

DOI: 10.11992/tis.201905059

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20190719.1623.006.html>

不完备决策系统下的多特定类广义决策约简

唐玉凯^{1,2}, 张楠^{1,2}, 童向荣^{1,2}, 张小峰³

(1. 烟台大学数据科学与智能技术山东省高校重点实验室, 山东烟台 264005; 2. 烟台大学计算机与控制工程学院, 山东烟台 264005; 3. 鲁东大学信息与电气工程学院, 山东烟台 264025)

摘要: 属性约简是粗糙集理论研究中最重要的领域之一。经典的不完备决策系统广义决策约简关注决策系统中的所有决策类, 而在实际应用中, 决策者往往只关注一个或者几个特定决策类。针对以上问题, 提出基于多特定类的不完备决策系统广义决策约简理论框架。首先, 定义了单特定类的不完备决策系统广义决策约简的相关概念, 提出并证明相关定理, 构造相应差别矩阵和区分函数。其次, 将单特定类的广义决策约简推广到多特定类, 提出基于差别矩阵的多特定类的不完备决策系统广义决策约简算法。最后, 采用 6 组 UCI 数据集进行实验。实验结果表明, 相对全部决策类数量, 当选定特定类数量较少时, 平均约简长度有不同程度的缩短, 占用空间有所减小, 约简效率有不同程度的提升。

关键词: 粗糙集; 属性约简; 不完备; 决策系统; 相容关系; 多特定类; 广义决策约简; 差别矩阵

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2019)06-1199-10

中文引用格式: 唐玉凯, 张楠, 童向荣, 等. 不完备决策系统下的多特定类广义决策约简 [J]. 智能系统学报, 2019, 14(6): 1199-1208.

英文引用格式: TANG Yukai, ZHANG Nan, TONG Xiangrong, et al. The multi-class-specific generalized decision preservation reduction in incomplete decision systems[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1199-1208.

The multi-class-specific generalized decision preservation reduction in incomplete decision systems

TANG Yukai^{1,2}, ZHANG Nan^{1,2}, TONG Xiangrong^{1,2}, ZHANG Xiaofeng³

(1. Key Lab for Data Science and Intelligence Technology of Shandong Higher Education Institutes, Yantai University, Yantai 264005, China; 2. School of Computer and Control Engineering, Yantai University, Yantai 264005, China; 3. School of Information and Electrical Engineering, Ludong University, Yantai 264025, China)

Abstract: Attribute reduction has an important place in rough set theory. The method of classical generalized decision preservation reduction in incomplete decision systems is to find the reducts of all decision classes. In practical applications, however, the decision makers may focus on one or several decision classes. To fill this gap, the theoretical framework of multi-class-specific generalized decision preservation reduction in incomplete decision systems is proposed. First, the single-class-specific generalized decision preservation reduction in incomplete decision systems is defined. Related theorems are proposed and proven, and the corresponding discernibility matrix and function are constructed. Then, the single-class-specific generalized decision preservation reduction is extended to the multi-class-specific generalized decision preservation reduction in incomplete decision systems. The algorithm of the multi-class-specific generalized decision preservation reduction based on discernibility matrix in incomplete decision systems (MGDRDM) is proposed. Finally, six datasets from UCI were used for experiments. The experimental results show that when the number of selected specific classes is less than all the decision classes, the average length of reducts will be shortened to varying degrees, the space used will be reduced, and the time efficiency will be roughly improved.

Keywords: rough sets; attribute reduction; incomplete; decision systems; tolerance relation; multi-class-specific; generalized decision preservation reduction; discernibility matrix

收稿日期: 2019-05-28. 网络出版日期: 2019-07-19.

基金项目: 国家自然科学基金项目 (61572418, 61572419, 61873117, 61403329); 山东省自然科学基金项目 (ZR2018BA004, ZR2016FM42).

通信作者: 张楠. E-mail: zhangnan0851@163.com.

粗糙集理论^[1-5] 是于 1982 年由波兰科学家 Pawlak 提出的一种用于分析和处理不确定或不精确数据的数学工具。目前, 粗糙集理论正被广泛

应用于人工智能、机器学习、模式识别、数据挖掘等领域,并取得了丰硕的研究成果。属性约简^[6-12]是粗糙集理论的核心研究内容之一。属性约简的本质是在保持原决策系统某种分辨能力不变的前提下,删除冗余属性,获取最小属性子集。

在决策系统中,若某个条件属性值存在缺失,则称该决策系统为不完备决策系统。目前,国内外相关学者针对不完备决策系统做了大量的研究工作。2002年,Liang等^[13]提出了基于粗糙熵的不完备决策系统约简算法。2010年,周献中等^[14]介绍了在不完备决策系统下基于相容关系和相似关系的粗糙集模型及其拓展模型,通过引入相容矩阵和相似矩阵,系统研究了相应的粗计算、属性约简以及决策规则提取的矩阵算法。2010年Qian等^[15]提出了基于极大相容块的不协调不完备决策系统下的上下近似约简概念并给出构造相应差别矩阵方法。2014年,Shu等^[16]提出了不完备决策系统下基于候选属性重要度的快速求属性约简方法。2015年,Qian等^[17]提出了基于紧凑差别矩阵的动态不完备决策系统下的特征选择方法。2018年,Xie等^[18]提出了不完备决策系统下不协调度的概念,证明了基于不协调度与基于正域的属性约简等价。

