

DOI:10.13232/j.cnki.jnju.2023.05.009

## 属性集变化下序决策信息系统的增量属性约简算法

张义宗<sup>1,2</sup>, 王磊<sup>1,2\*</sup>, 徐阳<sup>1,2</sup>

(1. 南昌工程学院信息工程学院, 南昌, 330099; 2. 江西省水信息协同感知与智能处理重点实验室, 南昌, 330099)

**摘要:** 当序决策信息系统中的属性集不断变化时, 基于优势关系的现有静态算法无法高效地更新其属性约简, 为此, 从属性增加和属性删除两个角度出发, 以知识粒度表征的属性重要度为启发信息, 提出两种新的增量属性约简算法. 首先介绍优势粗糙集方法的相关基础知识, 并将经典粗糙集中基于知识粒度的属性约简算法扩展到优势粗糙集方法, 得到可处理序决策信息系统的属性约简算法; 然后, 给出劣势属性矩阵的定义, 并基于知识粒度的矩阵计算方法分析属性增删时属性约简的增量式更新机制, 进一步设计了两种增量属性约简算法. 最后, 分析比较三种算法的时间复杂度, 选取了六个不同的 UCI 数据集进行算法性能的测试, 结果表明, 提出的算法比静态的属性约简算法更高效.

**关键词:** 序决策信息系统, 知识粒度, 优势粗糙集方法, 劣势属性矩阵, 增量属性约简

中图分类号: TP391

文献标志码: A

## Incremental attribute reduction algorithm for ordered decision information systems with the change of attribute set

Zhang Yizong<sup>1,2</sup>, Wang Lei<sup>1,2\*</sup>, Xu Yang<sup>1,2</sup>

(1. School of Information Engineering, Nanchang Institute of Technology, Nanchang, 330099, China;

2. Jiangxi Province Key Laboratory of Water Information Cooperative Sensing and Intelligent Processing,  
Nanchang, 330099, China)

**Abstract:** When attribute set in the ordered decision information system is constantly changing, existing static algorithms that have been studied based on dominance relationship cannot efficiently update its attribute reduction. To this end, this paper proposes two new incremental attribute reduction algorithms from the perspective of both attribute addition and attribute deletion, respectively, using the attribute importance of knowledge granularity representations as the heuristic information. Firstly, the relevant basic knowledge of the dominance rough set method are introduced, and the attribute reduction algorithm based on knowledge granularity in the classical rough set is extended to the dominance rough set method to obtain an attribute reduction algorithm that can handle ordered decision information systems; Then, the definition of the inferior attribute matrix is given, and the incremental update mechanism of attribute reduction during attribute addition and deletion is analyzed by the matrix calculation method of knowledge granularity. From there, two incremental attribute reduction algorithms are further designed; Finally, time complexity of the three algorithms is analyzed and compared, and six different UCI datasets are selected to test the algorithm performance. The test results show that the algorithm proposed in this paper is more efficient than the static attribute reduction algorithm.

**Key words:** ordered decision information system, knowledge granularity, the dominance rough set approach, inferior attribute matrix, incremental attribute reduction

基金项目: 国家自然科学基金(61562061), 江西省教育厅科技项目(GJJ2202005, GJJ211920)

收稿日期: 2023-07-03

\* 通讯联系人, E-mail: ezhoulel@163.com

粗糙集理论是 1982 年波兰数学家 Pawlak<sup>[1]</sup>提出的一种能够有效处理带有不确定性、模糊性、不精确性数据的数学工具,由于其解决实际问题的有效性,被大量应用于机器学习<sup>[2-3]</sup>、模式识别<sup>[4]</sup>、数据挖掘<sup>[5]</sup>和知识发现<sup>[6]</sup>等领域.属性约简是粗糙集理论中的核心与热点研究问题之一,旨在找出与决策信息系统分类能力一致的属性子集,达到决策信息系统属性降维的目的.

经典粗糙集理论以等价关系对论域进行划分形成的等价类为研究基础,但其不适合处理属性值具有偏序关系的数据.为此, Greco et al<sup>[7-8]</sup>基于优势关系对经典粗糙集方法进行推广,提出优势粗糙集方法(Dominance Rough Set Approach, DRSA),而 DRSA 近似集的运算对象是决策类的上(下)向联合集,由此给出了上(下)向联合集的上、下近似集的定义.此后,众多学者对优势粗糙集方法进行了一系列的研究和推广.李艳等<sup>[9]</sup>深入研究了不协调目标信息系统的属性约简,在优势关系上给出了浓缩布尔矩阵的概念来计算属性约简. Li et al<sup>[10]</sup>针对区间值序信息系统提出一种基于优势关系的特征选择(属性约简)方法. Du and Hu<sup>[11]</sup>针对一致不完备序信息系统的属性约简问题,基于可分辨矩阵和可分辨函数,提出一种计算所有属性约简的方法. Yang et al<sup>[12]</sup>将优势关系拓展到区间值决策系统上,提出基于  $\alpha$ -优势的粗糙集模型,并给出了上、下近似的计算方法. Zhang and Yang<sup>[13]</sup>将  $\alpha$ -优势推广至集值信息系统,结合合取和析取给出了两种新的优势关系,并基于此提出集值决策表的特征选择(属性约简)方法.以上研究均是对优势粗糙集模型的扩展和推广,可解决不同类型数据集的属性约简问题.

然而,实际生活中数据的属性集会发动态变化,相应的属性约简结果也随之变化,使用非增量属性约简算法对属性集动态变化的数据进行属性约简是不可行的,因为这会重复计算未变动前的部分属性,增加计算成本,在大数据集上甚至无法满足其属性约简的效率要求.因而,众多学者针对属性集动态变化的序决策信息系统展开了一系列研究,如 Luo et al<sup>[14]</sup>将知识粒度引入序集值决策信息系统,对系统中属性增加和删除的情况,基于矩阵提出一种增量更新近似集的方法.

Wang et al<sup>[15]</sup>针对有序信息系统多维变化的动态更新近似集的问题,基于矩阵提出一种可以在对象和属性同时增加的情况下有效更新的增量方法. Huang et al<sup>[16]</sup>将近似集的布尔矩阵表示方法引入动态模糊信息系统,考虑属性和对象增加的情况,通过更新布尔矩阵达到动态更新近似集的目的. Sang et al<sup>[17]</sup>将条件熵引入序决策信息系统,在系统中添加或删除多个对象时,基于矩阵提出两种增量属性约简算法,还针对动态变化的有序模糊信息系统的特征选择(属性约简)问题,提出一种新的模糊优势邻域粗糙集模型,并基于条件熵提出一种启发式增量特征选择(属性约简)算法<sup>[18]</sup>.上述研究推动了序决策信息系统近似集的动态更新在众多领域的发展,但是,针对属性集动态变化的序决策信息系统,基于矩阵计算其属性约简的增量算法并不常见.

