Filomat 32:5 (2018), 1817–1822
https://doi.org/10.2298/FIL1805817L

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Local View Based Cost-Sensitive Attribute Reduction

Jingzheng Li[a], Xiangjian Chen[a], Pingxin Wang[b], Xibei Yang[a,c]

[a] *School of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212003, P.R. China*
[b] *School of Science, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212003, P.R. China*
[c] *Intelligent Information Processing Key Laboratory of Shanxi Province, Shanxi University, Taiyuan, Shanxi, 030006, P.R. China*

Abstract. In traditional cost-sensitive attribute reduction, the variation of decision cost is referred to as a global difference of costs because the considered decision cost is the variation of sum of decision costs over all objects. However, such reduction does not take the variation of decision costs of each object into account. To solve this problem, a local view based cost-sensitive attribute reduction is introduced. Firstly, through considering the variation of decision costs of single object if the used attributes change, a local difference of costs is presented. Secondly, on the basis of the fuzzy decision-theoretic rough set model, a new significance function is given to measure the importance of attribute. Finally, the experimental results illustrate that by comparing the traditional reduction, the proposed local view can decreases both global and local differences of costs effectively on several UCI data sets.

1. Introduction

Up to now, there has been increasing attention to the issue of cost sensitive learning in rough set theory [2, 3, 6]. For example, Yao [7, 8] presented the concept of the Decision-Theoretic Rough Set (DTRS) by considering the delayed-decision cost and misclassification cost. Unfortunately, the DTRS is not suitable for processing numerical data. Therefore, the Fuzzy Decision-Theoretic Rough Set model (FDTRS) is proposed by Song et al. [4] through the use of the Gaussian kernel based similarity. Similar to DTRS, FDTRS provides the cost based semantic explanations for fuzzy positive, boundary and negative decisions, from which we can see that decision cost is also an important issue in the research of FDTRS.

To minimize the decision cost by removing redundant attributes, the concept of attribute reduction can be defined. However, such type of reduction does not take the variations of decision costs of each object into consideration. The variations of decision costs of single object may be increasing or decreasing because the monotonic property [5] does not hold in FDTRS. In order to display the variations of decision costs of single object in detail, a local view is proposed to compute reduct. Our local view considers the changings of decision costs of each object instead of the simple changings of whole decision cost. Therefore, a new significance function is presented which reflects the variations of decision costs of each object if attributes vary. Such significance function is introduced into the heuristic algorithm to derive reduct.

2010 *Mathematics Subject Classification*. 97E60.
Keywords. Attribute reduction; Fuzzy decision-theoretic rough set; Global view; Local view.
Received: 23 October 2017; Accepted: 30 January 2018
Communicated by Hari M. Srivastava
Research supported by the Natural Science Foundation of China (Nos. 61572242, 61502211, 61503160), Open Project Foundation of Intelligent Information Processing Key Laboratory of Shanxi Province (No. 2014002).
Corresponding author: Xibei Yang
Email addresses: maxlijingzheng@163.com (Jingzheng Li), cxj831209@163.com (Xiangjian Chen), wpxin8012@hotmail.com (Pingxin Wang), yangxibei@hotmail.com (Xibei Yang)

2. Fuzzy Decision-Theoretic Rough Set

For FDTRS, the state set consists of two classes [1, 9] such that $\Omega = \{C, \sim C\}$, it means that an object x_i is in class C or not in class C. The actions set is given by $\mathscr{A} = \{\tau_P, \tau_B, \tau_N\}$, in which τ_P, τ_B, τ_N express three actions: τ_P indicates that x_i is assigned to the positive region of C; τ_B indicates that x_i is assigned to the boundary region of C; τ_N indicates that x_i is assigned to the negative region of C. Table 1 describes the decision costs which three actions are divided into two classes respectively.

Table 1: Cost matrix

	C	$\sim C$
τ_P	λ_{PP}	λ_{PN}
τ_B	λ_{BP}	λ_{BN}
τ_N	λ_{NP}	λ_{NN}

In Table 1, λ_{PP}, λ_{BP} and λ_{NP} are decision costs for taking actions τ_P, τ_B and τ_N, respectively, if the state of object x_i is in C; λ_{PN}, λ_{BN} and λ_{NN} are decision costs for taking actions τ_P, τ_B and τ_N, respectively, if the state of object x_i is not in C.

