A Test Cost Sensitive Heuristic Attribute Reduction Algorithm for Partially Labeled Data

Shengdan Hu[1,2,3], Duoqian Miao[1,2](✉), Zhifei Zhang[1,4], Sheng Luo[1,2], Yuanjian Zhang[1,2], and Guirong Hu[2]

[1] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
dqmiao@tongji.edu.cn, hushengdan@163.com

[2] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, China

[3] Department of Computer Science, Shanghai Normal University Tianhua College, Shanghai 201815, China

[4] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

Abstract. Attribute reduction is viewed as one of the most important topics in rough set theory and there have been many researches on this issue. In the real world, partially labeled data is universal and cost sensitivity should be taken into account under some circumstances. However, very few studies on attribute reduction for partially labeled data with test cost have been carried out. In this paper, based on mutual information, the significance of an attribute in partially labeled decision system with test cost is defined, and for labeled data, a heuristic attribute reduction algorithm TCSPR is proposed. Experimental results show the impact of test cost on reducts for partially labeled data and comparative experiments of classification accuracy indicate the effectiveness of the proposed method.

Keywords: Attribute reduction · Uncertainty · Rough set
Test cost sensitive · Partially labeled data

1 Introduction

Uncertainty is a common phenomenon in the world. Reasoning and knowledge acquisition with uncertain or incomplete information is always a core subproblem of artificial intelligence. There have been plenty of theories on the problem of uncertainty, for example, probability theory [2], possibility theory [4,5], fuzzy set [3], rough set [6,7], evidence theory [8,9], cloud model [1]. As an extension of set theory, rough set which was proposed by Polish computer scientist Zdzislaw Pawlak [6] in 1982, is a soft computing tool to model imperfect knowledge. In rough set, it is assumed that knowledge is based on the ability to classify

H. S. Nguyen et al. (Eds.): IJCRS 2018, LNAI 11103, pp. 257–269, 2018.
https://doi.org/10.1007/978-3-319-99368-3_20

objects and tabular representation of knowledge is often employed. Uncertainty in this theory is represented by the boundary region of a set, and the boundary region can be specified in terms of a pair of crisp sets which give the lower and the upper approximation of the original set.

Recently, a deluge of data from a variety of sources has reached an unprecedented volume. However, there may exist incomplete data due to various reasons. In decision systems of rough set, some values of the decision attributes may be missing and the systems are actually partially labeled data. This could often occur in reality. For example, for the information system of patients in a hospital, some diagnoses of diseases may be missing due to the patients stop doing further examinations. Knowledge acquisition from partially labeled data is akin to semi-supervised learning which attracts plenty of researchers. To deal with partially labeled data, some methods based on rough set have been proposed [10–13], and incremental methods in dynamic system are studied [14–17].

The existing rough set-based methods for partially labeled data mentioned above seek low classifying error rates or high accuracy and implicitly assume that all classes or features have the same cost, nevertheless, this assumption may not be suitable in real-world scenarios. For example, in a clinical diagnosis system, it may cause some damage to a patient who is misclassified as cancer class, but may result in serious damage if a patient who has cancer is misclassified as non-cancer class and could not get treatment timely. Also, a patient often needs to undertake a number of medical tests, in this case, money and/or time for these tests are regarded as test costs and the costs may be various according to different tests. From these two examples, we can infer that cost sensitivity should be considered in some problems. In the cost sensitive settings, it is aimed to minimize the total cost, rather than simply minimize the error rate. Turney [18] concluded nine types of costs in inductive concept learning and in decision systems, some researchers have done much research on decision cost [19–22] and test cost [23,24]. However, there have been few studies about cost sensitive in decision system with missing decision values. Motivated by these analysis, this paper focuses on tackling the problem of attributes reduction for partially labeled data with test cost sensitive. We first define the significance of an attribute in the partially labeled decision system with test cost based on mutual information. Next, for labeled data, a heuristic algorithm TCSPR for attribute reduction is proposed. Then some attribute reduction experiments are conducted on several data sets to find out the impact of test cost on reducts. In order to verify the effectiveness of the proposed method, the quality of reducts are compared.

The remainder of the paper is organized as follows. Some preliminary concepts and uncertainty measures based on information entropy in rough set are briefly reviewed in Sect. 2. In Sect. 3, the definition of partially labeled decision system with test cost is given and an attribute reduction algorithm TCSPR is proposed. Section 4 illustrates some experiments and results. Section 5 concludes the paper with some discussions.

2 Preliminary

In this section, we present a review of some basic rough set concepts related to this article. One can refer to references [6,7,25] for detail of the theory.

2.1 Rough Set

Definition 1. *An information system is a tuple $IS = (U, A, V, f)$, where $U = \{x_1, x_2, \ldots, x_n\}$ is a finite nonempty set of objects, and $A = \{a_1, a_2, \ldots, a_m\}$ is a finite nonempty set of attributes, V is a nonempty set of values of $a_i \in A (i = 1, 2, \ldots, m)$, $f : U \to V$ is a nonempty set of information functions each of which maps an object in U to a exact value in V.*

If $A = C \cup D$, where C is a set of condition attributes and D is a decision attributes set, the information system is called decision information system or decision table and denoted as $DS = (U, A = C \cup D, V, f)$.

