Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RULES-BASED CLASSIFICATION WITH LIMITED COST

[1]GUIYING PAN, ZHONGMEI ZHOU

[1] Department of Compute Science and Engineering, Zhangzhou Normal University, Zhangzhou, 363000

E-mail: kentwingpan@163.com

ABSTRACT

In test cost-sensitive decision systems, it is difficulty for us to find an optimal attribute set and construct a quality classifier with limited cost. The minimal test cost-sensitive attribute reduction is proposed to address the former problem. However, it is inevitable to remove some good even better attributes in the minimal test cost-sensitive attribute reduction. As a result, the classification accuracy is not high in the minimal test cost attribute reduct. Suppose we have limited cost more than the minimal test cost, we can select addition important attributes to improve the classification accuracy. Therefore, our work includes two aspects. 1) Find an optimal attribute set with limited cost. We put forward a good method to find an attribute set based on the limited cost. 2) We improve the decision trees to build a quality classifier. We construct the root node of the decision tree by several best attribute values. These values just cover the entire dataset. So we can generate more quality rules than ID3. Experimental results indicate that the improved decision tree gets higher accuracy than ID3, and in comparison to the minimal test cost reduct, attribute selection based on the limited cost is feasible to improve the quality of classifiers.

Keywords: *Date Mining, Cost-Sensitive Learning, Test Cost-Sensitive Attribute Reduction, Limited Cost Classification*

1. INTRODUCTION

In data mining and machine learning, cost-sensitive learning is one of challenging problems. It is widely applied in many areas, such as medical diagnosis [1], Internet-based distributed systems [2], machine fault diagnosis, etc. In the cost-sensitive decision systems, our aim is to decrease cost and meanwhile keep the high quality of classifiers at the same time. Recently, it attracts much attention in many fields, such as decision trees [3, 4], neural networks [5], rough sets [6].

Test cost is one main type of costs. In test cost-sensitive systems, every conditional attribute has a characteristic value called test cost, so attribute is also called test. It means that one need afford test cost, while getting the value of an attribute. For an instance, in clinic systems, a patient need to be tested a series of body examining. Then each examination needs cost. However, if a patient can not afford all test cost, what can he/she do? In this situation, we must give up some tests. This problem is called limited cost problem. The minimal test cost reduct (MTR) has been proposed in [7, 8]. It aims at minimizing the test cost and meanwhile keeping the simplest information of the decision system. Some heuristic algorithms, such as genetic algorithms [9, 10], ant colony optimization algorithms [11], are used to find the minimal test cost reduct.

In test cost-sensitive systems, classification is another key issue. Classification aims to construct a high quality classifier to predict some unseen instances. A larger number of approaches have been proposed to build classifiers, for example, decision trees [12], support vectors machine [13], associative classification [14], etc. For the MTR problem, test cost-sensitive attribute reduction can find a minimal test cost reduct. However, the minimal test cost-sensitive attribute reduction removes some important attributes. As a result, the number of the rules generated by this reductive decision system may reduce considerably. It drops the quality of classifiers. Therefore if the resource is more than the minimal test cost, we can select addition important attributes to improve the quality of classifiers.

In this paper, confronting limited cost problem, we design a new approach to construct the classifiers. This approach includes two aspects: one is selecting an optimal attribute set and the other is building a quality classifier. We put forward a method to find an optimal attribute set based on limited cost and a greed strategy. This method can preserve more information of decision systems than

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the minimal test cost attribute reduct. We also propose another new attribute selection method to compare with the above-mentioned method. The second method selects attributes based on the minimal test cost ruduct, limited cost and a greed strategy. Then we improve decision trees to build high quality classifiers. In comparison with ID3 [12], the root node of our method consists of a set of best attribute values. These attribute values just cover the entire training dataset. Our method has following advantages: 1) The quality of internal nodes at the first level is higher than ID3; 2) It can generate more rules than ID3; 3) An instance may be covered by several rules in our method while it is covered only once in ID3. Experimental results indicate that the improved decision tree gets higher accuracy than ID3, and in comparison to the minimal test cost reduct, selecting addition important attributes can improve the classification accuracy obviously.