Without loss of generality, a decision system is described as $DS = \{U, AT \cup D\}$: U is called the universe which contains all objects; AT refers to the set of the condition attributes; $D = \{d\}$ is the set of decision attributes, $d(x_i)$ shows the label or classified information of object x_i. $\forall A \subseteq AT$, the lower and upper approximations of C are

$$\underline{A}(C) = \{x_i \in U : \Pr(C|[x_i]_A) \geq \alpha\}; \tag{1}$$

$$\overline{A}(C) = \{x_i \in U : \Pr(C|[x_i]_A) > \beta\}, \tag{2}$$

where $[x_i]_A$ is the fuzzy information granule derived by Gaussian kernel based similarity. The positive region of C is $\mathrm{POS}_A(C) = \underline{A}(C)$, the boundary region of C is $\mathrm{BND}_A(C) = \overline{A}(C) - \underline{A}(C)$, and the negative region of C is $\mathrm{NEG}_A(C) = U - \overline{A}(C)$. The parameters α and β can be systematically computed based on minimizing the overall decision cost, i.e., $\alpha = (\lambda_{PN}-\lambda_{BN})/((\lambda_{BP}-\lambda_{PP})+(\lambda_{PN}-\lambda_{BN}))$, $\beta = (\lambda_{BN}-\lambda_{NN})/((\lambda_{NP}-\lambda_{BP})+(\lambda_{BN}-\lambda_{NN}))$.

It is known that FDTRS is closely related to costs, we then obtain the decision costs which object x_i is classified into positive region, boundary region and negative region, respectively. Decision cost of positive region: $\mathrm{price}_A^{\mathrm{POS}}(x_i) = \Pr(C|[x_i]_A) \cdot \lambda_{PP} + (1 - \Pr(C|[x_i]_A) \cdot \lambda_{PN}$; Decision cost of boundary region: $\mathrm{price}_A^{\mathrm{BND}}(x_i) = \Pr(C|[x_i]_A) \cdot \lambda_{BP} + (1 - \Pr(C|[x_i]_A) \cdot \lambda_{BN}$; Decision cost of negative region: $\mathrm{price}_A^{\mathrm{NEG}}(x_i) = \Pr(C|[x_i]_A) \cdot \lambda_{NP} + (1 - \Pr(C|[x_i]_A) \cdot \lambda_{NN}$. Immediately, the whole decision cost is denoted by COST_A as follows

$$\mathrm{COST}_A = \sum_{x_i \in \mathrm{POS}_A(C)} \mathrm{price}_A^{\mathrm{POS}}(x_i) + \sum_{x_i \in \mathrm{BND}_A(C)} \mathrm{price}_A^{\mathrm{BND}}(x_i) + \sum_{x_i \in \mathrm{NEG}_A(C)} \mathrm{price}_A^{\mathrm{NEG}}(x_i). \tag{3}$$

Obviously, the whole cost consists of decision cost of positive region,boundary region and negative region.

3. Cost-Sensitive Attribute Reduction and Problem Description

3.1. Cost-Sensitive Attribute Reduction

Since the FDTRS is proposed based on the costs of classifications. It is reasonable for us to research cost-sensitive based attribute reduction which aims to minimize the decision cost via deleting redundant attributes. The following presents the definition of Cost-Sensitive Attribute Reduction (CSAR).

Definition 1. *Given a decision system DS, A is regarded as a reduct if and only if*

1. $\mathrm{COST}_A \leq \mathrm{COST}_{AT}$;
2. $\forall B \subset A$, $\mathrm{COST}_B > \mathrm{COST}_{AT}$.

The definition of CSAR shown in Definition 1 is to generate the minimal subset of attributes which keeps the decision cost unchanged or decreases the decision cost.

3.2. Problem Description

Definition 1 actually reflects the variation of the sum of decision costs of whole objects with respect to the decreasing of number of used attributes. It can be regarded as a global difference of costs. Nevertheless, such global view does not take the detailed variations of decision costs of single object into consideration. In other words, if the global view is employed, then the variations of decision costs may not be reflected effectively. The following Example 1 shows us the limitation of global view.

Example 1. *By Table 2, we can know that the decision cost of object x_i increases while that of object x_j decreases with reduction. The global difference of costs is zero.*

Nevertheless, it is observed that the variations of decision costs of single object do exist and then it follows that the gloabl view of computing the difference of decision costs is not good enough. Furthermore, we can compute the sum of the absolute variations of decision costs of each objects. It can be regarded as a local difference of costs. As a result, the local difference of costs is four in Table 2.

