International Journal of Machine Learning and Cybernetics (2023) 14:1445–1464
https://doi.org/10.1007/s13042-022-01708-2

ORIGINAL ARTICLE

Semi-supervised attribute reduction via attribute indiscernibility

Jianhua Dai[1,2] · Weisi Wang[1,2] · Chucai Zhang[1,2] · Shaojun Qu[1,2]

Received: 23 March 2022 / Accepted: 31 October 2022 / Published online: 29 November 2022

Abstract
Attribute reduction based on rough sets plays an important role in data preprocessing. Discernibility pair, as an effective information measurement, has received extensive attention in attribute reduction. Unfortunately, the existing attribute importance measurement strategies based on discernibility pairs do not apply well to partially labeled data. Meanwhile, most of the existing attribute reduction algorithms focus on the relationships between objects and neglect the relationships between attributes, which may bring highly redundant attributes. Under the background of rough set theory, this paper studies the issue of semi-supervised attribute reduction, i.e. attribute reduction for partially labeled data. Firstly, we introduce the concept of discernibility pair based on object indiscernibility and propose a semi-supervised attribute reduction algorithm via the maximum discernibility pair by combining supervised and unsupervised discernibility pair strategies. Secondly, considering the relationships between attributes, we put forward new methods to define the similarity and distinction between attributes by discernibility pairs. Thirdly, we propose a semi-supervised attribute reduction algorithm by indiscernible attribute classes. Finally, comparative experiments indicate that the proposed algorithms are effective.

Keywords Rough sets · Semi-supervised attribute reduction · Discernibility pair · Indiscernible attribute class

1 Introduction

Rough set theory, proposed by Pawlak [1, 2], is a new mathematical tool to deal with fuzzy and uncertain information. One of the core contents of rough set theory is attribute reduction, also called feature selection, which is a very important data preprocessing procedure. Attribute reduction removes irrelevant attributes to reduce the difficulty of learning tasks [3, 4].

The most of rough set methods partition the universe based on the indiscernibility relation between objects. Chen et al. [5] proposed a selection method of sample pairs in rough sets. Dai et al. [6] defined the discernibility relationship between objects in the framework of fuzzy rough sets, namely the maximal discernibility pair. Dai and Xu [7] proposed an attribute reduction method based on information gain ratio. Wang et al. [8] presented a new information term denoted as independent classification information. Susmaga [9] applied both indiscernibility relation and discernibility relation to attribute reduction. Based on the relative indiscernibility relation and relative discernibility relation of decision systems, Qin and Jing [10] proposed the concepts of λ reduction and μ reduction. Dai et al. [11] constructed two feature selection methods through label symmetric uncertainty correlation learning and feature redundancy evaluation. Qian et al. [12] proposed a mixed attribute reduction algorithm based on indiscernibility relation and differential relation.

It is worth noting that relationships exist not only between objects, but also between attributes. When studying attribute reduction algorithms, most researchers only consider the indiscernibility relation between objects and neglect the relationships between attributes. Actually, relationships between attributes are also very important for attribute reduction. Based on the relationships between attributes, attribute redundancy can be reduced by removing needless attributes and attribute independence can be improved by selecting high-quality attributes. Relationships between attributes have attracted the attention of some scholars in the machine learning community [13, 14]. Mitra et al.

✉ Shaojun Qu
qshj@hunnu.edu.cn

1 Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha 410081, China

2 College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China

[13] proposed the definition of attribute similarity based on the maximum information compression index to carry out attribute clustering. Restrepo and Cornelis [15] adopted functional dependency relations to get the reducts. Jia et al. [16] defined an attribute reduction method based on similarity from the perspective of clustering. Kudo and Murai [17] proposed a binary relationship between different attribute values of the same object.

Attribute reduction methods for complete data have been extensively studied in the field of rough sets. Unfortunately, because it is not easy to obtain the labels of objects, partially labeled data are more common in the real world. Some scholars have studied semi-supervised attribute reduction [18]. Dai et al. [19] proposed a semi-supervised attribute reduction method based on dual-dependency. In [20], Saha et al. used the simulated annealing method to deal with the multi-objective optimization problem consisting of feature selection and semi-supervised clustering. Chang and Yang [21] improved the performance of feature selection by mining the correlation among multiple tasks. Mi et al. [22] achieved dynamic classification learning over the semi-supervised data through concept-cognitive computing system. Xu et al. [23] studied a semi-supervised feature selection algorithm based on Pearson's correlation coefficient. Mi et al. [24] proposed a semi-supervised concept learning method from the perspective of concept space. Dai et al. [25] proposed two semi-supervised attribute reduction algorithms by applying discernibility pairs to partially labeled data sets.

Based on the above analysis, the motivations of this paper are as follows:

1. When studying semi-supervised attribute reduction algorithms, some researchers [25, 26] regarded the importance of attribute subsets as the sum of the importance on labeled data and unlabeled data. However, these methods have a certain shortcoming, that is, they cannot well represent the change of the importance of attribute subsets at different missing rates of datasets. In view of this, in order to measure the discernibility ability of attributes in partially labeled decision information systems, we introduce the concept of discernibility pair based on object indiscernibility. Inspired by the supervised and unsupervised discernibility pair strategies proposed by Dai et al. [25], we combine discernibility pairs from labeled data with discernibility pairs from unlabeled data, and propose a semi-supervised attribute reduction algorithm based on the maximum discernibility pair. Therefore, the combined discernibility pairs can better explain the discernibility ability of attributes and adapt to the change of missing rate.
2. In the field of rough sets, most studies regard a single attribute as a granularity, and many attribute reduction algorithms are proposed. In this process, however, attributes that contain similar discernibility information may be selected into the attribute subset. In other words, redundant attributes are selected into the attribute subset. To handle this issue, this paper proposes a new attribute reduction method. Firstly, we introduce the concept of discernibility pair as a criterion for quantifying the discernibility ability of attributes. Considering the relationship between attributes, we define new fuzzy similarity relation and fuzzy discernibility relation between attributes based on discernibility pairs. Further, inspired by the attribute granulation, we regard the fuzzy discernibility relationship as a distance function and divide the attribute set into indiscernible attribute classes. For the attributes in an indiscernible attribute class, a representative attribute is selected. Therefore, we can delete redundant attributes by selecting representative attributes and propose a semi-supervised attribute reduction algorithm based on attribute indiscernibility.

The rest of this paper is organized as follows. Section 2 reviews some basic notations of rough set theory. In Section 3, new methods to measure the similarity and distinction between attributes are advanced and attribute reduction algorithms based on the maximum discernibility pair and attribute indiscernibility are proposed. Experiments are carried out to verify the performance of the proposed algorithms in Section 4. Conclusion is presented in Section 5.

2 Preliminary knowledge

2.1 Some basic concepts of rough sets

Rough set theory is an effective mathematical tool for processing uncertain information. This section reviews some basic symbols of rough set theory [1, 2].

Definition 1 An information system and a decision information system are respectively represented as $IS =< U, C, V, f >$ and $DS =< U, C \cup \{d\}, V, f >$, where $U = \{x_1, x_2, \dots, x_n\}$ is a nonempty finite set of objects; $C = \{a_1, a_2, \dots, a_m\}$ is a nonempty finite set of conditional attributes, and d is a label of an object; $V = \cup_{a \in C \cup \{d\}} V_a$ and V_a is the domain of attribute a; f is a mapping function, which maps an object in U to exactly one value from domains of an attribute such as $\forall a \in C \cup \{d\}, x \in U, f(a, x) \in V_a$ and $f(a, x)$ is the value of the object x on attribute a. If $\exists x \in U$ such that $f(d, x)$ is equal to a missing value denoted as $*$, then we call it a partially labeled decision information system $PLDS =< U, C \cup \{d\}, V^*, f >$.

For simplicity, we use $a(x)$ to denote $f(a, x)$ and $d(x)$ to denote $f(d, x)$ in this paper.

Definition 2 Let $IS =< U, C, V, f >$ be an information system. $S \subseteq C$ is a subset of the conditional attribute set C. The indiscernibility relation on S is defined as follows:

$$IND(S) = \{(x, y) \in U \times U \mid \forall a \in S, a(x) = a(y)\}. \tag{1}$$

The indiscernibility relation $IND(S)$ means that x and y are indistinguishable by the attribute subset S. Thus, the partition on U induced from $IND(S)$ is expressed as either $U/IND(S)$ or U/S.

Based on the indiscernibility relation, for $X \subseteq U$, the lower approximation and the upper approximation are defined as follows.

Definition 3 Let $IS =< U, C, V, f >$ be an information system, $X \subseteq U$ and $S \subseteq C$. The lower approximation and the upper approximation of X with respect to S can be defined as follows:

$$\begin{aligned} &\underline{S}(X) = \{x \in U \mid [x]_S \subseteq X\}; \\ &\overline{S}(X) = \{x \in U \mid [x]_S \cap X \neq \emptyset\}. \end{aligned} \tag{2}$$

The lower approximation of a set X with respect to $IND(S)$ is the set of the objects which certainly belongs to X. The upper approximation of a set X with respect to $IND(S)$ is the set of the objects which possibly belongs to X.

Definition 4 Let $IS =< U, C, V, f >$ be an information system. The discernibility matrix IM is a $|U| \times |U|$ matrix. For $S \subseteq C$, each item $IM(x, y)$ represents the attribute set that can distinguish object x from object y.

$$IM(x, y) = \{a \in S \mid a(x) \neq a(y)\}. \tag{3}$$

Definition 5 Let $IS =< U, C, V, f >$ be an information system. $S \subseteq C$ is a reduct of IS if and only if:

(1) $\forall x, y \in U$, *if* $IM(x, y) \neq \emptyset$, *then* $S \cap IM(x, y) \neq \emptyset$;
(2) $\forall S' \subset S, \exists IM(x, y) \neq \emptyset, S' \cap IM(x, y) = \emptyset$.

Definition 6 Let $DS =< U, C \cup \{d\}, V, f >$ be a decision information system. The discernibility matrix DM is a $|U| \times |U|$ matrix. For $S \subseteq C$, each item $DM(x, y)$ represents the attribute set that can distinguish object x from object y.

$$DM(x, y) = \{a \in S \mid a(x) \neq a(y) \ \wedge \ d(x) \neq d(y)\}. \tag{4}$$

Obviously, the discernibility matrix is anti-reflexive and symmetric, namely $DM(x, x) = \emptyset$ and $DM(x, y) = DM(y, x)$. Therefore, in order to reduce the computational complexity, we only need to calculate the upper or lower trigonometric part of the discernibility matrix.

Definition 7 Let $DS =< U, C \cup \{d\}, V, f >$ be a decision information system. $S \subseteq C$ is a reduct of DS with respect to d if and only if:

(1) $\forall x, y \in U$, *if* $DM(x, y) \neq \emptyset$, *then* $S \cap DM(x, y) \neq \emptyset$;
(2) $\forall S' \subset S, \exists DM(x, y) \neq \emptyset, S' \cap DM(x, y) = \emptyset$.