在不完备决策系统中,具有相同条件属性的对象可能决策出多个不同的决策值,每个对象所有可能决策出的决策值称为该对象的广义决策值,广义决策约简旨在保持约简前后每个对象的广义决策值不变。1993年,Skowron^[19]提出了广义决策的概念。1998年,Kryszkiewicz^[20]讨论了不完备决策系统下的广义决策约简问题,给出相关决策规则的提取方法,并提出了基于差别矩阵^[21]的广义决策约简方法。上述研究考虑决策属性的所有决策类,而在实际应用中,决策者往往只关注部分决策类。为此相关学者针对局部约简即特定类约简做了大量研究。尹继亮等^[22]讨论了区间值系统下的局部属性约简。Qian等^[23]为解决经典粗糙集无法处理具有有限标记的大数据集的问题,引入了局部粗糙集的概念。Liu等^[24]提出了第 l 决策类下近似约简概念并提出相应差别矩阵构造方法。

综合上述研究,文献[20]研究讨论了不完备决策系统所有决策类的广义决策约简,文献[22]只在区间值系统下对局部约简进行了研究,文献[24]在完备系统下对单特定类的正域约简进行了研究。然而,基于多特定类的不完备决策系统下广义决策约简未见报道。为此,本文提出基于多特定类的不完备决策系统广义决策约简的理论

框架,定义了单特定类的不完备决策系统广义决策约简的相关概念,提出并证明相关定理,构造相应差别矩阵和区分函数,并将单特定类推广到多特定类,提出基于差别矩阵的多特定类不完备决策系统广义决策约简算法并通过实验验证了算法的有效性。

1 基本概念

为方便进一步研究基于多特定类的不完备决策系统广义决策约简,本节将给出不完备决策系统及广义决策约简的相关基本概念。

1.1 不完备决策系统及其上下近似

定义 1^[16] 四元组 $DS = (U, AT = C \cup D, V, f)$ 为一个决策系统,其中 U 表示所有对象的非空有限集合,称之为论域; C 表示条件属性的非空有限集合; D 表示决策属性的非空有限集合, $C \cap D = \emptyset$; $V = \bigcup_{a \in AT} V_a$, V_a 表示属性 $a \in AT$ 的值域; $f: U \times AT \rightarrow V$ 是一个信息函数,它使得任意对象的任意一个属性都有一个信息值, $f(x, a)$ 表示对象 $x \in U$ 在属性 $a \in AT$ 上的取值。

定义 2^[16] 四元组 $DS = (U, AT = C \cup D, V, f)$ 为一个决策系统,对任意非空属性集 $B \subseteq AT$ 导出论域 U 上的不可区分关系 $IND(B)$ 定义为

$$IND(B) = \{(x, y) \in U \times U \mid \forall b \in B, f(x, b) = f(y, b)\} \quad (1)$$

不可区分关系是一个满足自反性、对称性和传递性的等价关系。由不可区分关系 $IND(B)$ 导出对论域 U 的划分为 $U/IND(B) = \{[x]_B \mid x \in U\}$, 简记为 U/B , 其中 $[x]_B$ 表示包含 x 的等价类。

定义 3^[16] 四元组 $DS = (U, AT = C \cup D, V, f)$ 为一个决策系统,若存在条件属性 $c \in C$ 使得 V_c 含有缺失值,则称该决策系统为不完备决策系统,用 $IDS = (U, AT = C \cup D, V, f)$ 表示, V_c 中的缺失值使用 * 表示。

在不完备决策系统 IDS 中,决策属性 $d \in D$ 的值域 V_d 不含有缺失值。若决策属性集 D 中决策属性数目为 1, 则称 IDS 为不完备单决策系统;若 D 中决策属性数目大于 1, 则称 IDS 为不完备多决策系统。为表述方便,本文只考虑不完备单决策系统的情况,即 $IDS = (U, AT = C \cup \{d\}, V, f)$ 的情况。

定义 4^[20] 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统,对任意条件属性集 $A \subseteq C$, 定义其相容关系为

$$SIM(A) = \{(x, y) \in U \times U \mid \forall a \in A, f(x, a) = f(y, a) \vee f(x, a) = * \vee f(y, a) = *\} \quad (2)$$

其中 $f(x, a)$ 表示对象 $x \in U$ 在属性 $a \in A$ 上的取值。

相容关系 $\text{SIM}(A)$ 具有以下性质:

性质 1^[20] 四元组 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$ 为一个不完备决策系统, 对 $\forall A \subseteq C$, 有:

$$\text{SIM}(A) = \bigcap_{a \in A} \text{SIM}(\{a\}) \quad (3)$$

性质 2 四元组 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$ 为一个不完备决策系统, 对 $\forall A \subseteq C$, $\text{SIM}(A)$ 满足:

- 1) 自反性, 对 $\forall x \in U$, 有 $(x, x) \in \text{SIM}(A)$;
- 2) 对称性, 对 $\forall x, y \in U$, 若 $(x, y) \in \text{SIM}(A)$, 则 $(y, x) \in \text{SIM}(A)$;
- 3) 非传递性, 对 $\forall x, y, z \in U$, 若 $(x, y) \in \text{SIM}(A)$ 且 $(y, z) \in \text{SIM}(A)$, $(x, z) \in \text{SIM}(A)$ 不一定成立。

设 $S_A(x) = \{y \in U | (x, y) \in \text{SIM}(A)\}$, 其中条件属性集 $A \subseteq C$, $S_A(x)$ 表示通过条件属性集 A 与 x 可能不可区分的对象的最大集合。设 $D_A(x) = \{y \in U | (x, y) \notin \text{SIM}(A)\}$, 其中条件属性集 $A \subseteq C$, $D_A(x)$ 表示通过条件属性集 A 与 x 可能可区分的对象的最大集合。对 $\forall x \in U$, 有 $S_A(x) \cap D_A(x) = \emptyset$ 且 $S_A(x) \cup D_A(x) = U$ 。

在 IDS 中, 对 $\forall A \subseteq C$, 有 $U/\text{SIM}(A)$ 表示分类且 $U/\text{SIM}(A) = \{S_A(x) | x \in U\}$ 。 $U/\text{SIM}(A)$ 中的每一个元素 $S_A(x) (x \in U)$ 被称为相容类, 且满足 $\cup U/\text{SIM}(A) = U$ 。