本文以序决策信息系统为研究对象,首先将王磊和李天瑞<sup>[19]</sup>的知识粒度的矩阵表示与计算方法推广到序决策信息系统中,同时,将经典粗糙集中以知识粒度表征的属性重要度为启发信息的属性约简方法引入优势粗糙集方法,作为本文的非增量对照实验算法.然后,以矩阵分析序决策信息系统中知识粒度、劣势元素矩阵和优势关系矩阵在属性数目变化条件下的变化过程,并给出属性增删条件下属性约简的增量更新机制,基于此提出两种新的增量属性约简算法.最后,在 UCI 数据集上测试了属性约简算法的性能,实验结果证实了本文提出的增量属性约简算法的可行性和高效性.

本文提出的劣势元素矩阵和优势关系矩阵的增量更新机制能降低属性约简子集选取属性的计算成本,进一步增加算法的增量属性约简效率,同时,其不仅适用于序决策信息系统,还可扩展到其他模型的属性约简算法.究其本质,劣势元素矩阵和优势关系矩阵的增量更新机制的提出,给出了基于满足模型关系和不满足模型关系相结合的属性约简方法.

## 1 基本知识

首先介绍优势粗糙集方法的相关基础知识,然后,将知识粒度的矩阵表示与计算方法扩展到

优势粗糙集方法中,介绍了以知识粒度为启发信息的启发式属性约简算法.

### 1.1 优势粗糙集方法的相关基础知识

**定义 1** 决策信息系统  $S=(U, A=C \cup D, V, f)$ , 其中,  $U$  为非空有限对象集合,  $U=\{x_1, x_2, \dots, x_m\}$ ;  $A$  是一个有限非空属性集,  $C$  是条件属性集,  $D$  是决策属性集, 并且  $C \cap D = \emptyset$ ; 对于  $\forall a \in A$ , 均存在  $a$  的一个属性值集  $V_a$ ,  $V = \bigcup V_a$  为属性集  $A$  的属性值集;  $f$  为信息系统的信息函数, 能给出属性集  $A$  在  $U$  上的属性值集. 若属性值集  $V$  具有偏序关系, 则称该系统为序决策信息系统.

**定义 2** 给定一个序决策信息系统  $S=(U, A=C \cup D, V, f)$ , 对于  $\forall P \subseteq C$  均会确定一个优势关系  $R_P^{\leq}$ , 表示为:

$$R_P^{\leq} = \{(x_i, x_j) \in U \times U \mid \forall a \in P, f(x_i, a) \leq f(x_j, a)\} \quad (1)$$

此外, 对于  $\forall P \subseteq C$  也会确定一个劣势关系  $R_P^>$ , 可表示为:

$$R_P^> = \{(x_i, x_j) \in U \times U \mid \forall a \in P, f(x_i, a) > f(x_j, a)\} \quad (2)$$

其中,  $f(x_i, a)$  为对象  $x_i$  在属性  $a$  上的属性值,  $f(x_j, a)$  为对象  $x_j$  在属性  $a$  上的属性值.

**定义 3** 给定一个序决策信息系统  $S=(U, A=C \cup D, V, f)$ ,  $P \subseteq C$ , 那么属性子集  $P$  在  $U$  上的优势关系矩阵为  $M_U^{R_P^{\leq}}$ .  $R(i, j)$  表示  $M_U^{R_P^{\leq}}$  第  $i$  行第  $j$  列的元素, 其值为:

$$R(i, j) = \begin{cases} 1, & f(x_i, a) \leq f(x_j, a), \forall a \in P \\ 0, & f(x_i, a) > f(x_j, a), \exists a \in P \end{cases} \quad (3)$$

**定义 4**  $S=(U, A=C \cup D, V, f)$ ,  $P \subseteq C$ , 则属性集  $P \cup D$  在  $U$  上的优势关系矩阵为  $M_U^{R_{P \cup D}^{\leq}}$ .  $R(i, j)$  表示  $M_U^{R_{P \cup D}^{\leq}}$  第  $i$  行第  $j$  列的元素, 其值为:

$$R(i, j) = \begin{cases} 1, & f(x_i, a) \leq f(x_j, a) \wedge f(x_i, d) = f(x_j, d), \\ & \forall a \in P \\ 0, & \text{其他} \end{cases} \quad (4)$$

**定义 5** 给定一个序决策信息系统  $S=(U, A=C \cup D, V, f)$ ,  $P \subseteq A$ , 那么  $P$  的知识粒度记

为  $GK(P)$ , 定义为:

$$GK(P) = \frac{\sum_{i=1}^{|U|} \sum_{j=1}^{|U|} R(i, j)}{|U|^2} = \text{mean}(M_U^{R_P^{\leq}}) \quad (5)$$

其中,  $\text{mean}(M_U^{R_P^{\leq}})$  为优势关系矩阵  $M_U^{R_P^{\leq}}$  的算术平均值.

**定义 6**  $S=(U, A=C \cup D, V, f)$ ,  $P, Q \subseteq A$ , 知识  $Q$  关于知识  $P$  在  $U$  上的相对知识粒度为  $GK(Q|P)$ , 可表示为:

$$GK(Q|P) = GK(P) - GK(P \cup Q) = \text{mean}(M_U^{R_P^{\leq}}) - \text{mean}(M_U^{R_{P \cup Q}^{\leq}}) \quad (6)$$

**定义 7**  $S=(U, A=C \cup D, V, f)$ ,  $P \subseteq C$ ,  $\forall a \in C - P$ , 那么属性  $a$  相对于属性集  $P$  的外部属性重要度可定义为:

$$\text{Sig}_U^{\text{outer}}(a, P, D) = \text{mean}(M_U^{R_P^{\leq}}) - \text{mean}(M_U^{R_{P \cup \{a\}}^{\leq}}) - \text{mean}(M_U^{R_{P \cup \{a\}}^{\leq}}) + \text{mean}(M_U^{R_{P \cup \{a\} \cup D}^{\leq}}) \quad (7)$$

同理, 若  $\forall a \in P$ , 属性  $a$  相对于属性集  $P$  的内部属性重要度可定义为:

$$\text{Sig}_U^{\text{inner}}(a, P, D) = \text{mean}(M_U^{R_{P-\{a\}}^{\leq}}) - \text{mean}(M_U^{R_{P-\{a\} \cup D}^{\leq}}) - \text{mean}(M_U^{R_P^{\leq}}) + \text{mean}(M_U^{R_{P \cup D}^{\leq}}) \quad (8)$$

**定义 8**  $S=(U, A=C \cup D, V, f)$ ,  $R \subseteq C$ ,  $R$  是条件属性集  $C$  相对于决策属性集  $D$  的一个属性约简, 则  $R$  必须满足以下两个条件:

- (1)  $GK(D|R) = GK(D|C)$ ;
- (2)  $\forall a \in R$ , 使得  $GK(D|R - \{a\}) \neq GK(D|C)$ .

