we conduct a statistical analysis on these four UMs. The Friedman test is a nonparametric test used to determine if there are significant differences among the four UMs by ranking them. The data in Table 5 are input into the software SPSS, and the calculated results are shown in Table 6. If the significance level is taken as $\alpha = 0.05$, then $p = 0.0307 < \alpha$ in Table 6, indicating that there are significant differences among the four

Table 4 CV-values of four measures

Data set	$imp^{(1)}_{\lambda,\theta}(P)$	$imp^{(2)}_{\lambda,\theta}(P)$	$imp^{(3)}_{\lambda,\theta}(P)$	$imp^{(4)}_{\lambda,\theta}(P)$
Aba	0.7182	0.6458	0.2624	0.2642
Ann	0.5896	0.3292	0.7023	0.7023
Arr	0.2450	0.1233	0.1216	0.2616
ACA	0.4631	0.4511	0.3562	0.4036
Ban	0.1690	0.0632	0.2231	0.2485
SPF	0.3950	0.3738	0.2816	0.2827
Ion	0.1192	0.2092	0.2181	0.2637
DRD	0.6742	0.4631	0.2555	0.2832
QSA	0.2414	0.1881	0.1408	0.1791
Spa	0.6945	0.9711	0.3235	0.4142
Thy	1.0050	0.9968	0.9914	1.0052
EES	0.3374	0.2278	0.2204	0.2781
Average	0.4608	0.4201	0.3517	0.3822

Table 5 rank of CV-values with four measures

Data set	$imp^{(1)}_{\lambda,\theta}(P)$	$imp^{(2)}_{\lambda,\theta}(P)$	$imp^{(3)}_{\lambda,\theta}(P)$	$imp^{(4)}_{\lambda,\theta}(P)$
Aba	4	3	1	2
Ann	2	1	3	4
Arr	2	1	3	4
ACA	4	3	1	2
Ban	2	1	3	4
SPF	4	3	1	2
Ion	1	2	3	4
DRD	4	3	1	2
QSA	4	3	1	2
Spa	3	4	1	2
Thy	3	2	1	4
EES	4	2	1	3
Average	3.0833	2.3333	1.6667	2.9167

Table 6 Friedman test for four UMs

Source	SS	df	MS	χ^2	p
Groups	14.8333	3	4.9444	8.9	0.0307
Error	45.1667	33	1.3687		
Total	60	47			

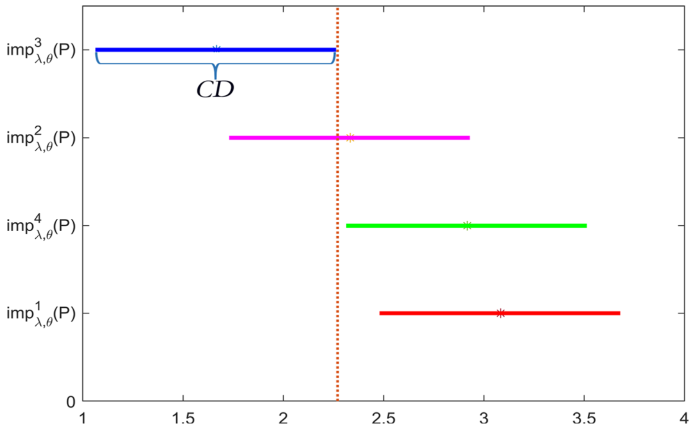


Fig. 3 Nemenyi test

UMs. Then, multiple comparisons are implemented by Nemenyi test. The critical range of the difference between the average ordinal values calculated is $CD = q_\alpha\sqrt{\frac{k(k+1)}{6N}}$.

It is known that $k = 4$ represents four UMs and $N = 12$ represents 12 datasets. Let $\alpha = 0.1$, then look up table $q_\alpha = 2.291$ and calculate $CD = 1.2075$. The test results are plotted in Fig. 3 for visual comparison. It can be seen from Fig. 3 that:

(1) The blue line is closer to the y-axis than the other lines;
(2) In the horizontal direction, the blue line does not overlap the green and red lines;
(3) In the horizontal direction, the blue line partially overlaps the magenta line.

It can be concluded that:

(a) $imp^{(3)}_{\lambda,\theta}(P)$ is statistically better than $imp^{(1)}_{\lambda,\theta}(P)$ and $imp^{(4)}_{\lambda,\theta}(P)$;
(b) There is no significant difference between $imp^{(3)}_{\lambda,\theta}(P)$, $imp^{(2)}_{\lambda,\theta}(P)$;
(c) There is no significant difference between $imp^{(1)}_{\lambda,\theta}(P)$, $imp^{(2)}_{\lambda,\theta}(P)$ and $imp^{(4)}_{\lambda,\theta}(P)$.

Therefore, $imp^{(3)}_{\lambda,\theta}(P)$ performs better than the other three UMs and next we will focus on it.

6 Semi-supervised attribute reduction for hybrid data

In this section, we study Semi-supervised attribute reduction for hybrid data.

6.1 Semi-supervised attribute reduction in a p-HIS

Definition 6.1 Let (U, C, d) be a p-HIS with $\lambda = \frac{|U^u|}{|U|}$ and $P \subseteq C$. Then P is called a coordinate subset of C with respect to d in a p-HIS (U, C, d), if $POS^{l,\theta}_P(d) = POS^{l,\theta}_C(d)$, $ind^u_\theta(P) = ind^u_\theta(C)$.

All coordinate subsets of C with respect to d is denoted by $co^p_{\lambda,\theta}(C)$.

Definition 6.2 Let (U, C, d) be a p-HIS with $\lambda = \frac{|U^u|}{|U|}$ and $P \subseteq C$. Then P is called a reduct of C with d, if $P \in co^p_{\lambda,\theta}(C)$ and for each $a \in P$, $P - \{a\} \notin co^p_{\lambda,\theta}(C)$.

All reducts of C with d is denoted by $red^p_{\lambda,\theta}(C)$.

Theorem 6.3 *(U, C, d) is a p-HIS with $\lambda = \frac{|U^u|}{|U|}$. The following conditions are equivalent:*

(1) $P \in co^p_{\lambda,\theta}(C)$;

(2) $imp^{(1)}_{\lambda,\theta}(P) = 1$;

(3) $imp^{(2)}_{\lambda,\theta}(P) = 1$;

Proof (1) $\Rightarrow$ (2). $P \in co^p_{\lambda,\theta}(C)$. Then $POS^{l,\theta}_P(d) = POS^{l,\theta}_C(d)$, $ind^u_\theta(P) = ind^u_\theta(C)$. Thus $\Gamma^{l,\theta}_P(d) = \Gamma^{l,\theta}_C(d), |ind^u_\theta(P)| = |ind^u_\theta(C)|$.

By Proposition 4.5, $imp^{(1)}_{\lambda,\theta}(P) = 1$.

(2) $\Rightarrow$ (3). Suppose $imp^{(1)}_{\lambda,\theta}(P) = 1$. Then by Proposition 4.5,

$$\Gamma^{l,\theta}_P(d) = \Gamma^{l,\theta}_C(d),\ |ind^u_\theta(P)| = |ind^u_\theta(C)|.$$

By Proposition 2.6, $dis_d^{l,\theta}(P) = dis_d^{l,\theta}(C)$. Then $|dis_d^{l,\theta}(P)| = |dis_d^{l,\theta}(C)|$.

By Proposition 4.8, $imp_{\lambda,\theta}^{(2)}(P) = 1$.

(3) $\Rightarrow$ (1). Suppose $imp_{\lambda,\theta}^{(2)}(P) = 1$. Then by Proposition 4.8,

$$|dis_d^{l,\theta}(P)| = |dis_d^{l,\theta}(C)|,\ |ind_\theta^u(P)| = |ind_\theta^u(C)|.$$

Note that $dis_d^{l,\theta}(P) \subseteq dis_d^{l,\theta}(C)$ and $ind_\theta^u(P) \supseteq ind_\theta^u(C)$. Then $dis_d^{l,\theta}(P) = dis_d^{l,\theta}(C)$ and $ind_\theta^u(P) = ind_\theta^u(C)$.

By Proposition 2.6, $POS_P^{l,\theta}(d) = POS_C^{l,\theta}(d)$.

Thus $P \in co_{\lambda,\theta}^p(C)$. $\square$

Corollary 6.4 *(U, C, d) is a p-HIS with $\lambda = \frac{|U^u|}{|U|}$ and $P \subseteq C$. The following conditions are equivalent*:

(1) $P \in red_{\lambda,\theta}^p(C)$;

(2) $imp_{\lambda,\theta}^{(1)}(P) = 1$ *and* $\forall\, a \in P$, $imp_{\lambda,\theta}^{(1)}(P - \{a\}) < 1$;

(3) $imp_{\lambda,\theta}^{(2)}(P) = 1$ *and* $\forall\, a \in P$, $imp_{\lambda,\theta}^{(2)}(P - \{a\}) < 1$;

Proof It follows from Theorem 6.3. $\square$

Lemma 6.5 *(U, C, d) is a p-HIS and P is a subset of C. If $R_P^{l,\theta} \subseteq R_d^{l,\theta}$, then $\forall\, u \in U^l$ and j,*

$$[u]_P^{l,\theta} \cap D_j = \begin{cases} [u]_P^{l,\theta} & u \in D_j \\ \emptyset & u \notin D_j \end{cases}.$$

Proof (1) $u \in D_j$, $D_j = [u]_d^{l,\theta}$. Since $R_P^{l,\theta} \subseteq R_d^{l,\theta}$, therefore $[u]_P^{l,\theta} \subseteq [u]_d^{l,\theta}$. Thus $[u]_P^{l,\theta} \cap D_j = [u]_P^{l,\theta}$.