根据相容类, 可得到对象集 $X \subseteq U$ 有关于条件属性集 $A \subseteq C$ 的下近似集和上近似集。

定义 5^[20] 四元组 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$ 为一个不完备决策系统, 对 $\forall X \subseteq U, \forall A \subseteq C$, X 关于条件属性集 A 的下近似集和上近似集分别定义为

$$\underline{A}X = \{x \in U | S_A(x) \subseteq X\} = \{x \in X | S_A(x) \subseteq X\} \quad (4)$$

$$\overline{A}X = \{x \in U | S_A(x) \cap X \neq \emptyset\} = \cup \{S_A(x) | x \in X\} \quad (5)$$

例 1 不完备决策系统 IDS 如表 1 所示, 论域 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $C = \{a_1, a_2, a_3, a_4\}$ 为条件属性集, 决策属性集为 $\{d\}$ 。

表 1 不完备决策系统
Table 1 An incomplete decision system

U	a_1	a_2	a_3	a_4	d
x_1	*	*	2	2	1
x_2	1	*	1	2	1
x_3	1	1	2	*	2
x_4	2	2	2	1	2
x_5	*	1	1	2	3
x_6	2	1	1	*	3

由定义 4 得 IDS 在条件属性集 C 上相容类集合 $U/\text{SIM}(C) = \{S_C(x_1), S_C(x_2), S_C(x_3), S_C(x_4), S_C(x_5), S_C(x_6)\}$, 其中 $S_C(x_1) = \{x_1, x_3\}$, $S_C(x_2) = \{x_2, x_5\}$, $S_C(x_3) = \{x_1, x_3\}$,

$S_C(x_4) = \{x_4\}$, $S_C(x_5) = \{x_2, x_5, x_6\}$, $S_C(x_6) = \{x_5, x_6\}$ 。决策属性 d 对论域 U 的划分为 $U/\{d\} = \{D_1, D_2, D_3\}$, 其中 $D_1 = \{x_1, x_2\}$, $D_2 = \{x_3, x_4\}$, $D_3 = \{x_5, x_6\}$ 。由 IDS 中条件属性集 C 的相容类集合 $U/\text{SIM}(C)$ 及决策属性 d 对论域 U 的划分 $U/\{d\}$ 可得决策类 $D_i \in U/\{d\} (i = 1, 2, 3)$ 在 C 上的下近似集 $\underline{C}D_i$ 和上近似集 $\overline{C}D_i$ 分别为:

$$\underline{C}D_1 = \emptyset, \overline{C}D_1 = \{x_1, x_2, x_3, x_5\};$$

$$\underline{C}D_2 = \{x_4\}, \overline{C}D_2 = \{x_1, x_3, x_4\};$$

$$\underline{C}D_3 = \{x_6\}, \overline{C}D_3 = \{x_2, x_5, x_6\}。$$

1.2 不完备决策系统广义决策约简

定义 6^[20] 四元组 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$ 为一个不完备决策系统, 对 $\forall x \in U$, 关于条件属性集 $A \subseteq C$ 的广义决策值定义为

$$\partial_A(x) = \{f(y, d) | y \in S_A(x)\} \quad (6)$$

在 IDS 中, $\partial_A(x)$ 表示对象 x 关于条件属性集 A 的所有可能决策值的集合。

属性约简旨在保证决策系统的某种分类能力不变的情况下, 删除冗余属性, 获得最小属性子集。给出不完备决策系统的广义决策约简定义如下。

定义 7^[20] 四元组 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$ 为一个不完备决策系统, 若条件属性集 $A \subseteq C$ 为属性集 C 的一个广义决策约简, 当且仅当满足以下两个条件:

- 1) 对 $\forall x \in U$, 有 $\partial_A(x) = \partial_C(x)$;
- 2) 对 $\forall A' \subset A, \exists x' \in U$, 有 $\partial_{A'}(x') \neq \partial_A(x')$ 。

条件 1) 保证了条件属性联合的充分性, 即约简前后决策系统广义决策值的一致性; 条件 2) 保证了条件属性个体的必要性, 即缺少任意一个必要属性, 则无法保持决策系统的广义决策值不变。

由广义决策约简定义可以构造相应的差别矩阵及区分函数, 定义如下。

定义 8^[20] 四元组 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U = \{x_1, x_2, \dots, x_n\}$, $\forall x, y \in U$, IDS 广义决策约简的差别矩阵为 $n \times n$ 的矩阵, 记为 $M_{\text{GEN}} = (M_{\text{GEN}}(x, y))_{n \times n}$, 其中矩阵元素 $M_{\text{GEN}}(x, y)$ 为

$$M_{\text{GEN}}(x, y) = \begin{cases} \{a \in C | f(x, a) \neq f(y, a)\} \wedge \\ f(x, a) \neq * \wedge f(y, a) \neq * \} & y \in \Pi_{\text{GEN}} \\ \emptyset, & \text{其他} \end{cases} \quad (7)$$

其中 $\Pi_{\text{GEN}} = \{z \in U | f(z, d) \notin \partial_C(x)\}$ 。

定义 9^[20] 四元组 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$ 为一个不完备决策系统, M_{GEN} 为 IDS 广义决策约简的差别矩阵, 若 $M_{\text{GEN}}(x, y) = \{a_1, a_2, \dots, a_k\} \neq \emptyset$, 其中 k 表示 $M_{\text{GEN}}(x, y)$ 中属性的数量, 用 $\vee M_{\text{GEN}}(x, y)$

