第 33 卷　第 5 期　　计 算 机 工 程　　2007 年 3 月
Vol.33　No.5　　Computer Engineering　　March 2007

·博士论文·　文章编号：1000—3428(2007)05—0016—04　文献标识码：A　中图分类号：TP311

# 不一致决策表的 k 阶分配序约简

黄　兵[1,2]，周献中[2]，胡作进[1]

(1. 南京审计学院计算机科学与技术系，南京 210029；2. 南京大学工程管理学院，南京 210093)

**摘　要：**在不一致决策表中定义了 k 阶分配序约简，给出了 k 阶分配序一致集的判定定理。通过定义 k 阶分配序区分矩阵，给出了求 k 阶分配序约简的区分矩阵法。为了克服区分矩阵法时间复杂度过高的缺陷，通过定义属性的相对重要性，提出了一种求 k 阶分配序约简的启发式算法，分析得到该算法的时间复杂度是多项式的结论。实例验证了算法的有效性。

**关键词：**信息系统；粗糙集；不一致决策表；k 阶分配序约简；区分矩阵

# k-ordered Assignment Reduction in Inconsistent Decision Tables

**HUANG Bing[1,2], ZHOU Xianzhong[2], HU Zuojin[1]**

(1. Department of Computer Science and Technology, Nanjing Audit University, Nanjing 210029;
2. School of Engineering & Management, Nanjing University, Nanjing 210093)

**【Abstract】**Knowledge reduction is one of the most important tasks in rough set theory. The k-ordered assignment reduction is defined in inconstant decision tables, which maintains the first k membership orders of the equivalence classes determined by attributes set to the decision classes determined by decision attributes set. The judgment theorem and discernibility matrix with respect to consistent ordered assignment set are obtained, by means of which, the method of k-ordered assignment reduction is presented. To overcome the disadvantage of ordered assignment reduction based on discernibility matrix because the time complexity is exponential, a heuristic algorithm based on the significance of attributes is proposed, which aims at acquiring one of the minimal k-ordered assignment reduction. The time complexity of the algorithm is analyzed. The experimental result shows that the algorithm is valid.

**【Key words】**Information systems; Rough sets; Inconstant decision table; k-ordered assignment reduction; Discernibility matrix

## 1 概述

粗糙集理论是一门处理不精确、不确定信息的数学理论。经过 20 余年的发展，粗集理论的发展已取得了长足进步，并成功应用于智能信息处理、模式识别、智能控制和数据挖掘等领域。

粗集理论的研究对象是一个二维信息表，称为信息系统。信息系统由一些对象通过在一些属性上的取值来构成。若属性集合分为条件集和决策集，则信息系统称为决策表。若决策表中的对象满足：在条件属性上取值相同，则在决策属性上取值也相同，则称该决策表是一致决策表；否则是不一致的。在现实中，由于数据采集能力的不足，因此往往很难保证决策表的一致性。

众所周知，信息系统中的属性并非都是必要的，在保持信息系统分类或决策能力不变的前提下，消除冗余属性和属性值，从而获取简洁有效的分类或决策规则，称为信息系统的知识约简。知识约简是粗集理论的精髓所在，一般分为属性约简和值约简。王国胤[1,2]分析了属性约简的代数观点和信息观点，得到在一致决策表中，两种约简是等价的；而在不一致决策表中，信息观点下的约简要求更为严格的结论。张文修等[3,4]将最大分布约简引入到不一致决策表中，分析了最大分布约简、分布约简、近似约简和分配约简的关系，并给出了相应的求约简的区分矩阵方法。米据生[5]进一步提出了上下近似约简，给出了相应的约简方法。黄兵等人[6]将决策序引入不一致决策表，定义了分配序约简并给出了相应的区分矩阵方法。

本文针对不一致决策表，提出了一种新的约简—k 阶分配序约简。它保持条件属性确定的等价类对不同决策类的隶属度从大到小排列的前 k 个顺序不发生变化。现实中决策有时不能仅凭少数服从多数的原则。囿于客观条件的限制，这个决策在主观上是最好的，但却可能代价过高或者风险过大，甚至根本不可能实施该决策。此时可考虑选择得票第二的决策，依此类推。k 阶分配序约简具有重要的现实意义。