Definition 2. *Let $DS = (U, A = C \cup D, V, f)$ be a decision information system, $B \subseteq A$ be an equivalence relation (also called B-indiscernibility relation), for an arbitrary set $X \subseteq U$, the lower approximation and upper approximation of X with respect to B respectively are defined as:*

$$\underline{B}(X) = \{x \in U | [x]_B \subseteq X\},$$
$$\overline{B}(X) = \{x \in U | [x]_B \cap X \neq \emptyset\},$$

where $[x]_B$ is the equivalence class including x with respect to B, and $[x]_B = \{y \in U | f(x, a) = f(y, a), \forall a \in B\}$. If $\underline{B}(X) = \overline{B}(X)$, X is B-definable, and if $\underline{B}(X) \neq \overline{B}(X)$, X is rough with respect to B.

Definition 3. *Let $DS = (U, A = C \cup D, V, f)$ be a decision information system, and the objects in U are partitioned into r disjoint crisp subsets by decision attributes set D, namely, $U/D = \{D_1, D_2, \ldots, D_r\}$, then C-positive region of D is defined as:*

$$POS_C(D) = \bigcup_{i=1}^{r} \underline{C}(D_i),$$

and the boundary region of D w.r.t. C is defined as:

$$BN_C(D) = \bigcup_{i=1}^{r} \overline{C}(D_i) - \bigcup_{i=1}^{r} \underline{C}(D_i).$$

For any $B \subseteq A$ and $X \subseteq U$, the positive region $POS_B(X)$ is the collection of the objects that can be certainly classified as members of X with respect to relation B. The boundary region $BN_B(X)$, in a sense, is the undecidable area of the universe and none of the objects in this region can be certainly classified into X or $\sim X$. In rough set theory, uncertainty can be represented by the boundary region of a set.

2.2 Uncertainty Measure Based on Entropy and Reduct

In rough set theory, there are some algebraic measurement methods to express the inexactness of object or set, such as accuracy, roughness, attribute dependency degree. Inspired by Shannon's information entropy, Miao gave the information representation of the concepts and operations about rough set theory, and proposed the heuristic reduction algorithm based on mutual information [26].

Definition 4. *Let $DS = (U, A = C \cup D, V, f)$ be a decision information system, $B \subseteq A$ and the objects in U are partitioned into m disjoint crisp subsets $\{B_1, B_2, \ldots, B_m\}$ by B, then rough entropy of B is defined as:*

$$H(B) = -\sum_{i=1}^{m} \frac{|B_i|}{|U|} log_2 \frac{|B_i|}{|U|},$$

where $|B|$ denotes the cardinality of B, and $\sum_{i=1}^{m} |B_i| = |U|$ holds.

Definition 5. *Let $DS = (U, A = C \cup D, V, f)$ be a decision information system, $U/C = \{X_1, X_2, \cdots, X_m\}$ and $U/D = \{Y_1, Y_2, \cdots, Y_n\}$, the entropy of D conditioned on C is defined as:*

$$H(D|C) = -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \sum_{j=1}^{n} \frac{|X_i \cap Y_j|}{|X_i|} log_2 \frac{|X_i \cap Y_j|}{|X_i|}. \quad (1)$$

Let $I(x; y)$ be the mutual information of x and y, the increment of mutual information, which is defined as:

$$I(B \cup \{a\}; D) - I(B; D) = H(D|B) - H(D|B \cup \{a\}), \quad (2)$$

can be used to measure the attribute significance.

Definition 6. *Let $DS = (U, A = C \cup D, V, f)$ be a decision information system, $B \subseteq C$, for $\forall a \in C - B$, the significance measure of a on B can be defined by mutual information as:*

$$SGF(a, B, D) = H(D|B) - H(D|B \cup \{a\}).$$

If $B = \emptyset$, the significance measure of a is:

$$SGF(a, D) = H(D) - H(D|\{a\}).$$

$SGF(a, B, D)$ expresses the importance of attribute a to decision D conditioned on the given attributes B.

Reduct is a subset of attributes that maintains some particular properties as the original data. For a given decision table, there may be multiple reducts. Based on the definitions above, relative reduct can be defined as follows.

Definition 7. *Let $DS = (U, A = C \cup D, V, f)$ be a decision information system and $B \subseteq C$, B is a reduct of C relative to D iff:*
(1) $H(D|B) = H(D|C)$;
(2) $\forall a \in B, H(a|B - \{a\}) > 0$.

In a given decision table, the intersection of all attribute reducts is core, and each element of a core should be in every reduct. The core may be an empty set.

2.3 Test Cost Sensitive Rough Set

Definition 8 ([27]). *A test cost sensitive decision system is a tuple $TDS = (U, A = C \cup D, V, f, c)$, where U, A, C, D, V and f have the same meanings as in definition 1, $c : C \rightarrow R^+ \cup \{0\}$ is the test cost function and R^+ is the set of positive real numbers.*

Assuming that the test cost of every attribute is independent, test cost function can be represented by a vector $c = [c(a_1), c(a_2), \cdots, c(a_{|C|})]$, where $c(a_i)(i = 1, 2, \ldots, |C|)$ is the test cost of attribute a_i, and for $\forall B \subseteq C, c(B) = \sum_{a \in B} c(a)$.

3 Attribute Reduction for Partially Labeled Data

3.1 Partially Labeled Decision System

In a partially labeled decision system, some values of the decision attributes are missing. In the light of test cost, partially labeled decision system can be defined as:

Definition 9. *A partially labeled decision system with test cost is a tuple $TPDS = (U = L \cup N, A = C \cup D, V, f, c)$, where U, A, C, D, V, f and c have the same meanings as in definition 8. L denotes the set of labeled objects, and N denotes the set of unlabeled objects.*

Then, we can define the significance of attribute a on B in a partially labeled decision system with test cost as follows:

$$SGF(a, B, D, c(a), \lambda) = (H(D|B) - H(D|B \cup \{a\}))c(a)^{\lambda}, \tag{3}$$

where $H(D|B)$ and $H(D|B \cup \{a\})$ are the entropy of D conditioned on B and $B \cup \{a\}$ respectively, and they can be calculated by equation (1) in which the number of objects $|U|$ should be replaced by the number of labeled objects $|L|$. $c(a)$ is the test cost of attribute a, and $c(a) \geq 0$. λ is a parameter that can adjust the weight of test cost and $\lambda \leq 0$. If $\lambda = 0$, the significance of attribute a on B is based on conditional entropy as shown in definition 6. $c(a_1), c(a_2), \cdots, c(a_{|C|})$ and λ can be specified in real application by domain experts.

3.2 Attribute Reduction Algorithm

It has been proved that finding a minimal reduct of a decision table with exhaustive algorithm is NP-hard in rough set [28], and correspondingly, computing the minimal test cost of a reduct will have the same complexity. Actually, some heuristic algorithms have been proposed, and most of them are greedy.

In this paper, a heuristic algorithm (TCSPR) for attribute reduction of partially labeled data based on test cost sensitive is as Algorithm 1. In the algorithm, based on the objects with labeled, we first find the core of the attributes set. Then in each iterative step of the while loop, after the computation of significance of every attribute in the unselected attributes subset, choose the attribute with highest significance and add it to the reduct set, until the end condition holds. The significance is computed based on Eq. (3).

Algorithm 1. A heuristic attribute reduction algorithm for partially labeled data with test cost sensitive, called TCSPR

Input: $TPDS = (U = L \cup N, A = C \cup D, V, f, c), \lambda$
Output: An attributes subset B as a relative reduct
$U \leftarrow L$;
$B \leftarrow \emptyset$;
for all $a \in C$ **do**
  if $POS_{C-\{a\}}(D)! = POS_C(D)$ **then**
    $B \leftarrow B \bigcup \{a\}$;
  end if
end for
$tempA \leftarrow C - B$;
while $H(D|B)! = H(D|C)$
  for all $a \in tempA$ **do**
    compute $SGF(a, B, D, c(a), \lambda)$;
  end for
  select a' with maximal $SGF(a', B, D, c(a'), \lambda)$;
  $B \leftarrow B \bigcup \{a'\}$;
  $tempA \leftarrow tempA - \{a'\}$;
end while
return B;

3.3 Complexity Analysis

If the core is a reduct of the attributes set, then it is the minimal reduct. Let m be the number of condition attributes, l be the number of labeled objects, namely $m = |C|$, $l = |L|$, the computational complexity of finding the core is $O(ml)$, and this is the best case of finding a reduct. In the worst case, the reduct is the whole condition attributes, correspondingly the computational complexity is $O(m^2 l^2)$.

4 Experiments

In this section, some experiments are conducted on several data sets from UCI repository [29] with the following purposes: (1) to find out the impact of parameter λ on the reducts of partially labeled data, (2) to find out the impact of test cost on the reducts of partially labeled data, (3) to compare the classification accuracy of classifiers trained from partially labeled data.

4.1 Data Sets and Experiment Environment

According to the experimental requirements of attribute reduction, we adopt 4 data sets with task of classification, as shown in Table 1. The datasets are preprocessed as follows: (1) we delete the eleventh attribute of dataset mushroom because of missing attribute values, (2) the continuous attributes in the dataset wine and ionosphere are discreted by Weka using 3 bins. All the attributes are

Table 1. Summary of datasets

Name	#attributes	#objects	#classes
wine	13	178	3
zoo	16	101	7
mushroom	21	8124	2
ionosphere	34	351	2

identified by natural numbers for convenience. All the experiments are implemented in MATLAB on a PC with CPU 2.60 GHz and 4 GB memory.

4.2 Impact of Parameter λ on Reduct

Because there are no existent test costs of datasets in Table 1, we first assume some values of them. The test costs can be produced by many methods, and here we adopt normal distribution with the mean is 0, and the variance is 1. Then we scale the costs between 1 and 100. Let the test costs of all the attributes be the numbers in Table 2.

Table 2. Test costs of datasets

dataset	#attributes	test cost
wine	13	42 49 54 59 64 71 1 61 100 46 93 79 71
zoo	16	61 100 67 41 1 52 52 28 70 55 60 35 31 36 21 34
mushroom	21	48 70 1 54 45 17 32 45 100 86 16 91 52 38 51 36 37 65 63 63 51
ionosphere	34	84 29 69 39 47 39 24 27 51 1 78 54 81 67 74 8 24 43 61 96 48 100 74 97 62 36 8 63 85 48 40 47 21 41

With the test costs in Table 2, we let labeled ratio be 0.2, 0.4, 0.6, 0.8, and 1.0 respectively, the reducts produced by Algorithm 1 with different λ ($\lambda = 0, -0.5, -1, -2, -4$) are shown in Table 3. When $\lambda = 0$, we do not consider the test costs of attributes and the significance of attribute is based on conditional entropy in reality. Here, we assume the objects of different class in the labeled data are of the same proportion as the objects of different class in the whole dataset. When the labeled ratio is 1.0, that is the dataset and there are no unlabeled data.