表示布尔函数 $a_1 \vee a_2 \vee \dots \vee a_k$, 对 $\forall M_{GEN}(x,y) \neq \emptyset$, IDS 广义决策约简的区分函数为

$$DF(M_{GEN}) = \wedge(\vee M_{GEN}(x,y)) \quad (8)$$

例 2 不完备决策系统 IDS 如表 1 所示, 论域 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $C = \{a_1, a_2, a_3, a_4\}$ 为条件属性集, 决策属性集为 $\{d\}$ 。

由定义 6 得 IDS 在条件属性集 C 上广义决策值 $\partial_C(U) = \{\partial_C(x_1), \partial_C(x_2), \partial_C(x_3), \partial_C(x_4), \partial_C(x_5), \partial_C(x_6)\}$, 其中 $\partial_C(x_1) = \{1, 2\}$, $\partial_C(x_2) = \{1, 3\}$, $\partial_C(x_3) = \{1, 2\}$, $\partial_C(x_4) = \{2\}$, $\partial_C(x_5) = \{1, 3\}$, $\partial_C(x_6) = \{3\}$ 。IDS 中所有决策类的广义决策约简差别矩阵为

$$M_{GEN} = \begin{bmatrix} \emptyset & \emptyset & \emptyset & \emptyset & \{a_3\} & \{a_3\} \\ \emptyset & \emptyset & \{a_3\} & \{a_1, a_3, a_4\} & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \{a_3\} & \{a_1, a_3\} \\ \{a_4\} & \{a_1, a_3, a_4\} & \emptyset & \emptyset & \{a_2, a_3, a_4\} & \{a_2, a_3\} \\ \emptyset & \emptyset & \{a_3\} & \{a_2, a_3, a_4\} & \emptyset & \emptyset \\ \{a_3\} & \{a_1\} & \{a_1, a_3\} & \{a_2, a_3\} & \emptyset & \emptyset \end{bmatrix}$$

根据 IDS 所有决策类的广义决策约简的差别矩阵 M_{GEN} 可得对应的区分函数为

$$DF(M_{GEN}) = (a_3) \wedge (a_1 \vee a_3 \vee a_4) \wedge (a_1 \vee a_3) \wedge (a_4) \wedge (a_2 \vee a_3 \vee a_4) \wedge (a_2 \vee a_3) \wedge (a_1) = a_1 \wedge a_3 \wedge a_4$$

因此, IDS 的所有决策类的广义决策约简为 $\{a_1, a_3, a_4\}$ 。

2 多特定类广义决策约简

在实际应用中, 相较经典广义决策约简中关注全部决策类, 决策者往往只关注决策属性中的一种或者几种决策类。因此, 对于某些特定决策类的约简可能更有意义。本节将讨论不完备决策系统下多特定类的广义决策约简。

2.1 基于差别矩阵的单特定类的广义决策约简

定义 10 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 对 $\forall D_i \in U/\{d\}, i = 1, 2, \dots, |U/\{d\}|$, $\forall A \subseteq C, \forall x \in \{z \in U | D_i \cap S_C(z) \neq \emptyset\}$, 若满足 $\partial_A(x) = \partial_C(x)$, 则称条件属性集 A 为单特定类 D_i 的广义决策协调集。

定义 11 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 对 $\forall D_i \in U/\{d\}, i = 1, 2, \dots, |U/\{d\}|$, 若 $\forall A \subseteq C$ 为单特定类 D_i 的广义决策约简当且仅当满足以下两个条件:

- 1) $\forall x \in \{z \in U | D_i \cap S_C(z) \neq \emptyset\}, \partial_A(x) = \partial_C(x)$;
- 2) $\forall A' \subset A, \exists x' \in \{z \in U | D_i \cap S_C(z) \neq \emptyset\}$, 使得 $\partial_{A'}(x') \neq \partial_C(x')$ 。

定理 1 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 对 $\forall D_i \in U/\{d\}, i = 1, 2, \dots, |U/\{d\}|$, 条件属性集 $A \subseteq C$ 为单特定类 D_i 的广义决策协调集当且仅当 $D_i \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$ 时, 有 $(x, y) \notin SIM(A)$ 。

证明:

充分性: 因条件属性集 A 是 IDS 单特定类 D_i 的广义决策协调集, 对 $\forall x \in \{z \in U | D_i \cap S_C(z) \neq \emptyset\}$, 有 $\partial_A(x) = \partial_C(x)$ 。当 $D_i \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$, 即 $f(y, d) \notin \partial_A(x)$, 即 $(x, y) \notin SIM(A)$ 。

必要性: 当 $D_i \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$ 时, 有 $(x, y) \notin SIM(A)$, 所以有 $f(y, d) \notin \partial_A(x)$ 成立。当满足 $D_i \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$ 时, 必有 $f(y, d) \notin \partial_A(x)$, 即 $\partial_C(x) \supseteq \partial_A(x)$, 又因为 $A \subseteq C, \partial_A(x) \supseteq \partial_C(x)$, 所以当 $\forall x \in \{z \in U | D_i \cap S_C(z) \neq \emptyset\}$ 时, 有 $\partial_A(x) = \partial_C(x)$ 。因此条件属性集 $A \subseteq C$ 为单特定类 $D_i \in U/\{d\}$ 的广义决策协调集。证毕。

定义 12 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, $U = \{x_1, x_2, \dots, x_n\}, \forall x, y \in U$, IDS 单特定类 $D_i \in U/\{d\}$ 广义决策约简的差别矩阵为 $n \times n$ 的矩阵, 记为 $M_i = (M_i(x, y))_{n \times n}$, 其中矩阵元素 $M_i(x, y)$ 为

$$M_i(x, y) = \begin{cases} \{a \in C | f(x, a) \neq f(y, a) \wedge f(x, a) \neq * \wedge f(y, a) \neq *\} & (x, y) \in \Pi_i \\ \emptyset, & \text{其他} \end{cases} \quad (9)$$

