Knowledge-Based Systems 91 (2016) 71–83

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Decision-theoretic rough set: A multicost strategy

Huili Dou [a], Xibei Yang[a,b,*], Xiaoning Song[c], Hualong Yu[a], Wei-Zhi Wu [d], Jingyu Yang[e]

[a] School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, PR China
[b] School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, PR China
[c] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China
[d] Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province, Zhejiang Ocean University, Zhoushan 316022, PR China
[e] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, PR China

ARTICLE INFO

Article history:
Received 11 January 2015
Revised 8 September 2015
Accepted 9 September 2015
Available online 14 September 2015

Keywords:
Cost reduct
Decision-monotocity reduct
Decision-theoretic rough set
Multiple cost matrices

ABSTRACT

By introducing the misclassification and delayed decision costs into the probabilistic approximations of the target, the model of decision-theoretic rough set is then sensitive to cost. However, traditional decision-theoretic rough set is proposed based on one and only one cost matrix, such model does not take the characteristics of multiplicity and variability of cost into consideration. To fill this gap, a multicost strategy is developed for decision-theoretic rough set. Firstly, from the viewpoint of the voting fusion mechanism, a parameterized decision-theoretic rough set is proposed. Secondly, based on the new model, the smallest possible cost and the largest possible cost are calculated in decision systems. Finally, both the decision-monotocity and cost criteria are introduced into the attribute reductions. The heuristic algorithm is used to compute decision-monotonicity reduct while the genetic algorithm is used to compute the smallest and the largest possible cost reducts. Experimental results on eight UCI data sets tell us: 1. compared with the raw data, decision-monotocity reduct can generate greater lower approximations and more decision rules; 2. the smallest possible cost reduct is much better than decision-monotocity reduct for obtaining smaller costs and more decision rules. This study suggests new research trends concerning decision-theoretic rough set theory.

1. Introduction

Cost-sensitive learning is a valuable problem. It has been addressed by many researchers from a variety of fields including decision making [22], machine learning [8,9,55,41], pattern recognition [56] and so on. Note that in recent years, cost-sensitive learning has also attracted much attention of researchers in rough set theory [32]. In broad terms, rough set is one of the most important tools of Granular Computing (GrC) [33,44,46] and then introduction of cost sensitivity into rough set is bound to bring GrC a worthwhile topic.

Roughly speaking, driven by many practical applications, two main types of costs have been discussed in rough set: test cost and misclassification cost. On the one hand, each test (i.e., attribute, measurement, feature) may have an associated cost which is regarded as test cost. For example, in medical diagnosis, a blood test has a cost which may be the time or money spent on testing blood. On the other hand, misclassification cost is basically the loss when classifying data into a specific outcome. For example, it may be troublesome if a healthy subject is misclassified as a patient, but it could result in a more serious consequence if a patient is misclassified as healthy subject.

In rough set theory, two crucial aspects should be covered in terms of costs. Firstly, since the basic model, i.e., lower and upper approximations are not sensitive to cost, then how to introduce cost into modeling is an interesting issue. Such aspect will provide us a new perspective to characterize the uncertainty, mainly because if lower and upper approximations are sensitive to cost, then inevitably, uncertainty [1,7] in rough set is sensitive to cost. Secondly, it is well known that attribute reduction is one of the key topics in rough set, from which we can see that finding reducts with some particular requirements of cost (e.g., to find reduct with the smallest cost) is also a challenge. In recent years, unremitting efforts have led to great progress around the two aspects we mentioned above.

- As far as the modeling is concerned, both test cost and misclassification cost have been explored.

  For example, by considering test cost of the attributes, Yang

* Corresponding author at: School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, PR China. Tel.: +86 511 84433445.

E-mail addresses: douhuili@163.com (H. Dou), yangxibei@hotmail.com (X. Yang), xnsong@yahoo.com.cn (X. Song), yuhualong@just.edu.cn (H. Yu), wuwz@zjou.edu.cn (W.-Z. Wu), yangjy@mail.njust.edu.cn (J. Yang).

http://dx.doi.org/10.1016/j.knosys.2015.09.011
0950-7051/

et al. [43] introduced test cost into the construction of the multi-granulation rough set [35,37,13]; by considering misclassification cost, Yao [47] proposed the concept of the Decision-Theoretic Rough Set (DTRS). Decision-theoretic rough set is actually a probabilistic rough set. Since the determination for a pair of the thresholds used in probabilistic rough set is a substantial challenge, then the pair of thresholds presented in decision-theoretic rough set is calculated by loss function. It must be noticed that for Yao's loss function matrix, not only misclassification cost is used, but also the delayed decision cost is considered. Therefore, decision-theoretic rough set is corresponding to a three-way decision procedure [45]. For more generalizations and applications of decision-theoretic rough set, please refer to Refs. [2,4,14,15, 17–21,23,24,26,27,38,39,42,53].

- As far as the attribute reduction is concerned, both test cost and misclassification cost have also been studied by some researchers. For example, Min et al. [28,31] studied the approaches to test cost based attribute reduction. Their goals are to find reducts with a smaller (obtained by a competition strategy) or the smallest (obtained by a backtracking approach) test costs. Following decision-theoretic rough set, Jia et al. [11,12] formulated an optimization problem. They aimed to minimize the cost of decision. With respect to different requirements, Yao and Zhao [48] studied different definitions of attribute reductions in decision-theoretic rough set.