Definition 8 Let $IS =< U, C, V, f >$ and $DS =< U, C \cup \{d\}, V, f >$ be an information system and a decision information system, respectively. The core of attribute reduction can be defined as:

$$\begin{aligned} &Core_I(C) = \{a \in C \mid a \in IM(x, y) \wedge |IM(x, y)| = 1\}; \\ &Core_D(C \cup \{d\}) = \{a \in C \mid a \in DM(x, y) \wedge \\ &\qquad\qquad |DM(x, y)| = 1\}. \end{aligned} \tag{5}$$

The intersection of all the reducts is called the core. Therefore, we can first find the core to reduce the complexity of subsequent calculations.

2.2 Discernibility pair

In this section, we introduce the concept of discernibility pair which will be used as a criterion for quantifying the discernibility ability of attributes. The number of discernibility pairs produced by an attribute set reflects the discernibility ability of this attribute set.

Definition 9 [25] Let $IS =< U, C, V, f >$ be an information system. $S \subseteq C$ is a subset of the conditional attribute set C. The discernibility pair set with respect to attribute subset S is defined as follows:

$$DisP_I(S) = \{(x, y) \mid \exists a \in S, a(x) \neq a(y)\}. \tag{6}$$

Proposition 1 *Let* $IS =< U, C, V, f >$ *be an information system. Given* $S' \subseteq S$, *then*

(1) $DisP_I(S') \subseteq DisP_I(S)$;
(2) $DisP_I(S) - DisP_I(S') \subseteq DisP_I(S - S')$.

Proof

(1) For any $(x, y) \in DisP_I(S'), \exists a \in S', s.t.\ a(x) \neq a(y)$. Since $S' \subseteq S$, $\exists a \in S, a(x) \neq a(y)$. Hence, $\subseteq DisP_I(S)$ $(x, y) \in DisP_I(S), DisP_I(S') \subseteq DisP_I(S)$.
(2) According to (1), we know $DisP_I(S') \subseteq DisP_I(S)$. For any $(x, y) \in [DisP_I(S) - DisP_I(S')]$, $DisP_I(S') \subseteq DisP_I(S)$,

thus $(x, y) \in DisP_I(S)$ but $(x, y) \notin DisP_I(S')$. That is, for any $a \in S', a(x) = a(y)$. Since $S' \subseteq S$, $\exists b \in S - S', b(x) \neq b(y)$. Hence, $(x, y) \in DisP_I(S - S')$.

□

Definition 10 [25] Let $IS =< U, C, V, f >$ be an information system. $S \subseteq C$ is a reduct of IS if and only if:

(1) $|DisP_I(S)| = |DisP_I(C)|$;
(2) $\forall S' \subset S, |DisP_I(S')| < |DisP_I(S)|$.

In Definition 10, condition (1) indicates that the attribute subset S can generate the same discernibility pairs as the attribute set C. That is, the attribute subset S has the same discernibility ability as the attribute set C. Condition (2) is used to ensure that the attribute subset S is the minimum, that is, the number of attributes is the least.

Definition 11 [25] Let $DS =< U, C \cup \{d\}, V, f >$ be a decision information system. $S \subseteq C$ is a subset of the conditional attribute set C. The discernibility pair set with respect to attribute subset S relative to decision attribute d is defined as follows:

$$DisP_D(S, d) = \{(x, y) \mid \exists a \in S, a(x) \neq a(y) \wedge d(x) \neq d(y)\}. \tag{7}$$

Proposition 2 *Let $DS =< U, C \cup \{d\}, V, f >$ be a decision information system. Given $S' \subseteq S$, then*

(1) $DisP_D(S', d) \subseteq DisP_D(S, d)$;
(2) $DisP_D(S, d) - DisP_D(S', d) \subseteq DisP_D(S - S', d)$.

Proof

(1) For any $(x, y) \in DisP_D(S', d), \exists a \in S', s.t.\ a(x) \neq a(y), d(x) \neq d(y)$. Since $S' \subseteq S$, $\exists a \in S, a(x) \neq a(y), d(x) \neq d(y)$. Hence, $(x, y) \in DisP_D(S, d), DisP_D(S', d) \subseteq DisP_D(S, d)$.
(2) According to (1), we know $DisP_D(S', d) \subseteq DisP_D(S, d)$. For any $(x, y) \in [DisP_D(S, d) - DisP_D(S', d)]$, $DisP_D(S', d) \subseteq DisP_D(S, d)$, thus $(x, y) \in DisP_D(S, d)$ but $(x, y) \notin DisP_D(S', d)$. That is, for any $a \in S', a(x) = a(y) \wedge d(x) \neq d(y)$. Since $S' \subseteq S$, $\exists b \in S - S', b(x) \neq b(y) \wedge d(x) \neq d(y)$. Hence, $(x, y) \in DisP_D(S - S', d)$.

□

Based on Propositions 1 and 2, we can find that discernibility pairs satisfy monotonicity with respect to attribute subset S. Moreover, as the same as information entropy and positive domain, the number of discernibility pairs produced by an attribute can also be considered as the amount of information carried by the attribute. The more discernibility pairs an attribute generates, the more information that attribute carries, and vice versa.

Definition 12 [25] Let $DS =< U, C \cup \{d\}, V, f >$ be a decision information system. $S \subseteq C$ is a reduct of DS with respect to d if and only if:

(1) $|DisP_D(S, d)| = |DisP_D(C, d)|$;
(2) $\forall S' \subset S, |DisP_D(S', d)| < |DisP_D(S, d)|$.

In Definition 12, condition (1) indicates that relative to decision attribute d, the attribute subset S can generate the same discernibility pairs as the attribute set C. That is, the attribute subset S has the same discernibility ability as the attribute set C relative to decision attribute d. Condition (2) is used to ensure that the attribute subset S is the minimum, that is, the number of attributes is the least.

3 Semi-supervised attribute reduction based on relations of objects and attributes

3.1 Semi-supervised attribute reduction by the maximum discernibility pair

From the above discussion, we know that discernibility pairs can be used to evaluate the discernibility ability of an attribute subset $S \subseteq C$ in IS and DS. In this section, we propose a semi-supervised attribute reduction algorithm by dividing partially labeled data into a labeled part and an unlabeled part. In order to deal with partially labeled data, the discernibility pair set induced by the attribute subset $S \subseteq C$ is defined as follows:

Definition 13 Let $PLDS =< U, C \cup \{d\}, V^*, f >$ be a partially labeled decision information system. $S \subseteq C$ is a subset of the conditional attribute set C. The discernibility pair set with respect to attribute subset S relative to decision attribute d is defined as follows:

$$DisP_P(S, d) = DisP_I(S, d) \cup DisP_D(S). \tag{8}$$

Based on Definition 13, the discernibility pair set with respect to attribute subset S relative to decision attribute d combines discernibility pairs from labeled data with discernibility pairs from unlabeled data. It can better express the discernibility ability of attribute subset S than Definitions 9 and 11 in partially labeled decision information system.

Proposition 3 *Let $PLDS = < U, C \cup \{d\}, V^*, f >$ be a partially labeled decision information system. Given $S' \subseteq S$, then*

(1) $DisP_P(S', d) \subseteq DisP_P(S, d)$;
(2) $DisP_P(S, d) - DisP_P(S', d) \subseteq DisP_P(S - S', d)$.

Proof

(1) For any $(x, y) \in DisP_P(S', d)$,

 (a) if $d(x) =* \wedge d(y) =*$, $\exists a \in S', s.t.\ a(x) \neq a(y)$. Since $S' \subseteq S$, $\exists a \in S, a(x) \neq a(y)$. Hence, $(x, y) \in DisP_P(S, d), DisP_P(S', d) \subseteq DisP_P(S, d)$;
 (b) if $d(x) \neq * \wedge d(y) \neq *$, $\exists a \in S', s.t.\ a(x) \neq a(y), d(x) \neq d(y)$. Since $S' \subseteq S, \exists a \in S, a(x) \neq a(y), d(x) \neq d(y)$. Hence, $(x, y) \in DisP_P(S, d), DisP_P(S', d) \subseteq DisP_P(S, d)$;
 (c) if $d(x) \neq * \wedge d(y) =*$ or $d(x) =* \wedge d(y) \neq *$, $(x, y) \notin DisP_P(S', d)$. In summary, $DisP_P(S', d) \subseteq DisP_P(S, d)$.

(2) According to (1), we know $DisP_P(S', d) \subseteq DisP_P(S, d)$. For any $(x, y) \in [DisP_P(S, d) - DisP_P(S', d)]$, $DisP_P(S', d) \subseteq DisP_P(S, d)$, thus $(x, y) \in DisP_P(S, d)$ but $(x, y) \notin DisP_P(S', d)$.

 (a) If $d(x) =* \wedge d(y) =*$, for any $a \in S', a(x) = a(y)$. Since $S' \subseteq S$, $\exists b \in S - S', b(x) \neq b(y)$. Hence, $(x, y) \in DisP_P(S - S')$;
 (b) If $d(x) \neq * \wedge d(y) \neq *$, for any $a \in S', a(x) = a(y) \wedge d(x) \neq d(y)$. Since $S' \subseteq S$, $\exists b \in S - S'$, $b(x) \neq b(y) \wedge d(x) \neq d(y)$. Hence, $(x, y) \in DisP_P(S - S', d)$;
 (c) If $d(x) \neq * \wedge d(y) =*$ or $d(x) =* \wedge d(y) \neq *$, $(x, y) \notin [DisP_P(S, d) - DisP_P(S', d)]$. In summary, $DisP_P(S, d) - DisP_P(S', d) \subseteq DisP_P(S - S', d)$.

□

Based on Proposition 3, we can find that discernibility pairs satisfy monotonicity with respect to the attribute subset S in partially labeled decision information systems. Therefore, we can use discernibility pairs to study semi-supervised attribute reduction method.

Theorem 1 $Dis_I(S) \cap DisP_D(S, d) = \emptyset$.

Proof For any $(x, y) \in Dis_D(S, d), d(x) \neq *$ *and* $d(y) \neq *$. However, for any $(x, y) \in Dis_I(S), d(x) =*$ *and* $d(y) =*$. Hence, $(x, y) \notin Dis_I(S)$. □

Based on Theorem 1, we can find that discernibility pair set from labeled data and discernibility pair set from unlabeled data have no intersection. Because discernibility pair set satisfies monotonicity with respect to attribute subset S in both labeled data and unlabeled data, Theorem 1 also indicates that the union of discernibility pair sets induced by the attribute subset S is monotonically increasing.