The outline of this paper is as follows. In section 2, we introduce some preliminary knowledge of the limited cost problem in test cost-sensitive decision systems. The section 3 presents the idea of attribute selection and improved decision trees. We report the analyses on the experimental results in section 4. Finally, we conclude in section 5.

2. TEST COST-SENSITIVE DECISION SYSTEM WITH LIMITED COST

This section reviews some basic knowledge about attribute selection with limited cost in test cost-sensitive decision systems.

A simple test cost-sensitive decision system (TC-DS) is often denoted as $S = (U, C, D, \{V_a \mid a \in C \cup D\}, \{I_a \mid a \in C \cup D\}, c)$, where U is a finite set of objects called the universe, C is the set of conditional attributes, D is the set of decision attributes, V_a is the set of values for each $a \in C \cup D$, $I_a : U \rightarrow V_a$ is an information function for each $a \in C \cup D$, and $c : C \rightarrow R^+ \cup \{0\}$ is an attribute cost function which is denoted by a vector $c = [c(a_1), c(a_2), \ldots, c(a_{|C|})]$. Given a set $A \subset C$, $c(A) = \Sigma_i c(a_i)$, $a_i \in A$. Table 1 and Table 2 represent a TC-DS.

Attribute reduction is an important issue in rough sets. It is widely used in data mining and machine learning. Much research on rough sets constantly comes forth, such as covering-based rough set [15], decision-theoretical rough set [16], variable-precision rough set [17] and test cost-sensitive rough set [7]. We give some definitions on the test cost-sensitive decision system with limited cost.

Definition 1: [7] Any $B \in C$ is called a decision relative reduct of S iff:

1) $POS_B(D) = POS_C(D)$, and
2) $\forall\, a \in B$, $POS_{B-\{a\}}(D) \subset POS_B(D)$.

Definition 3: [7] Let $Red(S)$ denote the set of all decision relative reducts of a TC-DS S. Any $R \in Red(S)$ where $c(R) = min\{c(R') \mid R' \in Red(S)\}$ is call a minimal test cost reduct.

Given a decision system S, let u_i denote the number of instances with class cls_i, and $|U| = \Sigma_i u_i$ be total number of U. Then $p(u_i) = u_i / |U|$ denotes the probability of class cls_i in U. The information entropy of U is $H(U) = -\Sigma_i\, p(u_i)\log p(u_i)$. Let t_i denote the number of instances with q_i of conditional attribute Q, u_j denote the number of instances with di of the decision attribute D. The conditional information entropy of Q respect to D is $H(D \mid Q) = -\Sigma_i\, p(q_i)\, \Sigma_j\, p(u_j \mid q_i) \log p(u_j \mid q_i)$, where $p(q_i) = |q_i| / |U|$, $p(u_j \mid q_i) = |u_j \cap q_i| / |q_i|$, $i, j = 1, 2, \ldots, k$. The information gain is denoted as follows.

Definition 4: [7] Let $B \subset C$, $a_i \in C$ and $a_i \notin B$, $d \in D$. The information gain of a_i respect to B is

$$f(B, a_i) = H(\{d\} \mid B) - H(\{d\} \mid B \cup \{a_i\}). \quad (1)$$

The information gain-based λ-weighted reduction algorithm [7] is an efficient heuristic algorithm to find the minimal test cost reduct for small datasets. The heuristic function is denoted by

$$f(B, a_i, c_i) = f\,(B, a_i)c_i^{\lambda}, \quad (2)$$

where λ is a non-positive number.

Definition 5: Let lc denote limited cost. Any $A \subset C$, $a_i \in A$, $c(A) <= lc$ is called test cost attribute selection with the limited cost problem.

In test cost-sensitive decision systems, our task is to find an optimal test cost attribute set with the limited cost and construct a high quality classifier.

