

区间值序信息系统中差别信息树的属性约简*

杨 蕾¹, 张晓燕²⁺, 徐伟华²

1. 重庆理工大学 理学院, 重庆 400054

2. 西南大学 数学与统计学院, 重庆 400715

+ 通讯作者 E-mail: zxy19790915@163.com

摘 要: 属性约简是粗糙集领域的一个热门研究课题, 而差别矩阵是获得属性约简的有效方法。然而, 差别矩阵含有重复元素, 增加了获得约简所需要的时间。差别信息树的提出解决了差别矩阵含有重复元素的问题, 实现了对差别矩阵中非空元素的压缩存储。但是差别信息树是在等价关系下的差别矩阵的基础上提出的, 并没有考虑序决策信息系统的情况。在区间值序信息系统的背景下提出了基于可分辨矩阵的差别信息树, 解决了可分辨矩阵中存在冗余元素的问题, 实现了对可分辨矩阵中非空元素的压缩存储。得到了该树的相关性质定理并对其进行了验证, 并在此基础上给出区间值序信息系统的基于差别信息树的完备的属性约简方法。最后给出了实证分析, 验证了该方法的可行性以及有效性。

关键词: 区间值序信息系统; 差别信息树; 压缩储存; 属性约简

文献标志码: A **中图分类号:** TP18

杨蕾, 张晓燕, 徐伟华. 区间值序信息系统中差别信息树的属性约简[J]. 计算机科学与探索, 2019, 13(6): 1062-1069.
YANG L, ZHANG X Y, XU W H. Attribute reduction of discernibility information tree in interval-valued ordered information system[J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(6): 1062-1069.

Attribute Reduction of Discernibility Information Tree in Interval-Valued Ordered Information System *

YANG Lei¹, ZHANG Xiaoyan²⁺, XU Weihua²

1. School of Science, Chongqing University of Technology, Chongqing 400054, China

2. School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

Abstract: Attribute reduction is a hot topic in rough set, and discernibility matrix is an effective method to get attribute reduction. However, the discernibility matrix contains duplicate elements, which increases the time needed to get the reduction. The discernibility information tree solves the issue, and realizes the compression storage of the

* The National Natural Science Foundation of China under Grant Nos. 61472463, 61402064, 61772002 (国家自然科学基金); the Natural Science Foundation of Chongqing under Grant No. cstc2015jcyjA40053 (重庆市自然科学基金); the Science and Technology Project of Chongqing Municipal Education Committee under Grant No. KJ1709221 (重庆市教委科技项目).

Received 2018-05-17, Accepted 2018-07-13.

CNKI 网络出版: 2018-07-24, <http://kns.cnki.net/KCMS/detail/11.5602.TP.20180723.1614.006.html>

non-empty elements in the discernibility matrix. Unfortunately, the discernibility information tree is based on the discernibility matrix under the equivalence relation, and does not consider the situation of the ordered decision information system. In this paper, a discernibility information tree is proposed based on distinguishable matrix in interval-valued ordered information systems, which solves the problem of redundant elements in the distinguishable matrix and realizes the compression storage of non-empty elements in the distinguishable matrix. Furthermore, the related theorem of the tree is obtained and verified carefully. And, a complete attribute reduction method is given based on the discernibility information tree of the interval-valued ordered information system. Finally, an empirical analysis is addressed, which verifies the feasibility and effectiveness of the method.

Key words: interval value information system; discernibility information tree; compressed storage; attribute reduction

1 引言

波兰数学家 Pawlak^[1]在 1982 年提出了粗糙集理论用以解决不精确、不完备数据问题, 它的一个主要课题是属性约简^[2-8]。属性约简即去掉知识库中冗余的属性之后其数据的分类能力保持不变, 从而达到数据表降维的目的, 减少了冗余数据, 简化了规则。不少学者已经在不同的粗糙集背景下根据不同的属性约简思想提出了不同的属性约简方法^[9]。自从 Skowron 等人^[10]提出通过差别矩阵来求取属性约简的思想之后, 差别矩阵因为其自身的直观简明性被广大学者关注和研究。蒋瑜、王燮等人^[11]提出了基于差别矩阵的 Rough 集属性约简算法, 他们主要分析了不同的差别矩阵对求取属性约简效率的影响, 然后定义一种新的差别矩阵来减少矩阵中非空差别信息的个数从而达到提高效率的目的。蒋瑜、王鹏等人^[12]提出了基于差别矩阵的属性约简的完备算法, 它们结合的属性重要度以及迭代的思想来求取决策表的最小约简; 王兵和陈善本^[13]也给出了一种基于差别矩阵的属性约简完备算法, 该算法的特别之处在于迭代过程中对各类属性的处理。但是在通过差别矩阵来求取属性约简的这些算法当中, 要想获得约简就必须使用差别矩阵中所有的非空元素。Skowron 等^[10]提出在求决策表的属性约简时将决策表所对照的差别矩阵中的所有非空元素进行合取即可, 但在使用合取运算的时候相同(重复)元素和父集元素对属性约简时没有任何作用, 但这些重复元素和父集元素却占用了大量的存储空间, 同时也增加了求取约简的时间。为了消除差别矩阵中重复元素的出现, 文