Proposition 4 *Let S and d be a non-empty attribute subset and a decision attribute, respectively. For any $x, y \in U$, $DisP_P(S, d)$ meets the following properties*:

(1) *Irreflexivity. If $x = y$, then $(x, y) \notin DisP_P(S, d)$;*
(2) *Symmetric. If $(x, y) \in DisP_P(S, d)$, then $(y, x) \in DisP_P(S, d)$.*

Proof

(1) Since $x = y$, for any $a \in S$, $a(x) = a(y)$. Thus, $(x, y) \notin Dis_I(S)$ and $(x, y) \notin Dis_D(S, d)$. Hence, $(x, y) \notin Dis_P(S, d)$.
(2) If $(x, y) \in DisP_P(S, d)$, it satisfies that $\exists a \in S, a(x) \neq a(y)$ or $a(x) \neq a(y) \wedge d(x) \neq d(y)$. Thus, $\exists a \in S, a(y) \neq a(x)$ or $a(y) \neq a(x) \wedge d(y) \neq d(x)$. Hence, $(y, x) \in DisP_P(S, d)$.

□

Based on Proposition 4, we can find that discernibility pairs satisfy irreflexivity and symmetry. It can be used to reduce the computational complexity of the algorithm by only calculating the upper or lower triangle of the matrix.

Definition 14 Let $PLDS = < U, C \cup \{d\}, V^*, f >$ be a partially labeled decision information system. $S \subseteq C$ is a reduct of $PLDS$ with respect to d if and only if:

(1) $|DisP_P(S)| = |DisP_P(C)|$;
(2) $\forall S' \subset S, |DisP_P(S')| < |DisP_P(S)|$.

In Definition 14, condition (1) indicates that the attribute subset S can generate the same discernibility pairs as the attribute set C relative to decision attribute d. That is, the attribute subset S has the same discernibility ability as the attribute set C. Condition (2) is used to ensure that the attribute subset S is the minimum, that is, the number of attributes is the least.

$DisP_P(S, d)$ combines discernibility pair set from labeled data with discernibility pair set from unlabeled data, and reflects the discernibility ability of S in partially labeled data. According to Definition 14, a semi-supervised attribute reduction algorithm based on the maximum discernibility pair is proposed as Algorithm 1.

Algorithm 1 Semi-supervised attribute reduction algorithm based on the maximum discernibility pair (Semi-DP)

Input: A partially labeled decision information system $PLDS =< U, C \cup \{d\}, V, f >$
Output: A reduct R

Compute $|DisP_P(C,d)|$, set $R = \varnothing$, $maxnum = 0$, $selAtt = 0$.
while $|DisP_P(R,d)| < |DisP_P(C,d)|$ **do**
  for $a \in C - R$ **do**
    if $|DisP_P(R \cup \{a\}, d)| > maxnum$ **then**
      $maxnum = |DisP_P(R \cup \{a\}, d)|$;
      $selAtt = a$.
    end if
  end for
  $R := R \cup \{selAtt\}$.
end while
for $a \in R$ **do**
  if $DisP_P(R - \{a\}, d) = DisP_P(R, d)$ **then**
    $R := R - \{a\}$.
  end if
end for

Classical supervised or unsupervised algorithms consider either labeled information or unlabeled information alone, which often leads to insufficient use of information. Furthermore, it may influence the evaluation of the discernibility ability of attribute and increase the number of core attributes. Semi-DP uses both labeled information and unlabeled information by the metric of discernibility pairs, which fully reflects the amount of information carried by the attribute subset S.

In order to better understand the algorithm Semi-DP, we give an example as shown in Table 1. We only need to calculate the upper or lower triangle of the matrix because of the irreflexivity and symmetry of discernibility pairs.

Example 1 Considering a partially labeled decision information system $PLDS =< U, C \cup \{d\}, V^*, f >$ shown in Table 1, $U = \{x_1, x_2, \dots x_7\}$ is the object set, $C = \{a_1, a_2, a_3, a_4\}$ is the attribute set , and d is the decision attribute.

Table 1 The first partially labeled decision information system

U	a_1	a_2	a_3	a_4	d
x_1	1	1	2	1	d_1
x_2	1	2	1	2	d_2
x_3	1	2	2	1	d_2
x_4	2	2	2	1	*
x_5	1	2	1	2	*
x_6	2	2	1	2	*
x_7	2	2	1	2	*

(1) We calculate the discernibility pair set defined in Definition 13 $DisP_P(C,d) = \{(x_1,x_2),(x_1,x_3),(x_4,x_5),(x_4,x_6),(x_4,x_7),(x_5,x_6),(x_5,x_7)\}$.

(2) In the first round, we set $R = \varnothing$. For any $a_k \in C - R$ and calculate the number of discernibility pairs $|DisP_P(a_k,d)|$ as follows: $|DisP_P(a_1,d)| = 3$, $|DisP_P(a_2,d)| = 2$, $|DisP_P(a_3,d)| = 4$ and $|DisP_P(a_4,d)| = 4$. We can conclude that attribute a_3 should be selected, $R = R \cup \{a_3\}$, and $|DisP_P(R,d)| < |DisP_P(C,d)|$.

(3) In the second round, we can get $|DisP_P(R \cup \{a_1\},d)| = 6$, $|DisP_P(R \cup \{a_2\},d)| = 5$ and $|DisP_P(R \cup \{a_4\},d)| = 4$. We can conclude that attribute a_1 should be selected, $R = R \cup \{a_1\}$, and $|DisP_P(R,d)| < |DisP_P(C,d)|$.

(4) In the third round, we can get $|DisP_P(R \cup \{a_2\},d)| = 7$ and $|DisP_P(R \cup \{a_4\},d)| = 6$. We can conclude that attribute a_2 should be selected, $R = R \cup \{a_2\}$, and $|DisP_P(R,d)| = |DisP_P(C,d)|$.

(5) Finally, because $|DisP_P(R - a_3,d)| < |DisP_P(R,d)|$, $|DisP_P(R - a_1,d)| < |DisP_P(R,d)|$ and $|DisP_P(R - a_2,d)| < |DisP_P(R,d)|$, there are no redundant attributes in R. We get the attribute subset $R = \{a_3, a_1, a_2\}$.

3.2 Semi-supervised attribute reduction by attribute indiscernibility

Considering the importance of the relationships between attributes such as the implication relation and the indiscernibility relation in the attribute reduction, we define the concept of similarity for measuring the relationships between attributes as follows:

Definition 15 Let $PLDS =< U, C \cup \{d\}, V^*, f >$ be a partially labeled decision information system. For any $a, b \in C$, the similarity between attribute a and attribute b with respect to decision attribute d is defined as follows:

$$Simatt(a,b,d) = \frac{|DisP_P(a,d) \cap DisP_P(b,d)|}{|DisP_P(a,d) \cup DisP_P(b,d)|}. \tag{9}$$

If $DisP_P(a,d) = \varnothing$ and $DisP_P(b,d) = \varnothing$, which means we cannot distinguish between any two objects through attribute a and attribute b, then we default $Simatt(a,b,d) = 1$.

Definition 15 is different from the definition of the classical distance-based similarity. The reason is that the similarity between attribute a and attribute b is determined by the number of discernibility pairs induced by attribute a and attribute b. As shown in Fig. 1, the shadowed part indicates the redundant information induced by attribute a and attribute b with respect to decision attribute d (formulate

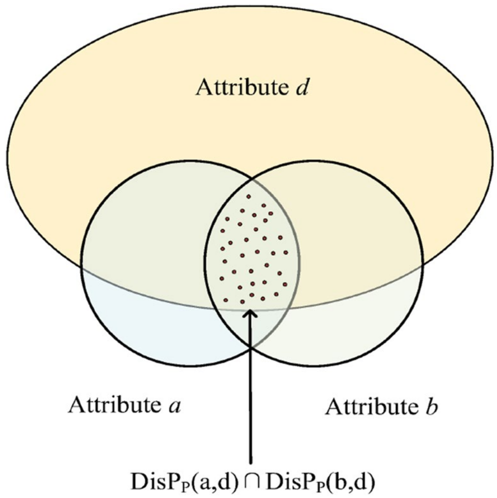


Fig. 1 The amount of common information between attribute a and attribute b with respect to d

$|DisP_P(a,d) \cap DisP_P(b,d)|$). From another perspective, this part of redundant information is also a reflection of the similarity between attribute a and attribute b with respect to attribute d. After getting the definition of similarity, we can easily get the definition of the distinction between attribute a and attribute b with respect to attribute d.

Proposition 5 *Let S and d be a non-empty attribute subset and a decision attribute, respectively. For any* $a, b \in S$, *Simatt*(a, b, d) *meets the following properties*:

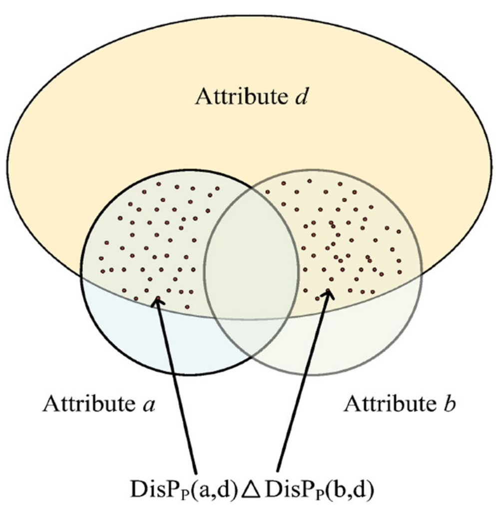


Fig. 2 The amount of differentiated information between attribute a and attribute b with respect to d

(1) *Non-negativity*. $Simatt(a,b,d) \geqslant 0$;
(2) *Reflexivity*. $Simatt(a,a,d) = 1$;
(3) *Symmetric*. $Simatt(a,b,d) = Simatt(b,a,d)$.

Proof

(1) Since $|DisP_P(a,d) \cap DisP_P(b,d)| \geqslant 0$, $|DisP_P(a,d) \cup DisP_P(b,d)| \geqslant 0$. Thus, $Simatt(a,b,d) \geqslant 0$.

$$Simatt(a,a,d) = \frac{|DisP_P(a,d) \cap DisP_P(a,d)|}{|DisP_P(a,d) \cup DisP_P(a,d)|} = \frac{|DisP_P(a,d)|}{|DisP_P(a,d)|} = 1. \tag{10}$$

(2)

$$Simatt(a,b,d) = \frac{|DisP_P(a,d) \cap DisP_P(b,d)|}{|DisP_P(a,d) \cup DisP_P(b,d)|} = \frac{|DisP_P(b,d) \cap DisP_P(a,d)|}{|DisP_P(b,d) \cup DisP_P(a,d)|} = Simatt(b,a,d). \tag{11}$$

(3)

□

Based on Proposition 5, we can find that *Simatt*(a, b, d) satisfies reflexivity and symmetry. It can be used to reduce the computational complexity of the algorithm by only calculating the upper or lower triangle of the matrix. At the same time, it also indicates that *Simatt*(a, b, d) satisfies the properties of fuzzy similarity relation which can be used to measure the similarity between attributes.