**1.2 优势粗糙集方法中的启发式非增量属性约简算法** 根据 Jing et al<sup>[20]</sup> 的经典粗糙集中基于知识粒度的属性约简算法, 将其引入优势粗糙集方法, 构造以知识粒度表征的属性重要度为启发信息的属性约简算法, 如算法 1 所示.

**算法 1** 优势关系下基于知识粒度的属性约简算法 (Attribute Reduction Algorithm Based on Knowledge Granularity under Dominance Relationship, ARKG-DR)

输入: 序决策信息系统  $S=(U, A=C \cup D, V, f)$ .

输出: 一个属性约简  $RED$ .

步骤 1. 计算  $GK(D|C)$ .

步骤 2. 初始化  $RED \leftarrow \emptyset$ ,  $RED\_Pool \leftarrow C$ .

步骤 3. 对于  $\forall a \in C$ , 计算  $\text{Sig}_U^{\text{inner}}(a, C, D)$ , 如果  $\text{Sig}_U^{\text{inner}}(a, C, D) > 0$ , 则  $RED \leftarrow RED \cup \{a\}$ ,  $RED\_Pool \leftarrow$

$RED\_Pool - \{a\}$ .

步骤 4. 计算  $GK(D|RED)$ , 如果  $GK(D|C) = GK(D|RED)$ , 则转至步骤 8, 否则转至步骤 5.

步骤 5. 对于  $\forall a \in RED\_Pool$ , 计算  $Sig_U^{outer}(a, RED, D)$ , 得出  $Sig_U^{outer}(a_m, RED, D) = \max(Sig_U^{outer}(a, RED, D))$ , 则  $RED \leftarrow RED \cup \{a_m\}$ ,  $RED\_Pool \leftarrow RED\_Pool - \{a_m\}$ .

步骤 6. 计算更新后的  $GK(D|RED)$ , 若  $GK(D|C) = GK(D|RED)$ , 则转至步骤 5, 否则转至步骤 7.

步骤 7. 对于  $\forall a \in RED$ , 如果  $GK(D|C) = GK(D|RED - \{a\})$ , 则  $RED \leftarrow RED - \{a\}$ .

步骤 8. 输出属性约简结果  $RED$ .

然而, 对属性集动态变化的数据集进行属性约简时, 算法 1 会重新计算变化后的数据集, 这会重复计算上次的属性约简结果, 极大地提升属性集动态变化数据集的属性约简时间复杂度. 此外, 在步骤 5 和步骤 7 中增加或者删除某个属性后, 需要重新计算更新对应的优势关系矩阵, 重新判断对象在未变化属性上的优势关系, 很大程度上降低了算法属性约简的效率. 针对这些问题, 本文研究了属性集变化下属性约简的增量更新机制, 包括知识粒度的矩阵增量更新方法、优势关系矩阵的增量更新方法和劣势元素矩阵的增量更新方法, 基于此设计了启发式增量属性约简算法.

## 2 属性集变化下序决策信息系统的增量属性约简算法

为了提高属性集动态变化的数据集的属性约简效率, 通过矩阵分析序决策信息系统中知识粒度在属性数目变化条件下的增量更新机制, 进一步分析优势关系矩阵和劣势元素矩阵的增量更新机制, 最后在属性集增加和属性集删除的条件下, 分别设计了两种启发式增量属性约简算法.

### 2.1 属性集增加条件下的增量更新机制

**定义 9** 给定一个序决策信息系统  $S = (U, A = C \cup D, V, f)$ ,  $P \subseteq A$ ,  $M_U^{R_P} = (M_{ij}^{R_P})_{|U| \times |U|}$  是属性集  $P$  在论域  $U$  上的优势关系矩阵. 假设  $Q$  是新增的属性集,  $M_U^{R_Q} = (M_{ij}^{R_Q})_{|U| \times |U|}$  是属性集  $Q$  在论域  $U$  上的优势关系矩阵, 那么  $P$  上的增量关系矩阵为  $\Delta M_U^{R_P}$ , 表示为:

$$\Delta M_U^{R_P} = (r_{ij})_{|U| \times |U|} = \begin{cases} 1, & M_{ij}^{R_P} = 1 \text{ 且 } M_{ij}^{R_Q} = 0 \\ 0, & \text{其他} \end{cases} \quad (9)$$

根据文献[19]可知, 数据集  $P \cup Q$  的优势关系矩阵  $M_U^{R_{P \cup Q}}$  是在数据集  $P$  的优势关系矩阵  $M_U^{R_P}$  上变化而来, 增加属性集  $Q$  只会令  $M_U^{R_P}$  中的 1 变为 0, 对应变化的位置会标记为 1, 从而构成了  $P$  的增量关系矩阵.

**定理 1** 给定一个序决策信息系统  $S = (U, A = C \cup D, V, f)$ ,  $P \subseteq C$ ,  $Q$  是新增的属性集,  $GK(P)$  为属性集  $P$  的知识粒度, 那么新增属性集后更新的知识粒度为:

$$GK(P \cup Q) = \text{mean}(M_U^{R_{P \cup Q}}) = GK(P) - \text{mean}(\Delta M_U^{R_P}) \quad (10)$$

**定理 2** 给定一个序决策信息系统  $S = (U, A = C \cup D, V, f)$ ,  $P \subseteq C$ ,  $Q$  是新增的属性集,  $GK(D|P)$  为属性集  $P$  相对于属性集  $D$  的相对知识粒度, 则新增属性集后更新的相对知识粒度为:

$$GK(D|P \cup Q) = GK(D|P) - \text{mean}(\Delta M_U^{R_P}) + \text{mean}(\Delta M_U^{R_{P \cup D}}) \quad (11)$$

证明如下:

$$\begin{aligned} GK(D|P \cup Q) &= GK(P \cup Q) - GK(P \cup Q \cup D) = \\ &= \text{mean}(M_U^{R_{P \cup Q}}) - \text{mean}(M_U^{R_{P \cup Q \cup D}}) = \\ &= GK(P) - \text{mean}(\Delta M_U^{R_P}) - \\ &= GK(P \cup D) + \text{mean}(\Delta M_U^{R_{P \cup D}}) = \\ &= GK(D|P) - \text{mean}(\Delta M_U^{R_P}) + \text{mean}(\Delta M_U^{R_{P \cup D}}) \end{aligned}$$