针对不一致决策表，给出 k 阶分配序约简的数学定义和判定定理，依此给出 k 阶分配序约简的区分矩阵法。为克服区分矩阵法的时间复杂度过高的缺陷，提出一种求 k 阶分配序约简的启发式算法。

## 2 不一致决策表

下面给出一些基本概念：

**定义 1** 称 $S=(U, A=C\cup D, V, f)$ 是一个信息系统，其中 U 是非空有限对象集合，$U=\{x_1, x_2, \ldots, x_n\}$；C 是非空有限条件属性集合，D 是非空有限决策属性集合，且 $C\cap D=\phi$。一般地，D 是单元素集合。f 是一个 $U\times A$ 到属性值集合 V 上的一个映射，它表示每个对象在每个属性上对应一个值，称为信

**基金项目：**国家自然科学基金资助项目(70571032)；江苏省教育厅高校自然科学研究指导性计划基金资助项目(05JKD520102)；中国博士后科学基金资助项目(20060390916)江苏省博士后科研计划基金资助项目；南京审计学院科学研究基金资助项目(NSK2006/A03)

**作者简介：**黄　兵(1972－)，男，讲师、博士后，研究方向：粗糙集理论文应用；周献中，教授、博导；胡作进，副教授

**收稿日期：**2006-03-23　**E-mail：**hbhuangbing@126.com

息函数。

例如: $f(x_1, a_1) = 1$ 表示对象 $x_1$ 在属性 $a_1$ 上的对应取值是 1。

对任意 $B \subseteq C$，记

$R_B = \{(x_i, x_j) : f(x_i, b) = f(x_j, b),\ \forall \quad b \in B\}$；

$R_D = \{(x_i, x_j) : f(x_i, d) = f(x_j, d),\ \forall \quad d \in D\}$

则 $R_B, R_D$ 都是 U 上的等价关系，分别称为由条件属性子集 $B \subseteq C$ 和决策属性集 D 确定的不可区分关系。它们构成对 U 的划分，记为：$U/R_B = \{[x]_B : x \in U\}$， $U/R_D = \{[x]_D : x \in U\}$。其中：

$[x]_B = \{y \in U : (x, y) \in R_B\}$ ， $[x]_D = \{y \in U : (x, y) \in R_D\}$

分别表示 x 关于条件属性子集 B 和决策属性集 D 的等价类。

**定义 2** 对 $X \subseteq U$，由条件属性子集 $B \subseteq C$ 确定的 X 的下近似和上近似定义为

$\underline{R_B}(X) = \{x \in U : [x]_B \subseteq X\} = \bigcup\{[x]_B : [x]_B \subseteq X\}$；

$\overline{R_B}(X) = \{x \in U : [x]_B \cap X \neq \phi\} = \bigcup\{[x]_B : [x]_B \cap X \neq \phi\}$

**定义 3** 对决策表 $S = (U, A = C \cup D, V, f)$，若 $R_C \subseteq R_D$，则称决策表是一致的，否则称决策表是不一致的。

在一致决策表中，当对象在条件属性集上取值相同时，决策属性值也必定相同；而在不一致决策表中，至少存在两个对象，在条件属性集上取值相同，但它们的决策值却不相等。因此，从一致决策表中得到的决策规则都是确定的，而在不一致决策表中必定存在不确定性规则。

## 3 k 阶分配序约简

对不一致决策表 $S = (U, A = C \cup D, V, f)$，设 $U/R_D = \{D_1, D_2, \ldots, D_m\}$， $B \subseteq C$， $\forall \quad x \in U$, 记

$$\rho_B(x) = D_{i_1} \geq D_{i_2} \geq \ldots \geq D_{i_j}$$

其中，

$D_{i_l} \in U/R_D (1 \leq l \leq j)$ ， $\left|[x]_B \cap D_{i_1}\right| \geq \left|[x]_B \cap D_{i_2}\right| \geq \ldots \geq \left|[x]_B \cap D_{i_j}\right| > 0$ ，