Definition 16 Let $PLDS =< U, C \cup \{d\}, V^*, f >$ be a partially labeled decision information system. For any $a, b \in C$, the distinction between attribute a and attribute b with respect to decision attribute d is defined as follows:

$$Disatt(a,b,d) = \frac{|DisP_P(a,d) \vartriangle DisP_P(b,d)|}{|DisP_P(a,d) \cup DisP_P(b,d)|}. \tag{12}$$

$\vartriangle$ means symmetrical difference. $DisP_P(a,d) \vartriangle DisP_P(b,d) = (DisP_P(a,d) - DisP_P(b,d)) \cup (DisP_P(b,d) - DisP_P(a,d))$. If $DisP_P(a,d) = \varnothing$ *and* $DisP_P(b,d) = \varnothing$, which means we cannot distinguish between any two objects through attribute a and attribute b, then we default $Disatt(a,b,d) = 0$.

Definition 16 is different from the definition of the classical distance-based distinction. The reason is that the distinction between attribute a and attribute b is determined by the number of discernibility pairs induced by attribute a and attribute b. As shown in Fig. 2, the shadowed parts indicate the unique classification information induced by attribute a and attribute b with respect to decision attribute d (formulate $|DisP_P(a,d) \vartriangle DisP_P(b,d)|$). From another perspective, the amount of differentiated information in this part is also the reflection of the distinction between attribute a and attribute b with respect to attribute d.

Proposition 6 *Let S and d be a non-empty attribute subset and a decision attribute, respectively. For any $a, b \in S$, $Disatt(a, b, d)$ meets the following properties:*

(1) *Non-negativity.* $Disatt(a, b, d) \geqslant 0$;
(2) *Irreflexivity.* $Disatt(a, a, d) = 0$;
(3) *Symmetric.* $Disatt(a, b, d) = Disatt(b, a, d)$;
(4) *Triangle inequality.* $Disatt(a, b, d) + Disatt(b, c, d) \geqslant Disatt(a, c, d)$.

Proof

(1) Since $|DisP_P(a, d) \triangle DisP_P(b, d)| \geqslant 0$, $|DisP_P(a, d) \cup DisP_P(b, d)| \geqslant 0$. Thus, $Disatt(a, b, d) \geqslant 0$.

$$Disatt(a,a,d) = \frac{|DisP_P(a,d) \triangle DisP_P(a,d)|}{|DisP_P(a,d) \cup DisP_P(a,d)|} = \frac{0}{|DisP_P(a,d)|} = 0. \quad (13)$$

(2)

$$\begin{aligned} Disatt(a,b,d) &= \frac{|DisP_P(a,d) \triangle DisP_P(b,d)|}{|DisP_P(a,d) \cup DisP_P(b,d)|} \\ &= \frac{|DisP_P(b,d) \triangle DisP_P(a,d)|}{|DisP_P(b,d) \cup DisP_P(a,d)|} \\ &= Disatt(b,a,d). \end{aligned} \quad (14)$$

(3)

(4) As can be seen from the Venn diagram in Fig. 3, $Disatt(a, b, d) + Disatt(b, c, d) \geqslant Disatt(a, c, d)$ which can be written as

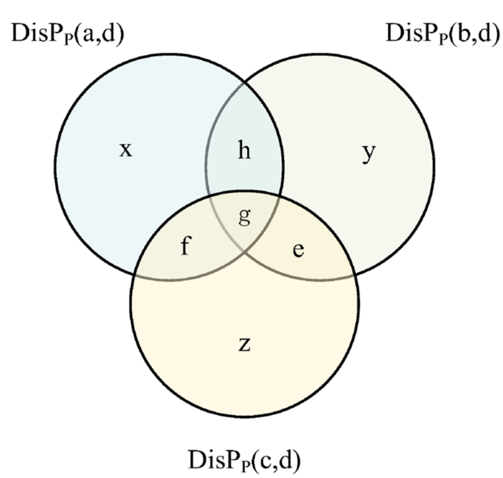


Fig. 3 The Venn diagram of triangle inequality

$$\frac{x+y+e+f}{x+y+h+e+f+g} + \frac{y+z+h+f}{y+z+h+e+f+g} \geqslant \frac{x+z+h+e}{x+z+h+e+f+g}. \quad (15)$$

By reduction of fractions to a common denominator, we have

$$\begin{aligned} &(x^2y + xy^2 + x^2z + 2xyz + y^2z + xz^2 + yz^2 + x^2h \\ &+ 2xyh + y^2h + 2xzh + 2yzh + xh^2 + yh^2 + 2xye \\ &+ y^2e + 2xze + 2yze + z^2e + xhe + 2yhe + zhe \\ &+ ye^2 + ze^2 + x^2f + 4xyf + 2y^2f + 4xzf + 4yzf \\ &+ z^2f + 4xhf + 5yhf + 3zhf + 2h^2f + 3xef \\ &+ 5yef + 4zef + 4hef + 2e^2f + 3xf^2 + 4yf^2 \\ &+ 3zf^2 + 4hf^2 + 4ef^2 + 2f^3 + 2xyg + 2y^2g \\ &+ 2xzg + 2yzg + xhg + 3yhg + 3yeg + zeg \\ &+ 3xfg + 6yfg + 3zfg + 4hfg + 4efg + 4f^2g \\ &+ 2yg^2 + 2fg^2)/[(x + y + h + e + f + g) \\ &(x + z + h + e + f + g)(y + z + h + e + f + g)] \geqslant 0. \end{aligned}$$

Since $x, y, z, h, e, f, g \geqslant 0$, the above inequality is always true.

□

Based on Proposition 6, we can find that $Disatt(a, b, d)$ satisfies non-negativity, irreflexivity, symmetry and triangle inequality. It can be used to reduce the computational complexity of the algorithm by only calculating the upper or lower triangle of the matrix. At the same time, it also indicates that $Disatt(a, b, d)$ satisfies the properties of the distance function which can be used to measure the distinction degree between attributes .

Theorem 2 $Simatt(a, b, d) + Disatt(a, b, d) = 1$.

Proof It can be easily proved by Definitions 15 and 16.

□

Theorem 2 shows that the sum of $Simatt(a, b, d)$ and $Disatt(a, b, d)$ is equal to 1, just like other fuzzy similarity relations and fuzzy distinction relations. Thus, similarity (distinction) can be obtained by 1- distinction (similarity).

At this point, we have a definition of distinction (also known as distance) between any two attributes. Next, we propose an attribute reduction algorithm based on attribute indiscernibility. This will be divided into two steps. In the first step, we divide the attribute set C into p ($p \leqslant |C|$) different classes. In the second step, we select an attribute from each class that can generate the most discernibility pairs as the representative attribute of the corresponding class.

Before clustering, core attributes need to be preprocessed. When calculating the discernibility matrix, an entry in the discernibility matrix contains a core attribute if it has only one attribute. In order to prevent multiple core attributes from being divided into the same class, we incorporate the attributes into the reduct in the first place. For the remaining attributes, the attribute clustering algorithm [13] is used.

Through the attribute clustering algorithm, we can divide the attribute set into indiscernible attribute classes $IndF(IndF_1, IndF_2, \dots IndF_m, m \leqslant |C|)$, which are similar to the division of the equivalence classes in rough set theory.

Proposition 7 *Let $PLDS = < U, C \cup \{d\}, V^*, f >$ be a partially labeled decision information system. Given $C' = C - [Core_I(C) \cup Core_D(C \cup \{d\})]$, then*

(1) $\cup_{F_i \in F} IndF_i = C'$;
(2) $IndF_i \cap IndF_j = \varnothing$.

Proof

(1) For any $a \in \cup_{F_i \in F} IndF_i$, since $a \in C', \cup_{F_i \in F} IndF_i \subseteq C'$. For any $a \in C'$, $\exists IndF_i \subseteq \cup_{F_i \in F} IndF_i$, s.t. $a \in IndF_i$. Thus, $C' \subseteq \cup_{F_i \in F} IndF_i$. In conclusion, $\cup_{F_i \in F} IndF_i = C'$.
(2) Assuming that $\exists a \in IndF_i \cap IndF_j$, then $a \in IndF_i$ and $a \in IndF_j$. Since the clustering process only puts each attribute into a certain class, the assumption is wrong. Thus, $IndF_i \cap IndF_j = \varnothing$.

□

Based on Proposition 7, we can find that indiscernible attribute classes have the same properties as equivalence class, that is, the intersection is empty and the union is the universal set. When we choose representative attributes, it ensures that we will consider all attributes and will not choose duplicate attributes.

The indiscernible attribute classes are obtained by the attribute clustering algorithm. Then, we propose a semi-supervised attribute reduction algorithm based on attribute indiscernibility as shown in Algorithm 2. It mainly contains the following parts. Firstly, for each indiscernible attribute class $IndF_i$, we get the attribute that generates the most discernibility pairs. Secondly, we get a new attribute subset FS (each element in FS is from different indiscernible attribute classes, that is $|FS| = |IndF|$). Finally, we select the attributes with the most discernibility information in the FS.

Algorithm 2 Semi-supervised attribute reduction algorithm based on attribute indiscernibility(Semi-AI)

Input: A partially labeled decision information system $PLDS = < U, C \cup \{d\}, V^*, f >$
Output: A reduct R
for $\forall a_i \in C$ **do**
  Compute $DisP_P(a_i, d)$.
end for
Compute $DisP_P(C, d)$, $Core_I(C)$ and $Core_D(C \cup \{d\})$.
Set $R = \varnothing$, $R := R \cup Core_I(C) \cup Core_D(C \cup \{d\})$, $C := C - [Core_I(C) \cup Core_D(C \cup \{d\})]$.
for $\forall a, b \in C$ **do**
  Compute $Disatt(a, b, d)$.
end for
Get the indiscernible attribute class $IndF(IndF_1, IndF_2, \dots IndF_m, m < D)$.
Set $FS = \varnothing, count = 1, maxnum = 0, selAtt = 0$.
while $count \leqslant m$ **do**
  for $a_i \in IndF_{count}$ **do**
    if $|DisP_P(a_i, d)| > maxnum$ **then**
      $maxnum = |DisP_P(a_i, d)|$;
      $selAtt = a_i$.
    end if
  end for
  $FS := FS \cup \{selAtt\}$, $count = count + 1$.
end while
Set $maxnum = 0, selAtt = 0$.
while $|DisP_P(R, d)| < |DisP_P(C, d)|$ and $FS \neq \varnothing$ **do**
  for $a_j \in FS - R$ **do**
    if $|DisP_P(R \cup \{a_j\}, d)| > maxnum$ **then**
      $maxnum = |DisP_P(R \cup \{a_j\}, d)|$;
      $selAtt = a_j$.
    end if
  end for
  $R := R \cup \{selAtt\}$.
end while
Set $maxnum = 0, selAtt = 0$.
while $|DisP_P(R, d)| < |DisP_P(C, d)|$ **do**
  for $a_k \in C - R$ **do**
    if $|DisP_P(R \cup \{a_k\}, d)| > maxnum$ **then**
      $maxnum = |DisP_P(R \cup \{a_k\}, d)|$;
      $selAtt = a_k$.
    end if
  end for
  $R := R \cup \{selAtt\}$.
end while
for $\forall a \in R$ **do**
  if $DisP_P(R - \{a\}, d) = DisP_P(R, d)$ **then**
    $R := R - \{a\}$.
  end if
end for
Return R.