(2) $u \notin D_j$, then $[u]_d^{l,\theta} \cap D_j = \emptyset$. Since $R_P^{l,\theta} \subseteq R_d^{l,\theta}$, therefore $[u]_P^{l,\theta} \subseteq [u]_d^{l,\theta}$. Thus $[u]_P^{l,\theta} \cap D_j = \emptyset$. $\square$

Lemma 6.6 *(U, C, d) is a p-HIS. $\forall\, P \subseteq C$. If $R_P^{l,\theta} \subseteq R_d^{l,\theta}$, $\forall\, u \in U^l$,*

$$\sum_{j=1}^{r} \frac{|[u]_P^{l,\theta} \cap D_j|}{n}\left(1 - \frac{|[u]_P^{l,\theta} \cap D_j|}{n}\right) = \frac{|[u]_P^{l,\theta}|}{n}\left(1 - \frac{|[u]_P^{l,\theta}|}{n}\right),$$

$$\sum_{j=1}^{r} \frac{|[u]_P^{l,\theta} \cap D_j|}{n} \log_2 \frac{|[u]_P^{l,\theta} \cap D_j|}{n} = \frac{|[u]_P^{l,\theta}|}{n} \log_2 \frac{|[u]_P^{l,\theta}|}{n}.$$

Proof Since $R_P^{l,\theta} \subseteq R_d^{l,\theta}$, by Lemma 6.5, we have

$$[u]_P^{l,\theta} \cap D_j = \begin{cases} [u]_P^{l,\theta} & j = j^* \\ \emptyset & j \neq j^* \end{cases}.$$

Thus

$$\sum_{j=1}^{r}\frac{|[u]_P^{l,\theta}\cap D_j|}{n}(1-\frac{|[u]_P^{l,\theta}\cap D_j|}{n})=\frac{|[u]_P^{l,\theta}|}{n}(1-\frac{|[u]_P^{l,\theta}|}{n}),$$
$$\sum_{j=1}^{r}\frac{|[u]_P^{l,\theta}\cap D_j|}{n}\log_2\frac{|[u]_P^{l,\theta}\cap D_j|}{n}=\frac{|[u]_P^{l,\theta}|}{n}\log_2\frac{|[u]_P^{l,\theta}|}{n}.$$

□

Proposition 6.7 *(U, C, d) is a p-HIS and P is a subset of C. the following conditions are equivalent*:

(1) $R_P^{l,\theta}\subseteq R_d^{l,\theta}$;

(2) $H_\theta^l(P\cup d)=H_\theta^l(P)$;

(3) $H_\theta^l(P|d)=0$.

Proof "(1) ⇒ (2)" is proved by Lemma 6.6.

"(2) ⇒ (3)" follows from Proposition 4.14.

(3) ⇒ (1). Suppose $H_\theta^l(P|d)=0$. Then

$$\sum_{i=1}^{n}\sum_{j=1}^{r}\frac{|[u_i]_P^{l,\theta}\cap D_j|}{n}\log_2\frac{|[u_i]_P^{l,\theta}|}{|[u_i]_P^{l,\theta}\cap D_j|}=0.$$

Suppose $R_P^{l,\theta}\nsubseteq R_d^{l,\theta}$. Then $\exists\, i^*\in\{1,\cdots,n\}, [u_{i^*}]_P^{l,\theta}\nsubseteq[u_{i^*}]_d^{l,\theta}$. Denote

$$[u_{i^*}]_d^{l,\theta}=D_{j^*}\ (j^*\in\{1,\cdots,r\}).$$

We have

$$|[u_{i^*}]_P^{l,\theta}|>|[u_i]_P^{l,\theta}\cap D_{j^*}|.$$

This follows that

$$\frac{|R_P^{l,\theta}(u_{i^*})\cap D_{j^*}|}{n}\log_2\frac{|[u_{i^*}]_P^{l,\theta}|}{|R_P^{l,\theta}(u_{i^*})\cap D_{j^*}|}>0.$$

Note that

$$\forall\, i,j,\quad |[u_i]_P^{l,\theta}|\geq|[u_i]_P^{l,\theta}\cap D_j|.$$

Then

$$\forall\, i,j,\quad \frac{|[u_i]_P^{l,\theta}\cap D_j|}{n}\log_2\frac{|[u_i]_P^{l,\theta}|}{|[u_i]_P^{l,\theta}\cap D_j|}\geq 0.$$

So

$$\sum_{i=1}^{n}\sum_{j=1}^{r}\frac{|[u_i]_P^{l,\theta}\cap D_j|}{n}\log_2\frac{|[u_i]_P^{l,\theta}|}{|[u_i]_P^{l,\theta}\cap D_j|}>0.$$

This is a contradiction.

Thus $R_P^{l,\theta}\subseteq R_d^{l,\theta}$.

□

Proposition 6.8 *(U, C, d) is a p-HIS. Given $P\subseteq C$. If $R_P^{l,\theta}\subseteq R_d^{l,\theta}$, then $E_\theta^l(P|d)=0$.*

Proof It is directly proved by Proposition 4.24 and Lemma 6.6.

□

Corollary 6.9 *(U, C, d) is a p-HIS and θ-consistent. Then $H_\theta^l(C|d)=E_\theta^l(C|d)=0$.*

Proof It follows from and Propositions 3.16, 6.7 and 6.8. □

Theorem 6.10 *(U, C, d) is a θ-consistent p-HIS with $\lambda=\frac{|U^u|}{|U|}$. The following conditions are equivalent*:

(1) $P\in co_{\lambda,\theta}^p(C)$;

(2) $imp_{\lambda,\theta}^{(3)}(P)=1$.

Proof (1) ⇒ (2). $P\in co_{\lambda,\theta}^p(C)$, we have $POS_P^{l,\theta}(d)=POS_C^{l,\theta}(d)$, $ind_\theta^u(P)=ind_\theta^u(C)$. Thus $\Gamma_P^{l,\theta}(d)=\Gamma_C^{l,\theta}(d), |ind_\theta^u(P)|=|ind_\theta^u(C)|$.

By Proposition 4.2, we have

$$\sum_{j=1}^{r}(|\underline{R_C^{l,\theta}}(D_j)|-|\underline{R_P^{l,\theta}}(D_j)|)=0.$$

Obviously, $\forall\, j$, $\underline{R_C^{l,\theta}}(D_j)\supseteq\underline{R_P^{l,\theta}}(D_j)$, which implies

$$|\underline{R_C^{l,\theta}}(D_j)|-|\underline{R_P^{l,\theta}}(D_j)|\geq 0.$$

Thus,

$$|\underline{R_C^{l,\theta}}(D_j)|-|\underline{R_P^{l,\theta}}(D_j)|=0.$$

Therefore $\forall\, j$,

$$\underline{R_C^{l,\theta}}(D_j)=\underline{R_P^{l,\theta}}(D_j).$$

This means that

$$[u]_C^{l,\theta}\subseteq D_j\ \Leftrightarrow\ [u]_P^{l,\theta}\subseteq D_j.$$

(U, C, d) is θ-consistent, from Proposition 3.16, we have $R_C^{l,\theta}\subseteq R_d^{l,\theta}$.

Then $\forall\, u\in U^l$,

$$[u]_C^{l,\theta} \subseteq [u]_d^{l,\theta}.$$

Let $[u]_d^{l,\theta} = D^u$. $\forall\, u \in U^l$, $[u]_P^{l,\theta} \subseteq D^u = [u]_d^{l,\theta}$, which implies $R_P^{l,\theta} \subseteq R_d^{l,\theta}$.

By Proposition 6.7, $H_\theta^l(P|d) = 0$.

(U, C, d) is θ-consistent, from Corollary 6.9, we have $H_\theta^l(C|d) = 0$. Then $H_\theta^l(P|d) = H_\theta^l(C|d)$.

By Proposition 4.18, $imp_{\lambda,\theta}^{(3)}(P) = 1$.

(2) $\Rightarrow$ (1). Suppose $imp_{\lambda,\theta}^{(3)}(P) = 1$. Then by Proposition 4.18, $H_\theta^l(P|d) = H_\theta^l(C|d)$, $|ind_\theta^u(P)| = |ind_\theta^u(C)|$.