From discussions above, we can see that decision-theoretic rough set takes the cost into consideration (model is sensitive to cost) and then attribute reduction of decision-theoretic rough set is bound to close to cost. For Yao's classical decision-theoretic rough set, the loss function is a 3×2 cost matrix which includes both misclassification and delayed decision costs. However, one and only one cost matrix may not be good enough for problem solving. We have some practical examples to illustrate the limitations of one cost matrix.

1. Take for instance one medical accident, it is reasonable to assume that different regions may execute different criteria for making compensation, i.e., the costs of the same medical accident may be different in different regions. The developed regions may pay more than that paid by a developing region for economic factors.
2. Here is a China old saying: "Everyone has a steelyard in his heart" and "It is hard to please all", that is, for a given decision, different individuals may have different views which will inevitably generate different costs.
3. Suppose that Mr. X wants to buy a house, if the prices show an up trend, then Mr. X faces different costs during different periods. In other words, one and only one cost is not enough to characterize the variability of cost.

Based on the above examples, it is noticeable that in Yao's decision-theoretic rough set, one and only one cost matrix does not take the characteristics of multiplicity and variability of cost into consideration. Therefore, multiple cost matrices are required. From this point of view, a voting fusion mechanism will be used and then three models are constructed when facing multiple cost matrices. They are referred to as θ (parameterized), optimistic and pessimistic decision-theoretic rough sets in this paper. Note that optimistic and pessimistic models are two limits of parameterized approach.

To facilitate our discussions, we present the basic knowledge about rough set and decision-theoretic rough set in Section 2. In Section 3, not only three multicost based decision-theoretic rough sets are proposed, but also the computations of smallest and largest possible costs are discussed. In Section 4, decision-monotocity criterion based attribute reduction and cost criterion based attribute reduction are presented. The heuristic algorithm is used to compute decision-monotocity reduct while the genetic algorithm is used to compute cost reduct. In Section 5, the theoretical results shown in this paper are tested on eight UCI data sets. The paper ends with conclusions and outlooks for further research in Section 6.

2. Preliminary knowledge on rough sets

2.1. Rough set

An information system can be considered as a pair $I = < U, AT >$, in which U is a non-empty finite set of the objects called the universe; AT is a non-empty finite set of the attributes. $\forall a \in AT, V_a$ is the domain of attribute a. $\forall x \in U, a(x)$ denotes the value that x holds on a ($\forall a \in AT$). Given an information system $I, \forall A \subseteq AT$, an indiscernibility relation $IND(A)$ may be defined as $IND(A) = \{(x,y) \in U^2 : \forall a \in A, a(x) = a(y)\}$.

Obviously, $IND(A)$ is an equivalence relation. $\forall X \subseteq U$, one can construct the lower and upper approximations of X by $IND(A)$ such that

$$\underline{A}(X) = \{x \in U : [x]_A \subseteq X\}, \tag{1}$$

$$\overline{A}(X) = \{x \in U : [x]_A \cap X \neq \emptyset\}; \tag{2}$$

where $[x]_A = \{y \in U : (x,y) \in IND(A)\}$ is the equivalence class of x. The pair $(\underline{A}(X), \overline{A}(X))$ is a Pawlak's rough set of X with respect to A. The positive region of X is $POS^A(X) = \underline{A}(X)$, the boundary region of X is $BND^A(X) = \overline{A}(X) - \underline{A}(X)$, and the negative region of X is $NEG^A(X) = U - \overline{A}(X)$.

2.2. Decision-theoretic rough set

For a Bayesian decision procedure, a finite set of the states can be denoted by $\Omega = \{\omega_1, \omega_2, \ldots, \omega_s\}$. A finite set of t possible actions can be denoted by $\mathcal{A} = \{a_1, a_2, \ldots, a_t\}$. $\forall x \in U$, let $Pr(\omega_j|x)$ be the conditional probability of object x being in state $\omega_j, \lambda(a_i|\omega_j)$ be the loss, or cost for taking action a_i when state is ω_j. Suppose that we take the action a_i for object x, then the expected loss is

$$\mathcal{R}(a_i|x) = \sum_{j=1}^{s} \lambda(a_i|\omega_j) \cdot Pr(\omega_j|x). \tag{3}$$

For Yao's decision-theoretic rough set model, the set of states is composed by two classes such that $\Omega = \{X, \sim X\}$. It indicates that an object is in class X or out of class X; the set of actions is given by $\mathcal{A} = \{a_P, a_B, a_N\}$, in which a_P, a_B and a_N expresses three actions: a_P indicates that x is classified into decision-theoretic positive region of X, i.e., $POS_{DT}^A(X)$; a_B indicates that x is classified into decision-theoretic boundary region of X, i.e., $BND_{DT}^A(X)$; a_N indicates that x is classified into decision-theoretic negative region of X, i.e., $NEG_{DT}^A(X)$. The loss function regarding the costs of three actions in two different states is given in Table 1. Obviously, Table 1 is a 3×2 matrix. It is denoted by **M** in this paper.

In Table 1, $\lambda_{PP}, \lambda_{BP}$ and λ_{NP} are the losses for taking actions of a_P, a_B and a_N, respectively, when stating x is included into X; $\lambda_{PN}, \lambda_{BN}$ and λ_{NN} are the losses for taking actions of a_P, a_B and a_N, respectively, when stating x is out of X. $\forall x \in U$, by using the conditional probability $Pr(X|[x]_A)$, the expected losses associated with three actions are: