Minimal decision cost reduct in fuzzy decision-theoretic rough set model

Jingjing Song [a,*], Eric C.C. Tsang [a], Degang Chen [b], Xibei Yang [c]

[a] *Faculty of Information Technology, Macau University of Science and Technology, Macau, PR China*
[b] *School of Mathematics and Physics, North China Electric Power University, Beijing 102206, PR China*
[c] *School of Computer Science and Technology, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu Province 212003, PR China*

ARTICLE INFO

ABSTRACT

Decision-theoretic rough set model interprets the parameters of existing probabilistic rough sets by introducing Bayesian theory to minimize the risks of classification. It is a representative probabilistic rough set model. As an important extension of decision-theoretic rough set model, fuzzy decision-theoretic rough set model can deal with numerical data directly. To investigate the attribute reduction in fuzzy decision-theoretic rough set model, we propose two attribute reduction approaches: global reduction and local reduction. Global reduction can keep the cost of all the decision classes unchanged or decreased. Moreover, local reduction can make the cost of single decision class unchanged or decreased. In this paper, we use two different heuristic functions to implement heuristic algorithms. Heuristic algorithms are used to find the minimal subset of condition attributes which minimizes the (P) cost, (B) cost and (N) cost in fuzzy decision theoretic-rough set model. Six UCI datasets are employed to test our heuristic algorithms. The experimental results indicate that global reduction and local reduction can minimize the decision cost in fuzzy decision-theoretic rough set model effectively. After reduction, the numbers of condition attributes are reduced greatly. On the other hand, global reduction and local reduction both can decrease the numbers of (P) rule and (B) rule on the whole. The results show that when we reduce the decision cost, we will also reduce the rule numbers as well.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The decision-theoretic rough set model(DTRS) proposed by Yao et al. [1] was a representative probabilistic rough set model [2–5]. DTRS introduced Bayesian theory to minimize the risks of classification, which interprets the parameters from existing probabilistic approaches to rough sets. For DTRS, the pair of thresholds can be calculated by the loss function with minimal risk, where the losses are related to the risk of decision-making.

Up to now, DTRS has gained increasing attentions from the rough set community. For example, Sun et al. [6] proposed the decision-theoretic rough fuzzy set model to approximate the fuzzy object in probabilistic approximation space. By considering the new evaluation format of loss function with intuitionistic fuzzy numbers [7], Liang et al. [8] gave the model of intuitionistic fuzzy decision-theoretic rough set. By combining multigranulation rough set model [9] with DTRS, Qian et al. [10] proposed multigranulation decision-theoretic rough sets. Feng et al. [11] com-

bined the variable precision rough set model with multigranulation fuzzy rough set model by considering the noise attributes, and then proposed the variable precision multigranulation decision-theoretic fuzzy rough sets. Li et al. [12] introduced the concept of neighborhood based decision-theoretic rough set model to deal with numerical data directly. By considering relative and absolute quantitative information between the class and concept, Xu et al. [13] proposed two kinds of generalized multigranulation double-quantitative decision-theoretic rough sets, which are feasible when making decisions in real life. Fan et al. [14] proposed a couple of double-quantitative decision-theoretic rough fuzzy set models based on logical operations. Within the framework of interval-valued fuzzy probabilistic approximation space, Zhao et al. [15] established the interval-valued fuzzy decision-theoretic rough set model. Jia et al. [16] proposed an optimization problem to minimize the decision cost and then defined a minimal cost attribute reduction based on DTRS.

Similar to classical rough set [17], attribute reduction [18,19] also plays a crucial role in the development of DTRS, which can be regarded as finding the minimal condition attribute subset to preserve one or several criteria unchanged or improved.

* Corresponding author.
  E-mail address: songjingjing108@163.com (J. Song).

In DTRS, the lower approximation and upper approximation are decided by the conditional probability and the thresholds (α, β) together. Many traditional region-related attribute reductions can not be employed in DTRS as the monotonic property of regions does not hold [1]. DTRS encountered some challenges to compute attribute reduction. Yao et al. [20] discussed various criteria on how to systematically perform attribute reduction using DTRS, which provided a new insight into the problem of attribute reduction. Li et al. [21] redefined the attribute reduction in DTRS by replacing the positive region unchanged with the positive region being expanded. Ma et al. [22] proposed a solution to attribute reduction in DTRS based on decision region preservation. Jia et al. [23] regarded attribute reduction as an optimization problem in DTRS, and then gave a heuristic approach to minimize the decision cost. Presently, attribute reductions of DTRS have been applied to many fields successfully, both in theories and methodologies [24–31]. The results of these researches increase our understanding of DTRS and its applications.

In rough set theory, there are two kinds of costs, i.e., test cost and decision cost. On one hand, in many practical applications, data are not free, and there is a test cost for each data item. For example, in a clinic system, a patient is often required to undergo a number of medical tests, money or time are involved in performing these tests. On the other hand, DTRS is a widely used model which considers the decision cost and provides a semantic explanation for positive, boundary and negative regions. Therefore, decision cost is an important notion in DTRS which can be regarded as the objective criterion for defining an attribute reduction. In other words, decision cost is a kind of classification cost [23]. In real life, there exist many issues with decision cost. For example, bank policymakers need to decide whether the bank should accept loan applications or not. Then the bank policymakers have to consider the decision cost which can minimize the loss of the bank. The decision cost that the bank loans to the eligible applicants is less than the decision cost the bank loans to the applicants who do not meet the application requirements.

Notably, many numerical data existed in practice. However, Yao's classical DTRS can not deal with numerical data directly. To solve this problem, Wang et al. [19] redefined the conditional probability from the viewpoint of fuzzy membership degree, and then proposed the fuzzy decision-theoretic rough set model(FDTRS). The motivation of this paper is to minimize the decision cost [18,23,32] of FDTRS. By employing a decreasing cost attribute reduction, we ensure the decision cost will be decreased or unchanged. With this purpose, we propose two attribute reductions in FDTRS: global reduction and local reduction. Global reduction considers all decision classes together, while local reduction [33] mainly considers key conditional attributes for special decision class. In the context of our paper, local reduction only considers single decision class. The innovation of this study includes three parts. Firstly, we research the attribute reduction in FDTRS from the viewpoint of cost. Secondly, local reduction which considers key condition attributes for special decision class is proposed to FDTRS. Thirdly, two different heuristic functions are employed to compute the local reduction for FDTRS.

The remainder of this paper is organized as follows. Section 2 briefly reviews some basic notions of FDTRS. In Section 3, we use two different heuristic functions to construct heuristic algorithms, and then propose two approaches, i.e., global reduction and local reduction, to minimize the decision cost of FDTRS. The experiments are carried out in Section 4, and then we analyze the experimental results. This paper ends with conclusions in Section 5.

2. Preliminaries

In this section, we present some basic definitions of FDTRS.

2.1. Relevant knowledge of FDTRS

Formally, an information system can be described as $S = \{U, AT, V, f\}$, where U is a nonempty finite set, called the universe of discourse; AT is the set of attributes; $\forall a \in AT$, V_a is the domain of attribute a; $V = \cup V_a$ is the set of all attribute values; f: $U \times AT \rightarrow V$ is an information function. Specially, the information systems $S = \{U, AT \cup D, V, f\}$, where AT is the set of condition attributes, $D = \{d\}$ is the decision attribute, then the information system S is called a decision system. $U/IND\{D\} = \{X_1, X_2, \ldots, X_n\}$ is the partition induced by the decision attribute D[17]. In the context of this paper, the system we considered is a decision system.

Let U be the universe of discourse, suppose that F is a mapping from U to the interval [0,1], that is, F: $U \rightarrow [0, 1]$ such that $F(x) \in [0, 1]$ for each $x \in U$. Then, F is called a fuzzy set [34] over U where $F(x)$ is the membership value of x. In the context of this paper, $F(U)$ is considered as the set of all fuzzy sets over U.

Definition 1. Given a decision system $S = \{U, AT \cup D, V, f\}$, $\forall R \subseteq AT$, if the binary fuzzy relation μ_R on U satisfies the following conditions:

1. Reflexivity: $\forall x \in U, \mu_R(x, x) = 1$.
2. Symmetric: $\forall x, y \in U, \mu_R(x, y) = \mu_R(y, x)$.
3. T-transitivity: $\forall x, y, z \in U, T(\mu_R(x, y), \mu_R(y, z)) \leq \mu_R(x, z)$.

then μ_R is called the fuzzy T-equivalence relation, where $\mu_R(x, y)$ is the membership function of fuzzy relation μ_R and T is a triangular norm.

Definition 2 [35]**.** Given a nonempty and finite set U, a real-valued function k: $U \times U \rightarrow R$ is said to be a kernel if it is symmetric, that is, $k(x, y) = k(y, x)$ for $\forall x, y \in U$, and positive-semidefinite.

Theorem 1 [36]**.** *Any kernel k: $U \times U \rightarrow [0, 1]$ with $k(x, x) = 1$ is (at least) $T_{\cos}$-transitivity, where*

$$T_{\cos}(a, b) = max\left(ab - \sqrt{1 - a^2}\sqrt{1 - b^2}, 0 \right). \tag{1}$$

Theorem 1 shows that any kernel satisfying reflexivity and symmetry is at least $T_{\cos}$-transitivity. Then, the relation computed with this kind of kernel functions is a fuzzy T-equivalence relation. In this paper, we employ a Gaussian kernel function [37,38] to construct the fuzzy relation of the objects.

Given a decision system $S = \{U, AT \cup D, V, f\}$, $\forall R \subseteq AT$, $\forall x, y \in U$, the fuzzy similarity between x and y is

$$\mu_R(x, y) = exp\left(-\frac{\|x - y\|_R^2}{2\sigma^2} \right), \tag{2}$$

where $\|x - y\|_R^2 = \sum_{i=1}^n (f(x, a_i) - f(y, a_i))^2$, $n = |R|$ is the cardinality of R.

It is obvious that the fuzzy relation constructed by Gaussian kernel function is a fuzzy T-equivalence relation by Theorem 1. Eq. (2) shows the fuzzy T-equivalence relation by measuring the similarity between all pairs of objects.

Definition 3 [19]**.** Given a decision system $S = \{U, AT \cup D, V, f\}$ and fuzzy T-equivalence relation μ_R, $\forall x, y \in U$, the fuzzy T-equivalence class of object x is

$$[x]_{\mu_R} = \sum_{y \in U} \mu_R(x, y)/y \tag{3}$$