Since $ind_\theta^u(P) \supseteq ind_\theta^u(C)$, we have $ind_\theta^u(P) = ind_\theta^u(C)$.

(U, C, d) is θ-consistent, from Corollary 6.9, we have $H_\theta^l(C|d) = 0$. Then $H_\theta^l(P|d) = 0$. From Proposition 6.7, $R_P^{l,\theta} \subseteq R_d^{l,\theta}$.

Suppose that $\exists\, j^*$,

$$\underline{R_C^{l,\theta}}(D_{j^*}) \nsubseteq \underline{R_P^{l,\theta}}(D_{j^*}).$$

Then $\underline{R_C^{l,\theta}}(D_{j^*}) - \underline{R_P^{l,\theta}}(D_{j^*}) \neq \emptyset$.

$$u^* \in \underline{R_C^{l,\theta}}(D_{j^*}) - \underline{R_P^{l,\theta}}(D_{j^*}).$$

Thus

$$u^* \in \underline{R_C^{l,\theta}}(D_{j^*}),\ u^* \notin \underline{R_P^{l,\theta}}(D_{j^*}).$$

$u^* \in \underline{R_C^{l,\theta}}(D_{j^*})$ implies that $u^* \in R_C^{l,\theta}(u^*) \subseteq D_{j^*}$. Therefore $D_{j^*} = [u^*]_d^l$. $u^* \notin \underline{R_P^{l,\theta}}(D_{j^*})$ implies that $[u^*]_P^l \nsubseteq D_{j^*}$. Thus $[u^*]_P^l \nsubseteq [u^*]_d^l$. So $R_P^{l,\theta} \nsubseteq R_d^{l,\theta}$. This is a contradiction.

Hence $\forall\, j$,

$$\underline{R_C^{l,\theta}}(D_j) \subseteq \underline{R_P^{l,\theta}}(D_j).$$

Obviously, $\forall\, j$, $\underline{R_C^{l,\theta}}(D_j) \supseteq \underline{R_P^{l,\theta}}(D_j)$.

Then $\underline{R_C^{l,\theta}}(D_j) = \underline{R_P^{l,\theta}}(D_j)$. Thus

$$POS_C^{l,\theta}(d) = \bigcup_{j=1}^{r} \underline{R_C^{l,\theta}}(D_j) = \bigcup_{j=1}^{r} \underline{R_P^{l,\theta}}(D_j) = POS_P^{l,\theta}(d).$$

Hence $P \in co_{\lambda,\theta}^p(C)$.

□

Corollary 6.11 *(U, C, d) is a θ-consistent p-HIS with $\lambda = \frac{|U^u|}{|U|}$. The following conditions are equivalent*:

(1) $P \in red_{\lambda,\theta}^p(C)$;

(2) $imp_{\lambda,\theta}^{(3)}(P) = 1$ *and* $\forall\, a \in P$, $imp_{\lambda,\theta}^{(3)}(P - \{a\}) < 1$.

Proof It is straightly proved by Theorem 6.10. □

Theorem 6.12 *(U, C, d) is a θ-consistent p-HIS with $\lambda = \frac{|U^u|}{|U|}$. If $P \in co^p_{\lambda,\theta}(C)$, then $imp^{(4)}_{\lambda,\theta}(P) = 1$.*

Proof Suppose $P \in co^p_{\lambda,\theta}(C)$. Then $POS^{l,\theta}_P(d) = POS^{l,\theta}_C(d)$, $ind^u_\theta(P) = ind^u_\theta(C)$. Thus $\Gamma^{l,\theta}_P(d) = \Gamma^{l,\theta}_C(d), |ind^u_\theta(P)| = |ind^u_\theta(C)|$.

It is easily proved that $R^{l,\theta}_P \subseteq R^{l,\theta}_d$. $E^l_\theta(P|d) = 0$ by Proposition 6.7.

(U, C, d) is θ-consistent, from Corollary 6.9, we have $E^l_\theta(C|d) = 0$. Therefore $E^l_\theta(P|d) = E^l_\theta(C|d)$.

By Proposition 4.28, $imp^{(4)}_{\lambda,\theta}(P) = 1$.

□

Corollary 6.13 *(U, C, d) is a θ-consistent p-HIS with $\lambda = \frac{|U^u|}{|U|}$. If $P \in red^p_{\lambda,\theta}(C)$, then $imp^{(4)}_{\lambda,\theta}(P) = 1$ and $\forall\, a \in P$, $imp^{(4)}_{\lambda,\theta}(P - \{a\}) < 1$.*

Proof It follows from Theorem 6.12. □

6.2 Semi-supervised attribute reduction algorithms in a p-HIS

The coefficient of variation for the four UMS are discussed in the above chapters. As we know that $imp^{(3)}_{\lambda,\theta}(P)$ work best, so it is selected to compile the attribute reduction algorithm.

Algorithm 1: Attribute reduction algorithm for a p-HIS based on the type 3 importance (T3I).

Input: A p-HIS (U, C, d), and $\lambda, \theta \in [0, 1]$.
Output: One attribute reduct P.

$P \leftarrow \emptyset$.
Calculate $imp^{(3)}_{\lambda,\theta}(P)$.
while $imp^{(3)}_{\lambda,\theta}(P) \leq 1 - 5\%$ **do**
  for *each* $a \in C - P$ **do**
    Calculate $imp^{(3)}_{\lambda,\theta}(P \cup \{a\})$.
  end
  Find $a^* \in C - P$ such that
    $imp^{(3)}_{\lambda,\theta}(P \cup \{a^*\}) = max\{imp^{(3)}_{\lambda,\theta}(P \cup \{a\}) : a \in C - P\}$.
  $P \leftarrow P \cup \{a^*\}$.
end
Return P.

The key of T3I is to calculate the importance of the attribute subsets. By traversing the importance of all attributes, the attributes with the greatest importance are found, and then they are put into the reduction set in turn until the stopping condition is satisfied. The specific process of this algorithm is shown in Fig. 4.

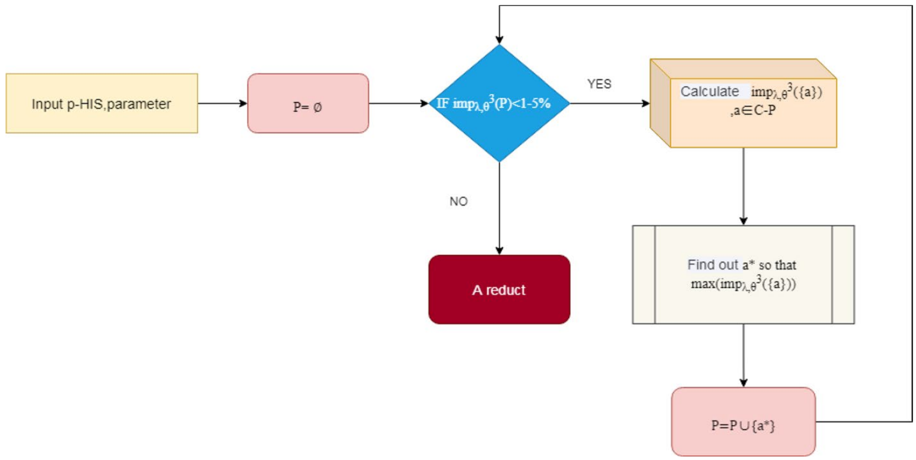


Fig. 4 T3I flow

We consider the algorithm in terms of time and space complexity. The complexity of step 1 is $O(|C||U|^2)$, and the complexity of steps 3-9 is $O(|C||U|^2 + (|C| - 1)|U|^2 + \cdots, +|P||U|^2)$. So the total complexity of the algorithm is $O((\frac{|C|^2}{2} + \frac{|C|}{2} - \frac{|P|^2}{2} + \frac{|P|}{2})|U|^2)$. When $P = \emptyset$, it means that the algorithm has not completed the reduction. At this time, the maximum complexity of the algorithm is $O(|C|^2|U|^2)$. The spatial complexity of the algorithm is $O(|C||U|^2)$.

7 Experimental analysis

7.1 Numerical experiments

There are two important parameters λ and θ in T3I algorithm. To facilitate the experiments, let $\lambda = 0.5$, which means that 50% of the labels are randomly missing. θ is a parameter that controls the distance between information values. In order to analyze the influence of θ on the reduction algorithm, two classifiers are used to analyze the accuracy of the reduced attribute subset. Two classification algorithms were used: one is gradient Boosting Decision Tree (BDT), the other is K-NearestNeighbor(KNN, K=3). An average performance measure was computed based on a 10-fold cross validation result repeated 10 times. We have plotted the relationship between θ and classification accuracy, as shown in Fig. 5. The experiments show that the θ can affect the classification accuracy of the model. Moreover, when the maximum classification accuracy is obtained, different datasets correspond to different parameters θ. This can provide a basis for us to find out the maximum classification accuracy.

In order to further study the performance of the algorithm, four algorithms are selected from other literatures for comparison to illustrate the effectiveness of the algorithm. They are MEHAR (Hu et al. 2021), SHIVAM (Shreevastava et al. 2019), SFSE (Wan et al. 2021), and RnR-SSFSM (Solorio-Fernndez et al. 2020). Algorithm SHIVAM has no abbreviation, so it is replaced by the author's name. MEHAR, SHIVAM and SFSE are

Table 7 Number of selected attributes

Data set	Raw data	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	8	5	7	7	8	5
Ann	38	11	8	15	11	4
Arr	279	13	14	25	20	16
ACA	14	8	9	7	7	5
Ban	39	10	13	11	5	10
SPF	27	11	18	5	5	4
Ion	33	7	11	11	17	15
DRD	19	8	9	14	9	8
QSA	41	15	14	15	13	16
Spa	57	7	12	15	14	11
Thy	29	1	14	13	8	11
EES	14	5	7	14	7	4
Average	49.83	8.42	11.33	12.67	10.33	9.08

Table 8 Comparison of classification accuracies of reduction set with BDT

Data set	Raw data	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	0.2607	0.2624	0.2643	0.2473	0.259	**0.2677**
Ann	0.9060	**0.9787**	0.7907	0.8421	0.985	0.8860
Arr	0.7080	0.5907	0.5088	0.5465	0.6018	**0.6261**
ACA	0.8696	0.7826	0.7667	0.8565	0.8101	**0.8609**
Ban	0.7829	0.7477	0.7514	0.6939	0.6252	**0.7829**
SPF	0.6909	0.6625	**0.6842**	0.5440	0.5641	0.6146
Ion	0.7436	0.7493	0.8803	0.8063	**0.9202**	0.8405
DRD	0.6620	0.6699	0.6525	0.6594	0.6690	**0.6846**
QSA	0.8607	0.8502	0.8512	0.8133	0.8436	**0.8673**
Spa	0.9400	0.8642	0.8366	0.8153	0.8583	**0.8883**
Thy	0.9846	0.9389	0.9379	0.9386	0.9789	**0.9811**
EES	0.7749	0.6743	0.6314	**0.7799**	0.6611	0.7187
Average	0.7653	0.7310	0.7130	0.7119	0.7314	**0.7516**

algorithms for partially labeled hybrid data. RnR-SSFSM is for supervised hybrid data. The five algorithms are restored by programming, and the same 12 datasets are used for numerical experiments.

Table 7 shows the attribute subset of each algorithm after reduction. Black bold type indicates optimal value, and the last line is the average number of reduction. It can be seen

Table 9 Comparison of classification accuracies of reduction set with KNN

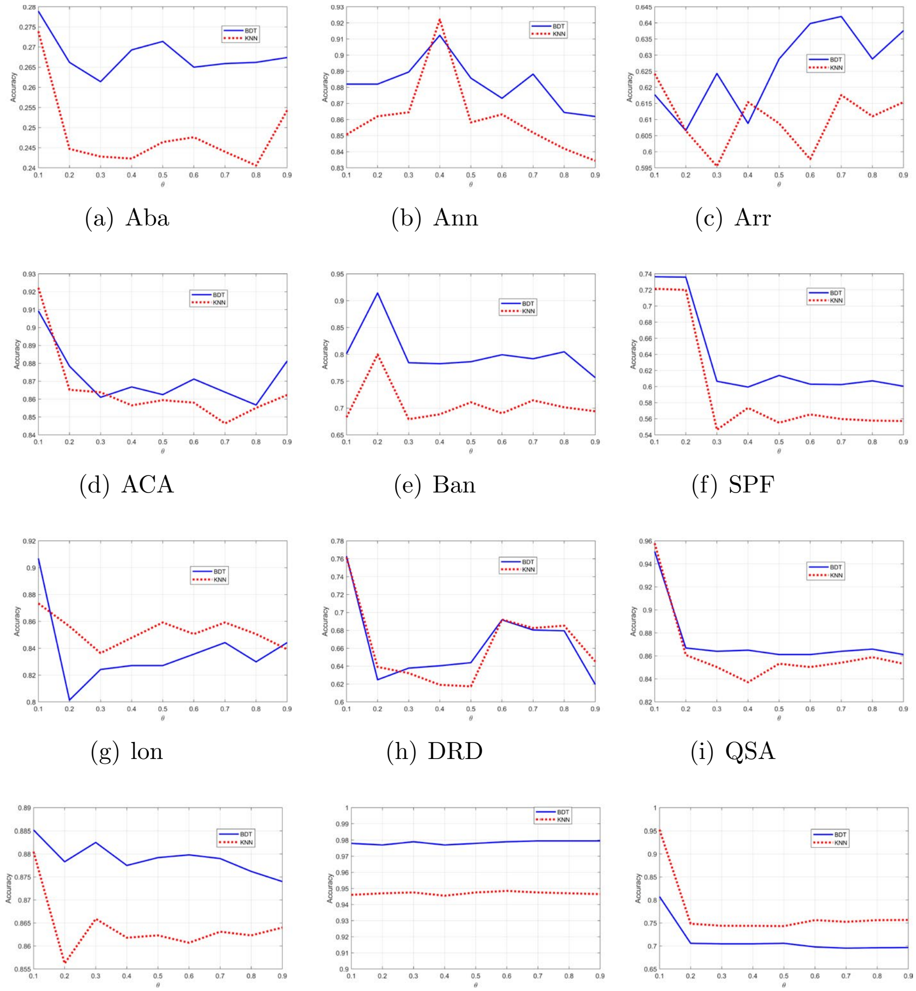
Data set	Raw data	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	0.2332	0.2205	0.2305	0.2375	0.2301	**0.2437**
Ann	0.9311	**0.9173**	0.7782	0.8321	0.8772	0.8496
Arr	0.5841	0.5863	0.4757	0.5487	0.5686	**0.5929**
ACA	0.8565	0.7246	0.7826	0.8522	0.7565	**0.8536**
Ban	0.7328	0.6883	0.6939	0.6698	0.5362	**0.7050**
SPF	0.7367	0.6728	**0.7223**	0.5703	0.5374	0.5600
Ion	0.8234	0.8120	0.8177	0.8234	0.8348	**0.8575**
DRD	0.6247	0.6646	0.6290	0.6551	0.6560	**0.6890**
QSA	0.8550	0.8512	0.8389	0.7981	0.8408	**0.8521**
Spa	0.9081	0.8618	0.7403	0.5008	0.8474	**0.8650**
Thy	0.9386	0.9371	0.9386	0.9389	0.9393	**0.9482**
EES	0.8579	0.6546	0.8070	**0.8703**	0.8150	0.7510
Average	0.7568	0.7159	0.7046	0.6914	0.7033	**0.7306**

that MEHAR performs best, which is 8.42. The proposed algorithm T3I takes the second place, which is 9.08. This shows that T3I reduction efficiency is above the average level.

Then, we analyze the classification accuracy of the reduction set. The comparison of classification accuracy calculated is shown in Tables 8 and 9. It can be seen that TI3 has reached its optimal value in many datasets, with a total of 17 tests ranking first in Tables 8 and 9. This is significantly more than the 2 first-place rankings achieved by MEHAR, the 2 first-place rankings achieved by SHIVAM, the 2 first-place rankings achieved by SFSE and the 1 first-place rankings achieved by RnR-SSFSM. In addition, T3I also get the best average classification accuracy in these two tables, with 0.7516 and 0.7306, respectively. So it can be concluded that T3I performs well in most cases.

To analyze these algorithms, it is not sufficient to compare classification accuracy. We also need to calculate the True Position Rate (*TPR*)and False Positive Rate (*FPR*) to evaluate the effect of data prediction. According to the real and predicted values it can be divided into: (1) True Positive (*TP*); (2) False Positive (*FP*); (3) True Negative (*TN*); (4) False Negative (*FN*). Then True Position Rate (*TPR*) and False Positive Rate (*FPR*) can be calculated respectively as follows: $TPR = TP/TP + FN$ and $FPR = FP/FP + TN$. The precision (*P*) and the recall (*R*) which is also call sensitivity can be expressed as $P = TP/(TP + FP)$ and $R = TP/(TP + FN)$. Geometric mean (G-mean) can be showed as $GM = \sqrt{(P * R)}$.

Therefore, we plotted the Receiver Operating Characteristic Curve (ROC) Narkhede (2018) and calculated the Area Under Curve(AUC) area. See Figs. 6 and 7 for details. The x-axis is *FPR*, and the y-axis is *TPR*. The blue line in the figures represents the T3I algorithm. From the Fig. 6, the blue line is not drawn on the outermost edge of the subgraphs Ann, SPF, lon and EES. It can be also seen from Fig. 7 that the blue line is not located on the outermost layer of subgraphs Ban, SPF, Ion and EES. This indicates that T3I does not have any advantages in these datasets, but it performs very well in the remaining datasets. From these two figures, we find that all curves of subgraph Aba are concave, and the area enclosed by the x-axis is less than 0.5, which indicates that the classification characteristics of dataset AD are not good, and key attributes are missing during data collection. Tables 10



(a) Aba (b) Ann (c) Arr

(d) ACA (e) Ban (f) SPF

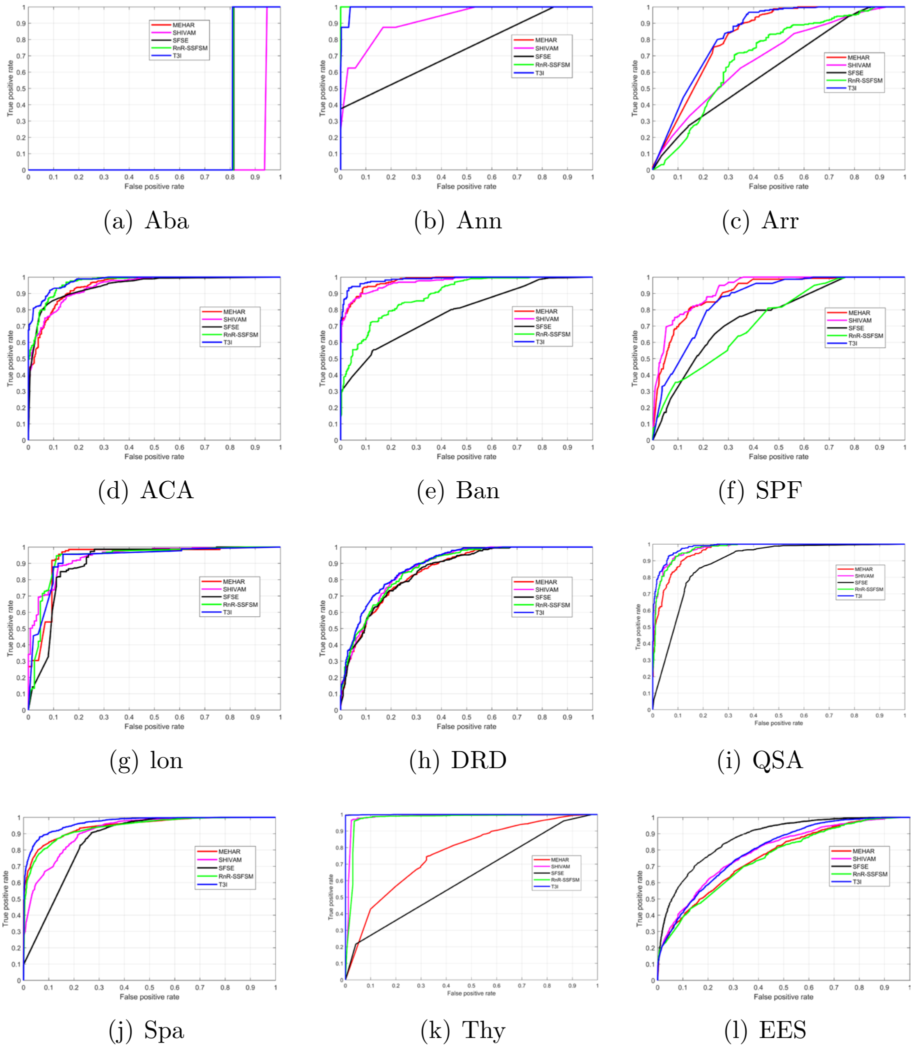
(g) lon (h) DRD (i) QSA

(j) Spa (k) Thy (l) EES

Fig. 5 Effect of parameter θ on classification accuracy

and 11 record the area enclosed by each curve in Figs. 6 and 7. The larger the area is, the better the effect of the algorithm classification is. As can be seen from the Tables, the average AUC of T3I is the largest, 0.8688 and 0.8402 respectively. Thus, the attribute subset of T3I after reduction has a nice classification effect.

Furthermore, the G-mean is employed as an additional indicator to evaluate the effectiveness of classification. Prior to that, precision and recall metrics are calculated for the reduction set of these algorithms using two classifiers(See Tables 12 and 13). From



(a) Aba (b) Ann (c) Arr

(d) ACA (e) Ban (f) SPF

(g) Ion (h) DRD (i) QSA

(j) Spa (k) Thy (l) EES

Fig. 6 ROC curve comparison of five algorithms on classifier BDT

Tables 12 and 13, it is apparent that while the computed *P* and *R* values from the reduction set of T3I may not always be optimal for each dataset individually, the average performance is the most favorable. However, these tables alone do not provide sufficient insight into the overall quality of these algorithms. For this reason, G-mean is calculated as shown in Tables 14 and 15 according to the data in Tables 12 and 13. Then, we can easily determine which algorithm performs better. It is evident that T3I has achieved a total of 18 top rankings in Tables 14 and 15, surpassing MEHAR's 2 first places, SHIVAM's 1 first place,

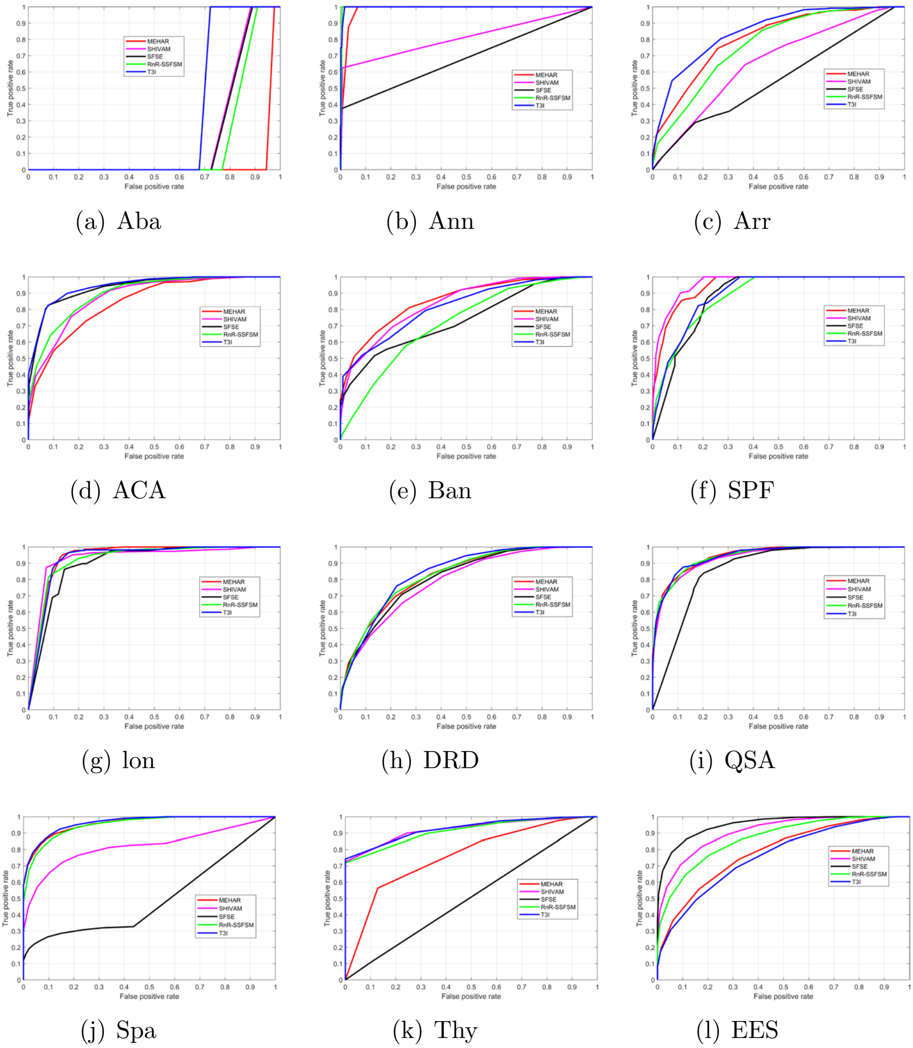


Fig. 7 ROC curve comparison of five algorithms on classifier KNN

SFSE's 2 first places, and RnR-SSFSM's 1 first place by a significant margin. It is not difficult to intuitively analyze from the above results that T3I performs better. However, further statistical analysis is needed to determine whether our judgment is accurate.