3. CLASSIFICATION WITH LIMITED COST

In this section, we describe the new algorithm which has two stages in detail. The first stage is an attribute selection. We propose two selection ways so that the experimental results can be compared with each other. In the second stage, we improve the decision trees to build a quality of classifiers. The detail classification strategy is described in the following.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1 Attribute Selection With Limited Cost

In test cost-sensitive decision systems with limited cost, test cost attribute selection is a key issue due to that the quality of selected attribute set is concerned with the quality of classifiers. Given a limited budget which is more than the test cost of the minimal test cost reduct, it is obvious that addition attribute selection based on the minimal test cost reduct and a greed strategy is fast and convenient. However, the classifier builded by this attribute set may be not high. Therefore, in comparison to above method, we propose another optimal attribute selection method based on the limited cost and a greed strategy.

The first attribute selection method is shown as Algorithm 1 and Algorithm 3. It selects addition attributes based on the minimal test cost reduct, limited cost and a greed strategy. The detailed measure is shown as the following Example 1.

Example 1 A training data set T is shown in Table 1 and the corresponding cost vector is shown in Table 2. In this training data set, the total test cost is 653, and suppose the budget is 288.

First, Algorithm 1 finds the minimal test cost reduct through Formula 1 and 2. We produce different test cost reducts in the different λ environments, shown in Table 3. Then we select the best test cost reduct of them. The total test cost of reduct (a_0, a_8) is 96, and the total test cost of reduct (a_0, a_4, a_5) is 118. So we select the reduct (a_0, a_8) as the minimal test cost reduct and add them to the solution set. Then the remainder budget is 192.

Second, we select other attributes CA according to information gain and test cost through a greed strategy. We compute the information gain $f(\varnothing, a_i)$ of each attribute a_i, where $a_i \in CA$ and $c(a_i)$ is less than remainder budget. Then we sort CA according to $f(\varnothing, a_i)$ in a descending order. If there are two attributes with the same information gain, we sort the two attributes according to test cost in a ascending order. The order of CA is (a_3, a_6, a_4, a_7, a_5, a_1, a_2). Then we select attributes from CA one by one until the budget is not enough. So we select a_3 and a_6 to the solution set, and the remainder budget is 3. None attribute in CA meets the condition. The final solution set is (a_0, a_3, a_6, a_8).

The second attribute selection method is shown as Algorithm 2 and Algorithms 3. It is based on limited cost and a greed strategy. It contains two main steps: Step 1, if the remainder budget allows, we add the current best attribute a one by one into the solution set according to the information gain $f(SS, a)$ until the solution set becomes a super-reduct, shown in Algorithm 2 from Line 2 to Line 14; Step 2, whatever the solution set is a super-reduct or not after Step 1, we consider other attributes CA of which test cost meets the remainder budget. We compute the information gain $f(\varnothing, a_i)$ of a_i, where $a_i \in CA$. The next attribute selection steps are the same to the second step of the combine of Algorithm 1 and Algorithm 3, using Algorithm 3 to select addition attributes one by one according to the information gain and test cost.

Table 1 A Training Data Set

*a*0	*a*1	*a*2	*a*3	*a*4	*a*5	*a*6	*a*7	*a*8	*d*
1	9	19	30	55	60	72	79	82	89
5	9	11	32	55	59	76	80	88	90
3	7	14	35	53	60	73	81	88	89
3	10	17	32	55	59	76	80	83	90
3	9	11	35	55	58	72	81	88	89
3	9	14	33	52	57	75	80	85	90
4	9	11	36	54	59	73	80	88	89
4	9	19	33	52	56	75	80	85	90
3	9	11	37	54	60	73	81	88	89
3	10	14	33	55	60	75	79	82	90
1	10	18	34	55	60	72	79	82	89

Table 2 The Test Cost Vector

a	*a*0	*a*1	*a*2	*a*3	*a*4	*a*5	*a*6	*a*7	*a*8
c(*a*)	14	81	94	91	47	57	98	89	82

Table 3 The Reducts With Different λ

λ	-0.50	-0.75	-1.00	-1.25	-1.50
reduct	(0, 8)	(0, 8)	(0, 4, 5)	(0, 4, 5)	(0, 4, 5)