**定理 3** 给定一个序决策信息系统  $S = (U, A = C \cup D, V, f)$ ,  $P \subseteq C$ ,  $Q$  是新增的属性集,  $\forall a \in (C - P) \cup Q$ ,  $Sig_U^{outer}(a, P, D)$  为属性  $a$  相对于属性集  $P$  关于属性集  $D$  的外部重要度, 可表示为:

$$\begin{aligned} Sig_U^{outer}(a, P, D) &= GK(D|P) - GK(D|P \cup \{a\}) = \\ &= \text{mean}(\Delta M_U^{R_P}) - \text{mean}(\Delta M_U^{R_{P \cup D}}) \end{aligned} \quad (12)$$

证明如下:

$$\begin{aligned} Sig_U^{outer}(a, P, D) &= GK(D|P) - GK(D|P \cup \{a\}) = \\ &= GK(D|P) - GK(D|P) + \\ &= \text{mean}(\Delta M_U^{R_P}) - \text{mean}(\Delta M_U^{R_{P \cup D}}) = \\ &= \text{mean}(\Delta M_U^{R_P}) - \text{mean}(\Delta M_U^{R_{P \cup D}}) \end{aligned}$$



**定义 10**  $S=(U, A=C \cup D, V, f), P \subseteq A$ ,  $Q$  是新增的属性集,  $M_U^{R_P^{\leq}} = (M_{ij}^{R_P^{\leq}})_{|U| \times |U|}$  是属性集  $P$  在论域  $U$  上的优势关系矩阵,  $\Delta M_U^{R_P^{\leq}} = (r_{ij})_{|U| \times |U|}$  是增加属性集  $Q$  后增量关系矩阵, 那么属性集  $P \cup Q$  的优势关系矩阵  $M_U^{R_{P \cup Q}^{\leq}}$  中的元素  $M_{ij}^{R_{P \cup Q}^{\leq}}$  可表示为:

$$M_{ij}^{R_{P \cup Q}^{\leq}} = \begin{cases} 1, & M_{ij}^{R_P^{\leq}} = 1 \text{ 且 } r_{ij} = 0 \\ 0, & \text{其他} \end{cases} \quad (13)$$

## 2.2 属性集删除条件下的增量更新机制

**定义 11** 给定一个序决策信息系统  $S=(U, A=C \cup D, V, f), P \subseteq C, P=\{c_1, c_2, \dots, c_n\} (n \leq |C|), \forall x_i, x_j \in U$ , 那么,  $U$  在  $P$  上的劣势属性矩阵为  $M_U^{R_P^{\geq}}, M_{ij}^{R_P^{\geq}}$  是  $M_U^{R_P^{\geq}}$  第  $i$  行第  $j$  列的元素, 其本质是一个条件属性子集, 表示为:

$$M_{ij}^{R_P^{\geq}} = \{c_i \in P \mid x_i, x_j \in U, f(x_i, c_i) > f(x_j, c_i)\} \quad (14)$$

同理,  $U$  在  $P \cup D$  上的劣势属性矩阵为  $M_U^{R_{P \cup D}^{\geq}}, M_{ij}^{R_{P \cup D}^{\geq}}$  是  $M_U^{R_{P \cup D}^{\geq}}$  第  $i$  行第  $j$  列的元素, 其本质也是一个条件属性子集, 表示为:

$$M_{ij}^{R_{P \cup D}^{\geq}} = \begin{cases} \{c_i \in P \mid x_i, x_j \in U, f(x_i, c_i) > f(x_j, c_i)\}, & f(x_i, d) = f(x_j, d) \\ P, & f(x_i, d) \neq f(x_j, d) \end{cases} \quad (15)$$

此外, 当删除或者增加属性集  $Q$  时,  $U$  在属性集  $P-Q, P \cup D-Q, P \cup Q$  和  $P \cup Q \cup D$  上的劣势属性矩阵元素都会发生相应的增量变化, 具体表示为:

$$M_{ij}^{R_{\Delta}^{\geq}} = \begin{cases} M_{ij}^{R_P^{\geq}} - Q, & P - Q \\ M_{ij}^{R_{P \cup D}^{\geq}} - Q, & P \cup D - Q \\ M_{ij}^{R_P^{\geq}} \cup M_{ij}^{R_Q^{\geq}}, & P \cup Q \\ M_{ij}^{R_{P \cup D}^{\geq}} \cup M_{ij}^{R_{Q \cup D}^{\geq}}, & P \cup D \cup Q \end{cases} \quad (16)$$

**定义 12** 给定一个序决策信息系统  $S=(U, A=C \cup D, V, f), Q \subseteq P \subseteq A, M_U^{R_P^{\leq}} = (M_{ij}^{R_P^{\leq}})_{|U| \times |U|}$  是属性集  $P$  在论域  $U$  上的优势关系矩阵,  $M_U^{R_P^{\geq}} = (M_{ij}^{R_P^{\geq}})_{|U| \times |U|}$  是属性集  $P$  在论域  $U$  上的劣势属性矩阵, 假设  $Q$  是  $S$  删除的属性集 (若  $P$  中包含  $D$ , 那么  $|Q| < |P-D|$ ), 那么  $P$  上的增量关系矩阵为  $\Delta M_U^{R_P^{\leq}}$ , 表示为:

$$\Delta M_U^{R_P^{\leq}} = (r_{ij})_{|U| \times |U|} = \begin{cases} 1, & M_{ij}^{R_P^{\leq}} = 0 \text{ 且 } M_{ij}^{R_Q^{\leq}} \subseteq Q \\ 0, & \text{其他} \end{cases} \quad (17)$$

根据文献[19]可知, 数据集  $P-Q$  的优势关系矩阵  $M_U^{R_{P-Q}^{\leq}}$  是在数据集  $P$  的优势关系矩阵  $M_U^{R_P^{\leq}}$  上变化而来, 删除属性集  $Q$  只可能令  $M_U^{R_P^{\leq}}$  中的 0 变为 1, 对应变化的位置会标记为 1, 构成  $P$  的增量关系矩阵.

**定理 4** 给定一个序决策信息系统  $S=(U, A=C \cup D, V, f), Q \subseteq P \subseteq C, Q$  是  $S$  删除的属性集,  $GK(P)$  为属性集  $P$  的知识粒度, 那么新增属性集后更新的知识粒度为:

$$GK(P-Q) = \text{mean}(M_U^{R_{P-Q}^{\leq}}) = GK(P) + \text{mean}(\Delta M_U^{R_P^{\leq}}) \quad (18)$$

**定理 5** 给定一个序决策信息系统  $S=(U, A=C \cup D, V, f), Q \subseteq P \subseteq C, Q$  是  $S$  删除的属性集,  $GK(D|P)$  为属性集  $P$  相对于属性集  $D$  的相对知识粒度, 那么新增属性集后的相对知识粒度为:

$$GK(D|P-Q) = GK(D|P) + \text{mean}(\Delta M_U^{R_P^{\leq}}) - \text{mean}(\Delta M_U^{R_{P \cup D}^{\leq}}) \quad (19)$$

证明过程与定理 2 的证明过程类似, 略.