献[14]提出了一种新的差别矩阵存储方法(C-Tree)实现了差别矩阵的压缩储存, 但是该方法还存在一定的弊端, 当父集元素和子集元素同时出现的时候它保留了父集元素, 从而使该树中还存在一定数量的冗余元素; 为了消除差别矩阵中冗余的父集元素, 蒋瑜^[15]给出了一种基于差别信息树的属性约简算法, 该算法不仅能消除差别矩阵中的重复元素, 在大多数情况下亦能消除父集元素的影响, 实现对差别矩阵的压缩储存。但是蒋瑜所提出的差别信息树是通过等价关系下的差别矩阵获得的, 而基于等价关系的差别矩阵是对称的, 因此在求约简时只用差别矩阵中的上三角元素或者下三角元素即可。现在在区间值序信息系统的背景下考虑属性约简, 其差别矩阵便失去了对称性, 因此基于差别信息树的属性约简并不一定适用于区间值序信息系统下的属性约简。

本文在区间值序信息系统的背景下, 结合了蒋瑜提出的差别信息树的优点, 提出了基于可分辨矩阵的差别信息树, 并在此基础上给出区间值序信息系统的基于差别信息树的完备的属性约简方法。最后在第 4 章给出了实证分析, 验证了该方法的可行性以及有效性。

2 预备知识

本章主要讲述了有关区间值序信息系统下求取属性约简的理论基础以及差别信息树等相关知识。

定义 1^[16] 称三元组 $I=(U, A, F)$ 为信息系统, 其中: U 是有限的对象集; A 是有限的条件属性集; F 是 U 与 A 的关系集; V_a 为 a 的有限值域。若 $\forall f \in F$,

$a \in A$ 和 $x_i \in U$ 都有 $f(x_i, a) = [a^L(x_i), a^U(x_i)]$ 。

那么称 $I = (U, A, F)$ 为区间值信息系统, 其中 $a^L(x_i)$ 和 $a^U(x_i)$ 都是实数且满足 $a^L(x_i) \leq a^U(x_i)$; $f(x_i, a)$ 为对象 x_i 在属性 a 下的属性值且 $f(x_i, a)$ 是一个区间数。当 $a^L(x_i) = a^U(x_i)$ 时, 属性值 $f(x_i, a)$ 为一个实数, 因此区间值信息系统是单值信息系统的推广。

定义 2^[16] 设 $I = (U, A, F)$ 是一个区间值信息系统, 对 $\forall a \in A$, 可以对区间值信息系统中的属性值进行比较, 定义:

$$f(x_i, a) \leq f(x_j, a) \Leftrightarrow (\forall a \in A)[a^L(x_i) \leq a^L(x_j), a^U(x_i) \leq a^U(x_j)]$$

$$f(x_i, a) \geq f(x_j, a) \Leftrightarrow (\forall a \in A)[a^L(x_i) \geq a^L(x_j), a^U(x_i) \geq a^U(x_j)]$$

其中, “ \leq ”和“ \geq ”分别在区间值信息系统中构建递增以及递减偏序, 若某属性在区间值信息系统中的值域是递增或者递减偏序, 则称该属性为准则。本文考虑由递增偏序构成优势关系的情况, 递减偏序的情况可同理给出。称 $I = (U, A, F)$ 为区间值序信息系统, 若区间值信息系统中所有的条件属性都为准则, 记作 I^\geq 。

定义 3^[16] 区间值序信息系统中, A 为准则集。对 $a \in A$, 存在优势关系“ \geq_a ”, $x_j \geq_a x_i$ 表示 x_j 关于 a 优于 x_i , 则 $x_j \geq_a x_i \Leftrightarrow (\forall a \in A)[x_j \geq_a x_i]$, 定义区间值优势关系 R_A^\geq 为:

$$R_A^\geq = \{(x_i, x_j) \in U \times U | x_j \geq_a x_i, \forall a \in A\} = \{(x_i, x_j) \in U \times U | (\forall a \in A)[a^L(x_i) \leq a^L(x_j), a^U(x_i) \leq a^U(x_j)]\}$$

由区间值优势关系 R_A^\geq 诱导的区间优势类 $[x_i]_A^\geq$ 为:

$$[x_i]_A^\geq = \{x_j \in U | (x_i, x_j) \in R_A^\geq\} =$$

$$\{x_j \in U | (\forall a \in A)[a^L(x_i) \leq a^L(x_j), a^U(x_i) \leq a^U(x_j)]\}$$

