

# 基于相对辨识关系的属性约简算法

孙 滨,毛建景

(郑州工业应用技术学院 信息工程学院,河南 郑州 451100)

**摘 要:**介绍了决策信息系统中的可辨识关系、相对辨识关系等相关概念,完成了属性集独立性判定工作。将属性集的可辨识能力和相对辨识能力与属性集所辨识的对象联系起来,研究了两种基于相对辨识关系的属性重要度求解方法,完成了属性集独立或依赖、是否为决策信息系统约简的判定。依据相对辨识关系,给出了相应的改进算法,利用该算法描述因属性集属性的增减而引起它的相对辨识能力的变化。该算法是从条件属性集中先判断各个属性的相对可辨识关系,其对象对个数最大者组成约简集,然后再在此约简集中逐渐添加属性,直至满足约简的条件。该算法是一种无核的属性约简算法,无论算法时间复杂度还是约简工作量在一定程度上都有所降低,并通过实例验证了其有效性。

**关键词:**粗糙集;相对可辨识关系;属性集;属性约简

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2018)04-0099-05

doi:10.3969/j.issn.1673-629X.2018.04.021

## An Attribute Reduction Algorithm Based on Relative Discernible Relation

SUN Bin, MAO Jian-jing

(School of Information Technology, Zhengzhou University of Industrial Technology, Zhengzhou 451100, China)

**Abstract:** We mainly introduce some related concepts in decision information system, such as discernible relation and relative discernible relation, and complete the attribute set independence judgment. Connecting the discernibility and relative discernibility of the attribute set with the object of identification of the property set, we also study the evaluation methods of attribute importance between the two kinds of relative identification and finish the judgment of attribute's independence or dependence and whether to do reduction of decision information system. Based on the relative recognition relation, the corresponding improvement algorithm is given, by which the variation of the relative recognition caused by the addition or subtraction of attribute set attributes is described. The algorithm determines the relatively discernible relation of each attribute from the condition attribute set, taking the largest number of object as the reduction set where the attributes are gradually added until meeting the conditions of reduction. It is a kind of seedless attribute reduction algorithm, which is reduced to a certain extent regardless of time complexity or reduction workload. Its validity is verified by an example.

**Key words:** rough sets; relative discernible relation; attribute set; attribute reduction

## 0 引 言

粗糙集理论可以有效地进行不精确、不完备数据分析与处理,国外在这方面的研究已经取得了不错的成绩<sup>[1-4]</sup>,其中决策信息系统中的属性约简是其核心研究内容。许多学者利用多种方法来度量决策信息系统中属性的重要度。文献[5]在分析信息熵度量不确定性数据的基础上,定义信息熵属性重要度概念,引入蚁群优化算法,提出基于信息熵与蚁群优化的最小属性约简算法。文献[6]通过减少约简过程中基数排序

次数来提升效率,设计了相对分辨能力的约简算法。文献[7]设计了一种启发式函数—决策重要度,这种启发式函数根据每个属性正决策对象集合的大小来定义其重要性,正决策对象集合越大表示重要性越高,由此构造了基于决策重要度的启发式属性约简算法。文献[8]的约简算法既考虑信息决策表的相对正域,也考虑以核属性为启发信息逐个增加条件属性时对边界域样本的影响。文献[9]定义了一种粒度差别矩阵和基于该差别矩阵的属性约简,并证明了该差别矩阵的

收稿日期:2017-04-27

修回日期:2017-08-25

网络出版时间:2017-12-05

基金项目:河南省重大科技专项(162102210319);河南省基础与前沿技术研究计划项目(142300410283);河南省科学技术研究重点项目(12B520063,14B520065);河南省高等学校青年骨干教师资助计划项目(2016GGJS-182)

作者简介:孙 滨(1983-),男,讲师,硕士,研究方向为云计算与知识发现、现代教育技术。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20171205.1433.098.html>

属性约简定义与基于知识粒度的属性约简定义等价。文献[10]给出了对象矩阵的属性约简定义,证明了属性约简与基于正区域的属性约简的等价性。

因此,对于如何从海量数据集中寻找一种高效的属性约简算法,不仅能保持分类能力不变,还能简化决策规则生成过程,便具有重要的研究意义。文中依据相对辨识关系给出了基于相对辨识关系的属性约简算法,从而为属性约简提供了新的研究思路。