Table 2 The second partially labeled decision information system

U	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	1	1	2	1	1	2	d_1
x_2	2	1	1	2	1	2	d_2
x_3	1	1	2	1	1	2	d_1
x_4	2	2	2	1	2	1	d_3
x_5	2	2	2	1	2	2	*
x_6	2	1	1	2	1	1	*
x_7	1	1	1	2	2	1	*

Semi-AI considers not only the indiscernibility relationship between objects, but also the indiscernibility relationship between attributes. In order to better understand the algorithm Semi-AI, we give an example as shown in Table 2. We only need to calculate the upper or lower triangle of the matrix because of the irreflexivity and symmetry of discernibility pairs.

Example 2 Considering a partially labeled decision information system $PLDS =< U, C \cup \{d\}, V^*, f >$ shown in Table 2, $U = \{x_1, x_2, \dots x_7\}$ is the object set, $C = \{a_1, a_2, \dots a_6\}$ is the attribute set and d is the decision attribute.

(1) We calculate the discernibility pair sets defined in Definition 13.

$$\begin{aligned}
DisP_P(C,d) = \{&(x_1,x_2),(x_1,x_4),(x_2,x_3),(x_2,x_4),\\
&(x_3,x_4),(x_5,x_6),(x_5,x_7),(x_6,x_7)\};\\
DisP_P(a_1,d) = \{&(x_1,x_2),(x_1,x_4),(x_2,x_3),(x_3,x_4),\\
&(x_5,x_7),(x_6,x_7)\};\\
DisP_P(a_2,d) = \{&(x_1,x_4),(x_2,x_4),(x_3,x_4),(x_5,x_6),\\
&(x_5,x_7)\};\\
DisP_P(a_3,d) = \{&(x_1,x_2),(x_2,x_3),(x_2,x_4),(x_5,x_6),(x_5,x_7)\};\\
DisP_P(a_4,d) = \{&(x_1,x_2),(x_2,x_3),(x_2,x_4),(x_5,x_6),(x_5,x_7)\};\\
DisP_P(a_5,d) = \{&(x_1,x_4),(x_2,x_4),(x_3,x_4),(x_5,x_6),(x_6,x_7)\};\\
DisP_P(a_6,d) = \{&(x_1,x_4),(x_2,x_4),(x_3,x_4),(x_5,x_6),(x_5,x_7)\};\\
&Core_I(C) = \emptyset, Core_D(C \cup \{d\}) = \emptyset.
\end{aligned}$$

(2) We calculate the distinction between attributes according to Definition 16 and obtain the distinction matrix DtM.

$$DtM = \begin{pmatrix}
0.0000 & 0.6250 & 0.6250 & 0.6250 & 0.6250 & 0.6250\\
0.6250 & 0.0000 & 0.5714 & 0.5714 & 0.3333 & 0.0000\\
0.6250 & 0.5714 & 0.0000 & 0.0000 & 0.7500 & 0.5714\\
0.6250 & 0.5714 & 0.0000 & 0.0000 & 0.7500 & 0.5714\\
0.6250 & 0.3333 & 0.7500 & 0.7500 & 0.0000 & 0.3333\\
0.6250 & 0.0000 & 0.5714 & 0.5714 & 0.3333 & 0.0000
\end{pmatrix} \tag{16}$$

(3) We set $k = 3$. By using the attribute clustering algorithm to the cluster attribute set $C = \{a_1, a_2, a_3, a_4, a_5, a_6\}$, we obtain the indiscernible attribute classes $\{a_2, a_6, a_5\}$, $\{a_3, a_4\}$ and $\{a_1\}$.

(4) We set $FS = \emptyset$ and select attribute a_k from each indiscernible attribute class by calculating $DisP_P(a_k, d)$ to represent the indiscernible attribute class. For $\{a_2, a_6, a_5\}$, we can get $|DisP_P(a_2, d)| = 5$, $|DisP_P(a_6, d)| = 5$ and $|DisP_P(a_5, d)| = 5$. We choose attribute a_2 as the representative attribute of $\{a_2, a_6, a_5\}$, and set $FS = FS \cup \{a_2\}$. For $\{a_3, a_4\}$, we can get $|DisP_P(a_3, d)| = 5$ and $|DisP_P(a_4, d)| = 5$. We choose attribute a_3 as the representative attribute of $\{a_3, a_4\}$, and set $FS = FS \cup \{a_3\}$. For $\{a_1\}$, we can get $|DisP_P(a_1, d)| = 6$. We choose attribute a_1 as the representative attribute of $\{a_1\}$, and set $FS = FS \cup \{a_1\}$. Finally, a new attribute subset $FS = \{a_2, a_3, a_1\}$ is obtained.

(5) In the first round, we set $R = \emptyset$. For any $a_j \in FS - R$ and calculate the number of discernibility pairs $|DisP_P(a_j, d)|$ as follows: $|DisP_P(a_2, d)| = 5$, $|DisP_P(a_3, d)| = 5$, $|DisP_P(a_1, d)| = 6$. We can conclude that attribute a_1 should be selected, $R = R \cup \{a_1\}$, and $|DisP_P(R, d)| < |DisP_P(C, d)|$.

(6) In the second round, we can get $|DisP_P(R \cup \{a_2\}, d)| = 8$ and $|DisP_P(R \cup \{a_3\}, d)| = 8$. We can conclude that attribute a_2 should be selected, $R = R \cup \{a_2\}$, and $|DisP_P(R, d)| = |DisP_P(C, d)|$.

(7) Finally, because $|DisP_P(R - a_1, d)| < |DisP_P(R, d)|$ and $|DisP_P(R - a_2, d)| < |DisP_P(R, d)|$, there are no redundant attributes in R. We get the attribute subset $R = \{a_1, a_2\}$.

4 Experiments

In this section, several experiments are performed to verify the performance of the proposed algorithms. The data sets used in the experiments are from the UCI Machine Learning Database [27]. The types of attribute values used in the experiments are nominal and numeric, and we adopt the equal width operation to discretize the attribute value into four values for numeric data. The characteristics of the data sets are summarized in Table 3.

In order to get partially labeled decision information systems, decision attribute values of original complete data sets have been randomly selected to delete. Then every processed data set is divided into a labeled part and an unlabeled part, and decision attribute values in the unlabeled part are denoted as $*$.

In order to verify the effectiveness of proposed algorithms in processing symbolic data, we compare our algorithms with some classical supervised attribute reduction algorithms, unsupervised attribute reduction algorithm and semi-supervised attribute reduction algorithms, listed as follows:

1. *Positive region POS.* An attribute reduction algorithm based on attribute dependence [28].
2. *Conditional entropy* H(D|A). An attribute reduction algorithm based on conditional entropy [29].

3. *LDP.* An attribute reduction algorithm for labeled data based on the maximum discernibility pair [25].
4. *UDP.* An attribute reduction algorithm for unlabeled data based on the maximum discernibility pair [25].
5. *DualPOS.* A semi-supervised attribute reduction algorithm using dual positive regions [19].
6. *Semi-mRMR.* A semi-supervised attribute reduction algorithm based on maximum relevance and minimum redundancy [30].
7. *Semi-rough-P.* A semi-supervised attribute reduction algorithm that processes labeled data with dependence and unlabeled data with discernibility pair [25].
8. *Semi-rough-D.* A semi-supervised attribute reduction algorithm that processes labeled data with discernibility pair and unlabeled data with discernibility pair [25].

The nearest neighbor classifier (KNN, K = 5) and decision tree classifier (CART) are used to evaluate the attribute reduction results. By using ten-fold cross-validation, the higher the accuracies under KNN classifier and CART classifier are, the more reasonable the selected attribute subset is.

4.1 Compared with supervised and unsupervised algorithms

In this section, the proposed algorithms are compared with classical supervised and unsupervised algorithms mentioned above. For the data sets used in the experiment, the missing rate is 20%.

The specific experimental results are shown in Tables 4 and 5. The optimal classification accuracy of each data set under different algorithms is shown in bold. In most cases, Semi-AI achieves the best accuracy compared with both supervised and unsupervised algorithms. The average accuracy of Semi-DP on all data sets is 0.49% higher than the average accuracy of the other algorithms by KNN classifier through ten-fold cross-validation. The average accuracy of Semi-AI on all data sets is 3.69% higher than the algorithm with the highest average accuracy on given data sets and 15.67% higher than the lowest average accuracy on given data sets by KNN classifier through ten-fold cross-validation. The average accuracy of Semi-AI on all data sets is 2.00% higher than the algorithm with the highest average accuracy on given data sets and 12.17% higher than the lowest average accuracy on given data sets by CART classifier through ten-fold cross-validation. Unsupervised algorithms can only process 20% of the data and ignore the remaining 80% of the data. Supervised algorithms can only process 80% of the data and ignore the remaining 20% of the data. The proposed algorithms take both labeled data and unlabeled data into consideration and the reducts can better describe the information contained in partially labeled data. From Tables 4 and 5, it can be seen that Semi-DP and Semi-AI can deal with the problem caused by missing labels well.

Table 3 A description of datasets

Index	Dataset	Objects	Attributes	Data type
1	Lung-cancer	31	55	Nominal
2	PersonGait	48	321	Numeric
3	SCADI	70	205	Nominal
4	Olitos	120	25	Numeric
5	Yale	165	1024	Numeric
6	Wine	178	13	Numeric
7	Sonar	208	60	Numeric
8	FeatMIAS	322	280	Numeric
9	Completed	390	38	Numeric
10	ORL	400	1024	Numeric
11	Musk2	707	166	Numeric
12	Hillvalley	1212	100	Numeric
13	Handwritten	1593	257	Nominal
14	Kr-vs-kp	3196	36	Nominal
15	Spambase	4601	57	Numeric

4.2 Compared with semi-supervised algorithms

In this section, the proposed algorithms are compared with classical semi-supervised algorithms mentioned above. For the data sets used in the experiment, the missing rate is 20%.