In the wine and zoo datasets, when the labeled ratio is up to 0.4, the changes of reducts based on different λ ($\lambda = -0.5, -1, -2, -4$) are small. In the mushroom dataset, the core of attributes is $\{5\}$ when the labeled ratio is less than or equal to 0.8, however the core is $\{1, 3, 5, 9, 13, 14\}$ when the labeled ratio is bigger than 0.85, owing to the huge difference between the core, the reducts are very different. In the ionosphere dataset, it seems that the difference between reducts are mainly caused by labeled ratio and λ has tiny impact on the reducts.

Table 3. Reducts on different λ with the same cost

dataset	ratio	core	reduct				
			$\lambda = 0$	$\lambda = -0.5$	$\lambda = -1$	$\lambda = -2$	$\lambda = -4$
wine	0.2	∅	{2,11,13}	{2,7,11,13}	{2,7,11,13}	{1,2,4,7,10}	{1,2,4,7,10}
	0.4	{13}	{2,5,10,11,12,13}	{1,2,7,9,11,13}	{1,2,4,7,9,10,13}	{1,2,3,4,7,8,10,13}	{1,2,3,4,7,8,10,13}
	0.6	{3,13}	{1,3,9,11,13}	{1,3,7,9,11,13}	{1,2,3,7,8,10,13}	{1,2,3,7,8,10,13}	{1,2,3,7,8,10,13}
	0.8	{1,3,13}	{1,3,4,9,11,12,13}	{1,3,7,8,10,11,13}	{1,2,3,4,7,8,10,13}	{1,2,3,4,7,8,10,13}	{1,2,3,4,7,8,10,13}
	1.0	{1,3,13}	{1,3,4,9,11,12,13}	{1,2,3,4,7,8,10,13}	{1,2,3,4,7,8,10,13}	{1,2,3,4,7,8,10,13}	{1,2,3,4,7,8,10,13}
zoo	0.2	∅	{1,6,13}	{4,5,6,13}	{4,5,6,13}	{4,5,6,13,14}	{4,5,6,13,14,15,16}
	0.4	{6}	{1,6,8,13}	{4,5,6,8,13}	{4,5,6,8,13}	{4,5,6,8,13}	{4,5,6,8,13,16}
	0.6	{6,13}	{3,4,6,11,13}	{4,5,6,8,12,13}	{4,5,6,8,12,13}	{4,5,6,8,12,13}	{4,5,6,8,12,13,16}
	0.8	{6,13}	{3,4,6,8,13}	{4,6,8,12,13}	{4,5,6,8,12,13}	{4,5,6,8,12,13}	{4,5,6,8,12,13,15,16}
	1.0	{6,13}	{3,4,6,8,13}	{4,6,8,12,13}	{4,5,6,8,12,13}	{4,5,6,8,12,13}	{4,5,6,8,12,13,16}
mushroom	0.2	{5}	{1,5}	{3,5}	{3,5}	{3,5}	{3,5}
	0.4	{5}	{5,19}	{3,5,11,21}	{3,5,7,8,11}	{3,5,7,8,11}	{3,5,7,8,11}
	0.6	{5}	{5,19}	{3,4,5,11}	{3,5,7,8,11}	{3,5,7,8,11}	{3,5,7,8,11}
	0.8	{5}	{3,5,19}	{1,3,4,5,11}	{3,5,7,11,12}	{1,3,5,7,8,11}	{1,3,5,7,8,11}
	1.0	{1,3,5, 9,13,14}	{1,3,4,5,9,13,14, 21}	{1,3,5,7,9,13,14, 21}	{1,3,5,7,9,11,13, 14,21}	{1,3,5,7,9,11,13, 14,17}	{1,3,5,7,9,11,13, 14,17}
ionosphere	0.2	∅	{4,15,20,22,34}	{4,7,8,10,16,27, 33}	{4,8,10,16,17,27, 33}	{4,7,8,10,16,27, 33}	{4,7,8,10,16,27, 33}
	0.4	{5}	{1,4,5,14,25,28, 34}	{5,6,8,10,12,16, 25,32,33}	{5,8,10,16,17,18, 25,27,32,33}	{5,8,10,16,17,18, 25,27,32,33}	{5,8,10,16,17,18, 25,26,27,32,33,34}
	0.6	{4,5,6,18, 23,26,34}	{1,4,5,6,8,9,14,18, 23,25,26,29,34}	{4,5,6,10,12,16,18, 23,25,26,27,32,33, 34}	{4,5,6,10,12,16,17, 18,23,25,26,27,32, 33,34}	{4,5,6,10,12,16,17, 18,23,25,26,27,32, 33,34}	{4,5,6,7,10,12,16, 17,18,23,25,26,27, 32,33,34}
	0.8	{4,5,6,8, 18,22,23, 26,32,34}	{1,4,5,6,8,9,18,22, 23,24,25,26,29,32, 34}	{4,5,6,7,8,10,12, 14,16,18,22,23,25, 26,27,32,33,34}	{4,5,6,7,8,10,12, 14,16,18,22,23,25, 26,27,32,33,34}	{4,5,6,7,8,10,12, 14,16,18,22,23,25, 26,27,32,33,34}	{4,5,6,7,8,10,12, 14,16,18,22,23,25, 26,27,32,33,34}
	1.0	{4,5,6,8, 18,22,23, 26,32,34}	{3,4,5,6,8,9,10,11, 18,22,23,24,26,27, 29,31,32,34}	{4,5,6,7,8,10,12, 14,16,18,22,23,25, 26,27,32,33,34}	{4,5,6,7,8,10,12, 14,16,18,22,23,25, 26,27,32,33,34}	{4,5,6,7,8,10,12, 14,16,18,22,23,25, 26,27,32,33,34}	{4,5,6,7,8,10,12, 14,16,18,22,23,25, 26,27,32,33,34}