## 1 可辨识关系与属性重要度

在决策信息系统中,对于论域  $U$ ,若存在属性集  $P, Q \subseteq C \cup D$  且  $P, Q \neq \emptyset$ ,使得  $U$  关于  $P, Q$  的划分相等,即  $U/P = U/Q$ ,则称  $P, Q$  具有相同的可辨识能力,否则若属性集  $P$  辨识论域  $U$  中的对象对越多,则称  $P$  的可辨识能力越强,反之则辨识能力越弱。下面给出属性集的可辨识关系。

### 1.1 可辨识关系

为了更好地引入可辨识关系,先给出一个属性集所对应的划分的性质,方便对可辨识关系的理解。

性质 1<sup>[11]</sup>: 设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,若  $P \subseteq Q \subseteq C$ ,则有  $U/Q \subseteq U/P$ ,此时称  $U$  关于  $Q$  的划分比  $P$  更精细,  $U$  关于  $P$  的划分比  $Q$  更粗糙。

性质 1 仅描述了属性集对应的划分之间辨识能力的高低,并未对属性集的可辨识能力进行精确描述,因此下面给出可辨识关系的形式化定义来加以阐述。

定义 1: 设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $n$  为所有对象个数,对  $P \subseteq C \cup D$ ,若令

$$\text{DIS}(P) = \{ \langle x_i, x_j \rangle \mid \text{存在属性 } a \in P, f(x_i, a) \neq f(x_j, a), 1 \leq i \leq j \leq n, x_i, x_j \in U \}$$

则称  $\text{DIS}(P)$  为属性集  $P$  在  $U$  上的可辨识关系。

由定义 1 可知,  $\text{DIS}(P)$  是属性集  $P$  能辨识的所有对象对,考察的是  $P$  的辨识能力,并且可知  $0 \leq |\text{DIS}(P)| \leq \lfloor |U| \cdot (|U| - 1) / 2 \rfloor$ 。当  $U/P = \{ \{x_1, x_2, \dots, x_n\} \}$  时,  $\text{DIS}(P)$  取得最小值 0; 当  $U/P = \{ \{x_1\}, \{x_2\}, \dots, \{x_n\} \}$  时,  $\text{DIS}(P)$  取得最大值  $\lfloor |U| \cdot (|U| - 1) / 2 \rfloor$ 。

显然  $U/P$  越精细,  $\text{DIS}(P)$  包含的对象对越多,说明利用  $P$  能辨识的对象对个数越多,越容易将论域中对象两两辨识,说明  $P$  的辨识能力越强,否则反之。

由定义 1 可得性质 2。

性质 2: 设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $P \subseteq C \cup D$ , 对任意的  $x_i, x_j \in U$  ( $1 \leq i \leq j \leq n, n$  为论域  $U$  中对象个数), 令  $[x_i]_P, [x_j]_P$  为  $x_i, x_j$  所对应的  $P$  等价类, 则

(1) 若  $[x_i]_P \cap [x_j]_P \neq \emptyset$ , 则有  $\langle x_i, x_j \rangle \in \text{DIS}(P)$ ; 万方数据

(2) 若  $[x_i]_P \cap [x_j]_P = \emptyset$ , 则有  $\langle x_i, x_j \rangle \notin \text{DIS}(P)$ 。

由性质 2 可知,  $\text{DIS}(U, P)$  能辨识不同  $P$  等价类中的对象, 而无法辨识同一  $P$  等价类中的对象。

由定义 1 可得定理 1。

定理 1: 设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $P \subseteq Q \subseteq C$ ,  $\text{DIS}(P), \text{DIS}(Q)$  分别为  $P, Q$  的可辨识关系, 则有  $\text{DIS}(P) \subseteq \text{DIS}(Q)$ 。

证明: 令任意的  $x_i, x_j \in U$ , 其中  $1 \leq i \leq j \leq n$ 。由定义 1 可知, 若  $\langle x_i, x_j \rangle \in \text{DIS}(P)$ , 则存在属性  $a \in P, f(x_i, a) \neq f(x_j, a)$ , 由于  $P \subseteq Q$ , 则有  $a \in P \subseteq Q$ , 按照定义 1 可知, 存在属性  $a \in Q, f(x_i, a) \neq f(x_j, a)$ , 即得  $\langle x_i, x_j \rangle \in \text{DIS}(Q)$ 。由于  $\langle x_i, x_j \rangle$  具有任意性, 所以可得  $\text{DIS}(P) \subseteq \text{DIS}(Q)$ 。证毕。