3.2 Improved Decision Trees

In traditional decision trees, such ID3/C4.5, the root node includes an attribute in general. This tree is too small to build a high quality classifier. In this paper, we propose an improved decision tree called multi-attribute values in root node decision trees (MVTD), shown as Algorithm 4 and Algorithm 5. This approach contains two steps: Step 1, construct the root node. This is the most important step in our approach. We select several best attribute values to construct the root node. These values can just cover the entire the training set. Given a training set T, we split T to sub-datasets, s_1, s_2, . . ., s_i, with each attribute value, v_1, v_2, . . ., v_i, respectively. An attribute value v_i is corresponding to a sub-dataset s_i which is covered by v_i. We use a measurement to compute the gain of v_i. Suppose s_i split by v_i from T, if the $s_i.classEntropy$ > 0, the gain of v_i is $|s_i|/s_i.classEntropy$, where $|s_i|$ is the number of

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

instances of s_i and s_i.*classEntropy* is the information entropy of s_i. As we know, the s_i.*classEntropy* is smaller, the dataset is clearer. And the $|s_i|$ is larger, it can increase more gain of v_i. If s_i.*classEntropy* equals to 0, we think the gain of v_i is infinity. We use this measurement to compute the gain of all v_i. Then we sort all v_i according to the gain with a descending order. We select the v_i according to the gain one by one to construct the root node until the union of all selected s_i covered by v_i just covers the entire T, shown as Algorithm 4 from Line 2 to Line 21. This process is the difference with the ID3, which is our main work; Step 2, we employ the selected s_i to build decision trees respectively. The process of building decision trees is the same to ID3 [12]. We produce a rule beginning at the rood node and following path and internal nodes until reaching a leaf node. If the class value in leaf node is *null*, this rule is pruned.

3.3 Classification

In this part, we give a measurement of the important rules. Then we employ the best rule to predict a unseen instance. For a rule, we use the number of instances which the rule covers as the importance of this rule.

For a testing instance, we select the best rule which is matched by this testing instance to predict this testing instance. Otherwise, we select the best rule which the first item is matched by this testing instance to predict it.

Algorithm 1 Attribute selection based on the minimal test cost reduct, limited cost and a greed strategy.

Input: $S = (U, C, D, V, I, c)$, limited cost lc,
L the set of λ.
Output: A solution set, SS.
1. $MTR(R) = \varnothing$, $SS = \varnothing$, $CA = \varnothing$;
2. for($\forall \lambda \in L$)
3. use information gain-based λ-weighted reduction algorithm to find a test cost reduct $R \rightarrow MTR(R)$;
4. end for
5. select the minimal test cost reduct $CBAS$, where $c(CBAS) = min\{c(R) \mid R \in MTR(R)\}$;
6. Candidate attributes $CA = C - CBAS$, $SS = CBAS$, $lc = lc - c(CBAS)$;
7. for($\forall\ a \in CA$)
8. if($c(a) <= lc$)
9. compute the $f(\varnothing, a)$;
10. end if
11. end for
12. B = selectAttributes(lc, CA);
13. $SS = SS \cup B$;
14. return SS;

Algorithm 2 Attribute selection based on limited cost and a greed strategy.

Input: $S = (U, C, D, V, I, c)$, limited cost lc,
Output: A solution set, SS.
1. $CBAS = \varnothing$, $SS = \varnothing$, $CA = C$;
2. while($POS_{CBAS}(D) \neq POS_C(D)$ and $CA \neq \varnothing$)
3. for any $a \in CA$
4. if $c(a) <= lc$
5. compute $f(CBAS, a)$;
6. else
7. $C = C - \{a\}$;
8. end if
9. end for
10. select a' with $c(a') = min\{c(a) \mid a \in max\ f(CBAS, a)\}$;
11. $CBAS = CBAS \cup \{a'\}$;
12. $lc = lc - c(a')$;
13. $CA = C - \{a'\}$;
14. end while
15. $CA = C$;
16. for any $a \in CA$, if $c(a) <= lc$, compute $f(\varnothing, a)$;
17. CA = selectAttributes(lc, CA);
18. $SS = CA \cup CBAS$;
19. return SS;

Algorithm 3 Select attributes according to information gain and test cost.