其中, $\Pi_i = \{(x, y) | x, y \in U, D_i \cap S_C(x) \neq \emptyset \wedge f(y, d) \notin \partial_C(x)\}, i = 1, 2, \dots, |U/\{d\}|$ 。

定理 2 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 论域 $U = \{x_1, x_2, \dots, x_n\}$, 对 $\forall D_i \in U/\{d\}, i = 1, 2, \dots, |U/\{d\}|$, 条件属性集 $A \subseteq C$ 是单特定类 D_i 的广义决策协调集当且仅当 $\forall (x, y) \in \Pi_i$ 时, 有 $A \cap M_i(x, y) \neq \emptyset$ 。

证明:

充分性: 设条件属性集 A 是单特定类 D_i 的广义决策协调集, 对于 $\forall (x, y) \in \Pi_i$, 若 $D_i \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$, 则有 $(x, y) \notin SIM(A)$, 所以一定存在 $a \in A$ 使得 $(x, y) \notin SIM(\{a\})$, 故 $a \in M_i(x, y)$, 所以 $A \cap M_i(x, y) \neq \emptyset$ 。

必要性: 设 $\exists (x, y) \in \Pi_i$, 使得 $A \cap M_i(x, y) = \emptyset$, 则有 $(x, y) \in SIM(A)$ 成立, 又因为对 $\forall x \in U$, 当 $D_i \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$ 时, 必有 $(x, y) \notin SIM(A)$, 这与 $(x, y) \in SIM(A)$ 矛盾。证毕。

定义 13 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 对 $\forall D_i \in U/\{d\}, i = 1, 2, \dots, |U/\{d\}|$, M_i 为 IDS 单特定类 D_i 的广义决策约简的差别矩阵, 若 $M_i(x, y) = \{a_1, a_2, \dots, a_k\} \neq \emptyset$, 其中 k 表示 $M_i(x, y)$ 中属性的数量, 用 $\vee M_i(x, y)$ 表示布尔函数 $a_1 \vee a_2 \vee \dots \vee a_k$, 对 $\forall M_i(x, y) \neq \emptyset$, IDS 单特定类 D_i 的广义决策约简的区分函数为

$$DF(M_i) = \wedge(\vee M_i(x, y)) \quad (10)$$

定理 3 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 对 $\forall D_i \in U/\{d\}, i = 1, 2, \dots, |U/\{d\}|$, IDS 单特定类 D_i 的广义决策约简区分函数 $DF(M_i) = \wedge(\vee M_i(x, y))$ 的极小析取范式为 $DF'(M_i) = \bigvee_{k=1}^t (\bigwedge_{s=1}^{q_k} a_s)$, t 表示 $DF'(M_i)$ 中的析取项数目, q_k 表示 $DF'(M_i)$ 中第 k 个析取项中属性数目。记 $A_k = \{a_s | s = 1, 2, \dots, q_k\}$, 则 $\{A_k | k = 1, 2, \dots, t\}$ 是单特定类 $D_i \in U/\{d\}$ 的所有广义决策约简形成的集合。

证明:

对 $\forall k \leq t, \forall (x, y) \in \Pi_i$, 由极小析取范式定义可知 $A_k \cap M_i(x, y) \neq \emptyset$, 由定理 2 可知 A_k 是广义决策协调集。

$DF'(M_i) = \bigvee_{k=1}^t (A_k)$, 若在 A_k 中删除一个元素形成 A'_k , 则 $\exists (x, y) \in \Pi_i$ 使得 $A'_k \cap M_i(x, y) = \emptyset$, 故 A'_k 不是广义决策约简, 因此 A_k 是广义决策约简。因为区分函数中包含所有 $M_i(x, y)$, 故不存在其余广义决策约简。证毕。

2.2 基于差别矩阵的多特定类的广义决策约简

根据以上讨论, 将单特定类的不完备决策系统广义决策约简扩展得到多特定类的不完备决策系统广义决策约简。给出多特定类的广义决策约简相关定义如下。

定义 14 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 则多特定类 D_{mcs} 定义为

$$D_{mcs} \subseteq U/\{d\}, D_{mcs} \neq \emptyset \quad (11)$$

由多特定类的定义可知, 多特定类 D_{mcs} 至少含有一个单特定类至多含有 $|U/\{d\}|$ 个互不相同的单特定类。

定义 15 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 对 $\forall D_{mcs} \subseteq U/\{d\}, \forall A \subseteq C, \forall x \in \{z \in U | \cup D_{mcs} \cap S_C(z) \neq \emptyset\}$, 若 $\partial_A(x) = \partial_C(x)$, 则称条件属性集 A 为多特定类 D_{mcs} 的广义决策协调集。

定义 16 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为

决策类集合, 对 $\forall D_{mcs} \subseteq U/\{d\}$, 条件属性集 $A \subseteq C$ 为多特定类 D_{mcs} 的广义决策约简当且仅当满足以下两个条件:

1) 对 $\forall x \in \{z \in U | \cup D_{mcs} \cap S_C(z) \neq \emptyset\}$, 都有 $\partial_A(x) = \partial_C(x)$;

2) $\forall A' \subset A, \exists x' \in \{z \in U | \cup D_{mcs} \cap S_C(z) \neq \emptyset\}$, 使得 $\partial_{A'}(x') \neq \partial_C(x')$ 。

由定义 16 可知, 当 $|D_{mcs}| = 1$ 时, IDS 多特定类的广义决策约简将退化为单特定类的广义决策约简; 当 $|D_{mcs}| = |U/\{d\}|$ 时, IDS 多特定类的广义决策约简将退化为全决策类的广义决策约简。