$|X|$ 表示集合 $X$ 的基数，且 $[x]_B = \bigcup\limits_{1 \leq l \leq j} ([x]_B \cap D_{i_l})$ ；若 $D_{i_1} \geq D_{i_2} \geq \ldots \geq D_{i_j}$ 中大于成立，则用大于号，若等号成立，则下标按在 $U/R_D = \{D_1, D_2, \ldots, D_m\}$ 中从小到大排列。

**定义 4** 对不一致决策表 $S = (U, A = C \cup D, V, f)$，$B \subseteq C, \forall \ x \in U$，有唯一决策序 $\rho_B(x) = D_{i_1} \geq D_{i_2} \geq \ldots \geq D_{i_j}$ 与 $x$ 对应，称 $\rho_B(x)$ 是 U 上关于 B 的决策序函数。

**定义 5** 对不一致决策表 $S = (U, A = C \cup D, V, f)$，$x_s, x_t \in U$ ， $\rho_E(x_s) = D_{i_1} \geq D_{i_2} \geq \ldots \geq D_{i_j}$， $\rho_F(x_t) = D_{p_1} \geq D_{p_2} \geq \ldots \geq D_{p_q}$ 分别是 $x_s, x_t$ 关于 $E, F \subseteq C$ 的决策序函数。若对正整数 $k$，当 $j, q < k$ 时， $\rho_E(x_s) = \rho_F(x_t)$ ；当 $j, q \geq k$ 时， $\forall\, l \leq k$ ，有 $D_{i_l} = D_{p_l}$，则称 $x_s$ 与 $x_t$ 关于 $E, F$ 的 $k$ 阶分配序函数相等，记作 $\rho_E^k(x_s) = \rho_F^k(x_t)$ 。

**定义 6** 对不一致决策表 $S = (U, A = C \cup D, V, f)$，对正整数 $k$，若 $\forall\, x \in U$，有 $\rho_B^k(x) = \rho_C^k(x)$，则称 B 是 k 阶分配序一致集。若 B 是 k 阶分配序一致集而它的任意真子集都不是 k 阶分配序一致集，则称 B 是 S 的 k 阶分配序约简。

k 阶分配序一致集保持原信息系统每个条件属性确定的等价类对不同决策类的前 k 个最大分布顺序不发生变化；若 B 是 S 的 k 阶分配序约简，则由 B 产生的规则与由 C 产生的规则前 k 个最大的可信度大小顺序相同，而去掉 B 中任意一个属性，至少有一个规则的前 k 个最大的可信度大小顺序发生改变。

例 1 下列不一致决策表中，$U = \{x_1, x_2, \ldots, x_{10}\}$，$C = \{c_1, c_2, c_3, c_4\}$，$D = \{d\}$。

$U/R_C = \{\{x_1, x_7, x_8\}, \{x_2 x_9, x_{10}\}, \{x_3, x_5, x_6\}, \{x_4\}\}$；

$U/R_D = \{D_1 = \{x_1, x_5, x_6, x_8, x_{10}\}, D_2 = \{x_2, x_3, x_4, x_7, x_9\}\}$。

于是得

$\rho_C(x_1) = \rho_C(x_3) = \rho_C(x_5) = \rho_C(x_6) = \rho_C(x_7) = \rho_C(x_8) = D_1 > D_2$；

$\rho_C(x_2) = \rho_C(x_9) = \rho_C(x_{10}) = D_2 > D_1$；

$\rho_C(x_4) = D_2$。

**表 1 一个不一致决策表**

| U | $c_1$ | $c_2$ | $c_3$ | $c_4$ | d |
|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | 0 | 1 |
| $x_2$ | 0 | 1 | 1 | 1 | 2 |
| $x_3$ | 1 | 1 | 0 | 0 | 2 |
| $x_4$ | 0 | 0 | 1 | 0 | 2 |
| $x_5$ | 1 | 1 | 0 | 0 | 1 |
| $x_6$ | 1 | 1 | 0 | 0 | 1 |
| $x_7$ | 1 | 0 | 0 | 0 | 2 |
| $x_8$ | 1 | 0 | 0 | 0 | 1 |
| $x_9$ | 0 | 1 | 1 | 1 | 2 |
| $x_{10}$ | 0 | 1 | 1 | 1 | 1 |