The specific experimental results are shown in Tables 6 and 7. The optimal classification accuracy of each data set under different algorithms is shown in bold. In most cases, Semi-AI achieves the best accuracy compared with semi-supervised algorithms. The average accuracy of Semi-DP on all data sets is 3.32% higher than the average accuracy of the other algorithms by KNN classifier through ten-fold cross-validation. The average accuracy of Semi-DP on all data sets is 2.09% higher than the average accuracy of the other algorithms by CART classifier through ten-fold cross-validation. The average accuracy of Semi-AI on all data sets is 5.04% higher than the algorithm with the highest average accuracy on given data sets and 18.94% higher than the algorithm with the lowest average accuracy on given data sets by KNN classifier through ten-fold cross-validation. The average accuracy of Semi-AI on all data sets is 5.32% higher than the algorithm with the highest average accuracy on given data sets and 14.00% higher than the algorithm with the lowest average accuracy on given data sets by CART classifier through ten-fold cross-validation. The reason why the accuracies of the proposed algorithm on given data sets are higher than the other algorithms is that the other algorithms only consider the relationships between objects, but rarely consider the relationships between attributes. From Tables 6

Table 4 Accuracy results (*mean*% ± *std*) under KNN classifier with 20% missing labels compared with supervised and unsupervised algorithms

Dataset	Size of k	Semi-DP	Semi-AI	POS	H(D A)	LDP	UDP
Lung-cancer	12	48.06 ± 5.37	**76.13 ± 1.67**	55.81 ± 3.42	62.90 ± 3.48	48.06 ± 4.67	40.00 ± 4.61
PersonGait	149	45.63 ± 1.54	**71.04 ± 2.07**	41.06 ± 1.44	48.33 ± 1.91	45.63 ± 1.54	16.46 ± 2.29
SCADI	133	79.57 ± 0.69	**87.14 ± 0.00**	68.29 ± 1.88	85.71 ± 0.00	85.29 ± 0.96	59.86 ± 1.71
Olitos	4	68.67 ± 0.90	**72.83 ± 1.89**	64.00 ± 1.10	64.83 ± 2.80	72.67 ± 2.25	56.83 ± 2.07
Yale	166	34.79 ± 1.08	**42.97 ± 1.01**	36.00 ± 1.38	41.64 ± 1.11	35.21 ± 1.65	23.82 ± 1.21
Wine	4	87.13 ± 0.86	**93.43 ± 0.96**	88.43 ± 1.10	91.91 ± 0.71	90.22 ± 0.39	89.21 ± 0.52
Sonar	47	72.45 ± 1.42	76.06 ± 1.52	74.86 ± 1.04	**78.08 ± 0.99**	73.22 ± 1.22	61.49 ± 0.95
FeatMIAS	80	63.26 ± 1.27	**69.66 ± 1.38**	66.06 ± 1.17	66.27 ± 1.15	63.26 ± 1.27	54.32 ± 1.22
Completed	33	82.10 ± 0.51	83.46 ± 0.72	**84.56 ± 0.47**	82.95 ± 0.39	81.21 ± 0.53	82.26 ± 0.65
ORL	928	35.05 ± 0.81	**52.47 ± 0.58**	45.00 ± 0.68	50.32 ± 0.46	46.87 ± 0.88	22.22 ± 0.75
Musk2	161	87.51 ± 0.48	**90.52 ± 0.63**	87.62 ± 0.43	89.92 ± 0.40	88.61 ± 0.52	88.49 ± 0.52
Hillvalley	44	50.72 ± 0.19	51.07 ± 0.32	50.26 ± 0.41	**51.08 ± 0.22**	50.23 ± 0.34	50.05 ± 0.34
Handwritten	220	69.97 ± 0.42	73.94 ± 0.33	71.43 ± 0.63	**73.99 ± 0.32**	69.65 ± 0.45	60.66 ± 0.58
Kr-vs-kp	5	94.34 ± 0.15	94.34 ± 0.15	91.66 ± 0.28	91.90 ± 0.30	92.17 ± 0.10	**94.40 ± 0.18**
Spambase	3	**55.06 ± 0.10**	**55.06 ± 0.10**	**55.06 ± 0.10**	**55.06 ± 0.10**	**55.06 ± 0.10**	54.88 ± 0.10
Avg	–	64.95	**72.68**	65.34	68.99	66.49	57.01

and 7, it can be seen that Semi-DP and Semi-AI are effective in dealing with the problem caused by missing labels.

4.3 Classification performance of the selected attribute subset

To further demonstrate the discernibility ability of attribute subset selected by Semi-AI, we show the intuitive results of the data sets Wine and Sonar through two-dimensional feature space in Figs. 4 and 5. Subgraph (a) and Subgraph (b) show the data distribution induced by the first two selected attributes and the last two selected attributes, respectively. Subgraph (c) and Subgraph (d) show the data distribution induced by two random attributes. As we can see from the figures, the attributes selected by Semi-AI achieve better discernibility ability than randomly selected attributes from original data.

Table 5 Accuracy results (*mean*% ± *std*) under CART classifier with 20% missing labels compared with supervised and unsupervised algorithms

Dataset	Size of k	Semi-DP	Semi-AI	POS	H(D A)	LDP	UDP
Lung-cancer	54	40.32 ± 4.37	**72.90 ± 4.35**	70.00 ± 3.74	65.81 ± 4.35	37.42 ± 6.31	44.52 ± 3.33
PersonGait	242	44.17 ± 2.74	**47.29 ± 4.19**	38.75 ± 1.10	46.88 ± 2.74	44.17 ± 2.24	15.83 ± 2.45
SCADI	134	79.00 ± 0.69	**82.29 ± 1.54**	74.57 ± 1.31	78.86 ± 0.90	80.86 ± 1.38	68.14 ± 1.18
Olitos	11	57.25 ± 3.77	**66.08 ± 2.55**	60.25 ± 3.12	58.08 ± 2.78	63.75 ± 2.76	57.50 ± 2.58
Yale	23	35.94 ± 1.98	40.24 ± 1.72	**44.79 ± 2.56**	38.67 ± 2.49	37.70 ± 1.78	20.30 ± 2.31
Wine	4	84.33 ± 0.77	**91.29 ± 0.89**	88.99 ± 1.22	90.51 ± 1.17	85.34 ± 1.90	89.72 ± 1.45
Sonar	6	76.25 ± 1.25	**78.22 ± 1.92**	73.27 ± 1.44	75.58 ± 1.65	74.47 ± 1.29	59.09 ± 1.99
FeatMIAS	80	57.70 ± 1.39	**67.67 ± 1.59**	66.02 ± 1.85	65.25 ± 1.42	57.70 ± 1.39	51.21 ± 1.67
Completed	7	79.49 ± 1.59	**83.31 ± 1.25**	81.72 ± 1.09	82.41 ± 1.61	79.56 ± 1.48	80.79 ± 0.96
ORL	928	37.17 ± 1.50	**46.57 ± 0.91**	40.80 ± 1.26	40.67 ± 1.20	37.87 ± 0.87	28.10 ± 0.81
Musk2	80	88.76 ± 0.74	89.86 ± 0.64	86.00 ± 0.74	**90.41 ± 0.85**	87.79 ± 1.02	85.71 ± 1.39
Hillvalley	44	50.44 ± 0.46	51.64 ± 0.55	50.14 ± 0.46	**51.70 ± 0.33**	51.34 ± 0.66	49.88 ± 0.28
Handwritten	221	63.66 ± 0.66	68.38 ± 0.85	66.25 ± 0.61	**70.98 ± 0.67**	66.47 ± 0.86	52.54 ± 0.66
Kr-vs-kp	5	**99.12 ± 0.13**	99.10 ± 0.15	99.04 ± 0.10	99.10 ± 0.10	99.07 ± 0.11	99.03 ± 0.13
Spambase	3	71.69 ± 0.11	**71.70 ± 0.10**	71.67 ± 0.11	**71.70 ± 0.10**	71.67 ± 0.11	71.68 ± 0.11
Avg	–	64.35	**70.44**	67.48	68.44	65.01	58.27

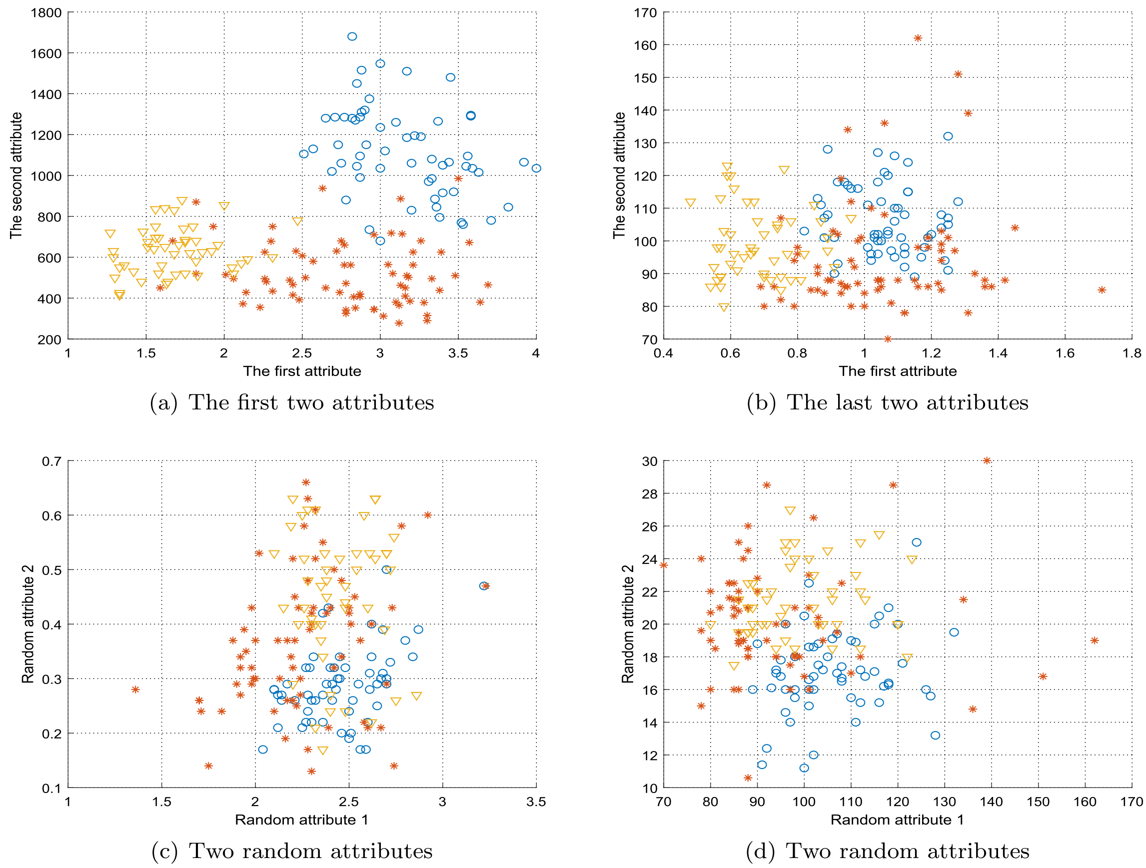
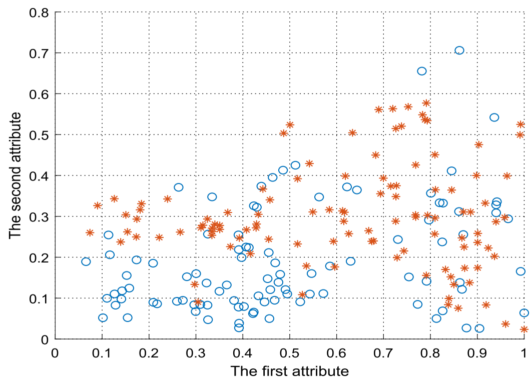