From Table 3, we can find that as the ratio raises, the core of a dataset expands, and this may result in the change of reduct. The changes are tiny in some datasets, such as wine and zoo, but huge in mushroom and ionosphere. The reducts are very different when $\lambda = 0$ compared to the reducts based on test costs ($\lambda = -0.5, -1, -2, -4$). Meanwhile, the attributes with the smallest test cost, namely the attribute 7 in wine dataset, attribute 5 in zoo dataset, attribute 3 in mushroom dataset and attribute 10 in ionosphere dataset respectively, are almost in all the reducts with different ratio and λ, but do not appear in the cores and reducts when $\lambda = 0$. The results in this table also indicates that the impact of λ on the reducts of partially labeled data may be limited, and the reducts almost the same when $\lambda = -1$ and $\lambda = -2$ in some datasets, however there may be some fluctuation when $\lambda = -0.5$ and $\lambda = -4$ compared to $\lambda = -1$.

4.3 Impact of Test Cost on Reduct

In the experiments above, we find the impact of λ on the reducts of partially labeled data based on the same test cost. Here, we let $\lambda = -1$, and conduct some experiments to show the impact of test costs on the reducts of partially labeled data. We let labeled ratio be 0.2, 0.4, 0.6, 0.8 and 1.0 respectively, test costs be produced randomly and satisfy normal distribution, the reducts produced by Algorithm 1 with different test costs are shown in Table 4.

In wine and ionosphere datasets, the reducts expand when labeled ratios raise in general. In zoo, the numbers of attributes in reduct change a little, mainly be 5 or 6. But the numbers change a lot in mushroom, for example, the reduct is {1,5} when the ratio is 0.2 and test cost is [1 40 51 76 1 78 17 63 85 73 100 97 38 70 27 14 94 61 59 94 82], and the reduct is {5,19} when the ratio is 0.6 and test cost is [31 53 57 77 39 1 26 70 22 62 45 71 4 39 19 100 59 70 22 34 38].