**定理 1** 对不一致决策表 $S = (U, A = C \cup D, V, f)$，$B \subseteq C$ 是 $k$ 阶分配序一致集的充分必要条件是：$\forall\, x, y \in U$, 当 $\rho_C^k(x) \neq \rho_C^k(y)$ 时，有 $[x]_B \neq [y]_B$。

**证明** 因为 $B$ 是 $k$ 阶分配序一致集，所以 $\forall\, x, y \in U$, 有 $\rho_B^k(x) = \rho_C^k(x)$, ，因为 $\rho_C^k(x) \neq \rho_C^k(y)$ ，所以 $\rho_B^k(x) \neq \rho_B^k(y)$，且 $[x]_B \neq [y]_B$；反之，$\forall\, x \in U$, 当 $y \in [x]_B$ 时，由条件知， $\rho_C^k(x) = \rho_C^k(y)$，即 $[x]_B$ 中所有对象关于 C 的 k 阶分配序函数相等。显然，将它们合在一起后的 $k$ 阶分配序不变。即 $\forall\ y \in [x]_B$，有 $\rho_C^k(y) = \rho_B^k(y)$，则 B 是 k 阶分配序一致集。

由定理 1 知，条件属性子集是分配序一致集的充分必要条件是它不会将原 $k$ 阶分配序函数不等的两对象划分到同一等价类中。这提供了判断 $k$ 阶分配序一致集的方法，同时也可利用区分矩阵来求不一致决策表的 $k$ 阶分配序约简。

当由条件属性集确定的每个对象的最大分布函数值唯一时，最大分布约简[3,4]与k阶分配序约简是等价的。因此，在一定情形下，k阶分配序约简是最大分布约简的推广。

## 4 k 阶分配序区分矩阵

**定义 7** 对不一致决策表 $S = (U, A = C \cup D, V, f)$，$U = \{x_1, x_2, \ldots, x_n\}$，k阶分配序区分矩阵定义为

$$D_C^k = (c_{st})_{n \times n} = \begin{cases} \{c \in C : f(x_s, c) \neq f(x_t, c)\} & \rho_C^k(x_s) \neq \rho_C^k(x_t) \\ \phi & \rho_C^k(x_s) = \rho_C^k(x_t) \end{cases}$$

分配序区分矩阵反映了不一致决策表中具有不同分配序对象间的全部区分关系。如在例 1 中给出的不一致决策表的 2 阶分配序区分矩阵为

$$\begin{bmatrix} \phi & c_1c_2c_3c_4 & \phi & c_1c_3 & \phi & \phi & \phi & \phi & c_1c_2c_3c_4 & c_1c_2c_3c_4 \\ c_1c_2c_3c_4 & \phi & c_1c_3c_4 & c_2c_4 & c_1c_3c_4 & c_1c_3c_4 & c_1c_2c_3c_4 & c_1c_2c_3c_4 & \phi & \phi \\ \phi & c_1c_3c_4 & \phi & c_1c_2c_3 & \phi & \phi & \phi & \phi & c_1c_3c_4 & c_1c_3c_4 \\ c_1c_3 & c_2c_4 & c_1c_2c_3 & \phi & c_1c_2c_3 & c_1c_2c_3 & c_1c_3 & c_1c_3 & c_2c_4 & c_2c_4 \\ \phi & c_1c_3c_4 & \phi & c_1c_2c_3 & \phi & \phi & \phi & \phi & c_1c_3c_4 & c_1c_3c_4 \\ \phi & c_3c_4 & \phi & c_2c_3 & \phi & \phi & \phi & \phi & c_1c_3c_4 & c_1c_3c_4 \\ \phi & c_1c_2c_3c_4 & \phi & c_1c_3 & \phi & \phi & \phi & \phi & c_1c_2c_3c_4 & c_1c_2c_3c_4 \\ \phi & c_1c_2c_3c_4 & \phi & c_1c_3 & \phi & \phi & \phi & \phi & c_1c_2c_3c_4 & c_1c_2c_3c_4 \\ c_1c_2c_3c_4 & \phi & c_1c_3c_4 & c_2c_4 & c_1c_3c_4 & c_1c_3c_4 & c_1c_2c_3c_4 & c_1c_2c_3c_4 & \phi & \phi \\ c_1c_2c_3c_4 & \phi & c_1c_3c_4 & c_2c_4 & c_1c_3c_4 & c_1c_3c_4 & c_1c_2c_3c_4 & c_1c_2c_3c_4 & \phi & \phi \end{bmatrix}$$