由定理 1 知, 对任意集  $P \subseteq Q \subseteq C$ , 则有  $\text{DIS}(P) \subseteq \text{DIS}(Q)$ , 说明  $Q$  比  $P$  所辨识的对象对相等或更多, 可辨识能力更强。特别地, 当  $P = Q$  时, 则有  $\text{DIS}(P) = \text{DIS}(Q)$ , 此时称  $P, Q$  可辨识能力相同。

### 1.2 相对可辨识关系

定义 2: 设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $n$  为所有对象个数, 对  $P \subseteq C$ , 若令

$$\text{DIS}(P, D) = \{ \langle x_i, x_j \rangle \mid \text{存在属性 } a \in P, f(x_i, a) \neq f(x_j, a) \text{ 并且 } f(x_i, D) \neq f(x_j, D), 1 \leq i \leq j \leq n, x_i, x_j \in U \}$$

则称  $\text{DIS}(P)$  为属性集  $P$  在  $U$  上关于  $D$  的相对可辨识关系。

由定义 2 可知,  $\text{DIS}(P, D)$  是属性集  $P$  相对于决策属性集  $D$  能辨识的所有相对对象对, 考察的是  $P$  关于  $D$  的相对辨识能力。按照定义 2, 可得性质 3。

性质 3: 设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $\text{DIS}(P), \text{DIS}(D)$  分别为  $P, D$  的可辨识关系,  $\text{DIS}(P, D)$  为  $P$  关于  $D$  的相对可辨识关系, 则有  $\text{DIS}(P, D) = \text{DIS}(P) \cap \text{DIS}(D)$ 。

结合定义 2 和性质 3, 可得定理 2。

定理 2: 设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $P \subseteq Q \subseteq C$ ,  $\text{DIS}(P), \text{DIS}(Q), \text{DIS}(D)$  分别为  $P, Q, D$  的可辨识关系,  $\text{DIS}(P, D), \text{DIS}(Q, D)$  分别为  $P, Q$  关于  $D$  的相对可辨识关系, 则有  $\text{DIS}(P, D) \subseteq \text{DIS}(Q, D)$ 。

证明: 由性质 3 可知,  $\text{DIS}(P, D) = \text{DIS}(P) \cap \text{DIS}(D)$ , 又知  $P \subseteq Q$ , 由定理 1 可知,  $\text{DIS}(P) \subseteq \text{DIS}(Q)$ 。故  $\text{DIS}(P, D) = \text{DIS}(P) \cap \text{DIS}(D) \subseteq \text{DIS}(Q) \cap \text{DIS}(D) = \text{DIS}(Q, D)$ , 即得  $\text{DIS}(P, D) \subseteq \text{DIS}(Q, D)$ 。证毕。

由定理 2 可知, 对任意集合  $P \subseteq Q \subseteq C$ , 则有  $\text{DIS}(P, D) \subseteq \text{DIS}(Q, D)$ , 说明  $Q$  比  $P$  关于  $D$  所辨识的

对象对相等或更多,相对可辨识能力更强。特别地,当  $P = Q$  时,则有  $\text{DIS}(P, D) = \text{DIS}(Q, D)$ , 此时称  $P, Q$  相对可辨识能力相同。

1.3 属性重要度

定义3:设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $P \subseteq C$ 。对任意的属性  $a \in C - P$ , 令

$$\text{Sig}_P(a) = (|\text{DIS}(P \cup \{a\}, D)| - |\text{DIS}(P, D)|) / |U|^2$$

则称  $\text{Sig}_P(a)$  为属性  $a$  相对于  $P$  的属性重要度。其中,  $|\text{DIS}(P \cup \{a\}, D)|$ 、 $|\text{DIS}(P, D)|$  分别表示相对可辨识关系  $\text{DIS}(P \cup \{a\}, D)$ 、 $\text{DIS}(P, D)$  所辨识的对象对个数,  $|U|$  为论域  $U$  的对象个数。