注 1 参考不同的实际意义, 可以定义不同的区间优势关系, 比如:

$$(1) R_A^{\geq 1} = \{(x_i, x_j) \in U \times U | a^U(x_i) \leq a^L(x_j), \forall a \in A\}$$

$$(2) R_A^{\geq 2} = \{(x_i, x_j) \in U \times U | a^L(x_i) \leq a^L(x_j), \forall a \in A\}$$

$$(3) R_A^{\geq 3} = \{(x_i, x_j) \in U \times U | a^U(x_i) \leq a^U(x_j), \forall a \in A\}$$

定义 4^[16] 设 $I^\geq = (U, A, F)$ 为区间值序信息系统, 记 $Dis_{\geq A}(x_i, x_j) = \{a \in A | (x_i, x_j) \notin R_a^\geq\}$, 称 $Dis_{\geq A}(x_i, x_j)$ 为 I^\geq 中对象 x_i, x_j 关于 R_A^\geq 的可分辨属性集。记 $Dis_{\geq A} = (Dis_{\geq A}(x_i, x_j))_{U \times U}$, 称 $Dis_{\geq A}$ 为 I^\geq 中对象关于 R_A^\geq 的可分

辨矩阵。特别地, 对 $\forall x_i, x_j \in U$ 有: $Dis_{\geq A}(x_i, x_i) = \emptyset$ 并且得到:

$$Dis_{\geq A}(x_i, x_j) \cap Dis_{\geq A}(x_j, x_i) = \emptyset$$

定义 5^[16] 设 $I^\geq = (U, A, F)$ 为区间值信息系统, 该区间值信息系统所对应的可分辨矩阵为 $Dis_{\geq A}$, 称 $M_{\geq A} = \bigwedge \{ \bigvee \{ a | a \in Dis_{\geq A}(x_i, x_j) \} \}$ 为 I^\geq 关于 R_A^\geq 的可辨识公式。

定义 6^[16] 设 $I^\geq = (U, A, F)$ 为区间值序信息系统, $M_{\geq A} = \bigwedge \{ \bigvee \{ a | a \in Dis_{\geq A}(x_i, x_j) \} \}$ 所对应的极小析取范式为:

$$M_{\geq \min} = \bigvee_{k=1}^p \left(\bigwedge_{s=1}^{q_k} a_s \right)$$

记 $B_k = \{a_s | s = 1, 2, \dots, q_k\}$, 则 $\{B_k | k = 1, 2, \dots, p\}$ 是区间值信息系统 I^\geq 的所有约简的集合。

定义 7^[15] 差别信息树是一棵有序树, 其中序体现在: 它的条件属性的排列顺序是按照原信息系统中条件属性的顺序从左到右进行排列, 顺序不能颠倒或者改变。差别信息树有如下几个特征:

(1) 差别信息树中的每个节点最多存在 $|A|$ 个子节点, 其中 $|A|$ 代表信息表中条件属性的个数。

(2) 差别信息树中的每一个节点涵盖了 4 个方面的信息, 分别为: 前缀指针、后继指针、节点名、同名指针。其中: 前缀指针指向该节点的父亲节点, 后继指针指向该节点的孩子节点, 节点名记录了该节点所对应的条件属性名称, 同名指针则指向差别信息树中与该节点具有相同条件属性名的其他路径当中的节点。

(3) 差别信息树的子树也是一棵有序树, 其中条件属性的排列顺序同样不能颠倒或者改变。

3 区间值序信息系统中基于差别信息树的属性约简方法

本章将通过差别信息树的概念, 给出区间值序信息系统中基于可分辨矩阵的差别信息树的算法, 并通过相关定理来证明由该差别信息树得到区间值序信息系统属性约简的合理性。

首先, 根据差别信息树的定义和蒋瑜提出的差别信息树的设计与实现过程给出区间值序信息系统中基于可分辨矩阵的差别信息树的构建过程:

算法 1 区间值序信息系统中基于可分辨矩阵的差别信息树的构建方法

输入: 区间值序信息系统 $I^{\geq} = (U, A, F)$ 。

输出: 区间值序信息系统中基于可分辨矩阵的差别信息树。

getDI-tree(information table I^{\geq})

{

1. 创建区间值序信息系统中基于可分辨矩阵的差别信息树的根节点 TN , 并令 TN 为 null;

2. 根据 $Dis_{\geq A}(x_i, x_j) = \{a \in A \mid (x_i, x_j) \notin R_a^{\geq}\}$ 求出任意 (x_i, x_j) 所对应的可分辨属性集 $Dis_{\geq A}(x_i, x_j)$, 从而得到可分辨矩阵 $Dis_{\geq A} = (Dis_{\geq A}(x_i, x_j))_{|U| \times |U|}$ 。设 $B \subseteq A$ 满足 $B = Dis_{\geq A}(x_i, x_j) (\forall x_i, x_j)$ 且 B 中元素的顺序按区间值信息系统中条件属性的顺序进行排列。

While($B \neq \emptyset$)

{

①选择 B 中最左边的元素 a ;

②If(TN 的所有孩子节点中存在节点名为 a 的孩子节点 N)

{

如果 N 是一叶子节点, 则不构建 B 中剩余属性对应的节点 (不扩展路径策略);