Input: limited cost lc, candidate attributes CA.
Output: a set of attribute SA.
Method: selectAttributes.
1. $SA = \varnothing$;
2. sort CA according to information gain in a descending order. If two attributes have the same information gain, sort the two attributes according to test cost in a ascending order;
3. while($CA \neq \varnothing$)
4. $CA = CA - \{a\}$, where a is the first element of CA;
5. if($f(\varnothing, a) <= 0$)
6. break;
7. else if($c(a) <= lc$)
8. $SA = SA \cup \{a\}$;
9. $lc = lc - c(a)$;
10. end if
11. end while
12. return SA;

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Algorithm 4 Multi-attribute values in root decision trees (MVDT).
Input: $S = (U, C, D, V, I)$.
Output: A classifier *cl.*

```
root = ∅, dataset = ∅, cl = ∅;
for any v ∈ V
   split S to s with v;
end for
for any v.s
    compute v.s.classEntropy;
    if(v.s.classEntroy == 0)
      root ∪ {v};
      dataset = dataset ∪ v.s;
    else
       gain(v) = |v.s| / v.s.classEntropy;
    end if
end for
sort V according to gain(v) in an descending
order;
while(V ≠ ∅)
    V = V - {v}, where v is the first
    element of V;
    root = root ∪ {v};
   dataset = dataset ∪ v.s ;
   if (dataset = S)
     break;
end while
for any v ∈ root
   v.s.makeTree();
   v.cl = v ∪ v.tree.
   cl = cl ∪ v.cl;
end for
return cl;
```

Algorithm 5 Make a decision tree.
Input: $S = (U, C, D, V, I)$
Method: makeTree.

```
if(|U| == 0)
  leaf = null;
  splitAttribute = null;
  return;
end if
if(classEntropy <= 0)
  leaf = classValue;
  splitAttribute = null;
return;
end if
for any a ∈ C, compute f(∅, a);
select a' with the maximal f(∅, a);
if ( f(∅, a') <= 0)
  leaf = bestDecisionValue;
  splitAttribute = null;
  return;
end if
splitAttribute = a';
split S to s_i by splitAttribute.v_i;
for any s_i,
  s_i.makeTree();
end for
```

4. EXPERIMENTS

The experiments perform in 7 UCI datasets which are digested in Table 4. The basic settings are as follows. Uniform distribution [7] is employed to produce test cost which ranges from 1 to 1000. The set of L ranges from -0.25 to -2.5 and the step length is -0.25. Experiments run ten-fold cross validation method for each dataset. The budget is produced by a function, but also can be set by users. The experiments run in a dynamic environment where both the training set and the test cost change.

In the Table 5 and Figure 1, we give the accuracy of ID3, MVDT, 0-MVDT, 1-MVDT, 2-MVDT. The accuracy of ID3 and MVDT is obtained without cost limit. 0-MVDT denotes MVDT based on the minimal test cost reduct. 1-MVDT denotes MVDT based on the minimal test cost reduct, limited cost and a greed strategy. 2-MVDT denotes MVDT based on the limited cost and a greed strategy. The *ave* denotes the average accuracy obtained by ID3, MVDT, 0-MVDT, 1-MVDT, 2-MVDT in the 7 datasets respectively. In the Table 6 and Figure 2, we give the limited cost LC and the total test cost of MVDT, 0-MVDT, 1-MVDT, 2-MVDT. The *ave* denotes the average total test cost of MVDT, 0-MVDT, 1-MVDT, 2-MVDT in the 7 datasets respectively.