**定理 2** k 阶分配序区分矩阵有如下性质：

(1) $c_{ss} = \phi, 1 \le s \le n$；

(2) $c_{st} = c_{ts}$, $1 \le s, t \le n$

即分配序区分矩阵是对称阵，且主对角元素都是 $\phi$。

**定理 3** 设不一致决策表 $S = (U, A = C \cup D, V, f)$ 的 k 阶分配序区分矩阵 $D_B^k = (c_{st})_{n \times n}$，则 $B \subseteq C$ 是 k 阶分配序一致集 $\Leftrightarrow \forall c_{st} \ne \phi$，有 $B \cap c_{st} \ne \phi$。

**证明** $\forall c_{st} \ne \phi$，则 $\rho_C^k(x_s) \ne \rho_C^k(x_t)$，而 B 是 k 阶分配序一致集，由定理 1 知，$[x_s]_B \ne [x_t]_B$，则 $\exists b \in B$，使得 $f(x_s, b) \ne f(x_t, b)$，即 $b \in c_{st}$，故 $B \cap c_{st} \ne \phi$。反之，若 $\rho_C^k(x_s) \ne \rho_C^k(x_t)$，则 $c_{st} \ne \phi$，因为 $B \cap c_{st} \ne \phi$，则 $\exists\ b \in B \cap c_{st}$，有 $f(x_s, b) \ne f(x_t, b)$，从而 $[x_s]_B \ne [x_t]_B$，由定理 1 知，B 是 k 阶分配序一致集。

**定义 8** 设不一致决策表 $S = (U, A = C \cup D, V, f)$ 的 k 阶分配序区分矩阵 $D_C^k = (c_{st})_{n \times n}$，称 $D^k = \underset{c_{st} \ne \phi}{\wedge} (\vee c_{st})$ 为 S 的 k 阶分配序辨识公式。

**定理 4** 辨识公式 $D^k = \underset{c_{st} \ne \phi}{\wedge} (\vee c_{st})$ 的极小范式中每个合取范式中所有属性组成的集合是 S 所有的 k 阶分配序约简集合。

**证明**：由极小范式和分配序矩阵的定义易证。

例 2 用 2 阶分配序辨识公式求表 1 给出的不一致决策表的 2 阶分配序约简，得约简结果 $\{c_1, c_2\}$，$\{c_1, c_4\}$，$\{c_2, c_3\}$，$\{c_3, c_4\}$；用 1 阶分配序区分矩阵和辨识公式求 1 阶分配序约简，得约简结果 $\{c_1\}$，$\{c_3\}$。

经过 k 阶分配序约简后，再经过值约简，可以得到 k 阶分配序规则。

## 5 k 阶分配序约简的启发式算法

用 k 阶分配序区分矩阵和 k 阶分配序辨识公式能找到不一致决策表的所有 k 阶分配序约简，但该方法的时间复杂度随不一致决策表的大小的增长而呈指数增长。同时，人们往往关心的是一个信息系统的最小约简而不是它的所有约简。因此，有必要给出一种以寻找不一致决策表的最小 k 阶分配序约简为目的的、时间复杂度较低的 k 阶分配序约简启发式算法。

**定义 9** 不一致决策表 $S = (U, A = C \cup D, V, f)$ 中，称 $b \in B \subseteq C$ 在 B 中是 k 阶分配序不必要的，若 $\rho_{B \setminus \{b\}}^k = \rho_B^k$；否则称 $b \in B \subseteq C$ 在 B 中是 k 阶分配序必要的。