由于  $P \subseteq P \cup \{a\}$ , 由定理2可知,  $\text{DIS}(P, D) \subseteq \text{DIS}(P \cup \{a\}, D)$ , 则存在  $|\text{DIS}(P \cup \{a\}, D)| \geq |\text{DIS}(P, D)|$ , 即  $\text{Sig}_P(a) \geq 0$ 。显然,  $\text{Sig}_P(a)$  描述的是添加属性  $a$  后引起  $P$  关于  $D$  所辨识对象对的变化,  $\text{Sig}_P(a)$  越大,  $P$  关于  $D$  所辨识对象对越多, 说明添加属性  $a$  后  $P$  的相对辨识能力提高, 从而属性  $a$  对  $P$  的重要性越大。

$\text{Sig}_P(a)$  度量的是在原有属性集  $P$  中添加不属于  $P$  的属性  $a$  的属性重要度, 下面给出一种度量原有属性集  $P$  中每个属性  $a$  (即  $a \in P$ ) 的属性重要度。

定义4:设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $P \subseteq C$ 。对任意属性  $a \in P$ , 令

$$\text{Sig}_{P-\{a\}}(a) = (|\text{DIS}(P, D)| - |\text{DIS}(P - \{a\}, D)|) / |U|^2$$

称  $\text{Sig}_{P-\{a\}}(a)$  为属性  $a$  相对于  $P$  的重要度。其中,  $|\text{DIS}(P, D)|$ 、 $|\text{DIS}(P - \{a\}, D)|$  分别为相对可辨识关系  $\text{DIS}(P, D)$ 、 $\text{DIS}(P - \{a\}, D)$  所辨识的对象对个数,  $|U|$  为论域  $U$  的对象个数, 显然  $\text{Sig}_{P-\{a\}}(a) \geq 0$ 。

$\text{Sig}_{P-\{a\}}(a)$  描述的是在  $P$  中删除属性  $a$  后引起  $P$  关于  $D$  所辨识对象对的变化,  $\text{Sig}_{P-\{a\}}(a)$  越大,  $P$  关于  $D$  所辨识对象对越多, 说明删除属性  $a$  后  $P$  的相对辨识能力下降, 从而属性  $a$  对  $P$  的重要性越大。特别地, 当  $\text{Sig}_{P-\{a\}}(a) = 0$  时, 由定义4可知,  $\text{DIS}(P, D) = \text{DIS}(P - \{a\}, D)$ , 即  $P, P - \{a\}$  的相对辨识能力相同, 说明删除属性  $a$  后,  $P$  的相对辨识能力并未发生变化, 此时称属性  $a$  为  $P$  的不必要属性或冗余属性。

结合定义4, 下面给出属性  $a$  相对于  $P$  是必要属性还是不必要属性。

定义5:设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $P \subseteq C$ 。若对任意的  $a \in P$ , 都有  $\text{Sig}_{P-\{a\}}(a) > 0$ , 则每个  $a$  相对于  $P$  均是必要的, 则称  $P$  为独立的, 否则称  $P$  为依赖的。

1.4 约简

定义6:设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $P \subseteq C$ 。若对任意的  $a \in P$ , 都有  $\text{Sig}_{P-\{a\}}(a) > 0$ , 则每个  $a$  相对于  $P$  均是必要的, 则称  $P$  为独立的, 否则称  $P$  为依赖的。

$\text{Sig}_{P-\{a\}}(a)$ , 定义5用来判断  $a$  在  $P$  中是不必要的还是必要的, 结合定义4和定义5便可给出基于相对辨识关系的属性约简方法。为了证明基于相对辨识关系的约简与文献[5]中的约简是等价的, 下面先给出文献[5]中约简的定义。

定义6:设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $P \subseteq C$ 。如果  $P$  满足:  $U/C = U/P$ ; 则对任意  $a \in P$ ,  $U/C \neq U/(P - \{a\})$ , 则称  $P$  是  $C$  的一个约简。

由定义4、5、6, 给出基于相对辨识关系的约简的判定方法, 可得定理3。

定理3:设  $S = (U, C \cup D, V, f)$  为一个决策信息系统,  $P \subseteq C$ 。若  $P$  是独立的且  $\text{DIS}(P, D) = \text{DIS}(C, D)$ , 则称  $P$  为  $C$  的一个约简。