7.2 Friedman and Nemenyi test of algorithms

Next, we analyze the differences between these five algorithms. Friedman test was used for significance analysis. The data in Tables 8 and 9 and Tables 14 and 15 are sorted in descending order, and the results are shown in Tables 16, 17, 18 and 19. Once the data is input into SPSS software for statistical analysis, then the calculated results are shown in Tables 20 and 21. It can be seen that the values of p in Tables 20 and 21 are

Table 10 AUC with classifier BDT

Data set	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	0.1936	0.0644	0.1918	0.1942	**0.2014**
Ann	**0.9995**	0.9192	0.7358	0.9976	0.9952
Arr	0.8202	0.6899	0.6267	0.7124	**0.8353**
ACA	0.9457	0.9421	0.9456	0.9675	**0.9779**
Ban	0.9776	0.9698	0.784	0.8893	**0.9868**
SPF	0.9046	**0.9270**	0.7588	0.7352	0.8615
Ion	0.9313	**0.9436**	0.9005	0.9343	0.9328
DRD	0.8563	0.8750	0.8556	0.8733	**0.8879**
QSA	0.9602	0.9745	0.8890	0.9728	**0.9834**
Spa	0.9464	0.9164	0.8609	0.9399	**0.9688**
Thy	0.7655	0.9762	0.6219	0.9718	**0.9996**
EES	0.7587	0.7892	**0.8748**	0.7493	0.7955
Average	0.8383	0.8323	0.7538	0.8281	**0.8688**

Table 11 AUC with classifier KNN

Data set	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	0.0450	0.1968	0.1938	0.1685	**0.2110**
Ann	0.9825	0.8104	0.6856	0.9962	**0.9968**
Arr	0.8065	0.6582	0.5548	0.7697	**0.8509**
ACA	0.8431	0.8796	0.9355	0.8963	**0.9432**
Ban	**0.8547**	0.8370	0.7364	0.7161	0.8173
SPF	**0.9469**	0.9614	0.8809	0.8852	0.8962
Ion	**0.9366**	0.9264	0.9043	0.9277	0.9289
DRD	0.8236	0.7889	0.8096	0.8270	**0.8366**
QSA	0.9459	0.9367	0.8649	0.9442	**0.9505**
Spa	0.9582	0.8199	0.4990	0.9524	**0.9642**
Thy	0.7606	0.9235	0.5089	0.9127	**0.9264**
EES	0.7844	0.9094	**0.9518**	0.8679	0.7601
Average	0.8073	0.8040	0.7105	0.8220	**0.8402**

0.0041, 0.0251, 0.0103 and 0.0116, respectively, all of which are less than the significance $\alpha = 0.05$. Consequently, it can be concluded that the five algorithms have significant differences.

Afterwards, the Nemenyi test was conducted as a post-hoc test with the purpose of further distinguishing the advantages and disadvantages of each algorithm. The calculation formula of the critical range CD is consistent with the description in section 5.3. It is known that $k = 5$ represents 5 algorithms, and $N = 12$ represents 12 datasets. Let $\alpha = 0.1$, $q_\alpha = 2.459$ was obtained by referencing the table, and $CD = 1.5873$ was calculated. Figure 8 was plotted based on the calculated CD value. By Fig. 8a, we have:

(a) The classification accuracy of T3I is significantly better than RnR-SSFSM, and SFSE;

Table 12 Precision and Recall with classifier BDT

Data set	MEHAR		SHIVAM		SFSE		RnR-SSFSM		T3I	
	P	R	P	R	P	R	P	R	P	R
P &R										
Aba	0.2246	0.2382	0.2130	0.3517	0.2033	0.3849	0.1979	0.3927	**0.2393**	**0.3628**
Ann	0.9885	0.9868	0.7955	0.9918	0.8292	**0.9984**	**0.9918**	0.9918	0.9245	0.9671
Arr	0.6706	0.9224	0.5569	**0.9592**	0.5545	0.9551	0.6385	0.8939	**0.6837**	0.9265
ACA	0.7667	0.8668	0.7644	0.8642	0.8684	0.8616	0.7737	**0.8747**	**0.8790**	0.8538
Ban	**0.7824**	0.5859	0.7964	0.5859	0.7006	0.4846	0.5542	0.4053	0.7705	**0.6211**
SPF	0.6732	0.7263	0.8359	**0.8579**	0.5568	0.7737	0.4029	0.5895	**0.8952**	0.8082
Ion	**0.9185**	0.9511	0.8841	0.9156	0.8938	0.8978	0.9174	0.9378	0.7782	**0.9822**
DRD	0.6124	0.7519	0.6254	0.7111	**0.6334**	0.7296	0.6293	0.7389	0.6203	**0.7685**
QSA	0.7557	0.7472	**0.8018**	0.7612	0.7026	0.7500	0.7988	0.7360	0.7982	**0.7669**
Spa	0.9138	0.8522	0.8294	0.9157	0.7899	0.9466	0.891	0.8734	**0.8925**	**0.9469**
Thy	0.9398	0.9985	**0.9908**	0.9863	0.9392	**0.9996**	0.9889	0.9855	**0.9908**	0.9871
EES	0.6413	0.8793	0.7253	0.768	**0.7663**	0.8518	0.6414	**0.8803**	0.6891	0.8484
Average	0.7406	0.7922	0.7349	0.8057	0.7032	0.8028	0.7022	0.7750	**0.7634**	**0.8200**

Table 13 Precision and Recall with classifier KNN

Data set	MEHAR		SHIVAM		SFSE		RnR-SSFSM		T3I	
	P	R	P	R	P	R	P	R	P	R
P &R										
Aba	0.2095	0.2902	0.2295	0.3139	**0.2437**	0.3060	0.2267	0.3028	0.2348	**0.3344**
Ann	0.9119	0.9539	0.7992	0.9490	0.8194	**1**	0.9392	0.9655	**0.9700**	0.9572
Arr	0.6057	0.9592	0.5668	0.8653	0.5561	**0.9918**	**0.6087**	0.9714	0.6020	0.9878
ACA	0.7130	0.8564	0.7689	0.8512	0.8469	0.8668	0.7361	**0.8956**	**0.8794**	0.8564
Ban	**0.6879**	0.4758	0.6650	**0.5771**	0.6726	0.4978	0.5125	0.3612	0.6437	0.4934
SPF	0.6116	0.7211	**0.7914**	0.8440	0.5278	0.6000	0.5105	0.5105	0.7650	**0.8737**
Ion	**0.8467**	**0.9822**	0.7739	0.9733	0.8036	**0.9822**	0.8066	**0.9822**	0.8333	0.9778
DRD	0.6280	0.7222	0.5821	0.6370	0.6175	0.6907	0.6279	0.7000	**0.6506**	**0.7241**
QSA	**0.7837**	0.7837	0.7730	0.7556	0.6652	**0.8371**	0.7600	0.7472	0.7769	0.8118
Spa	**0.9087**	0.8680	0.8343	0.7242	0.7730	0.2407	0.8897	0.8619	0.8958	**0.8999**
Thy	0.9402	0.9981	0.9387	0.9966	0.9389	**1**	0.9414	0.9958	**0.9421**	0.9970
EES	0.6468	0.7205	0.7923	0.8228	**0.8718**	**0.8840**	0.7514	0.7832	0.6671	0.7497
Average	0.7078	0.7776	0.7096	0.7758	0.69471	0.7414	0.6926	0.7564	**0.7384**	**0.8053**

(b) MEHAR, SHIVAM, RnR-SSFSM and SFSE have no significant statistical difference in classification accuracy.
(c) In terms of classification accuracy, there is no obvious difference among T3I, MEHAR and SHIVAM.

Table 14 G-mean with classifier BDT

Data set	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	0.2313	0.2737	0.2797	0.2788	**0.2946**
Ann	0.9877	0.8882	0.9099	**0.9918**	0.9456
Arr	0.7865	0.7309	0.7277	0.7555	**0.7959**
ACA	0.8152	0.8128	0.8650	0.8227	**0.8663**
Ban	0.6771	0.6831	0.5827	0.4739	**0.6918**
SPF	0.6992	0.8468	0.6564	0.4873	**0.8506**
Ion	**0.9347**	0.8997	0.8958	0.9275	0.8743
DRD	0.6786	0.6669	0.6798	0.6819	**0.6904**
QSA	0.7514	0.7812	0.7259	0.7668	**0.7824**
Spa	0.8825	0.8715	0.8647	0.8822	**0.9193**
Thy	0.9687	0.9885	0.9689	0.9872	**0.9889**
EES	0.7509	0.7463	**0.8079**	0.7514	0.7646
Average	0.7636	0.7658	0.7470	0.7339	**0.7887**

Table 15 G-mean with classifier KNN

Data set	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	0.2466	0.2684	0.2731	0.2620	**0.2802**
Ann	0.9327	0.8709	0.9052	0.9523	**0.9636**
Arr	0.7622	0.7003	0.7427	0.7690	**0.7711**
ACA	0.7814	0.8090	0.8568	0.8119	**0.8678**
Ban	0.5721	**0.6195**	0.5786	0.4302	0.5636
SPF	0.6641	0.8173	0.5627	0.5105	**0.8175**
Ion	**0.9119**	0.8679	0.8884	0.8901	0.9027
DRD	0.6735	0.6089	0.6531	0.6630	**0.6864**
QSA	0.7837	0.7643	0.7462	0.7536	**0.7942**
Spa	0.8881	0.7773	0.4313	0.8757	**0.8978**
Thy	0.9687	0.9672	0.9690	0.9682	**0.9692**
EES	0.6827	0.8074	**0.8779**	0.7671	0.7072
Average	0.7390	0.7399	0.7071	0.7211	**0.7684**

By Fig. 8b, we have:

(a) The classification accuracy of T3I is significantly better than MEHAR, SFSE and RnR-SSFSM;
(b) SHIVAM, MEHAR, SFSE and RnR-SSFSM have no significant statistical differences in classification accuracy;
(c) In terms of classification accuracy, there is no obvious difference between T3I and SHIVAM.