定理 4 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 条件属性集 $A \subseteq C$ 为多特定类 $D_{mcs} \subseteq U/\{d\}$ 的广义决策协调集当且仅当 $\cup D_{mcs} \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$ 时, 有 $(x, y) \notin SIM(A)$ 。

证明:

充分性: 因属性集 A 是 IDS 多特定类 D_{mcs} 的广义决策协调集, 对 $\forall x \in \{z \in U | \cup D_{mcs} \cap S_C(z) \neq \emptyset\}$ 有 $\partial_A(x) = \partial_C(x)$ 。 $\cup D_{mcs} \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$ 时, 必有 $f(y, d) \notin \partial_A(x)$ 成立, 所以 $(x, y) \notin SIM(A)$ 。

必要性: 当 $\cup D_{mcs} \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$ 时, 有 $(x, y) \notin SIM(A)$, 所以 $f(y, d) \notin \partial_A(x)$ 成立。因为当 $\cup D_{mcs} \cap S_C(x) \neq \emptyset$ 且 $f(y, d) \notin \partial_C(x)$ 时, 必有 $f(y, d) \notin \partial_A(x)$ 成立, 即 $\partial_C(x) \supseteq \partial_A(x)$ 。又因为当 $A \subseteq C$ 时, 必有 $\partial_A(x) \supseteq \partial_C(x)$ 成立, 所以对 $\forall x \in \{z \in U | \cup D_{mcs} \cap S_C(z) \neq \emptyset\}$, 有 $\partial_A(x) = \partial_C(x)$ 。因此条件属性集 $A \subseteq C$ 为多特定类 $D_{mcs} \subseteq U/\{d\}$ 的广义决策协调集。证毕。

定义 17 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, $U = \{x_1, x_2, \dots, x_n\}, \forall x, y \in U, IDS$ 多特定类 $D_{mcs} \subseteq U/\{d\}$ 广义决策约简的差别矩阵为 $n \times n$ 的矩阵, 记为 $M_{mcs} = (M_{mcs}(x, y))_{n \times n}$, 其中矩阵元素 $M_{mcs}(x, y)$ 为

$$M_{mcs}(x, y) = \begin{cases} \{a \in C | f(x, a) \neq f(y, a) \wedge f(x, a) \neq * \wedge f(y, a) \neq *\} & (x, y) \in \Pi_{mcs} \\ \emptyset, & \text{其他} \end{cases} \quad (12)$$

其中, $\Pi_{mcs} = \{(x, y) | x, y \in U, \cup D_{mcs} \cap S_C(x) \neq \emptyset \wedge f(y, d) \notin \partial_C(x)\}$ 。

定理 5 四元组 $IDS = (U, AT = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 对多特定类 $D_{mcs} \subseteq U/\{d\}$, 条件属性集 $A \subseteq C$ 是多特定类 D_{mcs} 的广义决策协调集当且仅当 $\forall (x, y) \in \Pi_{mcs}$ 时, 有 $A \cap M_{mcs}(x, y) \neq \emptyset$ 。

证明:

充分性: 设条件属性集 A 是多特定类 D_{mcs} 的广义决策协调集, 对 $\forall(x,y) \in \Pi_{mcs}$, 当 $\cup D_{mcs} \cap S_C(x) \neq \emptyset$ 且 $f(y,d) \notin \partial_C(x)$ 时, 则有 $(x,y) \notin \text{SIM}(A)$, 所以一定存在 $a \in A$ 使得 $(x,y) \notin \text{SIM}(\{a\})$, 故 $a \in M_{mcs}(x,y)$, 所以 $A \cap M_{mcs}(x,y) \neq \emptyset$ 。

必要性: 设 $\exists(x,y) \in \Pi_{mcs}$, 使 $A \cap M_{mcs}(x,y) = \emptyset$ 成立, 则有 $(x,y) \in \text{SIM}(A)$, 又因对 $\forall x \in U$, 当 $f(y,d) \notin \partial_C(x)$ 且 $\cup D_{mcs} \cap S_C(x) \neq \emptyset$ 时, 必有 $(x,y) \notin \text{SIM}(A)$, 这与 $(x,y) \in \text{SIM}(A)$ 矛盾。证毕。

定义 18 四元组 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $D_{mcs} \subseteq U/\{d\}$ 为多特定类, M_{mcs} 为 IDS 多特定类 D_{mcs} 的广义决策约简的差别矩阵, 若 $M_{mcs}(x,y) = \{a_1, a_2, \dots, a_k\} \neq \emptyset$, 其中 k 表示 $M_{mcs}(x,y)$ 中属性的数量, 用 $\vee M_{mcs}(x,y)$ 表示布尔函数 $a_1 \vee a_2 \vee \dots \vee a_k$, 对 $\forall M_{mcs}(x,y) \neq \emptyset$, IDS 多特定类 D_{mcs} 的广义决策约简的区分函数为

$$\text{DF}(M_{mcs}) = \wedge(\vee M_{mcs}(x,y)) \quad (13)$$

定理 6 四元组 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$ 为一个不完备决策系统, $U/\{d\} = \{D_1, D_2, \dots, D_{|U/\{d\}|}\}$ 为决策类集合, 多特定类 $D_{mcs} \subseteq U/\{d\}$ 的广义决策约简区分函数 $\text{DF}(M_{mcs}) = \wedge(\vee M_{mcs}(x,y))$ 的极小析取范式为 $\text{DF}'(M_{mcs}) = \bigvee_{k=1}^t (\bigwedge_{s=1}^{q_k} a_s)$, t 表示 $\text{DF}'(M_{mcs})$ 中的析取项数目, q_k 表示 $\text{DF}'(M_{mcs})$ 中第 k 个析取项中属性数目。设 $A_k = \{a_s | s = 1, 2, \dots, q_k\}$, $\{A_k | k = 1, 2, \dots, t\}$ 是多特定类 $D_{mcs} \subseteq U/\{d\}$ 的所有广义决策约简形成的集合。