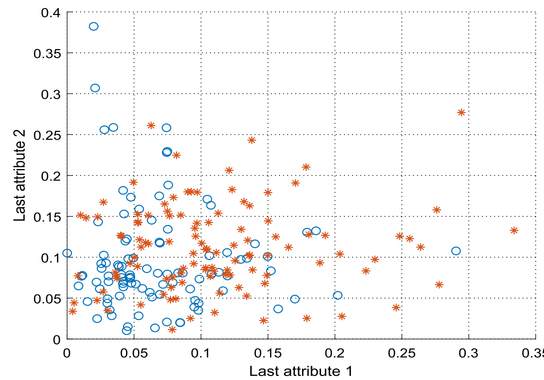
Fig. 4 Intuitive results of the data set Wine

Table 6 Accuracy results (*mean*% ± *std*) under KNN classifier with 20% missing labels compared with semi-supervised algorithms

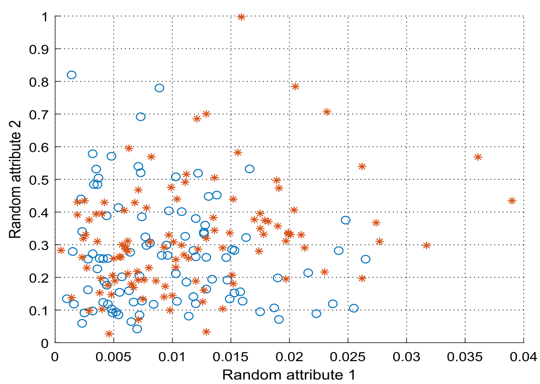
Dataset	Size of k	Semi-DP	Semi-AI	DaulPOS	Semi-mRMR	Semi-rough-P	Semi-rough-D
Lung-cancer	12	48.03 ± 5.37	**76.13 ± 1.67**	54.19 ± 3.66	36.77 ± 2.72	48.71 ± 4.67	42.26 ± 1.83
PersonGait	149	45.63 ± 1.54	**71.04 ± 2.07**	57.50 ± 1.46	31.25 ± 0.00	19.58 ± 1.76	42.92 ± 2.24
SCADI	133	79.57 ± 0.69	**87.14 ± 0.00**	84.29 ± 0.95	53.71 ± 1.00	67.14 ± 3.09	77.00 ± 1.42
Olitos	4	68.67 ± 0.90	**72.83 ± 1.89**	66.08 ± 1.52	51.17 ± 2.49	65.17 ± 2.63	68.75 ± 2.81
Yale	166	34.79 ± 1.08	42.97 ± 1.01	**47.03 ± 1.46**	27.27 ± 0.57	33.88 ± 1.47	24.61 ± 1.15
Wine	4	87.13 ± 0.86	**93.43 ± 0.96**	91.29 ± 0.85	79.61 ± 1.61	87.87 ± 0.71	87.13 ± 0.86
Sonar	47	72.45 ± 1.42	76.06 ± 1.52	**80.43 ± 1.22**	66.20 ± 0.96	69.33 ± 1.50	66.83 ± 0.75
FeatMIAS	80	63.26 ± 1.27	**69.66 ± 1.38**	66.30 ± 0.81	62.55 ± 1.09	63.82 ± 0.78	60.90 ± 0.98
Completed	33	82.10 ± 0.51	83.46 ± 0.72	81.79 ± 0.00	77.21 ± 0.60	**83.54 ± 0.67**	82.41 ± 0.78
ORL	928	35.05 ± 0.81	**52.47 ± 0.58**	44.35 ± 0.94	7.75 ± 0.24	40.72 ± 0.58	36.47 ± 0.62
Musk2	161	87.51 ± 0.48	**90.52 ± 0.63**	80.89 ± 0.61	84.58 ± 0.49	88.78 ± 0.49	87.54 ± 0.44
Hillvalley	44	50.72 ± 0.19	**51.07 ± 0.32**	50.03 ± 0.25	50.91 ± 0.00	50.36 ± 0.33	50.77 ± 0.37
Handwritten	220	69.97 ± 0.42	**73.94 ± 0.33**	67.58 ± 0.57	50.49 ± 0.55	70.23 ± 0.52	63.77 ± 0.32
Kr-vs-kp	5	**94.34 ± 0.15**	**94.34 ± 0.15**	92.23 ± 0.23	87.11 ± 0.25	93.34 ± 0.18	93.34 ± 0.18
Spambase	3	**55.06 ± 0.10**	**55.06 ± 0.10**	50.22 ± 0.02	44.25 ± 0.06	**55.06 ± 0.10**	**55.06 ± 0.10**
Avg	–	64.95	**72.68**	67.64	53.74	62.50	62.65



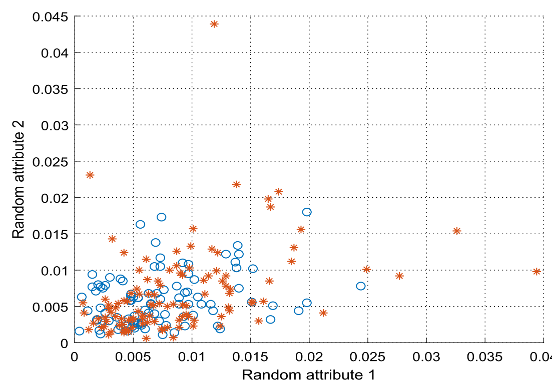
(a) The first two attributes



(b) The last two attributes



(c) Two random attributes



(d) Two random attributes

Fig. 5 Intuitive results of the data set Sonar

4.4 Robustness of the proposed method

In this section, we show the robustness of Semi-AI by setting different missing rates (10%, 20%, ⋯, 50%) of datasets. As shown in Figs. 6, 7, 8 and 9, the X-axis shows the different missing rates and the Y-axis represents the classification accuracy under certain classifier. Compared with the representative supervised, unsupervised and semi-supervised

Table 7 Accuracy results (*mean*% ± *std*) under CART classifier with 20% missing labels compared with semi-supervised algorithms

Dataset	Size of k	Semi-DP	Semi-AI	DaulPOS	Semi-mRMR	Semi-rough-P	Semi-rough-D
Lung-cancer	54	40.32 ± 4.37	**72.90 ± 4.35**	62.26 ± 4.57	37.74 ± 8.34	50.00 ± 4.09	69.03 ± 3.12
PersonGait	242	44.17 ± 2.74	**47.29 ± 4.19**	36.67 ± 2.45	26.04 ± 1.10	37.50 ± 2.95	37.05 ± 3.23
SCADI	134	79.00 ± 0.69	**82.29 ± 1.54**	75.71 ± 1.65	58.86 ± 2.59	75.57 ± 0.81	78.00 ± 2.45
Olitos	11	57.25 ± 3.77	**66.08 ± 2.55**	62.00 ± 3.73	48.67 ± 2.49	61.58 ± 3.13	58.83 ± 2.16
Yale	23	35.94 ± 1.98	40.24 ± 1.72	**42.67 ± 1.70**	25.15 ± 1.41	30.61 ± 2.16	25.82 ± 2.22
Wine	4	84.33 ± 0.77	**91.29 ± 0.89**	89.33 ± 1.32	82.87 ± 0.85	88.31 ± 1.40	84.33 ± 0.77
Sonar	6	76.25 ± 1.25	**78.22 ± 1.92**	75.48 ± 2.23	61.87 ± 2.04	69.95 ± 1.97	62.93 ± 2.69
FeatMIAS	80	57.70 ± 1.39	**67.67 ± 1.59**	61.55 ± 1.86	64.88 ± 1.94	61.80 ± 1.79	61.58 ± 1.50
Completed	7	79.49 ± 1.59	**83.31 ± 1.25**	82.03 ± 0.08	76.44 ± 0.88	81.85 ± 0.86	80.44 ± 1.05
ORL	928	37.17 ± 1.50	**46.57 ± 0.91**	33.80 ± 1.65	13.82 ± 0.26	30.77 ± 1.56	30.82 ± 1.41
Musk2	80	88.76 ± 0.74	**89.86 ± 0.64**	82.70 ± 0.63	84.61 ± 0.81	88.23 ± 0.83	87.79 ± 1.02
Hillvalley	44	50.44 ± 0.46	51.64 ± 0.55	51.24 ± 0.25	**51.65 ± 0.00**	50.73 ± 0.44	50.79 ± 0.46
Handwritten	221	63.66 ± 0.66	**68.38 ± 0.85**	58.56 ± 0.66	54.78 ± 0.73	61.37 ± 0.78	55.05 ± 0.70
Kr-vs-kp	5	99.12 ± 0.13	99.10 ± 0.15	96.31 ± 0.13	95.26 ± 0.21	99.11 ± 0.13	**99.14 ± 0.13**
Spambase	3	71.69 ± 0.11	71.70 ± 0.10	66.55 ± 0.02	64.03 ± 0.00	71.69 ± 0.11	**71.71 ± 0.11**
Avg	–	64.35	**70.44**	65.12	56.44	63.94	63.56

(a) lung-cancer (b) PersonGait (c) SCADI

(d) olitos (e) Yale (f) wine

(g) sonar (h) featMIAS (i) completed

(j) ORL (k) Musk2 (l) hillvalley

(m) handwritten (n) kr-vs-kp (o) spambase

Fig. 6 Classification accuracy line charts (5NN) compared with supervised and unsupervised algorithms

(a) lung-cancer (b) PersonGait (c) SCADI

(d) olitos (e) Yale (f) wine

(g) sonar (h) featMIAS (i) completed

(j) ORL (k) Musk2 (l) hillvalley

(m) handwritten (n) kr-vs-kp (o) spambase

Fig. 7 Classification accuracy line charts (CART) compared with supervised and unsupervised algorithms