Table 4. Reducts on different test cost with the same λ

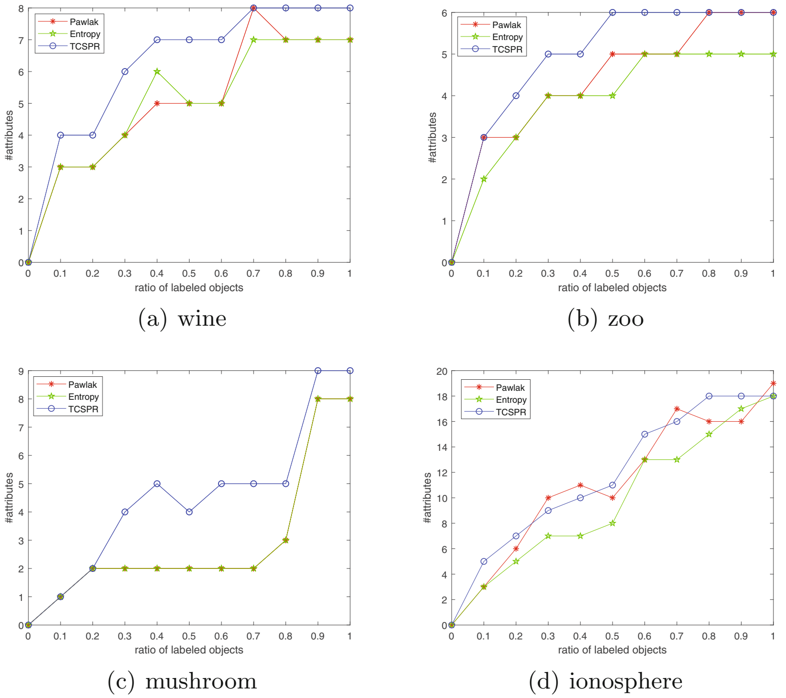
dataset	ratio	test cost	reduct
wine	0.2	48 70 1 54 45 17 32 45 100 86 16 91 52	{3,7,11,13}
		41 68 36 39 95 92 93 67 1 68 100 60 79	{3,4,8,9,12}
	0.4	1 100 58 73 14 38 45 60 14 61 24 52 65	{1,3,5,9,11,13}
		69 100 97 29 66 17 1 94 55 53 93 80 70	{4,5,6,7,9,11,13}
	0.6	85 97 64 63 100 1 52 28 56 89 94 36 53	{1,3,6,8,10,12,13}
		70 57 75 77 37 60 1 100 46 29 58 22 65	{2,3,5,7,10,11,12,13}
	0.8	42 100 78 1 63 24 48 57 70 48 87 44 61	{1,3,4,5,6,10,12,13}
		68 52 74 59 29 1 100 33 53 65 53 25 37	{1,3,5,6,9,11,12,13}
	1.0	46 100 21 77 61 42 15 41 47 1 57 23 47	{1,3,7,8,10,11,13}
		51 23 37 52 1 68 100 66 36 54 20 87 66	{1,2,3,5,9,10,11,12,13}
zoo	0.2	36 37 23 100 11 25 17 1 35 32 95 33 5 97 84 34	{3,7,8,13}
		36 61 68 78 65 81 80 42 49 54 59 64 71 1 61 100	{1,10,13,14}
	0.4	72 1 12 55 62 32 67 85 71 17 37 34 47 39 35 100	{2,3,6,10,14}
		51 57 72 4 24 78 74 23 100 60 67 66 44 1 58 10	{4,6,8,13,14}
	0.6	100 37 85 51 75 57 44 50 77 83 77 43 28 64 67 1	{4,6,8,12,13,16}
		48 100 42 31 21 1 34 55 22 53 67 32 59 61 46 96	{4,6,9,12,13}
	0.8	31 65 38 43 68 83 63 87 38 66 47 42 1 76 100 70	{1,3,6,9,12,13}
		90 51 13 23 25 82 35 52 10 44 1 28 72 100 64 52	{3,4,6,9,11,13}
	1.0	99 90 1 19 46 45 25 50 43 100 52 34 58 50 77 63	{3,4,6,9,13}
		3 58 33 26 20 63 39 1 100 33 49 52 8 75 49 46	{1,3,4,6,8,13}
mushroom	0.2	27 71 59 50 55 16 49 86 53 52 35 50 56 60 43 46 95 1 100 58 73	{1,5,18}
		1 40 51 76 1 78 17 63 85 73 100 97 38 70 27 14 94 61 59 94 82	{1,5}
	0.4	39 80 100 75 87 80 58 71 48 84 40 42 47 1 96 72 48 94 28 63 59	{5,14,19}
		56 56 14 44 39 67 84 85 14 48 1 5 45 100 17 58 37 85 6 46 65	{5,11,19}
	0.6	31 53 57 77 39 1 26 70 22 62 45 71 4 39 19 100 59 70 22 34 38	{5,19}
		99 56 87 1 54 39 16 81 74 66 23 100 76 56 66 57 10 56 39 34 29	{4,5,7,11,17}
	0.8	33 1 66 56 44 44 27 67 42 29 74 40 32 38 26 20 100 81 52 17 26	{2,5,10,20,21}
		60 87 29 1 25 74 76 78 62 70 52 89 28 78 42 56 80 94 34 100 83	{3,4,5,19}
	1.0	38 33 32 29 42 43 73 100 59 33 65 48 1 78 53 46 61 51 4 35 35	{1,3,5,9,13,14,19,21}
		31 100 21 33 26 11 42 39 95 40 15 98 86 41 1 34 43 56 40 62 60	{1,3,4,5,9,11,13,14}
ionosphere	0.2	76 100 92 73 49 50 98 24 66 15 94 90 67 65 1 82 9 5 69 42 78 85 90 49 79 70 89 81 91 63 63 94 11 54	{8,10,15,18,24,33}
		27 46 68 71 32 44 57 48 60 62 34 50 9 78 41 28 49 24 53 42 100 78 1 63 24 48 57 70 48 87 44 61 68 56	{8,13,16,18,20,23,34}
	0.4	13 28 54 17 75 27 100 14 52 1 25 29 58 74 55 35 70 70 50 63 58 17 69 27 28 51 37 42 64 77 40 56 39 53	{1,4,5,8,10,22,24,25,34}
		55 34 51 58 48 80 34 32 12 64 63 45 63 50 52 48 55 48 100 23 10 24 25 38 43 42 57 54 61 80 70 21 1 64	{5,8,9,12,20,21,24,25,32,33}
	0.6	22 66 44 53 67 46 57 61 75 55 82 42 40 90 26 41 68 45 82 71 49 19 57 1 48 64 47 66 100 68 49 60 53 3	{1,4,5,6,10,12,18,22,23,24, 25,26,34}
		45 87 54 70 30 57 71 21 64 51 31 55 7 1 25 30 25 11 63 15 58 76 55 58 46 18 54 80 100 44 14 60 65 50	{1,4,5,6,8,12,13,14,18,20,21, 23,25,26,31,34}
	0.8	26 90 1 39 22 36 55 52 35 65 62 62 40 70 62 100 71 73 50 23 41 34 38 60 38 31 41 46 57 30 55 51 76 45	{1,3,4,5,6,8,9,18,22,23,24, ,26,29,32,34}
		41 11 5 54 38 60 19 100 49 75 39 40 65 61 13 1 14 4 47 33 35 6 68 58 43 74 20 57 55 55 24 54 30 19	{3,4,5,6,8,11,12,16,18,20,22, 23,24,26,27,32,33,34}
	1.0	39 81 46 58 39 56 1 41 35 43 58 55 98 51 57 32 53 36 62 78 69 48 31 89 100 64 43 84 50 53 78 36 25 69	{1,4,5,6,7,8,9,10,11,14,18, 22,23,26,29,32,33,34}
		46 39 48 61 59 23 55 40 25 9 51 28 37 18 1 40 35 64 55 60 19 15 7 48 100 53 19 55 11 4 53 19 44 68	{4,5,6,8,9,10,11,14,15,18, 21,22,23,26,27,29,32,33,34}