By Fig. 9a, we have:

(a) The G-mean of T3I is significantly better than MEHAR, SHIVAM and SFSE;
(b) T3I and RnR-SSFSM have no significant statistical difference in G-mean.

Table 16 Ranking of classification accuracies of reduction set with BDT

Data set	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	3	2	5	4	1
Ann	2	5	4	1	3
Arr	3	5	4	2	1
ACA	4	5	2	3	1
Ban	3	2	4	5	1
SPF	2	1	5	4	3
Ion	5	2	4	1	3
DRD	2	5	4	3	1
QSA	3	2	5	4	1
Spa	2	4	5	3	1
Thy	3	5	4	2	1
EES	3	5	1	4	2
Average	2.9167	3.0000	3.9167	3.5833	1.5833

Table 17 Ranking of classification accuracies of reduction set with KNN

Data set	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	5	3	2	4	1
Ann	1	5	4	2	3
Arr	2	5	4	3	1
ACA	5	3	2	4	1
Ban	3	2	4	5	1
SPF	2	1	3	5	4
Ion	5	4	3	2	1
DRD	2	5	4	3	1
QSA	2	4	5	3	1
Spa	2	4	5	3	1
Thy	5	4	3	2	1
EES	5	3	1	2	4
Average	3.2500	3.1667	3.3333	3.5833	1.6667

By Fig. 9b, we have:

(a) The G-mean of T3I is significantly better than SFSE, RnR-SSFSM and SHIVAM;
(b) In terms of G-mean, there is no obvious difference between T3I and MEHAR.

It can be seen that the comprehensive ranking of T3I is higher than other algorithms, as shown in Figs. 8 and 9. In summary, the results confirm that the suggested T3I outperforms the other algorithms.

Table 18 Ranking of G-mean with classifier BDT

Data set	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	5	4	2	3	1
Ann	2	5	4	1	3
Arr	2	4	5	3	1
ACA	4	5	2	3	1
Ban	3	2	4	5	1
SPF	3	2	4	5	1
Ion	1	3	4	2	5
DRD	4	5	3	2	1
QSA	4	2	5	3	1
Spa	2	4	5	3	1
Thy	5	2	4	3	1
EES	4	5	1	3	2
Average	3.25	3.58	3.58	3	1.58

Table 19 Ranking of G-mean with classifier KNN

Data set	MEHAR	SHIVAM	SFSE	RnR-SSFSM	T3I
Aba	5	3	2	4	1
Ann	3	5	4	2	1
Arr	3	5	4	2	1
ACA	5	4	2	3	1
Ban	3	1	2	5	4
SPF	3	2	4	5	1
Ion	1	5	4	3	2
DRD	2	5	4	3	1
QSA	2	3	5	4	1
Spa	2	4	5	3	1
Thy	3	5	2	4	1
EES	5	2	1	3	4
Average	3.08	3.67	3.25	3.42	1.58

Table 20 Friedman test for classification accuracy on two classifiers

Source	SS	df	MS	χ^2	p
Groups(BDT)	38.3333	4	9.5833	15.33	0.0041
Error(BDT)	81.6667	44	1.85606		
Groups(KNN)	27.8333	4	6.9582	11.13	0.0251
Error(KNN)	92.1667	44	2.0947		

7.3 Experimental analysis of the parameter λ

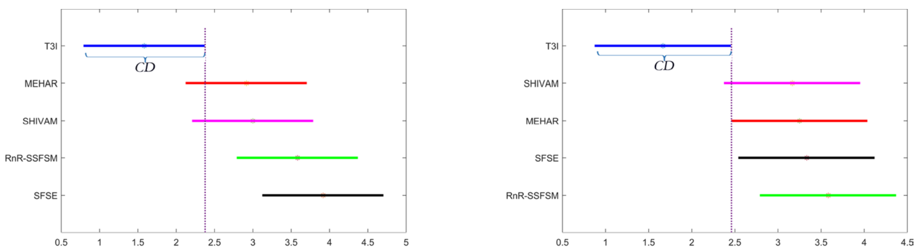
Let's next discuss another parameter λ of T3I. λ control the missing rate of labels. MEHAR, SHIVAM, SFSE and T3I are all algorithms dealing with semi-supervised information systems. When a takes different values, in order to compare the changes of the four

Table 21 Friedman test for G-mean on two classifiers

Source	SS	df	MS	χ^2	p
Groups(BDT)	33	4	8.25	13.2	0.0103
Error(BDT)	87	44	1.98		
Groups(KNN)	32	4	8.08	12.93	0.0116
Error(KNN)	87.67	44	1.99		

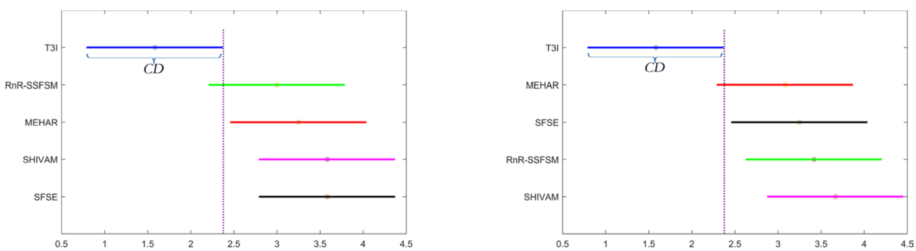
algorithms, Fig. 10 is drawn for observation. The x-axis is parameter λ, ranging from 0.1 to 0.9, indicating that the label missing rate ranges from 10% to 90%, and the y-axis is the classification accuracy. Each algorithm in the figure has two curves, and the solid line and dotted line respectively correspond to the accuracy under BDT and KNN classifiers. The blue line represents T3I, and other color lines are shown in the legend.

The classification accuracy curves of datasets Ann and SPF are relatively stable, indicating that they are not affected by missing labels. Other subgraphs shows that λ has a great impact on the classification accuracy, and the curve fluctuates obviously. We find that no matter what the value of λ is, the blue lines in the subgraphs Aba, Arr, QSA and Thy are closer to the line $y = 1$ than those in other colors. This confirms that the Aba, Arr,



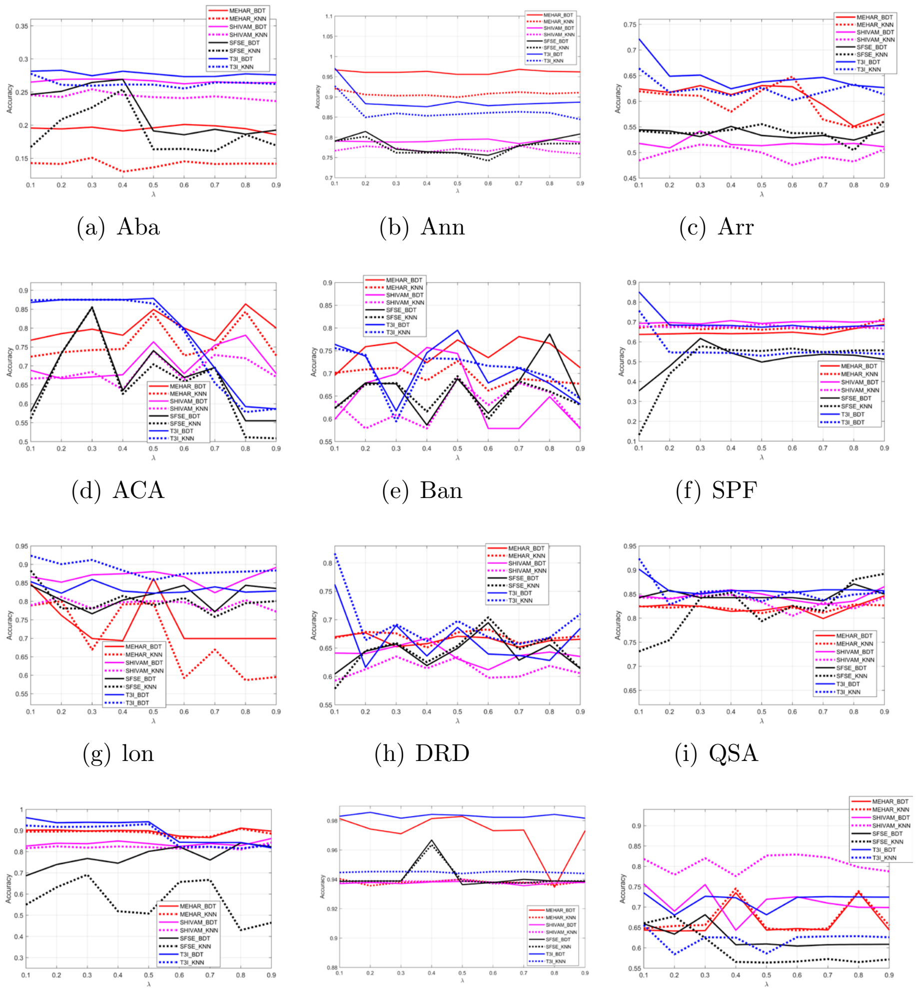
(a) BDT (b) KNN

Fig. 8 Nemenyi test for accuracy on two classifiers



(a) BDT (b) KNN

Fig. 9 Nemenyi test for G-mean on two classifiers



(a) Aba (b) Ann (c) Arr

(d) ACA (e) Ban (f) SPF

(g) Ion (h) DRD (i) QSA

(j) Spa (k) Thy (l) EES

Fig. 10 Influence of parameter λ on classification accuracy

QSA and Thy datasets of T3I algorithm experiment have higher classification accuracy. MEHAR performs better in dataset Ann, while SHIVAM performs better in dataset EES. The classification accuracy of other datasets does not show an obvious rule under different algorithms and λ.