证明:若  $P$  是独立的且  $\text{DIS}(P, D) = \text{DIS}(C, D)$ , 因  $P \subseteq C$ , 由定理1得  $\text{DIS}(P) \subseteq \text{DIS}(C)$ , 若  $U/P \neq U/C$ , 则  $\text{DIS}(P) \subset \text{DIS}(C)$ ,  $\text{DIS}(P) \cap \text{DIS}(D) \subset \text{DIS}(C) \cap \text{DIS}(D)$ , 由定理2可知,  $\text{DIS}(P, D) \subset \text{DIS}(C, D)$  与条件  $\text{DIS}(P, D) = \text{DIS}(C, D)$  矛盾, 所以  $U/P = U/C$ ,  $P$  满足定义6第一个条件。

又因为  $P$  是独立的, 则对任意的  $a \in P$ , 都有  $\text{Sig}_{P-\{a\}}(a) = (|\text{DIS}(P, D)| - |\text{DIS}(P - \{a\}, D)|) / |U| > 0$ , 则知  $|\text{DIS}(P, D)| > |\text{DIS}(P - \{a\}, D)|$ , 即  $\text{DIS}(P - \{a\}, D) \subset \text{DIS}(P, D)$ , 故  $U/P - \{a\} \neq U/P$ , 所以  $U/P - \{a\} \neq U/C$ ,  $P$  满足定义6第二个条件, 可知  $P$  是  $C$  的一个约简。

由定理3可知, 文中约简与文献[5]中约简等价。

2 基于相对辨识关系的属性约简算法

2.1 算法思想

基于相对辨识关系的属性约简算法思想是首先令约简集  $\text{Red}(S) = \emptyset$ , 对任意属性  $a \in C$ , 按照定义3依次计算各条件属性关于  $D$  的相对可辨识关系  $\text{DIS}(\{a\}, D)$ , 选择对象对个数最大的属性加入约简集  $\text{Red}(S)$  中。然后, 判断当前约简集  $\text{Red}(S)$  关于  $D$  的可辨识关系是否等于  $\text{DIS}(C, D)$ , 若是则输出  $\text{Red}(S)$ , 即为该决策信息系统的约简; 否则, 对任意属性  $a \in C - \text{Red}(S)$ , 按照定义3依次计算属性重要度  $\text{Sig}_{\text{Red}(S)}(a)$ , 选择对象对个数最大的属性加入约简集  $\text{Red}(S)$  中, 再次判断是否满足定理3关于约简的要求, 不断重复这个过程, 最终输出满足要求的  $\text{Red}(S)$ , 即为该决策信息系统的约简。

由此可知, 该算法是从条件属性集中先判断各个属性的相对可辨识关系, 其对象对个数最大者组成约简集, 然后再在此约简集中逐渐添加属性, 直至满足约简条件, 是一种无核的属性约简算法, 无论算法时间复杂度还是约简工作量在一定程度上都有所降低。

2.2 算法实现

定理 3 利用相对辨识关系约定了约简满足的要求,下面给出一种基于相对辨识度的属性约简算法—RDR 算法(an attribute reduction algorithm based on relative discernible relation)。

输入:一个决策信息系统  $S = (U, C \cup D, V, f)$ , 其中  $U = \{x_1, x_2, \dots, x_n\}, C = \{a_1, a_2, \dots, a_m\}$ ;

输出:决策信息系统的约简  $\text{Red}(S)$ 。

步骤 1:令  $\text{Red}(S) = \emptyset$ ,按定义 2 计算相对可辨识关系  $\text{DIS}(C, D)$ ;

步骤 2:令  $\text{Red}(S) = \{a\}$ , 其中  $|\text{DIS}(\{a\}, D)| = \max\{|\text{DIS}(\{a_i\}, D)| \mid a_i \in C, i = 1, 2, \dots, m\}$  (如果多个属性均满足, 则任选一个加入  $\text{Red}(S)$ ) 且  $|\text{DIS}(\{a_i\}, D)| \neq 0$ ;

步骤 3:若  $\text{DIS}(\text{Red}(S), D) = \text{DIS}(C, D)$ , 则算法终止,输出  $\text{Red}(S)$  (此时的  $\text{Red}(S)$  为  $S$  的一个约简), 否则执行步骤 4;

步骤 4:令  $\text{Red}(S) = \text{Red}(S) \cup \{b\}$ , 并且  $\text{Sig}_{\text{Red}(S)}(b) = \max\{\text{Sig}_{\text{Red}(S)}(b_i) \mid b_i \in C - \text{Red}(S), i = 1, 2, \dots, m\}$  (如果多个属性满足, 任选一个加入  $\text{Red}(S)$ ), 转步骤 3。