Table 2: A toy example.

Objects	Decision cost before reduction	Decision cost after reduction	Variation of costs	ϑ(global)	ϑ(local)
x_i	6	4	-2	0	4
x_j	8	10	+2		

In summary, there are two ways to describe the differences of decision costs with reduction by global view and local view, respectively. Given a decision system DS, the global difference of costs is

$$\vartheta(\text{global}) = |\text{COST}_{AT} - \text{COST}_{A}|, \tag{4}$$

while the local difference of costs is

$$\vartheta(\text{local}) = \sum_{i=1}^{m} |\text{price}_{AT}(x_i) - \text{price}_{A}(x_i)|, \tag{5}$$

where m is the cardinality of U.

3.3. Algorithm to Compute Reduct

By using the global and local differences of costs, the significance functions are presented as follows.

Definition 2. *Given a decision system DS, $\forall a \in AT - B$, the global significance of attribute a is*

$$\text{Sig}_{\text{global}}(a, B, D) = |\text{COST}_{B \cup \{a\}} - \text{COST}_{AT}|. \tag{6}$$

Definition 3. *Given a decision system DS. $\forall B \subseteq AT$, $\forall a \in AT - B$, $\forall x_i \in U$, the local significance of attribute a is*

$$\text{Sig}_{\text{local}}(a, B, D) = \sum_{i=1}^{m} |\text{price}_{B \cup \{a\}}(x_i) - \text{price}_{AT}(x_i)|. \tag{7}$$

The above two significance functions measure the variations of decision costs if a is introduced into B, and then determine the significance of attribute a, respectively. The smaller the value of significance of one attribute is, the more important such attribute will be. By using the above significance functions, a forward heuristic attribute reduction algorithm can be designed as follows.

Algorithm 1: Cost-Sensitive Attribute Reduction (CSAR).

Input: Decision system DS, Gaussian kernel parameter δ;
Output: One reduct A.
1. $A \leftarrow \emptyset$, calculate COST_{AT};
2. **Do**
   1) $\forall a \in AT - A$, calculate $\mathrm{Sig}_{\triangle}(a, A, D)$;
   2) Select a such that $\mathrm{Sig}_{\triangle}(a, A, D) = \min\{\mathrm{Sig}_{\triangle}(a, A, D) : \forall a \in AT - A\}$;
   3) $A = A \cup \{a\}$; // A is a temporary pool set
   4) Calculate COST_A;
   Until $\mathrm{COST}_A \leq \mathrm{COST}_{AT}$
3. Return A.

- If $\triangle$ = global, then Algorithm 1 is realized by global significance which is shown in Eq. (6), and CSAR is regarded as a Cost-Sensitive Attribute Reduction with Global Significance (CSARGS).
- If $\triangle$ = local, then Algorithm 1 is realized by local significance which is shown in Eq. (7), and CSAR is regarded as a Cost-Sensitive Attribute Reduction with Local Significance (CSARLS).

In Algorithm 1, for each looping in Step 2, the aim is to find an attribute with the highest importance and then add this attribute into reduct.

4. Experimental Analysis

Table 3 shows the UCI data sets which are normalized. Furthermore, 10 different cost matrices are randomly generated which satisfy the conditions of loss function. The final decision costs shown in this experiments are the average values of the experimental results over ten different cost matrices. In FDTRS, Gaussian kernel function is used to extract the fuzzy information granule. The Gaussian kernel parameters are set such that $\delta = 0.5, 1.0$ and 1.5, respectively. Furthermore, heuristic search strategy is employed to sort attributes, and measures these variations of the differences of costs if attributes are added into the pool set one by one.

Table 3: Data sets description.

ID	Data sets	Objects	Attributes	Classes
1	Breast Cancer Wisconsin (Diagnostic)	500	31	2
2	Congressional Voting	435	17	2
3	Connectionist Bench (Vowel Recognition - Deterding Data)	990	14	11
4	Diabetic Retinopathy Debrecen	1151	20	2
5	Forest Type Mapping	523	28	4
6	German	1000	25	2

4.1. The Comparison Between the Global Differences of Costs

First of all, we will compare the global differences of costs between CSARGS and CSARLS.