如果 a 为 B 中最后一个元素, 则从构建的差别信息树中删除以 N 为根的子树但保留 N (删除子树策略)

否则, 令 $TN = N$;

}

Else

{

创建一新的节点 N' 使其成为 TN 的子节点, 将 N' 的属性名初始化为 a , 并通过该节点的同名指针连接到具有与该节点有相同属性名的节点上, 从而构成了一个同名属性节点链;

令 $TN = N'$;

}

③令 $B \leftarrow B - \{a\}$;

}

}

从算法 1 可以看出: 在区间值序信息系统中构建基于可分辨矩阵的差别信息树时运用了不扩展路径策略以及删除子树策略, 即所构建的差别信息树具有如下几个特征:

(1) 将相同的分配可辨识属性集映射到同一条

路径当中;

(2) 将具有相同前缀的可分辨属性集映射到最小的可分辨属性集所对应的路径当中;

(3) 可分辨矩阵中的属性集存在共享前缀。

因此, 在区间值序信息系统中构建的基于可分辨矩阵的差别信息树实现了对分配可辨识矩阵的压缩储存, 从而减少了构建基于可分辨矩阵的差别信息树的时空复杂度。

区间值序信息系统中基于可分辨矩阵的差别信息树的相关定理如下:

定理 1 区间值序信息系统中基于可分辨矩阵的差别信息树中包含了获得区间值序信息系统的属性约简所需要的所有属性。

证明 设区间值序信息系统中基于可分辨矩阵的差别信息树中所有路径的可分辨属性集存放在集合 DS 中, 可分辨矩阵中所有元素存放在集合 $Dis_{\geq A}$ 中。由算法 1 所对应的基于可分辨矩阵的差别信息树的构建过程有: $DS \subseteq Dis_{\geq A}$, 对于 $\forall (x_i, x_j)$ 有 $Dis_{\geq A}(x_i, x_j) \in Dis_{\geq A}$, $\exists Dis_{\geq AT}(x_i, x_j) \in DS$, s.t. $Dis_{\geq AT}(x_i, x_j) \subseteq Dis_{\geq A}(x_i, x_j)$, 由运算公理可知: $Dis_{\geq AT}(x_i, x_j)$ 与 $Dis_{\geq AT}(x_i, x_j)$ 作合取运算得到的结果还是 $Dis_{\geq AT}(x_i, x_j)$ 。从而, 区间值序信息系统中基于可分辨矩阵的差别信息树中包含了获得区间值序信息系统的属性约简所需要的所有属性。 \square

定理 2 区间值序信息系统中基于可分辨矩阵的差别信息树中所有仅含一个节点的路径所对应的单元素可分辨属性集的并组成了该区间值序信息系统中条件属性集的核 $core(A)$ 。

证明 通过在区间值序信息系统中构建基于可分辨矩阵的差别信息树的过程可知: 假设该差别信息树中存在一个节点名为 a 的节点, 而且该树中存在仅包含节点名为 a 的节点的路径, 则存在可分辨属性集 $\{a\}$ 与该路径对应。在可分辨矩阵中, 若 $a \in A$ 并且 $\{a\}$ 为可分辨矩阵中的单元素集, 则称 a 为 A 中的必要属性。 A 中所有必要属性的集合构成了 A 的核, 即 $core(A)$ 。 \square

定理 3 设 R 是区间值序信息系统中基于可分辨

矩阵的差别信息树中根节点的所有子节点所构成的条件属性集合,则 $R_R^{\geq} = R_A^{\geq}$ 成立。

证明 令 DS 为包含了基于可分辨矩阵差别信息树的所有路径所代表的可分辨属性集的集合,则对于 $\forall Dis_{\geq A}(x_i, x_j) \in DS$, 有 $Dis_{\geq A}(x_i, x_j) \cap R \neq \emptyset$ 。因为 $R \subseteq A$, 则 $R_A^{\geq} \subseteq R_R^{\geq}$ 显然成立, 下证 $R_A^{\geq} \supseteq R_R^{\geq}$ 也成立。对于 $\forall (x_i, x_j) \notin R_A^{\geq}$, $\exists a \in A$, 使 $(x_i, x_j) \notin R_a^{\geq}$ 成立。因为有等式 $Dis_{\geq A}(x_i, x_j) \cap R \neq \emptyset$ 成立, 所以 $\exists a \in R$, 使 $a \in Dis_{\geq A}(x_i, x_j)$, 则 $a \in R$ 同时 $(x_i, x_j) \notin R_a^{\geq}$, 所以有 $(x_i, x_j) \notin R_R^{\geq}$ 。则 $R_A^{\geq} \supseteq R_R^{\geq}$ 得证, 综上, $R_R^{\geq} = R_A^{\geq}$ 。□