**定理 6** 给定一个序决策信息系统  $S=(U, A=C \cup D, V, f), Q \subseteq P \subseteq C, Q$  是  $S$  删除的属性集,  $\forall a \in (P-Q), \text{Sig}_U^{\text{inner}}(a, P, D)$  为属性  $a$  相对于属性集  $P$  关于属性集  $D$  的内部重要度, 可表示为:

$$\text{Sig}_U^{\text{inner}}(a, P, D) = GK(D|P-\{a\}) - GK(D|P) = \text{mean}(\Delta M_U^{R_P^{\leq}}) - \text{mean}(\Delta M_U^{R_{P \cup D}^{\leq}}) \quad (20)$$

证明过程与定理 3 的证明过程类似, 略.

**定义 13** 给定一个序决策信息系统  $S=(U, A=C \cup D, V, f), Q \subseteq P \subseteq A, M_U^{R_P^{\leq}} = (M_{ij}^{R_P^{\leq}})_{|U| \times |U|}$  是属性集  $P$  在论域  $U$  上的优势关系矩阵,  $\Delta M_U^{R_P^{\leq}}$  是删除属性集  $Q$  后的增量关系矩阵, 则属性集  $P-Q$  的优势关系矩阵  $M_U^{R_{P-Q}^{\leq}}$  中的元素  $M_{ij}^{R_{P-Q}^{\leq}}$  可表示为:

$$M_{ij}^{R_{P-Q}^{\leq}} = \begin{cases} 0, & M_{ij}^{R_P^{\leq}} = 0 \text{ 且 } r_{ij} = 0 \\ 1, & \text{其他} \end{cases} \quad (21)$$

## 2.3 属性集变化条件下的增量属性约简算法

基于 2.1 和 2.2 的知识粒度和优势关系矩阵的增

量更新机制,属性集变化下增量属性约简算法的过程主要分三个步骤:

(1)对属性集变化的数据集增量更新其相对知识粒度;

(2)向约简集中添加待约简集中外部属性重要度最大的属性;

(3)逐步向前删除约简集中的冗余属性,提高约简结果的准确性.

属性集变化下的增量属性约简算法如算法 2 和算法 3 所示.

**算法 2 优势关系下基于属性集增加的增量属性约简算法 (Incremental Attribute Reduction Algorithm Based on Attribute Set Increase under Dominance Relations, IARAI-DR)**

**输入:**序决策信息系统  $S=(U, A=C \cup D, V, f)$ ,  $S$  的属性约简  $RED_C$ , 增加的属性集  $P$ ,  $U$  在  $C, C \cup D$  上的劣势属性矩阵  $M_U^{R_C^{\leq}}, M_U^{R_{C \cup D}^{\leq}}$  和优势关系矩阵  $M_U^{R_C^{\geq}}, M_U^{R_{C \cup D}^{\geq}}$ .

**输出:**序决策信息系统更新后的属性约简  $RED_{C \cup P}$ .

步骤 1. 在计算优势关系矩阵  $M_U^{R_P^{\leq}}$  的同时计算  $M_U^{R_P^{\geq}}$ , 根据定义 9 求出更新后的增量关系矩阵  $\Delta M_U^{R_C^{\leq}}$  和  $\Delta M_U^{R_{C \cup D}^{\leq}}$ .

步骤 2. 根据定理 2 计算更新后的  $GK(D|C \cup P)$ .

步骤 3.  $B \leftarrow RED_C, RED\_Pool \leftarrow C \cup P - RED_C$ .

步骤 4. 根据定理 5 计算  $GK(D|B)$ , 并根据定义 13 计算  $M_U^{R_B^{\leq}}, M_U^{R_{B \cup D}^{\leq}}$ , 若  $GK(D|B) = GK(D|C \cup P)$ , 则执行步骤 7, 否则执行步骤 5.

步骤 5. 对于  $\forall a \in RED\_Pool$ , 根据定理 3 计算  $Sig_U^{outer}(a, B, D)$ , 取  $Sig_U^{outer}(a_m, B, D) = \max(Sig_U^{outer}(a, B, D))$ ,  $RED\_Pool \leftarrow RED\_Pool - \{a_m\}$ ,  $B \leftarrow B \cup \{a_m\}$ , 同时根据定义 10 更新  $M_U^{R_B^{\leq}}$  和  $M_U^{R_{B \cup D}^{\leq}}$ .

步骤 6. 若  $GK(D|C \cup P) = GK(D|B)$ , 则执行步骤 7, 否则执行步骤 5.

步骤 7. 对于  $\forall a \in B$ , 根据定理 6 计算  $Sig_U^{inner}(a, B, D)$ , 若  $Sig_U^{inner}(a, B, D) = 0$ , 则  $B \leftarrow B - \{a\}$ .

步骤 8.  $RED_{C \cup P} \leftarrow B$ , 输出  $RED_{C \cup P}$ .

**算法 3 优势关系下基于属性删除的增量属性约简算法 (Incremental Attribute Reduction Algorithm Based on Attribute Deletion under the Dominance Relation-ship, IARAD-DR)**

**输入:**序决策信息系统  $S=(U, A=C \cup D, V, f)$ ,  $S$  的属性约简  $RED_C$ , 删除的属性集  $P$ ,  $U$  在  $C, C \cup D$  上的劣势属性矩阵  $M_U^{R_C^{\leq}}, M_U^{R_{C \cup D}^{\leq}}$  和优势关系矩阵  $M_U^{R_C^{\geq}}, M_U^{R_{C \cup D}^{\geq}}$ .

**输出:**序决策信息系统更新后的属性约简  $RED_{C-P}$ .

步骤 1. 根据定义 12 求出更新后的增量关系矩阵  $\Delta M_U^{R_C^{\leq}}$  和  $\Delta M_U^{R_{C \cup D}^{\leq}}$ .

步骤 2. 根据定理 5 计算更新后的  $GK(D|C-P)$ .

步骤 3.  $B \leftarrow RED_C - P, RED\_Pool \leftarrow C - (P \cup RED_C)$ .

步骤 4. 根据定理 5 计算  $GK(D|B)$ , 并根据定义 13 计算  $M_U^{R_B^{\leq}}, M_U^{R_{B \cup D}^{\leq}}$ , 若  $GK(D|B) = GK(D|C-P)$ , 则执行步骤 7, 否则执行步骤 5.

步骤 5. 对于  $\forall a \in RED\_Pool$ , 根据定理 3 计算  $Sig_U^{outer}(a, B, D)$ , 取  $Sig_U^{outer}(a_m, B, D) = \max(Sig_U^{outer}(a, B, D))$ ,  $RED\_Pool \leftarrow RED\_Pool - \{a_m\}$ ,  $B \leftarrow B \cup \{a_m\}$ , 同时根据定义 10 更新  $M_U^{R_B^{\leq}}$  和  $M_U^{R_{B \cup D}^{\leq}}$ .

步骤 6. 若  $GK(D|C-P) = GK(D|B)$ , 则执行步骤 7, 否则执行步骤 5.

步骤 7. 对于  $\forall a \in B$ , 根据定理 6 计算  $Sig_U^{inner}(a, B, D)$ , 若  $Sig_U^{inner}(a, B, D) = 0$ , 则  $B \leftarrow B - \{a\}$ .

步骤 8.  $RED_{C \cup P} \leftarrow B$ , 输出  $RED_{C \cup P}$ .

**2.4 三种算法的时间复杂度分析** 表 1 给出了数据集属性增加时 ARKG-DR 和 IARAI-DR 两种算法的时间复杂度.

表 1 属性增加时两种算法的时间复杂度比较

Table 1 Time complexity of the two algorithms with the increase of attributes

算法	时间复杂度
ARKG-DR	$O( U ^2( C + P +( C + P )^2))$
IARAI-DR	$O( U ^2(2 C +3 P +7))$

IARAI-DR 算法中, 步骤 1~3 求新增属性集  $P$  后新数据集的知识粒度, 时间复杂度为  $O(|U|^2(|P|+2))$ ; 步骤 4~6 将每轮迭代中属性重要度最大的属性加入约简集, 由于可以根据优势关系矩阵和劣势属性矩阵的增量更新机制计算, 无须重新计算更新后的约简集的优势关系矩阵, 时间复杂度最差为  $O(|U|^2(2(|C|+|P|)+4))$ ; 步骤 7~8 向前删除约简集中的冗余属性, 时间复杂度为  $O(|U|^2)$ . 那么, IARAI-DR 算法总的时间复杂度为  $O(|U|^2(2|C|+3|P|+7))$ .

ARKG-DR 算法中, 步骤 1~4 求核属性, 其时

间复杂度为  $O(|U|^2(|C|+|P|+(|C|+|P|)^2))$ ;  
 步骤5~8加入必要属性,删除冗余属性,时间复杂度最差为  $O(|U|^2((|C|+|P|)^2+|C|+|P|))$ .  
 那么, ARKG-DR 算法总的复杂度为  $O(|U|^2(|C|+|P|+(|C|+|P|)^2))$ .

由以上分析可知,被约简集  $C$  的属性越多, IARAI-DR 和 ARKG-DR 算法相比,效率更高.

表2给出了数据集属性删除时 IARAD-DR 和 ARKG-DR 两种算法的时间复杂度.

表2 属性删除时两种算法的时间复杂度比较

Table 2 Time complexity of the two algorithms with the deletion of attributes

算法	时间复杂度
ARKG-DR	$O( U ^2( C - P +( C - P )^2))$
IARAD-DR	$O( U ^2(2 C -2 P +7))$

IARAD-DR 算法中,步骤1~3的时间复杂度为  $O(2|U|^2)$ ,步骤4~8的时间复杂度最差为  $O(|U|^2(2(|C|-|P|)+5))$ . IARAD-DR 总的复杂度为  $O(|U|^2(2(|C|-|P|)+7))$ .

ARKG-DR 算法中,步骤1~4的时间复杂度为  $O(|U|^2(|C|-|P|+(|C|-|P|)^2))$ ,步骤5~8的时间复杂度最坏为  $O(|U|^2(|C|-|P|+(|C|-|P|)^2))$ . ARKG-DR 算法总的复杂度为  $O(|U|^2(|C|-|P|+(|C|-|P|)^2))$ .

由以上的分析可知,被约简集  $C$  的属性越多, IARAD-DR 和 ARKG-DR 算法相比,效率更高.

### 3 实验与结果分析

为了验证算法2和算法3的有效性,及属性约简结果的准确性,从UCI机器学习数据库(<https://archive.ics.uci.edu/ml/datasets.php>)中选取六个数据集进行算法的性能测试.实验前对原始数据进行以下预处理:对属性值为字符串的

数据集,根据数据集属性信息将字符串属性值按从劣到优的趋势赋予从小到大的具体数值,且相同的字符串属性值赋予相同数值,对属性值为数值的数据集不作任何更改,此外,数据集的分类(决策)属性值若相同则赋予相同的具体数值.各UCI数据集信息的具体描述如表3所示.

表3 实验使用的UCI数据集

Table 3 UCI datasets used in experiments

序号	数据集	对象数	属性数	类别数
1	Post-operative Patient	90	8	3
2	Blood Transfusion Service	748	5	2
3	Absenteeism	740	20	3
4	Ionosphere	351	34	2
5	Cardiotocography	2126	36	3
6	Diabetic Retinopathy	1151	20	2

测试环境: Intel(R) Core(TM) i5-6300HQ CPU @ 2.30 GHz 处理器, 12 GB 内存, 操作系统为 Window10(64位), Pycharm 软件平台.

实验分三部分. 首先, 将各数据集的数据按属性分为五等份, 将第一份数据作为基础数据集, 每次添加一份数据直至将整个数据集添加进来进行属性约简, 最后比较分析 IARAI-DR 和 ARKG-DR 算法按此过程对数据集进行属性约简的运行时间和属性约简结果. 第二部分, 将第一部分实验得出的数据作为已知条件, 将整个数据集作为基础数据集, 每次从后往前删除一份数据直至还剩最后一份数据进行属性约简, 然后比较分析 IARAD-DR 算法按此过程对数据集进行属性约简和第一部分 ARKG-DR 算法属性约简的运行时间. 前两部分实验中的属性约简运行时间是十次运行时间的均值. 第三部分, 在 Weka 机器学习软件上测试 IARAI-DR 和 ARKG-DR 算法在六个数据集上属性约简的分类准确度.

从表3可知, 实验选取了六个不同规模的数据集. 规模最大的是 Cardiotocography 数据集, 有2126个对象、36维属性, 规模最小的是 Post-operative Patient 数据集, 仅有90个对象、8维属性; 属性维数最多的是 Cardiotocography 数据集, 有36维属性, 属性维数最少的是 Blood Transfusion Service 数据集, 具有5维属性.

图 1 给出了逐渐增加数据时, ARKG-DR 和 IARAI-DR 算法的运行时间. 由图可见, 数据量属性占比为 20% 时, ARKG-DR 和 IARAI-DR 算法处理各数据集的运行时间的差距不大, 这是因为属性较少时, 两者的时间复杂度差距不大; 随着数据量的增加, 增量算法 IARAI-DR 处理各数据集的运行时间明显少于静态算法 ARKG-DR. 总体上, 增量算法 IARAI-DR 计算数据集属性约简的运行效率优于 ARKG-DR 算法.