Therefore, we come to the conclusion that T3I has obvious advantages in attribute reduction of most hybrid datasets. However, with the different values of parameters θ and λ, the benefits are not clear, resulting in a small range of oscillations.

8 Conclusions

Based on the defined degrees of importance, we propose the semi-supervised attribute reduction algorithm. The algorithm can flexibly adapt to various missing rates of label. Experimental results and statistical tests on 12 datasets have shown that the degrees of importance are effective, the proposed algorithm is not prone to over-fitting and under-fitting, and can deal with various missing rates more effectively by the comparison with other state-of-the-art algorithms. The findings can enable us to effectively cope with all kinds of data with different missing rates. In the future, we will consider applying this idea to gene data.

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 (12261096), Natural Science Foundation of Guangxi Province (2020GXNSFAA159155) and Natural Science Foundation of Yulin (202125001).

Author contributions ZL: Investigation, Methodology; JH: Data curation, Investigation, Writing-original draft. PW: Investigation, Software, Validation; CFW: Writing-review & editing.

Declarations

Competing interest The authors declare no competing interests

References

Cament LA, Castillo LE, Perez JP, Galdames FJ, Perez CA (2014) Fusion of local normalization and Gabor entropy weighted features for face identification. Pattern Recognit 47(2):568–577

Dai JH, Hu H, Zheng GJ, Hu QH, Han HF, Shi H (2016) Attribute reduction in interval-valued information systems based on information entropies. Front Inform Technol Electron Eng 17(9):919–928

Dai JH, Tian HW (2013) Entropy measures and granularity measures for set-valued information systems. Inform Sci 240:72–82

Dai JH, Wang WT, Xu Q (2013) An uncertainty measure for incomplete decision tables and its applications. IEEE Trans Cybern 43(4):1277–1289

Dai JH, Hu QH, Zhang JH, Hu H, Zheng NG (2017) Attribute selection for partially labeled categorical data by rough set approach. IEEE Trans Cybern 47(9):2460–2471

Dai JH, Xu Q, Wang WT, Tian HW (2012) Conditional entropy for incomplete decision systems and its application in data mining. Int J General Syst 41(7):713–728

Ershadi MM, Seifi A (2022) Applications of dynamic feature selection and clustering methods to medical diagnosis. Appl Soft Comput 126:109293

Hu SD, Miao DQ, Yao YY (2021) Three-way label propagation based semi-supervised attribute reduction. Chin J Comput 44(11):2332–2343

Houssein EH, Saber E, Ali AA, Wazery YM (2022) Centroid mutation-based search and rescue optimization algorithm for feature selection and classification. Expert Syst Appl 191:116235

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

Hu M, Tsang ECC, Guo YT, Xu WH (2022) Fast and robust attribute reduction based on the separability in fuzzy decision systems. IEEE Trans Cybern 52(6):5559–5572
Hu QH, Yu DR, Liu J, Wu C (2008) Neighborhood rough set based heterogeneous feature subset selection. Inform Sci 178(18):3577–3594
Han YH, Yang Y, Yan Y, Ma ZG, Zhou XF (2015) Semisupervised feature selection via spline regression for video semantic recognition. IEEE Trans Neural Netw Learn Syst 26(2):252–264
Kryszkiewicz M (1999) Rules in incomplete information systems. Inform Sci 113:271–292
Liang JY, Qian YH (2008) Information granules and entropy theory in information systems. Sci China (Ser F) 51:1427–1444
Narkhede S (2018) Understanding auc-roc curve. Towards Data Sci 26(1):220–227
Navarrete J, Viejo D, Cazorla M (2016) Color smoothing for RGB-D data using entropy information. Appl Soft Comput 46:361–380
Pawlak Z (1982) Rough sets. Int J Comput Inform Sci 11:341–356
Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht
Pashaei E, Pashaei E (2022) An efficient binary chimp optimization algorithm for feature selection in biomedical data classification. Neural Computi Appl 34:6427–6451
Qian YH, Liang JY, Wu WZ, Dang CY (2011) Information granularity in fuzzy binary GrC model. IEEE Trans Fuzzy Syst 19:253–264
Sang BB, Chen HM, Yang L, Li TR, Xu WH (2021) Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets. IEEE Trans Fuzzy Syst 30:1683–1697
Shreevastava S, Tiwari A, Som T(2019) Feature subset selection of semi-supervised data: an intuitionistic fuzzy-rough set-based concept. Proceedings of International Ethical Hacking Conference 2018. Springer, Singapore, 2019: 303–315
Solorio-Fernndez S, Martnez-Trinidad JF, Carrasco-Ochoa JA (2020) A supervised filter feature selection method for mixed data based on spectral feature selection and information-theory redundancy analysis. Pattern Recogn Lett 138:321–328
Swiniarski RW, Skowron A (2003) Rough set methods in feature selection and recognition. Pattern Recogn Lett 24:833–849
Singh S, Shreevastava S, Som T, Somani G (2020) A fuzzy similarity-based rough set approach for attribute selection in set-valued information systems. Soft Comput 24:4675–4691
Tiwari A, Chaturvedi A (2022) A hybrid feature selection approach based on information theory and dynamic butterfly optimization algorithm for data classification. Expert Syst Appl 196:116621
UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets.html
Wu XP, Chen HM, Li TR, Wan JH (2021) Semi-supervised feature selection with minimal redundancy based on local adaptive. Appl Intell 51:8542–8563
Wan JH, Chen HM, Yuan Z, Li TR, Yang XL, Sang BB (2021) A novel hybrid feature selection method considering feature interaction in neighborhood rough set. Knowl-Based Syst 227:107–167
Wierman MJ (1999) Measuring uncertainty in rough set theory. Int J General Syst 28:283–297
Wang CZ, Huang Y, Shao MW, Hu QH, Chen DG (2020) Feature selection based on neighborhood self-information. IEEE Trans Cybern 50(9):4031–4042
Wang F, Liu JC, Wei W (2018) Semi-supervised feature selection algorithm based on information entropy. Comput Sci 45(S2):427–430
Wan L, Xia SJ, Zhu Y, Lyu ZH (2021) An improved semi-supervised feature selection algorithm based on information entropy. Stat Decis 17:66–70
Wang YB, Chen XJ, Dong K (2019) Attribute reduction via local conditional entropy. Int J Mach Learn Cybernet 10(12):3619–3634
Wang P, Zhang PF, Li ZW (2019) A three-way decision method based on Gaussian kernel in a hybrid information system with images: an application in medical diagnosis. Appl Soft Comput 77:734–749
Xu ZL, King I, Michael RTL, Jin R (2010) Discriminative semi-supervised feature selection via manifold regularization. IEEE Trans Neural Netw 21(7):1033–1047
Yuan Z, Chen HM, Li TR (2022) Exploring interactive attribute reduction via fuzzy complementary entropy for unlabeled mixed data. Pattern Recognit 127:108651
Yuan Z, Chen HM, Yang XL, Li TR, Liu KY (2021) Fuzzy complementary entropy using hybrid-kernel function and its unsupervised attribute reduction. Knowl-Based Syst 231:107398
Zhang W, Miao DQ, Gao C, Li F (2016) Semi-supervised attribute reduction based on rough-subspace ensemble learning. J Chin Comput Syst 37(12):2727–2732
Zhang QL, Qu LD, Li ZW (2022) Attribute reduction based on D-S evidence theory in a hybrid information system. Int J Approx Reason 148:202–234

Authors and Affiliations

Zhaowen Li[1] · Jiali He[2] · Pei Wang[2] · Ching-Feng Wen[3,4]

✉ Jiali He
jialihe100@126.com

✉ Ching-Feng Wen
comcfwen@kmu.edu.tw; chingfengwen100@126.com

Zhaowen Li
lizhaowen8846@126.com

Pei Wang
peiwang100@126.com

1 College of Information Engineering, Nanning University, Nanning 530200, Guangxi, People's Republic of China

2 Center for Applied Mathematics of Guangxi, Key Laboratory of Complex System Optimization and Big Data Processing in Department of Guangxi Education, Yulin Normal University, Yulin 537000, Guangxi, People's Republic of China

3 Center for Fundamental Science, Research Center for Nonlinear Analysis and Optimization, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

4 Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung 80708, Taiwan