接下来, 将分析算法 1 的时空复杂度。

对一个区间值序信息系统 $I^{\geq} = (U, A, F)$ 来说, 若存在 $|U|$ 个对象、 $|A|$ 个条件属性, 则在可分辨矩阵中最多可以得到 $|U|^2$ 个非空的条件属性的子集(即差别信息), 假设可分辨矩阵中实际的非空条件属性的子集数为 M (一般情况下, $M \ll |U|^2$)。由基于可分辨矩阵的差别信息树的构建过程可以看出, 一棵差别信息树最多可以有 M 条不同的路径并且每一条路径中最多可以有 $|A|$ 个节点, 因此, 一棵差别信息树中最多可以有 $M \times |A|$ 个节点, 又由于在基于可分辨矩阵的差别信息树中存在许多的路径都存在共享前缀导致差别信息树中的实际节点数远小于 $M \times |A|$ 。因此在最坏的情况下基于分配可辨识矩阵的差别信息树的空间复杂度为 $O(|A| \times |U|^2)$ 。

另外, 在构建基于分配可辨识矩阵的差别信息树的过程中, 向差别信息树中插入路径的次数最多为 $|U|^2$ 次, 并且在构建路径的过程中最多比较并插入 $|A|$ 个节点, 删除 M_i (其中 $i \in \{1, 2, \dots, |U|^2\}$) 个节点, 基于分配可辨识矩阵的差别信息树的时间复杂度为 $O(|A| \times |U|^2 + (M_1 + M_2 + \dots + M_{|U|^2}))$, 已知在一棵差别信息树当中最多可以有 $|A| \times |U|^2$ 个节点, 则 $M_1 + M_2 + \dots + M_{|U|^2}$ 的值至多为 $|A| \times |U|^2$, 综上所述, 基于分配可辨识矩阵的差别信息树的时间复杂度为 $O(2|A| \times |U|^2)$ 。

和文献[11]提出的差别矩阵存储方法(C-Tree)相比, 在基于可分辨矩阵的差别信息树的构建中, 当父集元素和子集元素同时出现的时候, 它保留了子集元素, 消除差别矩阵中冗余的父集元素, 更进一步对

差别矩阵进行了存储压缩。在空间复杂度的比较上, 基于可分辨矩阵的差别信息树的空间复杂度小于 C-Tree 的空间复杂度。

为了验证算法 1 所提出的区间值序信息系统中基于可分辨矩阵的差别信息树的合理性以及有效性, 给出了区间值序信息系统中基于该差别信息树的属性约简方法:

算法 2 区间值序信息系统中基于可分辨矩阵差别信息树的属性约简方法

输入: 区间值序信息系统中基于可分辨矩阵的差别信息树。

输出: 由差别信息树得到的属性约简。

1. 创建空集 R ;
2. 将基于可分辨矩阵差别信息树中只含单个节点的路径, 将这些节点对应属性名放在一个集合 R' 中;
3. 若 $R' \neq \emptyset$, 对所有 $a \in R'$, 在基于可分辨矩阵差别信息树中删掉所有含节点 $\{a\}$ 的路径;
4. 令 $R \leftarrow R'$;
5. 从基于可分辨矩阵差别信息树中选择其根节点的最右孩子节点并假设该子节点所对应的属性名为 b , 此时令 $R \leftarrow R \cup \{b\}$, 然后在得到的差别信息树中去掉所有含节点 b 的路径;
6. 若基于可分辨矩阵的差别信息树中仅含有根节点, 输出 R , 算法结束。

下面给出算法 2 的完备性证明:

易知, 对一个区间值序信息系统来说, 若 $B \subseteq A$ 是该信息系统的一个约简则需满足两个条件: (1) $R_B^{\geq} = R_A^{\geq}$; (2) $\forall b \in B, R_{B-\{b\}}^{\geq} \neq R_A^{\geq}$ 。

对区间值序信息系统的可分辨矩阵来说, 若 $R \subseteq A$ 是一个完备约简, 同样的 R 需要满足两个条件: (1) $\forall Dis_{\geq A}(x_i, x_j) \neq \emptyset$, 有 $Dis_{\geq A}(x_i, x_j) \cap R \neq \emptyset$; (2) $\forall r \in R, \exists Dis_{\geq A}(x_i, x_j)$ 满足: $Dis_{\geq A}(x_i, x_j) \cap (R - \{r\}) = \emptyset$ 。

由定理 1 知道: 区间值序信息系统中基于可分辨矩阵的差别信息树中包含了获得区间值序信息系统的属性约简所需要的所有属性。因此对基于可分辨矩阵的差别信息树来说, 若 $R \subseteq A$ 是一个完备约简, R 需要满足两个条件如下:

- (1) $\forall B \in DS, B \cap R \neq \emptyset$ (其中 DS 为包含了基于可分辨矩阵差别信息树的所有路径所代表的可分辨属性集的集合);

(2) $\forall r \in R, \exists B \in DS$ 满足 $B \cap (R - \{r\}) = \emptyset$ 。

若算法 2 完备只需证明算法 2 输出的 R 满足条件(1)、(2)均可。

由算法 2 知, 条件(1)显然成立, 下证条件(2)满足。

由定理 2 可知: 区间值序信息系统中基于可分辨矩阵的差别信息树中所有仅含一个节点的路径所对应的单元素可分辨属性集的并组成了该区间值序信息系统中条件属性集的核 $core(A)$ 。