From Table 4, we find that the attributes in reducts vary a lot according to different test costs except the attributes in core, which indicates that test cost has great impact of reduct. Furthermore, the attribute with low cost probably be a member of reduct and this is consistent with the conclusion in Sect. 4.2.

4.4 Quality of Reducts

From partially labeled data based on reduct, one can train classifier and use it to predict the classification of new objects. Here, some experiments are conducted to show the prediction performance of the classifiers. First, the numbers of attributes based on different reduction algorithms are shown in Fig. 1, where Pawlak stands for reduction based on attribute dependency degree, and Entropy stands for reduction based on entropy. Obviously, TCSPR gets more attributes than other algorithms in most cases.

Then, based on the reduced partially labeled data from three different methods (Pawlak, Entropy and TCSPR), we use decision tree model and CART algorithm to train classifier. To avoid randomness, 10-fold cross-validation is adopted



(a) wine

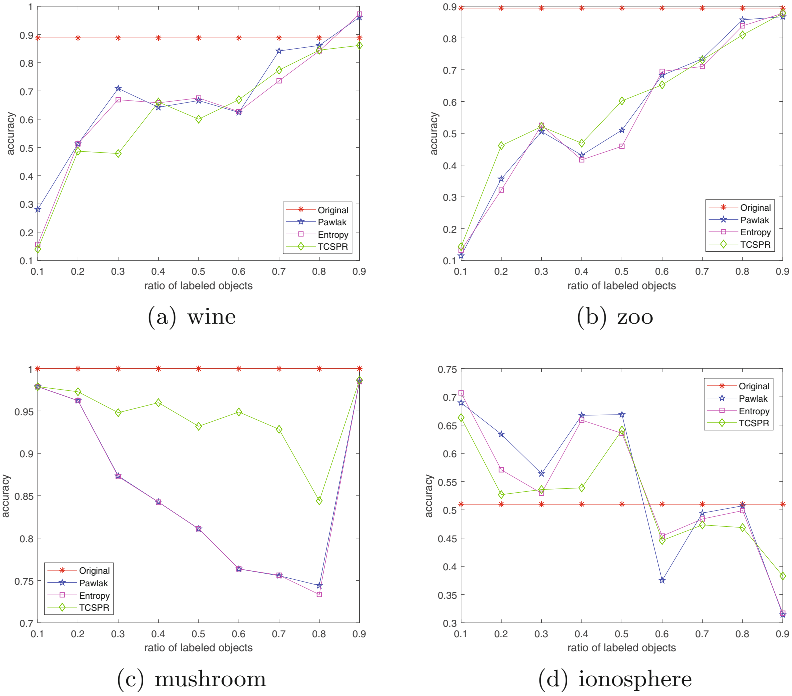
(b) zoo

(c) mushroom

(d) ionosphere

Fig. 1. Relationship between number of attributes and labeled ratio

and this is done 10 times. The relationships between classification accuracy of unlabeled data and ratio of labeled objects are shown in Fig. 2, where accuracy is the mean of the 100 experimental results, and "original" indicates the classification accuracy of the classifier trained by the whole dataset. Generally speaking, for all the three methods, there are some identical phenomenons in Fig. 2: in the wine and zoo datasets, the classification accuracy raises as the ratio of labeled data increases, and this accords with the common recognition; In the mushroom dataset, the classification accuracy is already near to 1 when the ratio is 0.1, and when the ratio increases, the accuracy decreases till the ratio is 0.8, then the accuracy goes up sharply when the ratio is 0.9; However, in the ionosphere dataset, the classification accuracy fluctuates according to the ratio. Figure 2 also shows that the accuracies of Pawlak and Entropy are closer, especially in the zoo and mushroom datasets, which indicates that the great impact of test cost on classification accuracy. In mushroom dataset, the classification accuracy of TCSPR is much superior than that of the other two methods, and in zoo dataset, the classification accuracy of TCSPR is higher than that of the other two methods when ratio is less than 0.6, however it approaches to other two in other settings. So, on the premise of not reducing the classification accuracy obviously, considering test cost for partially labeled data and finding the reduct with minimal test cost, which can be studied in future, are meaningful.



(a) wine (b) zoo

(c) mushroom (d) ionosphere

Fig. 2. Relationship between classification accuracy and labeled ratio

5 Conclusion

Based on rough set theory, this paper focuses on the attributes reduction of partially labeled data with test cost sensitive. Based on mutual information and the test cost of every condition attribute, we give the definition of attribute significance. Then a heuristic algorithm (TCSPR) for attribute reduction based on the significance is proposed. Experiments indicate the impact of labeled ratio and test cost on the reducts, and the effectiveness of our algorithm is verified too. In the future, more comparative experiments should be conducted to analyze the quality of the reducts, and further work can concentrate on incremental attribute reduction of partially labeled data with test cost sensitive.