表 4 给出了 ARKG-DR 和 IARAI-DR 算法在

各数据集上计算属性约简的结果. 由表可见, IARAI-DR 处理 Absenteeism 数据集的属性约简结果长度优于 ARKG-DR 算法, 处理其余数据集的属性约简结果的长度与 ARKG-DR 算法一致.

图 2 给出了从后往前逐渐删除数据时, ARKG-DR 和 IARAD-DR 算法处理数据集的运行时间. 由图可见, 当数据的删除量从 0 增加到 80% 时, 增量算法 IARAD-DR 处理各数据集的运行时间都少于 ARKG-DR 算法; 当数据的删除量接近或高于 80% 时, 增量算法 IARAD-DR 处理

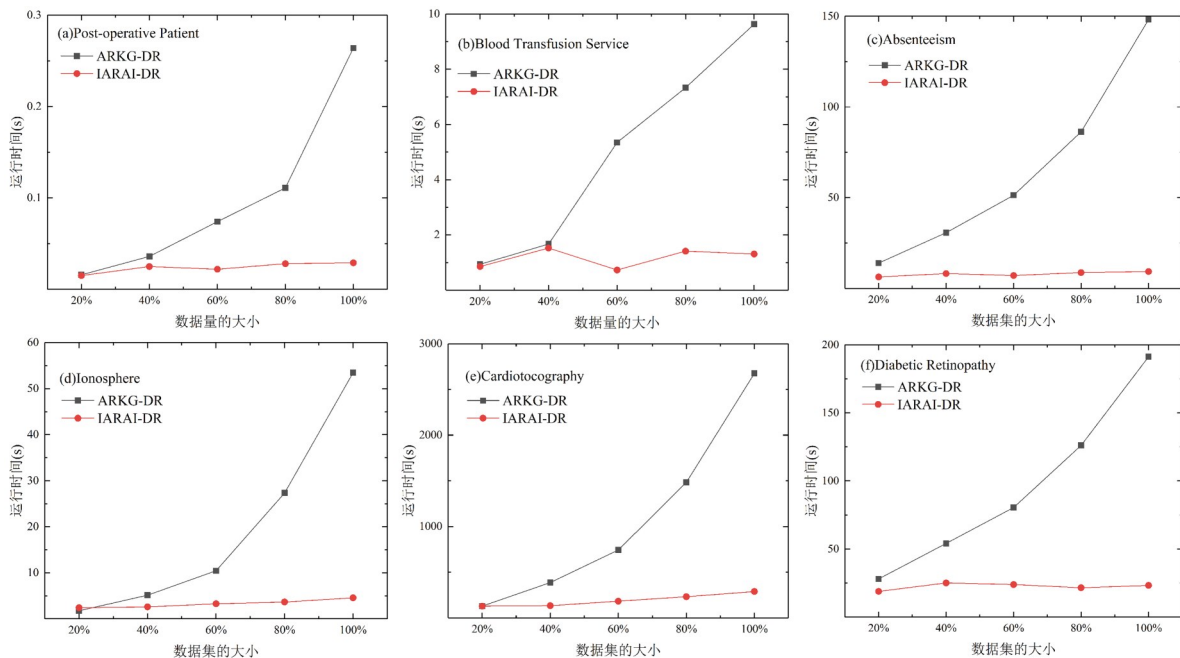


图 1 数据量逐渐增加时, ARKG-DR 和 IARAI-DR 算法运行时间的比较

Fig.1 Running time of ARKG-DR and IARAI-DR algorithms with the increase of data amount

表 4 两种算法在六个数据集上属性约简结果的比较

Table 4 Attribute reduction results of two algorithms on six datasets

属性集	属性数	ARKG-DR		IARAI-DR	
		属性约简	长度	属性约简	长度
Post-operative Patient	8	1, 2, 3, 4, 5, 6, 7, 8	8	1, 2, 3, 4, 5, 6, 7, 8	8
Blood Transfusion Service	5	1, 4, 5	3	1, 2, 4	3
Absenteeism	20	1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 19	15	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 15, 17, 19	14
Ionosphere	34	4, 6, 8, 9, 10, 11, 14, 16, 17, 18, 20, 22, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, 34	23	4, 6, 8, 9, 10, 11, 14, 16, 17, 18, 19, 20, 22, 23, 24, 26, 27, 29, 30, 31, 32, 33, 34	23
Cardiotocography	36	1, 5, 7, 8, 9, 10, 11, 12, 14, 16, 17, 20, 22, 26, 27, 28, 36	17	1, 5, 7, 8, 9, 10, 11, 12, 14, 16, 17, 20, 22, 26, 27, 28, 36	17
Diabetic Retinopathy	20	2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	18	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	18



各数据集的运行时间与ARKG-DR算法差不多. 总体上,增量算法IARAD-DR计算数据集属性约简的运行效率优于ARKG-DR算法.

为了验证ARKG-DR和IARAI-DR算法处理各数据集得到的属性约简结果的准确性,使用Weka机器学习软件上自带的贝叶斯分类器进行测试并使用十折交叉验证的方式计算ARKG-DR

和IARAI-DR算法得到的属性约简结果的分类准确度,结果如表5所示. 由表可见,IARAI-DR在大多数数据集上的分类准确度和ARKG-DR算法一致,在Absenteeism和Ionosphere数据集上稍优于ARKG-DR算法,可见IARAI-DR算法计算属性约简是有效且准确的.