2.3 实验验证

下面是采用文献[12]中的决策信息系统(见表 1), 其中文献[12]得到的约简是  $\{a_1, a_3\}$ 。

表 1 决策信息系统 S

$U$	$a_1$	$a_2$	$a_3$	$a_4$	$D$
$x_1$	0	1	0	1	0
$x_2$	1	0	0	1	1
$x_3$	0	0	1	0	0
$x_4$	1	0	1	1	0
$x_5$	1	1	0	1	1
$x_6$	1	1	1	0	0
$x_7$	0	1	1	1	0

首先,令  $\text{Red}(S) = \emptyset, \text{DIS}(C, D) = \{ \langle x_1, x_2 \rangle, \langle x_1, x_5 \rangle, \langle x_2, x_3 \rangle, \langle x_2, x_4 \rangle, \langle x_2, x_6 \rangle, \langle x_2, x_7 \rangle, \langle x_3, x_5 \rangle, \langle x_4, x_5 \rangle, \langle x_5, x_6 \rangle, \langle x_5, x_7 \rangle \}$ 。

对于任意的  $a \in C$ , 逐一计算其可辨识关系  $\text{DIS}(\{a\}, D)$ :

$\text{DIS}(\{a_1\}, D) = \{ \langle x_1, x_2 \rangle, \langle x_1, x_5 \rangle, \langle x_2, x_3 \rangle, \langle x_2, x_7 \rangle, \langle x_3, x_5 \rangle, \langle x_5, x_7 \rangle \}$   
 $\text{DIS}(\{a_2\}, D) = \{ \langle x_1, x_2 \rangle, \langle x_2, x_6 \rangle, \langle x_2, x_7 \rangle, \langle x_3, x_5 \rangle, \langle x_4, x_5 \rangle \}$

$\text{DIS}(\{a_3\}, D) = \{ \langle x_2, x_3 \rangle, \langle x_2, x_4 \rangle, \langle x_2, x_6 \rangle, \langle x_2, x_7 \rangle, \langle x_3, x_5 \rangle, \langle x_4, x_5 \rangle \}$

$\text{DIS}(\{a_4\}, D) = \{ \langle x_2, x_3 \rangle, \langle x_2, x_6 \rangle, \langle x_3, x_5 \rangle, \langle x_5, x_6 \rangle \}$

则可知,  $|\text{DIS}(\{a_1\}, D)| = |\text{DIS}(\{a_3\}, D)| > |\text{DIS}(\{a_2\}, D)|, |\text{DIS}(\{a_4\}, D)|$ , 按照步骤 2, 在属

性  $a_1, a_3$  中任意一个属性加入  $\text{Red}(S)$  (此处选择  $a_1$ ), 即此时  $\text{Red}(S) = \{a_1\}$ 。

按照步骤 3, 判断式子  $\text{DIS}(\text{Red}(S), D) = \text{DIS}(C, D)$  是否成立。显然, 由上述所求相对辨识关系, 可得  $\text{DIS}(\text{Red}(S), D) = \text{DIS}(\{a_1\}, D) \neq \text{DIS}(C, D)$ , 执行步骤 4。

先计算所有属性集  $\text{Red}(S) \cup \{b\} (b \in C - \text{Red}(S))$  关于  $D$  的相对辨识关系。

$\text{DIS}(\{a_1, a_2\}, D) = \{ \langle x_1, x_2 \rangle, \langle x_1, x_5 \rangle, \langle x_2, x_3 \rangle, \langle x_2, x_6 \rangle, \langle x_2, x_7 \rangle, \langle x_3, x_5 \rangle, \langle x_4, x_5 \rangle, \langle x_5, x_7 \rangle \}$   
 $\text{DIS}(\{a_1, a_3\}, D) = \{ \langle x_1, x_2 \rangle, \langle x_1, x_5 \rangle, \langle x_2, x_3 \rangle, \langle x_2, x_4 \rangle, \langle x_2, x_6 \rangle, \langle x_2, x_7 \rangle, \langle x_3, x_5 \rangle, \langle x_4, x_5 \rangle, \langle x_5, x_6 \rangle, \langle x_5, x_7 \rangle \}$   
 $\text{DIS}(\{a_1, a_4\}, D) = \{ \langle x_1, x_2 \rangle, \langle x_1, x_5 \rangle, \langle x_2, x_6 \rangle, \langle x_2, x_7 \rangle, \langle x_3, x_5 \rangle, \langle x_5, x_6 \rangle, \langle x_5, x_7 \rangle \}$