**定义 10** 不一致决策表 $S = (U, A = C \cup D, V, f)$ 中，C 中所有必要属性构成的集合称为 S 的 k 阶分配序核。记为：$Core^k(C) = \{c \in C : \rho_{C \setminus \{c\}}^k \ne \rho_C^k\}$。

**定理 5** $Core^k(C) = \cap \operatorname{Re} d^k(C)$，其中 $\operatorname{Re} d^k(C)$ 表示所有的 k 阶分配序约简。

**证明** $\forall\ c \in Core^k(C)$，则 $\exists\ x \in U, \rho_{C \setminus \{c\}}^k(x) \ne \rho_C^k(x)$，由定理 1 知，$\exists\ y \in U, \rho_C^k(x) \ne \rho_C^k(y)$，使得 $y \in [x]_{C \setminus \{c\}}$ 若 $\exists\ B \in \operatorname{Re} d^k(C)$，使 $c \notin B$，则 $B \subseteq C \setminus \{c\}$，从而 $y \in [x]_B$。由于 $B$ 是一个 $k$ 阶分配序约简，则 $\rho_C^k(x) = \rho_B^k(x) = \rho_B^k(y) = \rho_C^k(y)$，矛盾。故 $Core^k(C) \subseteq \cap \operatorname{Re} d^k(C)$。反之，$\forall\ c \in \cap \operatorname{Re} d^k(C)$，若 $c \notin Core^k(C)$，则 $\rho_{C \setminus \{c\}}^k = \rho_C^k$，即 $C \setminus \{c\}$ 是 k 阶分配序一致集，则 $C \setminus \{c\}$ 至少存在一个子集是 k 阶分配序约简，而这个约简不含 c，矛盾。故 $\cap \operatorname{Re} d^k(C) \subseteq Core^k(C)$。综上有 $Core^k(C) = \cap \operatorname{Re} d^k(C)$。

**定义 11** 在不一致决策表 $S = (U, A = C \cup D, V, f)$ 中，$b \in B \subseteq C$ 在 B 中的 k 阶分配序重要度定义为

$$Sig_{B \setminus \{b\}}^k(b) = \frac{\left|\{x \in U : \rho_{B \setminus \{b\}}^k(x) \ne \rho_B^k(x)\}\right|}{|U|}$$

显然，$Sig_{B \setminus \{b\}}^k(b) > 0 \Leftrightarrow b$ 在 $B$ 中是 $k$ 阶分配序必要的；$Core^k(C) = \{c \in C : Sig_{C \setminus \{c\}}^k(c) > 0\}$。

**定义 11** 不一致决策表 $S = (U, A = C \cup D, V, f)$ 中，$B \subset C, b \in C \setminus B$ 对 $B$ 的 $k$ 阶分配序重要度定义为

$$Sig_B^k(b) = \frac{\left|\{x \in U : \rho_{B \cup \{b\}}^k(x) \ne \rho_B^k(x)\}\right|}{|U|}$$

下面给出利用 $k$ 阶分配序属性重要度为启发知识的 $k$ 阶分配序约简算法。

**算法**

输入不一致决策表 $S = (U, A = C \cup D, V, f)$；

输出 $S = (U, A = C \cup D, V, f)$ 的一个 $k$ 阶分配序约简。

(1)计算 $Core^k(C) = \{c \in C : Sig_{C \setminus \{c\}}^k(c) > 0\}$；

(2)令 $R = Core^k(C)$，若 $Core^k(C) = \phi$，则选取 $c \in C$，使得 $U / R_{\{c\}}$ 不同等价类上的 $k$ 阶分配序函数值最多。

1)若 $\rho_R^k = \rho_C^k$，则终止；2)对所有 $c \in C \setminus R$，计算 $Sig_R^k(c)$；3)取 $c_0 \in \{c : Sig_R^k(c) = \max\limits_{c \in C \setminus R} Sig_R^k(c)\}$，若 $\{c : Sig_R^k(c) = \max\limits_{c \in C \setminus R} Sig_R^k(c)\}$ 中不止一个元素，则选取使 $U / R \cup \{c_0\}$ 构成的等价类最少的 $c_0$，作 $R = R \cup \{c_0\}$，转 1)。