From above tables and figures, we can achieve the following conclusions: 1) MVDT can get higher accuracy than ID3 expect the *zoo* dataset, and the average accuracy of MVDT is higher than one of ID3 by 2.5%. It indicates that MVDT can build a more quality classifier than ID3 in general; 2) 1-MVDT and 2-MVDT get higher accuracy than 0-MVDT by 0.6% and 1.1% respectively. It indicates that when limited cost is more than the minimal test cost, it is feasible for us to select addition important attributes to improve the classification accuracy. 3) 2-MVDT gets the same average accuracy to MVDT, but the average total test cost of 2-MVDT, which is 4842, is much less than one of MVDT, which is 7190. It indicates that our approach can save much money and get the accuracy as high as MVDT. 4) 2-MVDT is better than 1-MVDT. The cause may be that some more important attributes are not

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

selected in the minimal test reduct, duet to that their test costs are larger. However, 2-MVDT based on limited cost can select them.

5. CONCLUSIONS

In test cost-sensitive decision systems, test cost-sensitive attribute reduction can decrease the test cost. However, it removes some important attributes at the same time. As a result, it makes the accuracy lower in the classification. In addition, traditional decision trees generate few rules in general, such as ID3. Each of rules produced by ID3 only covers an instance. In reductive decision system, the accuracy is not as high as one produced by the entire dataset, even falls. In this paper, we propose a new approach to select an optimal attribute set if the limited cost is more than the cost of the minimal test cost reduct. We also propose a new method (MVDT) to improve the decision tree. Experimental results show that MVDT achieves higher accuracy than ID3 without cost limit. Confronting limited cost problem, building classifiers based on the minimal test cost reduct is not the best choice while selecting addition important attribute can improve the quality of classifiers obviously in our method. The MVDT with limited cost can get the accuracy as high as the MVDT without cost limit. Therefore, it is feasible for us to employ MVDT with attribute selection based on limited cost to build classifiers.

Table 4 Dataset Information

Dataset	|U|	|C|	|D|
zoo	101	16	7
lymph	148	18	4
iris	150	4	3
auto	193	21	6
voting	435	16	2
breast	683	9	2
austra	690	14	2

Table 5 The Accuracy Of ID3, MVDT, 0-MVDT, 1-MVDT, 2-MRDT

Ds	Accuracy				
	ID3	MVDT	0-MVDT	1-MVDT	2-MVDT
zoo	0.970	0.930	0.940	0.940	0.940
lymph	0.703	0.723	0.709	0.736	0.764
iris	0.780	0.873	0.847	0.847	0.840
auto	0.835	0.840	0.804	0.820	0.825
voting	0.936	0.958	0.952	0.952	0.956
breast	0.933	0.961	0.959	0.958	0.960
austra	0.757	0.803	0.800	0.801	0.803
ave	0.845	0.870	0.859	0.865	0.870

Table 6 The Limited Cost LC And The Total Test Cost Of MVDT, 0-MVDT, 1-MVDT, 2-MVDT

DS	LC	Total test cost			
		MVDT	0-MVDT	1-MVDT	2-MVDT
zoo	4367	8200	2116	4246	4292
lymph	4861	9093	2358	4735	4784
iris	1432	2005	1156	1167	1271
auto	6869	10841	3563	6765	6753
voting	5406	8200	4049	5219	5336
breast	4601	4737	4046	4390	4303
austra	7250	7250	6536	7109	7156
ave	4970	7190	3403	4804	4842