按照定义 3, 对于所有属性  $b \in C - \text{Red}(S)$ , 计算其重要度  $\text{Sig}_{\text{Red}(S)}(b)$ 。

$\text{Sig}_{|a_1|}(a_2) = (|\text{DIS}(\{a_1, a_2\}, D)| - |\text{DIS}(\{a_1\}, D)|) / |U| = (8 - 6) / 14 = 1/7$ ,  $\text{Sig}_{|a_1|}(a_3) = (|\text{DIS}(\{a_1, a_3\}, D)| - |\text{DIS}(\{a_1\}, D)|) / |U| = (10 - 6) / 14 = 2/7$ ,  $\text{Sig}_{|a_1|}(a_4) = (|\text{DIS}(\{a_1, a_4\}, D)| - |\text{DIS}(\{a_1\}, D)|) / |U| = (7 - 6) / 14 = 1/14$ 。显然,  $\text{Sig}_{|a_1|}(a_3)$  值最大。按照算法步骤 4, 将属性  $a_3$  加入当前  $\text{Red}(S)$  中, 此时  $\text{Red}(S) = \{a_1, a_3\}$ 。

按照步骤 3, 判断式子  $\text{DIS}(\text{Red}(S), D) = \text{DIS}(C, D)$  是否成立。显然, 由上述所求相对辨识关系, 可得  $\text{DIS}(\text{Red}(S), D) = \text{DIS}(\{a_1, a_3\}, D) = \text{DIS}(C, D)$ , 算法终止, 输出当前  $\text{Red}(S) = \{a_1, a_3\}$ , 即  $\{a_1, a_3\}$  为该决策信息系统的约简, 这与文献[12]所得到的结果一致, 说明该算法是有效的。

2.4 算法比较

本节给出该算法复杂度分析, 并与其他算法进行对比。

(1) 对于步骤 1, 计算  $\text{DIS}(C, D)$  的时间复杂度为  $O(|C||D|)$ ;

(2) 对于步骤 2, 要计算所有  $a \in C$  中的所有可辨识关系集合对象个数, 所以该步的时间复杂度为  $|C| \times O(|C||U|)$ ;

(3) 对于步骤 3, 需要判断  $\text{DIS}(\text{Red}(S), D) = \text{DIS}(C, D)$  的时间复杂度为  $O(|\text{Red}(S)||U|) ( \text{Red}(S) \subseteq C )$ ;

(4) 对于步骤 4, 计算  $\text{Sig}_{\text{Red}(S)}(b)$  的时间复杂度为  $O(|C|^2|U|)$ 。

综上所述, 文中的 RDR 算法时间复杂度为  $O(|C||U|) + |C| \times O(|C||U|) + O(|\text{Red}(S)||U|)$



$|U|) + O(|C|^2|U|) = O(|C|^2|U|)$ 。

文献[13]算法的时间复杂度为  $O(|C|^2|U|^2)$  , 文献[12]算法的时间复杂度为  $O(|U||A|2^{|A|} \log|U|)$  , 相比之下 RDR 算法至少也为  $O(|C|^2|U|)$  , 因此要优于文献[12-13]算法。又因为该算法是往约简集中逐渐添加属性, 不需要重复计算条件属性集的核集, 大大减少了计算量, 一定程度上降低了复杂度, 提高了算法效率。

3 结束语

给出了属性集的可辨识关系和相对辨识关系定义, 将属性集的可辨识能力和相对辨识能力与属性集所辨识的对象对联系起来, 并利用相对辨识关系来计算属性的属性重要度, 完成属性集独立或依赖、是否为决策信息系统约简的判定。通过改进算法的介绍与执行, 验证了该算法的有效性。

参考文献:

[1] 喻 昕, 邓文达, 彭久生. 粗糙集理论在处理不完全信息的应用[J]. 计算机工程与科学, 2005, 27(10): 65-68.

[2] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112(1-4): 39-49.