(3)设第(2)步加入到 $R = Core^k(C)$ 中的属性共有 r 个，记为 $\{c_{i_j} : j = 1, 2, \ldots, r\}$，其中 j 表示加入顺序。逆向检查：$Sig_{R \setminus \{c_{i_j}\}}^k(c_{i_j}) = 0$，若成立，则 $R = R \setminus \{c_{i_j}\}$；否则 R 不变;

(4)输出 R。

其中，第(2)步的 3)中 $c_0$ 的选取一方面是要满足重要程度最大，另一方面是为了最后得到的规则最简(条件属性确定的等价类最少)。

算法时间复杂度分析：

在第(1)步中，需要计算所有由条件属性集 C 确定的等价类，时间复杂度为 $O(|U|^2 |C|)$。同时对所有 $c \in C$，需要计算 $C \setminus \{c\}$ 确定的所有等价类，故时间复杂度为 $O(|U|^2 |C|^2)$。

在第(2)步中，步骤 1)的时间复杂度是 $O(|U|)$；在步骤 2)中，计算 $Sig_R(c)$ 的时间复杂度是 $O(|U|^2 |C|)$，在最坏情形下需

循环$|C|$次，因此时间复杂度为$O(|U|^2|C|^2)$；步骤(3)需要时间$|C|$。在最坏情形下，完成第(2)步需要循环$|C|$次。因此第(2)步的整个时间复杂度为$O(|U|^2|C|^3)$。

第(3)步的时间复杂度是$|C|$。

综上可知，算法的时间复杂度是$O(|U|^2|C|^3)$。

## 6 启发式算法的实例分析

下面利用 $k$ 阶分配序约简启发式算法来求表 1 给出的不一致决策表的 2 阶分配序约简。

第 1 步：计算 $Core^k(C)=\{c\in C: Sig^k_{C\setminus\{c\}}(c)>0\}=\phi$；

第 2 步：对每个计算 $c\in C$，计算 $U/R_c$ 不同等价类上的分配序函数值个数，$c_1,c_3,c_4$ 均有 2 个，而 $c_2$ 只有 1 个，于是在 $c_1,c_3,c_4$ 中任选一个，不妨选 $c_1$，令 $R=\{c_1\}$，判断 $\rho^k_R=\rho^k_C$，不成立。计算 $Sig^k_R(c_2)=1/10$、$Sig^k_R(c_3)=0$ 和 $Sig^k_R(c_4)=1/10$；因为 $U/\{c_1,c_2\}$ 有 4 个等价类，而 $U/\{c_1,c_4\}$ 有 3 个等价类，因此选 $c_4$，作 $R=R\cup\{c_4\}=\{c_1,c_4\}$，判断 $\rho^k_R=\rho^k_C$，成立。终止。输出 $R=\{c_1,c_4\}$ 即为一个 2 阶分配序约简。

如果在第 1 步选 $c_3$，则得到的约简是 $R=\{c_3,c_4\}$；而如果选 $c_4$，则得到 $R=\{c_1,c_4\}$ 或 $R=\{c_3,c_4\}$。这启示我们，可以用并行方式得到规则最简的所有最小 $k$ 阶分配序约简。

## 7 总结与讨论

传统的知识约简只保证确定性规则不变，即保持相对正域不发生变化，这是代数意义下的约简。在不一致决策表中，由于可能性规则(不确定性规则)的大量存在，因此提取这些简洁的可能性规则就需要对不一致决策表进行知识约简。信息观点的不一致决策表约简保持由条件属性集确定的等价类对不同决策类的隶属度不发生变化，即分布约简，这种约简的要求显然极为苛刻，实用性不大。