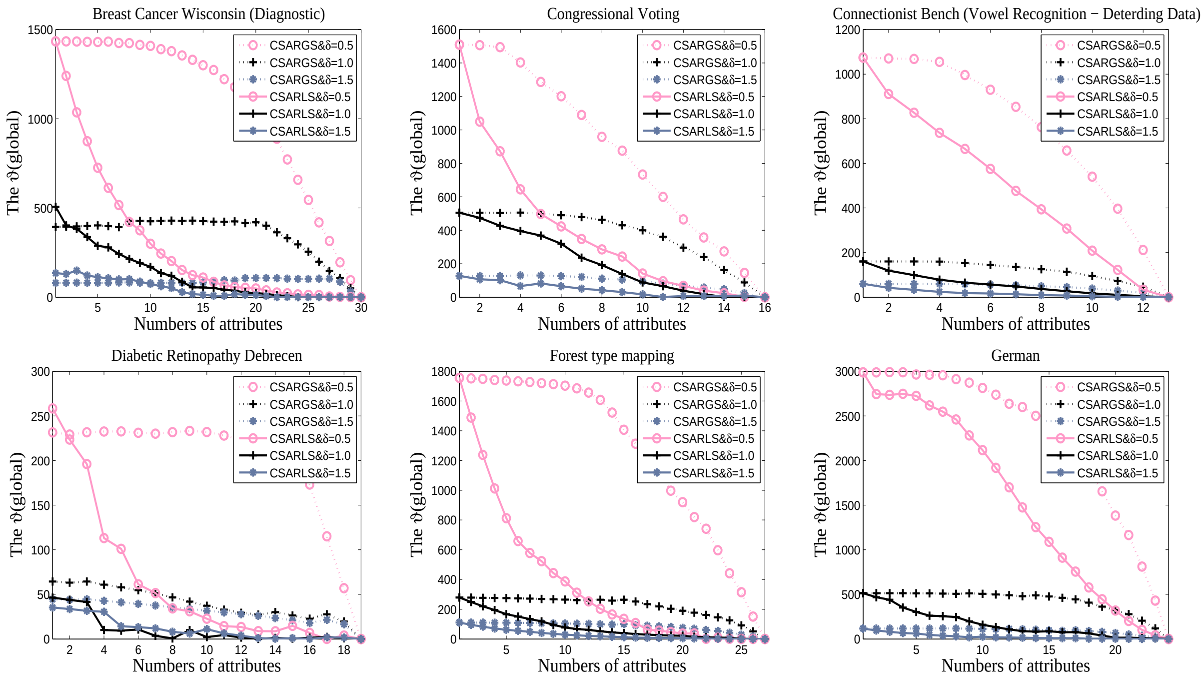


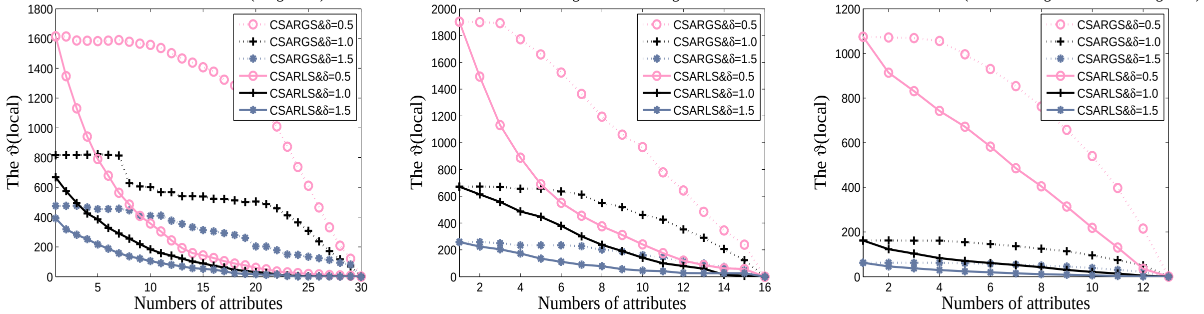
Figure 1: Comparisons between global differences of costs.

To observe Figure 1, we can draw the following conclusions.

1. If the attributes are added into the pool set one by one, then the differences of global costs derived by CSARGS are lower than those derived by CSARGS significantly.
2. The differences of global costs gradually decreases with the increasing of the number of attributes. Although the monotonic property of decision cost does not hold in FDTRS, as the number of attributes increases, the decision costs still tend to have a downward trending.