将算法 2 第二步得到的 $core(A)$ 看作 R 的一部分并在基于可分辨矩阵差别信息树中删掉所有含有 $core(A)$ 中元素的路径。

令 $R' = R - core(A)$ 。若 r 是 R' 中最右边的一个元素, 则在当前的差别信息树中, 根节点最右边的孩子节点的节点名一定是 r , 并且以该节点为根的子树中一定不包含 $R' - \{r\}$ 中任何属性所对应的节点, 则对 $r \in R, \exists B \in DS$ 满足 $B \cap (R - \{r\}) = \emptyset$ 。可证 R' 中所有元素都满足该条件。因此, 算法 2 得到的约简是完备约简。

4 实证分析

给定一个区间值序信息系统 $I^2 = (U, A, F)$ (表 1), 其中有限对象集 $U = \{x_1, x_2, \dots, x_{10}\}$, 有限属性集为 $A = \{a, b, c, d, e\}$ 。

Table 1 Interval valued sequence information systems

表 1 区间值序信息系统

U	a	b	c	d	e
x_1	1	[0,1]	2	1	[1,2]
x_2	[0,1]	0	[1,2]	2	1
x_3	[0,1]	0	[1,2]	1	1
x_4	0	0	1	2	1
x_5	2	[1,2]	3	[1,2]	[2,3]
x_6	[0,2]	[1,2]	[1,3]	[1,2]	[2,3]
x_7	1	1	2	1	2
x_8	[1,2]	[1,2]	[2,3]	2	[2,3]
x_9	[1,2]	2	[2,3]	[0,2]	3
x_{10}	2	2	3	[0,1]	3

先求取每个对象对的可分辨属性集, 从而获得区间值序信息系统的可分辨矩阵。从而, 表 1 所对应的可分辨矩阵如表 2 所示。

Table 2 Discernibility matrix of interval valued ordinal information systems in Table 1

表 2 表 1 对应的区间值序信息系统的可分辨矩阵

$Dis_{\geq A}$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	\emptyset	A	$abce$	$abce$	\emptyset	ac	\emptyset	\emptyset	d	d
x_2	d	\emptyset	d	ac	d	d	d	\emptyset	d	d
x_3	\emptyset	\emptyset	\emptyset	acd	\emptyset	\emptyset	\emptyset	\emptyset	d	d
x_4	d	\emptyset	d	\emptyset	d	d	d	\emptyset	d	d
x_5	A	$abce$	A	$abce$	\emptyset	ac	A	ac	acd	d
x_6	A	$abce$	A	$abce$	\emptyset	\emptyset	A	\emptyset	d	d
x_7	be	$abce$	$abce$	$abce$	\emptyset	ac	\emptyset	\emptyset	d	d
x_8	A	$abce$	A	$abce$	d	acd	A	\emptyset	d	d
x_9	A	$abce$	A	$abce$	be	$abce$	A	be	\emptyset	d
x_{10}	$abce$	$abce$	$abce$	$abce$	be	$abce$	$abce$	$abce$	ac	\emptyset

注 2 可分辨矩阵中每一对对象对的可分辨属性集均应该为集合的形式, 此处为了简便写成了如表 2 的形式。

根据极小析取范式可以求得表 1 所对应的区间值序信息系统的约简如下:

$$M_{\geq \min} = (a \vee b \vee c \vee d \vee e) \wedge (b \vee e) \wedge (a \vee b \vee c \vee e) \wedge d \wedge (a \vee c) \wedge (a \vee c \vee d) = (b \vee e) \wedge (a \vee c) \wedge d = (a \wedge b \wedge d) \vee (a \wedge d \wedge e) \vee (b \wedge c \wedge d) \vee (c \wedge d \wedge e)$$

根据算法 1 给出表 1 所对应的区间值序信息系统中基于可分辨矩阵的差别信息树的具体构建过程如下:

(1) 首先, 创建根节点; 然后求取各个对象对的可分辨属性集, 将可分辨属性集中的属性顺序按区间值序信息系统中条件属性从左至右的顺序排列; 最后得到可分辨矩阵如表 2 所示。

(2) 构建可分辨矩阵中第一个可分辨属性集 $A = \{a, b, c, d, e\}$ 所对应的路径 $\langle a, b, c, d, e \rangle$, 将此路径插入到要构建的基于可分辨矩阵的差别信息树当中。

(3) 为第二个可分辨属性集 $\{a, b, c, e\}$ 创建其对应的路径 $\langle a, b, c, e \rangle$, 其与路径 $\langle a, b, c, d, e \rangle$ 具有相同的前缀 $\langle a, b, c \rangle$ 。

(4) 对于第三个可分辨属性集 A , 因为基于可分辨矩阵的差别信息树中已经存在 A 所对应的路径 $\langle a, b, c, d, e \rangle$, 所以不构建新的路径。同理, 将所有相同的可分辨属性集映射到同一个子集中。

(5) 构建可分辨属性集 $\{a, c\}$ 对应的路径 $\langle a, c \rangle$, 它

与路径 $\langle a, b, c, d, e \rangle$ 和 $\langle a, b, c, e \rangle$ 具有相同的前缀 $\langle a \rangle$ 。

依此类推,最后得到区间值序信息系统的基于可分辨矩阵的差别信息树如图1所示。

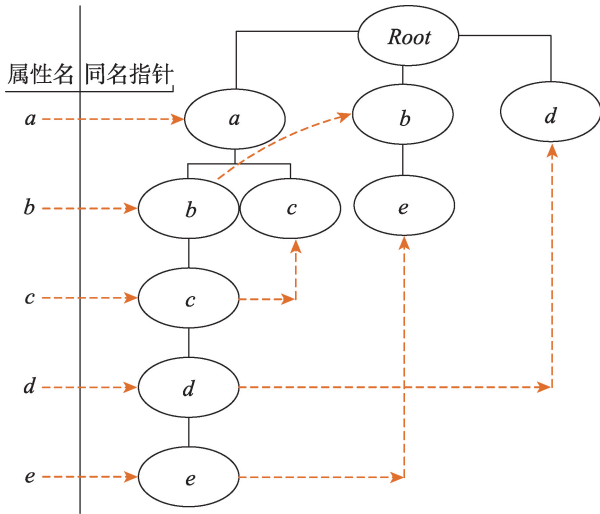


Fig.1 Difference information based on discernibility matrix (Table 2)

图1 基于可分辨矩阵(表2)的差别信息

根据如图1所示的区间值序信息系统的基于可分辨矩阵的差别信息树以及算法2给出求取表1所对应的区间值序信息系统约简的过程:

(1)创建空集 R ;

(2)从区间值序信息系统的基于可分辨矩阵的差别信息树中选择仅含单个节点的路径 $\langle d \rangle$,将属性 d 存放在集合 R 中,然后删去树中有 d 的所有路径;

(3)此时选择差别信息树的根节点的最右子节点 $\{b\}$ 令 $R=R \cup \{b\}$,然后删去信息树中含有节点名为 b 的全部路径;

(4)此时差别信息树只余下一条路径 $\langle a, c \rangle$,此时选择差别信息树的根节点的子节点 $\{a\}$,令 $R=R \cup \{a\}$,最后输出 $R=\{d, b, a\}$,算法结束。于是 $\{d, b, a\}$ 就是通过基于可分辨矩阵差别信息树求得的一个约简。

5 结束语

本文提出了区间值序信息系统的一种新的约简方法:首先构建区间值序信息系统的基于可分辨矩阵的差别信息树,然后通过该差别信息树求取该信

息系统的属性约简。通过该差别信息树,实现了对可分辨矩阵的压缩储存,从而缩短了求取区间值序信息系统属性约简的空间复杂度。但从差别信息树的构建过程可知,往该差别信息树中插入的可分辨属性集当中的属性顺序是按信息表中属性的原始顺序,没有考虑属性重要度对信息树构建的影响。因此接下来的工作可以考虑结合属性重要度去实现对差别信息树的构建,看是否能对可分辨矩阵进一步压缩储存。

References:

- [1] Pawlak Z. Rough sets[J]. International Journal of Parallel Programming, 1982, 11(5): 341-356.
- [2] Zhang X H, Pei D W, Dai J H. Fuzzy mathematics and rough set theory[M]. Beijing: Tsinghua University Press, 2013.
- [3] Xu W H, Zhang W X. Distribution reduction of information systems with inconsistent objectives based on dominance relations[J]. Fuzzy Systems and Mathematics, 2007, 21(4): 124-131.
- [4] Mi J S, Wu W Z, Zhang W X. A comparative study of knowledge reduction in inconsistent objective information systems[J]. Fuzzy Systems and Mathematics, 2003, 17(3): 54-60.
- [5] Sang B B, Xu W H. Distribution reduction of intuitionistic fuzzy ordered decision information systems[J]. Fuzzy Systems and Mathematics, 2017, 31(2): 164-170.
- [6] Zhang W X, Mi J S, Wu W Z. Inconsistent knowledge reduction of target information system[J]. Journal of Computer Science, 2003, 26(1): 12-18.
- [7] Xu W H, Zhang W X. Knowledge reduction in inconsistent information systems based on dominance relations[J]. Computer Science, 2006, 33(2): 182-184.
- [8] Sang B B, Xu W H. Distribution reduction of intuitionistic fuzzy decision making information system[J]. Computer Science, 2017, 44(S1): 164-170.
- [9] Thangavel K, Pethalakshmi A. Dimensionality reduction based on rough set theory: a review[J]. Applied Soft Computing Journal, 2009, 9(1): 1-12.
- [10] Skowron A, Rauszer C. The discernibility matrices and functions in information systems[J]. Intelligent Decision Support, 1992, 11: 331-362.
- [11] Jiang Y, Wang X, Ye Z. Attribute reduction algorithm of