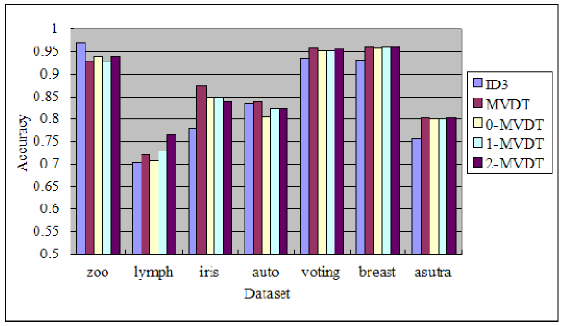

Figure 1 The Accuracy of ID3, MVDT, 0-MVDT, 1-MVDT, 2-MVDT

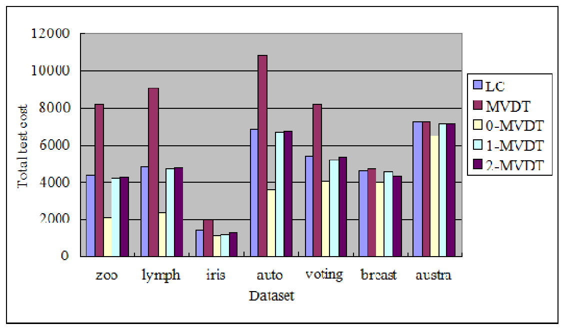

Figure 2 The Limited Cost LC And The Total Test Cost Of MVDT, 0-MVDT, 1-MVDT, 2-MVDT

ACKNOWLEDGEMENT

This work is funded by China NSF program (No.61170129) and the postgraduate education innovation base for Computer application technology, Signal and Information Processing of Fujian province (No. [2008]114, high education of Fujian).

REFRENCES:

[1] M. Nunez, "The use of background knowledge in decision tree induction", *Machine Learning*, Vol. 6, 1991, pp. 231-250.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[2] D. Ardagna, C. Francalanci, and M. Trubian, "A multi-model algorithm for the cost-oriented design of internet-based systems", *Information Sciences*, Vol. 176, 2006, pp. 3105-3131.

[3] V.S. Sheng, C.X. Ling, and A. Ni, and S. Zhang, "Cost-sensitive test strategies", *Proceedings of the National Conference on Artificial Intelligence*, Vol. 21, No. 1, 2006, pp. 482-487.

[4] W. Yi, M. Lu, and Z. Liu, "Multi-valued attribute and multi-labeled data decision tree algorithm", *International Journal of Machine Learning and Cybernetics*, Vol. 2, No.2, 2011, pp. 1-8.

[5] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Networks*, Vol. 2, No. 5, 1989, pp. 359-366.

[6] Z. Pawlak, "Rough sets", *International Journal of Computer and Information Sciences*, Vol. 11, 1982, pp. 341-356.

[7] F. Min, H. He, Y. Qian, and W. Zhu, "Test-cost-sensitive attribute reduction", *Information Sciences*, Vol. 181, 2011, pp. 492-4942.

[8] H. He and F. Min, "Accumulated cost based test-cost-sensitive attribute reduction", *RSFDGrC*, ser. LNAI, Vol. 6743, 2011, pp. 244-247.

[9] G. Pan, F. Min, and W. Zhu, "A genetic algorithm to the minimal test cost reduct problem", *Proceedings of IEEE International Conference on Granular Computing*, 2011, pp. 539-544.

[10] J. Liu, S. Liao, F. Min, and W. Zhu, "An improved genetic algorithm to minimal test cost reduction", accepted in *Proceedings of IEEE International Conference on Granular Computing*, 2012.

[11] Z. Xu, F. Min, J. Liu, and W. Zhu, "Ant colony optimization to minimal test cost reduction", accepted in *Proceedings of IEEE International Conference on Granular Computing*, 2012.

[12] J. R. Quinlan, "Induction on decision trees", *Machine Learning*, Vol. 1, 1986, pp. 81-106.

[13] V. Vapnik, A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities", *Theory of Probability and Its Application*, Vol. 16, No. 2, 1971, pp. 264-280.

[14] B. Liu, W. Hsu, and Y. Ma, "IntegratingClassification and Association Rule Mining", *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, 1998.

[15] W. Zhu and F. Wang, "Reduction and axiomization of cover-ing generalized rough sets", *Information Sciences*, Vol. 152, No. 1, 2003, pp. 217-230, 2003.

[16] Y. Yao and Y. Zhao, "Attribute reduction in decision-theoretic rough set models", *Information Sciences*, Vol. 178, No. 17, 2008, pp. 3356－3373.

[17] W. Ziarko, "Variable precision rough set model", *Journal of Computer and System Sciences*, Vol. 46, No. 1, 1993, pp. 39-59.