证明:

对 $\forall k \leq t, \forall(x,y) \in \Pi_{mcs}$, 由极小析取范式定义可知 $A_k \cap M_{mcs}(x,y) \neq \emptyset$, 由定理 5 可知 A_k 是 IDS 多特定类的广义决策协调集。

$\text{DF}'(M_{mcs}) = \bigvee_{k=1}^t (A_k)$, 若在 A_k 中删除一个元素得 $A'_k, \exists(x,y) \in \Pi_{mcs}$ 使 $A'_k \cap M_{mcs}(x,y) = \emptyset$, 故 A'_k 不是广义决策约简, 因此 A_k 是广义决策约简。因为区分函数中包含所有 $M_{mcs}(x,y)$, 故不存在其余广义决策约简。证毕。

根据不完备决策系统单特定类以及多特定类广义决策约简的相关定义, 可依据相应的差别矩阵以及区分函数进行广义决策约简的求解, 现给出实例如下。

例 3 不完备决策系统 IDS 如表 1 所示, 论域 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $C = \{a_1, a_2, a_3, a_4\}$ 为条件属性集, 决策属性集为 $\{d\}$ 。

由例 1 易得, IDS 在属性集 C 上的相容类集合 $U/\text{SIM}(C) = \{\{x_1, x_3\}, \{x_2, x_5\}, \{x_1, x_3\}, \{x_4\}, \{x_2, x_5, x_6\}, \{x_5, x_6\}\}$ 。决策属性 d 对论域 U 划分决策类为

$U/\{d\} = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}\}$ 。广义决策值为 $\partial_C(U) = \{\partial_C(x_1), \partial_C(x_2), \partial_C(x_3), \partial_C(x_4), \partial_C(x_5), \partial_C(x_6)\}$, 其中 $\partial_C(x_1) = \{1, 2\}, \partial_C(x_2) = \{1, 3\}, \partial_C(x_3) = \{1, 2\}, \partial_C(x_4) = \{2\}, \partial_C(x_5) = \{1, 3\}, \partial_C(x_6) = \{3\}$ 。

对于单特定类 $D_1 = \{x_1, x_2\}$, 依据定义 12 构造 IDS 单特定类 D_1 的差别矩阵为

$$M_1 = \begin{bmatrix} \emptyset & \emptyset & \emptyset & \emptyset & \{a_3\} & \{a_3\} \\ \emptyset & \emptyset & \{a_3\} & \{a_1, a_3, a_4\} & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \{a_3\} & \{a_1, a_3\} \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \{a_3\} & \{a_2, a_3, a_4\} & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \end{bmatrix}$$

根据 IDS 单特定类 D_1 的广义决策约简的差别矩阵 M_1 可得对应的区分函数为

$$\text{DF}(M_1) = (a_3) \wedge (a_1 \vee a_3 \vee a_4) \wedge (a_1 \vee a_3) \wedge (a_2 \vee a_3 \vee a_4) = a_3$$

因此 IDS 的单特定类 D_1 的广义决策约简为 $\{a_3\}$ 。

对多特定类 $D_{mcs} = \{D_1, D_2\} = \{\{x_1, x_2\}, \{x_3, x_4\}\}$, 依据定义 17 构造 IDS 多特定类 D_{mcs} 的差别矩阵为

$$M_{mcs} = \begin{bmatrix} \emptyset & \emptyset & \emptyset & \emptyset & \{a_3\} & \{a_3\} \\ \emptyset & \emptyset & \{a_3\} & \{a_1, a_3, a_4\} & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \{a_3\} & \{a_1, a_3\} \\ \{a_4\} & \{a_1, a_3, a_4\} & \emptyset & \emptyset & \{a_2, a_3, a_4\} & \{a_2, a_3\} \\ \emptyset & \emptyset & \{a_3\} & \{a_2, a_3, a_4\} & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \end{bmatrix}$$

根据 IDS 多特定类 D_{mcs} 的广义决策约简的差别矩阵 M_{mcs} 可得对应的区分函数为

$$\text{DF}(M_{mcs}) = (a_3) \wedge (a_1 \vee a_3 \vee a_4) \wedge (a_1 \vee a_3) \wedge (a_4) \wedge (a_2 \vee a_3 \vee a_4) \wedge (a_2 \vee a_3) = a_3 \wedge a_4$$

因此 IDS 的多特定类 D_{mcs} 的广义决策约简为 $\{a_3, a_4\}$ 。

因单特定类广义决策约简为多特定类广义约简的特例, 所以仅给出多特定类的不完备决策系统广义决策约简算法如下。

算法 1 基于差别矩阵的多特定类不完备决策系统广义决策约简 (multi-class-specific generalized decision preservation reduction based on discernibility matrix in incomplete decision systems, MG-DRDM)

输入 一个不完备决策系统 $\text{IDS} = (U, \text{AT} = C \cup \{d\}, V, f)$, 多特定类 D_{mcs} 。

输出 多特定类 D_{mcs} 的不完备决策系统广义决策约简。

1) 依据条件属性集 C 计算论域 U 中每个对象的相容类 $S_C(x_i) (x_i \in U)$, 根据决策属性 d 计算决策类集合 $U/\{d\}$;

2) 依据相容类 $S_C(x_i)$ 计算每个对象的广义

决策值 $\delta_C(x_i)$;

3) 依据决策系统中每个对象的广义决策值 $\delta_C(x_i)$ 和多特定类 D_{mcs} 构造相应差别矩阵 M_{mcs} ;

4) 依据差别矩阵 M_{mcs} 构造相应多特定类的广义决策约简区分函数 $DF(M_{mcs})$;