- rough sets based on discernibility matrix[J]. Journal of System Simulation, 2008, 20(14): 3717-3720.
- [12] Jiang Y, Wang P, Wang X, et al. Complete algorithm of attribute reduction based on discernibility matrix[J]. Computer Engineering and Applications, 2007, 43 (19): 185-187.
- [13] Wang B, Chen S B. A complete algorithm for attribute reduction based on discernibility matrix[J]. Journal of Shanghai Jiao Tong University, 2004, 38(1): 43-46.
- [14] Yang M, Yang P. A novel condensing tree structure for rough set feature selection[J]. Neurocomputing, 2008, 71(4): 1092-1100.
- [15] Jiang Y. Rough set attribute reduction algorithm based on differential information tree[J]. Control and Decision, 2015, 30(8): 1531-1536.
- [16] Xu W H. Preface information system and rough set[M]. Beijing: Science Press, 2013.
- 分布约简[J]. 模糊系统与数学, 2007, 21(4): 124-131.
- [4] 米据生, 吴伟志, 张文修. 不协调目标信息系统知识约简的比较研究[J]. 模糊系统与数学, 2003, 17(3): 54-60.
- [5] 桑彬彬, 徐伟华. 直觉模糊序决策信息系统的分布约简[J]. 模糊系统与数学, 2017, 31(2): 164-170.
- [6] 张文修, 米据生, 吴伟志. 不协调目标信息系统的知识约简[J]. 计算机学报, 2003, 26(1): 12-18.
- [7] 徐伟华, 张文修. 基于优势关系下不协调目标信息系统的知识约简[J]. 计算机科学, 2006, 33(2): 182-184.
- [8] 桑彬彬, 徐伟华. 直觉模糊序决策信息系统的分配约简[J]. 计算机科学, 2017, 44(S1): 164-170.
- [11] 蒋瑜, 王燮, 叶振. 基于差别矩阵的 Rough 集属性约简算法[J]. 系统仿真学报, 2008, 20(14): 3717-3720.
- [12] 蒋瑜, 王鹏, 王燮, 等. 基于差别矩阵的属性约简完备算法[J]. 计算机工程与应用, 2007, 43(19): 185-187.
- [13] 王兵, 陈善本. 一种基于差别矩阵的属性约简完备算法[J]. 上海交通大学学报, 2004, 38(1): 43-46.
- [15] 蒋瑜. 基于差别信息树的 Rough Set 属性约简算法[J]. 控制与决策, 2015, 30(8): 1531-1536.
- [16] 徐伟华. 序信息系统与粗糙集[M]. 北京: 科学出版社, 2013.

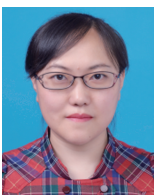
附中文参考文献:

- [2] 张小红, 裴道武, 代建华. 模糊数学与 Rough 集理论[M]. 北京: 清华大学出版社, 2013.
- [3] 徐伟华, 张文修. 基于优势关系下不协调目标信息系统的



YANG Lei was born in 1994. She is an M.S. candidate at Chongqing University of Technology. Her research interests include rough set, fuzzy set and the mathematical foundation of artificial intelligence, etc.

杨蕾(1994—),女,重庆人,重庆理工大学硕士研究生,主要研究领域为粗糙集,模糊集,人工智能的数学基础等。



ZHANG Xiaoyan was born in 1979. She received the Ph.D. degree from Northwest University in 2017. Now she is an associate professor and M.S. supervisor at Southwest University. Her research interests include fuzzy set, rough set and concept lattice, etc.

张晓燕(1979—),女,山西怀仁人,2017年于西北大学获得博士学位,现为西南大学副教授、硕士生导师,主要研究领域为模糊集,粗糙集,概念格等。在国内外重要学术期刊上发表论文近30篇,分别以独著、第一主编、第二主编出版3部著作;作为负责人主持两项国家自然科学基金、1项重庆市科委自然科学基金、1项重庆理工大学青年基金;作为主要研究人员参与多项国家自然科学基金及省部级课题。



XU Weihua was born in 1979. He received the Ph.D. degree from Xi'an Jiaotong University in 2007. Now he is a professor and Ph.D. supervisor at Southwest University. His research interests include artificial intelligence, granular computing, fuzzy mathematics and rough set, etc.

徐伟华(1979—),男,山西浑源人,2007年于西安交通大学获得博士学位,现为西南大学教授、博士生导师,主要研究领域为人工智能,粒计算,模糊数学,粗糙集等。在国内外重要学术期刊上发表论文70余篇,出版3部学术著作;先后承担5项国家自然科学基金、10余项省部级课题,在信息科学、模糊数学、人工智能与粒计算以及应用数学相关研究领域有丰富的经验积累。