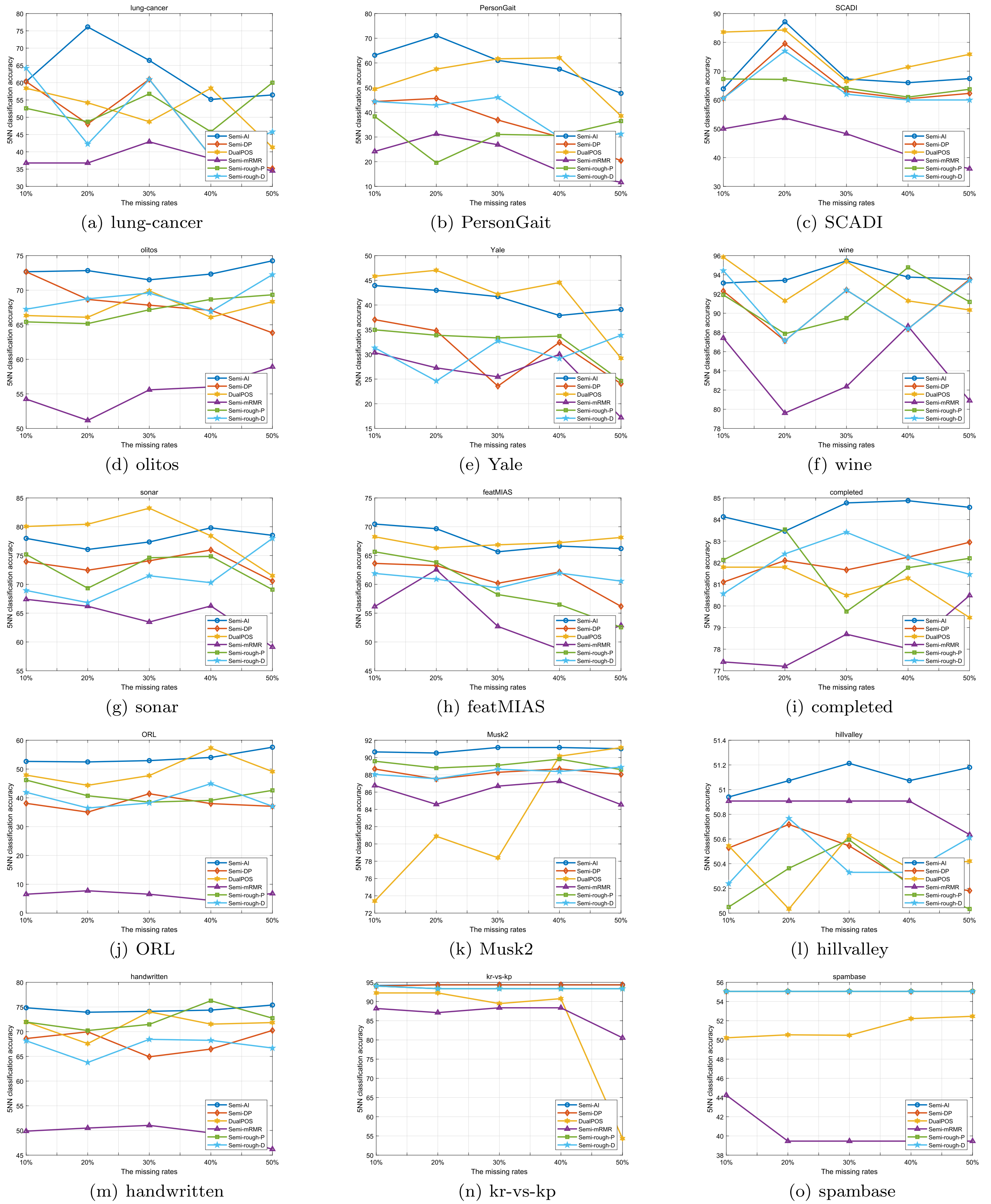


Fig. 8 Classification accuracy line charts (5NN) compared with semi-supervised algorithms

Fig. 9 Classification accuracy line charts (CART) compared with semi-supervised algorithms

algorithms, Semi-AI has achieved better classification accuracy results on most datasets. Moreover, Semi-AI has better adaptability to different missing rates, that is, it has better stability. Therefore, we can conclude that the proposed algorithm is effective and robust.

5 Conclusion

The work of this paper mainly has the following two contributions. On the one hand, we define discernibility pair set in partially labeled decision information system as a criterion for quantifying the discernibility ability of attributes. This provides researchers with a more intuitive perspective to evaluate the amount of information contained in attributes. On the other hand, we propose new fuzzy similarity relation and fuzzy discernibility relation between attributes by discernibility pairs. Further, we construct the concept of indiscernible attribute class which describes an indiscernibility relationship among attributes. This provides a good direction for researchers to explore the relationships between attributes. Experimental results demonstrate the effectiveness of the proposed algorithms compared with other representative algorithms.

In the future, there are two main directions worth exploring. On the one hand, influenced by the attribute clustering algorithm [13], the acquisition of the optimal indiscernible attribute class defined in this paper has the problem of high time complexity. Inspired by the gap neighborhood relation proposed by Zhou et al. [31] and the granular-ball rough set model proposed by Xia et al. [32, 33], we hope to explore an attribute reduction method in terms of relationships between attributes with adaptive neighborhood radius and low time complexity. On the other hand, since discernibility pairs can be regarded as the criterion for quantifying the discernibility ability of attributes, we hope to explore an attribute reduction method by combining discernibility pair with mutual information.

Acknowledgements This work is supported by the National Natural Science Foundation of China (61976089), the Major Program of the National Social Science Foundation of China (20&ZD047), the Natural Science Foundation of Hunan Province (2021JJ30451, 2022JJ30397), and the Hunan Provincial Science & Technology Project Foundation (2018RS 3065, 2018TP1018).

References

1. Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11(5):341–356
2. Pawlak Z (1991) Rough sets—theoretical aspects of reasoning about data. Kluwer, Dordrecht
3. Zhang C, Dai J, Chen J (2020) Knowledge granularity based incremental attribute reduction for incomplete decision systems. Int J Mach Learn Cybern 11(5):1141–1157
4. Wang C, He Q, Shao M, Hu Q (2018) Feature selection based on maximal neighborhood discernibility. Int J Mach Learn Cybern 9(11):1929–1940
5. Chen D, Zhao S, Zhang L, Yang Y, Zhang X (2012) Sample pair selection for attribute reduction with rough set. IEEE Trans Knowl Data Eng 24(11):2080–2093
6. Dai J, Hu H, Wu W, Qian Y, Huang D (2018) Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets. IEEE Trans Fuzzy Syst 26(4):2174–2187
7. Dai J, Xu Q (2013) Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification. Appl Soft Comput 13(1):211–221
8. Wang J, Wei J, Yang Z, Wang S (2017) Feature selection by maximizing independent classification information. IEEE Trans Knowl Data Eng 29(4):828–841
9. Susmaga R (2004) Reducts and constructs in attribute reduction. Fundam Inf 61(2):159–181
10. Qin K, Jing S (2017) The attribute reductions based on indiscernibility and discernibility relations. Lect Notes Comput Sci 10313:306–316
11. Dai J, Chen J, Liu Y, Hu H (2020) Novel multi-label feature selection via label symmetric uncertainty correlation learning and feature redundancy evaluation. Knowl Based Syst 207:106342
12. Qian J, Miao D, Zhang ZH, Li W (2011) Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation. Int J Approx Reason 52(2):212–230
13. Mitra P, Murthy CA, Pal SK (2002) Unsupervised feature selection using feature similarity. IEEE Trans Pattern Anal Mach Intell 24(3):301–312
14. Rezaei M, Cribben I, Samorani M (2021) A clustering-based feature selection method for automatically generated relational attributes. Ann Oper Res 303(1):233–263
15. Restrepo M, Cornelis C (2021) Attribute reduction using functional dependency relations in rough set theory. Lect Notes Comput Sci 12872:90–96
16. Jia X, Rao Y, Shang L, Li T (2020) Similarity-based attribute reduction in rough set theory: a clustering perspective. Int J Mach Learn Cybern 11(5):1047–1060
17. Kudo Y, Murai T (2012) Indiscernibility relations by interrelationships between attributes in rough set data analysis. In: Proceedings of 2012 IEEE international conference on granular computing, Hangzhou, China, August 11–13, p 220–225
18. Dai J, Liu Q (2022) Semi-supervised attribute reduction for interval data based on misclassification cost. Int J Mach Learn Cybern 13(6):1739–1750
19. Dai J, Han H, Hu H, Hu Q, Zhang J, Wang W (2016) Dualpos: a semi-supervised attribute selection approach for symbolic data based on rough set theory. Lect Notes Comput Sci 9659:392–402
20. Saha S, Alok AK, Ekbal A (2016) Use of semisupervised clustering and feature-selection techniques for identification of co-expressed genes. IEEE J Biomed Health Inform 20(4):1171–1177
21. Chang X, Yang Y (2017) Semisupervised feature analysis by mining correlations among multiple tasks. IEEE Trans Neural Netw Learn Syst 28(10):2294–2305
22. Mi Y, Quan P, Shi Y, Wang Z (2022) Concept-cognitive computing system for dynamic classification. Eur J Oper Res 301(1):287–299
23. Xu J, Tang B, He H, Man H (2017) Semisupervised feature selection based on relevance and redundancy criteria. IEEE Trans Neural Netw Learn Syst 28(9):1974–1984
24. Mi Y, Liu W, Shi Y, Li J (2022) Semi-supervised concept learning by concept-cognitive learning and concept space. IEEE Trans Knowl Data Eng 34(5):2429–2442
25. Dai J, Hu Q, Zhang J, Hu H, Zheng N (2017) Attribute selection for partially labeled categorical data by rough set approach. IEEE Trans Cybern 47(9):2460–2471

26. Li B, Xiao J, Wang X (2019) Feature selection for partially labeled data based on neighborhood granulation measures. IEEE Access 7:37238–37250
27. Dua D, Graff C (2019) UCI machine learning repository. http://archive.ics.uci.edu/ml
28. Hu X, Cercone N (1995) Learning in relational databases: a rough set approach. Comput Intell 11:323–338
29. Wang GY, Yu H, Yang D (2002) Decision table reduction based on conditional information entropy. Chin J Comput 25(7):759–766
30. Wang W (2014) Semi-supervised clustering and feature selection for symbolic data. MS thesis, Zhejiang University, Hangzhou, China
31. Zhou P, Hu X, Li P, Wu X (2019) Online streaming feature selection using adapted neighborhood rough set. Inf Sci 481:258–279
32. Xia S, Zhang H, Li W, Wang G, Giem E, Chen Z (2022) GBNRS: a novel rough set algorithm for fast adaptive attribute reduction in classification. IEEE Trans Knowl Data Eng 34(3):1231–1242
33. Xia S, Wang C, Wang G, Gao X, Giem E, Yu J (2022) GBRS: an unified model of pawlak rough set and neighborhood rough set. arXiv e-prints