Acknowledgements. The authors would like to thank the anonymous reviewers for their constructive comments that help improve the manuscript. This research was supported by the National Key R&D Program of China (213), National Natural Science Foundation of China (61673301), Major Project of Ministry of Public Security (20170004), and the Open Research Funds of State Key Laboratory for Novel Software Technology (KFKT2017B22).

References

1. Li, D.Y., Liu, C.Y., Du, Y., Han, X.: Artificial intelligence with uncertainty. Inn: International Conference on Computer and Information Technology, vol. 15, p. 2. IEEE (2008)
2. Dempster, A.P.: A Generalization of Bayesian Inference. Classic Works of the Dempster-Shafer Theory of Belief Functions. Springer, Heidelberg (2008)
3. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
4. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. **1**(1), 3–28 (1978)
5. Dubois, D., Prade, H.: Possibility Theory. Springer, US (1988)
6. Pawlak, Z.: Rough sets. Int. J. Comput. Inf. Sci. **11**(5), 341–356 (1982)
7. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht (1991)
8. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **38**(2), 325–339 (1967)
9. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
10. Miao, D.Q., Gao, C., Zhang, N., Zhang, Z.F.: Diverse reduct subspaces based co-training for partially labeled data. Int. J. Approx. Reasoning **52**(8), 1103–1117 (2011)
11. Jensen, R., Vluymans, S., Parthaláin, N.M., Cornelis, C., Saeys, Y.: Semi-supervised fuzzy-rough feature selection. In: Yao, Y., Hu, Q., Yu, H., Grzymala-Busse, J.W. (eds.) RSFDGrC 2015. LNCS (LNAI), vol. 9437, pp. 185–195. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25783-9_17
12. Zhang, W., Miao, D.Q., Gao, C., Li, F.: Rough set attribute reduction algorithm for partially labeled data. Comput. Sci. **44**(1), 25–31 (2017). (in Chinese)
13. Dai, J.H., Hu, Q.H., Zhang, J.H., Hu, H., Zheng, N.G.: Attribute selection for partially labeled categorical data by rough set approach. IEEE Trans. Cybern. **PP**(99), 1–12 (2017)
14. Ciucci, D.: Temporal dynamics in information tables. Fundamenta Informaticae **115**(1), 57–74 (2012)
15. Luo, C., Li, T.R., Chen, H.M., Fujita, H., Yi, Z.: Efficient updating of probabilistic approximations with incremental objects. Knowl.-Based Syst. **109**, 71–83 (2016)
16. Jing, Y.G., Li, T.R., Fujita, H., Yu, Z., Wang, B.: An incremental attribute reduction approach based on knowledge granularity with a multi-granulation view. Inf. Sci. **411**, 23–38 (2017)
17. Lang, G.M., Miao, D.Q., Yang, T., Cai, M.J.: Knowledge reduction of dynamic covering decision information systems when varying covering cardinalities. Inf. Sci. **346**(C), 236–260 (2016)
18. Turney, P.D.: Types of cost in inductive concept learning. In: 17th ICML Proceedings of the Cost-Sensitive Learning Workshop, California, pp. 1–7 (2000)
19. Yao, Y.Y., Wong, S.K.M.: A decision theoretic framework for approximating concepts. Int. J. Man-Mach. Stud. **37**, 793–809 (1992)
20. Yao, Y.Y., Zhao, Y.: Attribute reduction in decision-theoretic rough set models. Inf. Sci. **178**(17), 3356–3373 (2008)
21. Huang, J.J., Wang, J., Yao, Y.Y., Zhong, N.: Cost-sensitive three-way recommendations by learning pair-wise preferences. Int. J. Approx. Reasoning **86**(C), 28–40 (2017)

22. Li, H., Zhou, X., Zhao, J., Huang, B.: Cost-sensitive classification based on decision-theoretic rough set model. In: Li, T. (ed.) RSKT 2012. LNCS (LNAI), vol. 7414, pp. 379–388. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31900-6_47
23. Yang, X.B., Qi, Y.S., Song, X.N., Yang, J.Y.: Test cost sensitive multigranulation rough set: model and minimal cost selection. Inf. Sci. **250**(11), 184–199 (2013)
24. Ju, H.J., Li, H.X., Yang, X.B., Zhou, X.Z., Hang, B.: Cost-sensitive rough set: a multi-granulation approach. Knowl.-Based Syst. **123**(1), 137–153 (2017)
25. Zhang, W.X., Wu, W.Z., Liang, J.Y.: Rough Sets Theory and Methods. Science Press, Beijing (2003). (in Chinese)
26. Miao, D.Q., Hu, G.R.: A heuristic algorithm for reduction of knowledge. J. Comput. Res. Dev. **36**(6), 681–684 (1999). (in Chinese)
27. Min, F., He, H.P., Qian, Y.H., Zhu, W.: Test-cost-sensitive attribute reduction. Inf. Sci. **181**(22), 4928–4942 (2011)
28. Wong, S.K.M., Ziarko, W.: On optimal decision rules in decision tables. Bull. Polish Acad. Sci. Math. **33**(11–12), 693–696 (1985)
29. http://archive.ics.uci.edu/ml/index.php