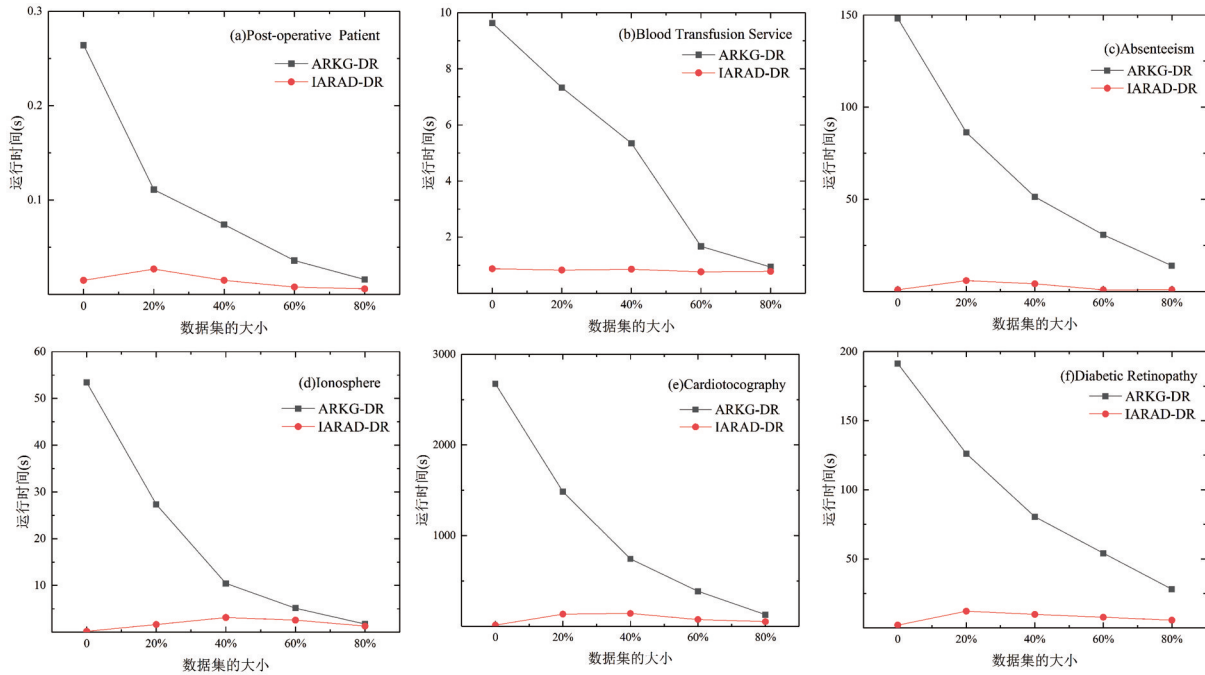


图2 从后向前删除数据时,ARKG-DR和IARAD-DR算法运行时间的比较

Fig 2 Running time of ARKG-DR and IARAD-DR algorithms with the deletion of data from back to front

表5 ARKG-DR和IARAI-DR算法在六个数据集上的分类准确度的比较

Table 5 Classification accuracy of ARKG-DR and IARAI-DR algorithms on six datasets

数据集	ARKG-DR	IARAI-DR
Post-operative Patient	96.76%	96.76%
Blood Transfusion Service	98.21%	98.21%
Absenteeism	89.65%	91.83%
Ionosphere	90.35%	90.87%
Cardiocography	86.54%	86.54%
Diabetic Retinopathy	93.35%	93.35%

## 4 结论

本文以知识粒度来度量序决策信息系统中的属性重要度,并以矩阵分析序决策信息系统中属

性增删条件下知识粒度和优势关系矩阵的增量更新机制,由此提出了两种以属性重要度为启发信息的增量属性约简算法. 为了验证算法的有效性和高效性,选择了六个UCI数据集进行实验. 实验结果表明:从属性约简效率的角度看,各数据集参与属性约简的属性数目越多,本文提出的两种增量属性约简算法明显优于非增量约简算法;从属性约简结果的准确性看,本文提出的两种增量属性约简算法与非增量约简算法相差不大. 另外,现实生活中的数据样本随时会发生动态变化,本文提出的算法无法对其进行属性约简. 未来研究的重点是在序决策信息系统中有多对象增加或者删除的情况下,设计以知识粒度表征属性重要度的快速增量式属性约简算法.

## 参考文献

- [1] Pawlak Z. Rough sets. *International Journal of Computer and Information Sciences*, 1982, 11(5): 341—356.
- [2] Jensen R, Shen Q. Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets and Systems*, 2004, 141(3): 469—485.
- [3] Swiniarski R W, Hargis L. Rough sets as a front end of neural-networks texture classifiers. *Neuro-computing*, 2001, 36(1—4): 85—102.
- [4] Swiniarski R W, Skowron A. Rough set methods in feature selection and recognition. *Pattern Recognition Letters*, 2003, 24(6): 833—849.
- [5] Pawlak Z. Rough sets and intelligent data analysis. *Information Sciences*, 2002, 147(1—4): 1—12.
- [6] Bazan J, Peters J F, Skowron A, et al. Rough set approach to pattern extraction from classifiers. *Electronic Notes in Theoretical Computer Science*, 2003, 82(4): 20—29.
- [7] Greco S, Matarazzo B, Slowinski R. A new rough set approach to multicriteria and multiattribute classification//*Proceedings of the 1<sup>st</sup> International Conference on Rough Sets and Current Trends in Computing*. Springer - Verlag Berlin Heidelberg, 1998: 60—67.
- [8] Greco S, Matarazzo B, Slowinski R. Rough approximation of a preference relation by dominance relations. *European Journal of Operational Research*, 1999, 117(1): 63—83.
- [9] 李艳, 郭娜娜, 吴婷婷, 等. 优势关系下基于浓缩布尔矩阵的属性约简方法. *计算机科学*, 2018, 45(10): 229—234. (Li Y, Guo N N, Wu T T, et al. Attribute reduction based on concentration Boolean matrix under dominance relations. *Computer Science*, 2018, 45(10): 229—234.)
- [10] Li W T, Zhou H X, Xu W H, et al. Interval dominance-based feature selection for interval-valued ordered data. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, DOI: 10.1109/TNNLS.2022.3184120.
- [11] Du W S, Hu B Q. Dominance-based rough set approach to incomplete ordered information systems. *Information Sciences*, 2016(346—347): 106—129.
- [12] Yang X B, Qi Y, Yu D J, et al.  $\alpha$ -dominance relation and rough sets in interval-valued information systems. *Information Sciences*, 2015(294): 334—347.
- [13] Zhang H Y, Yang S Y. Feature selection and approximate reasoning of large-scale set-valued decision tables based on  $\alpha$ -dominance-based quantitative rough sets. *Information Sciences*, 2017(378): 328—347.
- [14] Luo C, Li T R, Chen H M, et al. Fast algorithms for computing rough approximations in set-valued decision systems while updating criteria values. *Information Sciences*, 2015(299): 221—242.
- [15] Wang S, Li T R, Luo C, et al. Efficient updating rough approximations with multi-dimensional variation of ordered data. *Information Sciences*, 2016(372): 690—708.
- [16] Huang Y Y, Li T R, Luo C, et al. Matrix-based dynamic updating rough fuzzy approximations for data mining. *Knowledge-Based Systems*, 2017(119): 273—283.
- [17] Sang B B, Chen H M, Yang L, et al. Incremental attribute reduction approaches for ordered data with time-evolving objects. *Knowledge-Based Systems*, 2021(212): 106583.
- [18] Sang B B, Chen H M, Yang L, et al. Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets. *IEEE Transactions on Fuzzy Systems*, 2022, 30(6): 1683—1697.
- [19] 王磊, 李天瑞. 一种基于矩阵的知识粒度计算方法. *模式识别与人工智能*, 2013, 26(5): 447—453. (Wang L, Li T R. A matrix-based approach for calculation of knowledge granulation. *Pattern Recognition and Artificial Intelligence*, 2013, 26(5): 447—453.)
- [20] Jing Y G, Li T R, Fujita H, et al. An incremental attribute reduction method for dynamic data mining. *Information Sciences*, 2018(465): 202—218.

(责任编辑 杨可盛)