本文在文献[3]的基础上，定义了一种新的不一致决策表约简—— $k$ 阶分配序约简，并给出了相应的区分矩阵约简方法。为降低区分矩阵法复杂度过高的缺陷，提出了一种以寻找规则最简的最小约简为目的的、基于属性重要度的启发式 $k$ 阶分配序约简算法。通过实例分析可知，该算法能找到规则最简的最小约简。

如果将分配序函数中条件属性确定的等价类对决策属性确定的等价类隶属度相同的记为一个，则 1 阶分配序函数即为文献[3]定义的最大分布函数，1 阶分配序约简也为最大分布约简。

因此，本文提出的 $k$ 阶分配序约简可以很容易地类推为最大分布约简的更一般情形。$k$ 阶分配序约简可能造成规则可信度的降低，因此可以将 $k$ 阶分配序约简与变精度粗糙集模型[7,8]结合起来，定义满足一定精度并保持 $k$ 阶分配序不变的变精度 $k$ 阶分配序约简，并给出相应的约简算法。同时，由于客观原因，信息系统往往不是完备的（数据缺损），因此，建立适当的 $k$ 阶分配序变精度模型，更具实用价值。这些是我们下一步的工作。

### 参考文献


1 王国胤, 于 洪, 杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759-766.

2 王国胤. 决策表核属性的计算方法[J]. 计算机学报, 2003, 26(5): 611-615.

3 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.

4 张文修, 梁 怡, 吴伟志. 信息系统与知识发现[M]. 北京: 科学出版社, 2003-09.

5 Mi Jusheng, Wu Weizhi, Zhang Wenxiu. Approaches to Approximation Reducts in Inconsistent Decision Tables[C]//Proc. of the 9th International Conference on RSFDrC'03. 2003: 283-286.

6 黄 兵, 周献中. 不一致决策表的分配序约简[J]. 南京理工大学学报, 2005, 29(3): 360-362, 367.

7 Ziarko W. Variable Precision Rough Set Model[J]. Journal of Computer and System Sciences, 1993, 46(1): 39-59.

8 米据生, 吴伟志, 张文修. 基于变精度粗糙集理论的知识约简方法[J]. 系统工程理论与实践, 2004, 24(1): 76-82.


（上接第 15 页）

由图 2 可见，TPFAC 算法的运行时间和数据对象的个数略呈线性关系，与本文的时间复杂度推算大致一致。由于新算法采用两阶段聚类和调整 C 值控制后期收敛，在运行时间上要优于 LF 算法，且在处理大数据集时表现更为明显。同时，TPFAC 算法的分类错误率也优于 LF 算法，主要原因在于 TPFAC 算法第 1 阶段对相近似的数据进行了融合操作和设置了蚂蚁搜索禁忌表。

## 4 结束语

本文以 LF 聚类算法为基础，针对传统蚂蚁聚类算法计算效率低的问题，从“隐蔽疏散配置”这一陆战旅待机地域选取原则出发，提出了一种显著提高算法效率的两阶段模糊蚂蚁聚类算法 TPFAC 求解陆战旅待机地域选取问题。

实验证明，在无需特定的聚类先验信息的条件下，该算法以更好的聚类效果改善了传统蚂蚁聚类算法的性能，是一种高效率、鲁棒性好的算法。该方法实现了陆战旅待机地域选取方案的自动、准确、快速计算机辅助生成。

### 参考文献


1 刘小力. 登陆与抗登陆实践和经验[M]. 北京: 蓝天出版社, 2004.

2 Lumer E, Faieta B. Diversity and Adaptation in Populations of Clustering Ants[C]//Proc. of the 3rd International Conference on Simulation of Adaptive Behavior: from Animals to Animates, Cambridge. 1994.

3 Chen Yunfei, Liu Yushu, Fattah A, et al. HDACC: a Heuristic Density-based Ant Colony Clustering Algorithm[C]//Proc. of the 2004 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Beijing. 2004: 397-400.

4 杨 燕, 靳 蕃, Kamel M. 一种基于蚁群算法的聚类组合方法[J]. 铁道学报, 2004, 26 (4): 64-69.

5 彭喜元, 彭 宇, 戴毓丰. 群智能理论及应用[J]. 电子学报, 2003, 31(12A): 1982-1988.