(上接第98页)

完整、高效、安全的应用服务体系。通过数据中心采集存储企业科研生产和管理过程中所产生的大量数据, 同时通过知识工程平台的建设, 针对不同类型的数据, 使用相应的加工手段和方式, 提升数据的显性化、共享化和智慧化程度, 形成可利用的知识, 针对不同用户提供个性化服务。构建一个从感知—决策—行动—优化自完善的研发生态环境, 推动企业的自主知识创新及精益管控, 进而有效提升企业的核心竞争力。

参考文献:

[1] 李军宁, 陈 渭, 谢友柏. 现代设计知识服务平台框架体系及其知识流动规律[J]. 西安交通大学学报, 2012, 46(1): 60-65.

[2] XI W K, YUAN X Y. Design process control and knowledge service of complex product[J]. Applied Mechanics and Materials, 2011, 135-136: 421-427.

[3] 彭明峰, 李 玮, 李 华, 等. 基于 PLM 的三维数字化协同工艺系统研究与应用[J]. 航天制造技术, 2016(1): 56-60.

[4] 耿翠霞, 傅铅生, 姚 雄. 产品生命周期管理(PLM)技术研究[J]. 机械制造与自动化, 2004, 33(5): 86-88.

[5] 胡光忠, 柳忠伟, 兰 芳. 产品全生命周期管理(PLM)技

[3] LEUNG Y, LI D. Maximal consistent block technique for rule acquisition in incomplete information systems[J]. Information Sciences, 2003, 153: 85-106.

[4] SWINIARSKI R W, SKOWRON A. Rough set methods in feature selection and recognition[J]. Pattern Recognition Letters, 2003, 24(6): 833-849.

[5] 陈颖悦, 陈玉明. 基于信息熵与蚁群优化的属性约简算法[J]. 小型微型计算机系统, 2015, 36(3): 586-590.

[6] 葛 浩, 李龙澍, 杨传健. 基于相对分辨能力的属性约简算法[J]. 系统工程理论与实践, 2015, 35(6): 1595-1603.

[7] 常红岩, 蒙祖强. 一种新的决策粗糙集启发式属性约简算法[J]. 计算机科学, 2016, 43(6): 218-222.

[8] 梁海龙, 谢 珩, 续欣莹, 等. 新的基于区分对象集的邻域粗糙集属性约简算法[J]. 计算机应用, 2015, 35(8): 2366-2370.

[9] 张清国, 郑雪峰. 基于知识粒度的不完备决策表的属性约简的矩阵算法[J]. 计算机科学, 2012, 39(2): 209-211.

[10] 王 炜, 徐章艳, 李晓瑜. 不完备决策表中基于对象矩阵属性约简算法[J]. 计算机科学, 2012, 39(4): 201-204.

[11] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001: 1-12.

[12] 吴 静, 邹 海. 基于属性重要性的属性约简算法[J]. 计算机应用与软件, 2010, 27(2): 255-257.

[13] 吴尚智, 苟平章. 粗糙集和信息熵的属性约简算法及其应用[J]. 计算机工程, 2011, 37(7): 56-58.

术研究[J]. 机械设计与制造, 2006(9): 169-171.

[6] 魏正业. 企业产品全生命周期管理信息系统分析与集成[D]. 武汉: 武汉理工大学, 2003.

[7] 刘晓冰, 黄学文, 马 跃, 等. 面向产品全生命周期的XBOM研究[J]. 计算机集成制造系统, 2002, 8(12): 983-987.

[8] FERRARI F M, TOLEDO J C. Analyzing the knowledge management through the product development process[J]. Journal of Knowledge Management, 2004, 8(1): 117-129.

[9] 张 艳. 企业知识管理技术[J]. 情报探索, 2014(3): 99-102.

[10] 苏春萍, 石 建. 对知识管理及其技术应用系统的解析[J]. 情报技术, 2005, 24(9): 35-36.

[11] LU C Y, CAI J. STARS: a socio-technical framework for integrating design knowledge over the Internet[J]. IEEE Internet Computing, 2000, 4(5): 54-62.

[12] KINGSTON J, MACINTOSH A. Knowledge management through multi-perspective modelling: representing and distributing organizational memory[J]. Knowledge-Based Systems, 2000, 13(2-3): 121-131.

[13] 索柏民, 王家斌, 王伟光. 知识管理应用技术及实施中的问题[J]. 中国科技论坛, 2005(3): 112-116.

[14] 何二宝, 张太华. 基于 PLM 的云知识环境下知识服务模式研究[J]. 现代制造工程, 2015(1): 15-19.