4.2. *The Comparison Between Local Differences of Costs*

Similar to Section 3.1, we will further investigate the local differences of costs between CSARGS and CSARLS.



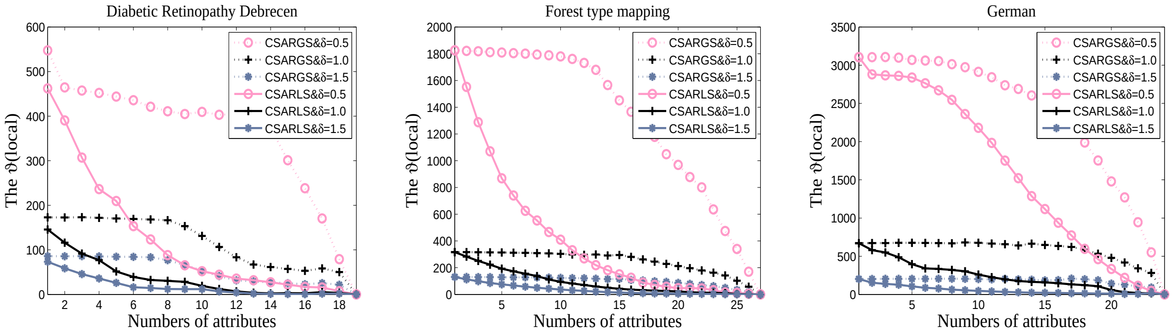


Figure 2: Comparisons between local differences of costs.

Following Figure 2, through comparing CSARLS with CSARGS, it is not difficult to observe that CSARLS can obtain lower local differences of costs than CSARGS can do.

Moreover, it should be noticed that though both CSARLS and CSARGS aim at keeping the decision cost unchanged or decreasing the decision cost, such two approaches employ different views to delete redundant attributes. In CSARGS, to minimize the value of ϑ(global) is regarded as the optimized objective while in CSARLS, to minimize the value of ϑ(local) is the optimized objective. Experiments demonstrate that the CSARLS is superior to the CSARGS on decreasing the differences of costs.

5. Conclusions

In this paper, we have investigated two approaches to compute the differences of decision costs in rough set based cost-sensitive attribute reduction, they are referred to as global view and local view, respectively. Accordingly, two significance functions are employed with respect to global and local differences of costs. The experimental results validated that by comparing global view, local view based cost-sensitive attribute reduction can produce lower differences of costs.

References

[1] H. L. Dou, X. B. Yang, X. N. Song, H. L. Yu, W. Z. Wu, J. Y. Yang, Decision-theoretic rough set: A multicost strategy, Knowledge-Based Systems, 91 (2016) 71–83.

[2] X. Y. Jia, W. H. Liao, Z. M. Tang, L. Shang, Minimum cost attribute reduction in decision-theoretic rough set models, Information Sciences, 219 (2013) 151–167.

[3] H. R. Ju, X. B. Yang, H. L. Yu, T. J. Li, D. J. Yu, J. Y. Yang, Cost-sensitive rough set approach, Information Sciences, 355 (2016) 282–298.

[4] J. J. Song, E. C. C. Tsang, D. G. Chen, X. B. Yang, Minimal decision cost reduct in fuzzy decision-theoretic rough set model, Knowledge-Based Systems, 124 (2017) 104–112.

[5] S. P. Xu, X. B. Yang, H. L. Yu, D. J. Yu, J. Y. Yang, E. C. C. Tsang, Multi-label learning with label-specific feature reduction, Knowledge-Based Systems, 104 (2016) 52–61.

[6] X. B. Yang, Y. Qi, H. L. Yu, J. Y. Yang, Want more? Pay more!, in: Proceedings of 9th International Conference on Rough Sets and Current Trends in Computing, Granada and Madrid, Spain, (2014) pp.144–151.

[7] Y. Y. Yao, Decision-theoretic rough set models, in: Proceedings of 2th International Conference on Rough Sets and Knowledge Technology, Toronto, Canada, (2007) pp.1–12.

[8] Y. Y. Yao, B. Zhou, Two Bayesian approaches to rough sets, European Journal of Operational Research, 251 (2016) 904–917.

[9] B. Zhou, Multi-class decision-theoretic rough sets, International Journal of Approximate Reasoning, 55 (2014) 211–224.