5) 利用吸收律和分配律将多特定类区分函数 $DF(M_{mcs})$ 转变为极小析取范式 $DF'(M_{mcs})$;

6) 依据极小析取范式 $DF'(M_{mcs})$ 输出多特定类 D_{mcs} 的广义决策约简, 算法结束。

算法 1 中, 步骤 1) 的时间复杂度为 $O(|C||U|^2)$, 步骤 2) 的时间复杂度为 $O(|U|^2)$, 步骤 3)、步骤 4) 的时间复杂度为 $O(|C||U|^2)$, 步骤 5) 的过程是一个 NP-hard 问题, 该步骤的时间复杂度为 $O(|C|^{|U|})$, 因此算法 1 的整体时间复杂度为 $O(|C|^{|U|})$ 。

3 实验结果与分析

针对本文提出的多特定类不完备决策系统广义决策约简算法进行实验验证与分析。本节实验选择与经典不完备决策系统广义决策约简针对约简结果、占用空间以及约简效率进行比较。

本节采用 6 组标准 UCI 数据集进行实验, 数据集从 <http://archive.ics.uci.edu/ml/datasets.php> 处下载。数据集使用 WEKA3.6 进行等频率离散化预处理, 因某些原始数据中存在缺失数据数量极少, 无法直接作为实验使用, 所以针对缺失数据使用相应属性下出现频率最高属性值作为替换, 针对名词性数据使用整数作为替换。数据集详细信息如表 2 所示。表 2 中, ID 表示数据集编号, “数据集”表示数据集名称, $|U|$ 表示数据集中对象数目, $|C|$ 表示数据集中条件属性数目, $|U/\{d\}|$ 表示数据集中决策类数目。本节实验环境: 操作系统 Windows7-64 bit; CPU Intel Core i5-6500(3.2 GHz); 内存 4 GB; 运行环境 Python 3.6; 开发工具 PyCharm。

表 2 数据集
Table 2 Data Sets

ID	数据集	$ U $	$ C $	$ U/\{d\} $
1	Zoo	101	16	7
2	Audiology	200	69	24
3	Glass	214	10	6
4	House	506	13	4
5	Connectionist	528	10	11
6	Yeast	1 484	8	10

因本节采用 UCI 标准数据集, 预处理之后数据集所表示的决策系统为完备决策系统, 所以需

要将其转为不完备决策系统。采用文献 [16] 中的方法对完备决策系统进行处理, 具体处理方式: 除决策属性之外, 每个条件属性对应列选取 10% 的属性值进行缺失处理, 缺失值使用 * 表示。

3.1 约简结果对比

两种约简算法所得平均约简长度对比如表 3 所示。表 3 中, “全决策类”表示经典全决策类约简算法所得约简结果, “单特定类”表示 MG-DRDM 算法选定一个特定类时所得约简结果, “多特定类”表示 MGDRDM 算法选定多个特定类时所得约简结果, “决策值”表示对应算法选取的一个或多个特定类所表现决策值, “平均约简长度”表示对应算法所得约简的平均约简长度。

由表 3 可知, 相比全部决策类, 当选定特定类数量较少时, 平均约简长度往往会比全决策类所得平均约简长度要短。若多特定类选择为所有决策类, 则算法将退化为全决策类约简, 平均约简长度不会缩短。

表 3 平均约简长度
Table 3 Average reduct length

ID	数据集	全决策类		单特定类		多特定类	
		平均约简长度	决策值	平均约简长度	决策值	平均约简长度	决策值
1	Zoo	10.3	2	7.6	2,4	8.0	
2	Audiology	24.1	4	16.7	1,4,6,11	20.0	
3	Glass	10.0	6	7.0	5,6	8.0	
4	House	13.0	0	10.0	0,2	12.0	
5	Connectionist	9.0	0	7.5	0,2	8.0	
6	Yeast	8.0	5	7.0	5,8	7.0	

3.2 占用空间对比

本节通过比较两种算法生成的差别矩阵中非空属性集占比进而比较两种算法占用空间大小。两种算法生成差别矩阵中非空属性集占比如表 4 所示, 其中“占比 (%)”表示对应不同算法构造差别矩阵中非空属性集所占比例。

由表 4 可知, 当选定一个或者几个特定类, 特定类数量相对全部决策类数量较少时, MG-DRDM 算法所构造差别矩阵中非空属性集占比要小于经典算法构造差别矩阵中的非空属性集占比。因此, 在保证约简目标不变的前提下, MG-DRDM 算法所构造差别矩阵占用空间要小于经典算法构造差别矩阵占用空间。除此之外, 利用差别矩阵中非空属性集构造区分函数以及求取极

小析取范式时, MGDRDM 算法也将占用更小空间。当选定特定类为全部决策类时, MGDRDM 算法将退化为全决策类约简, 占用空间不会减小。

表 4 差别矩阵非空属性集占比

Table 4 The proportion of non-empty attribute sets in the discernibility matrix

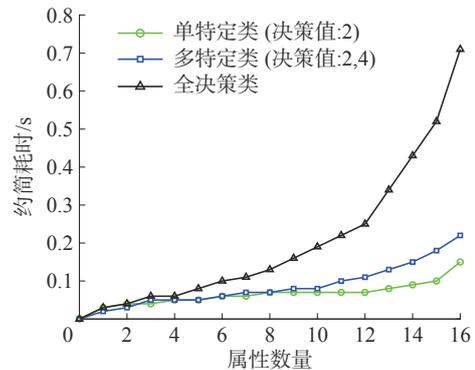
ID	数据集	全决策类		单特定类		多特定类	
		占比/%	决策值	占比/%	决策值	占比/%	决策值
1	Zoo	75.7	2	11.2	2,4	20.1	
2	Audiology	85.0	4	9.6	1,4,6,11	22.9	
3	Glass	71.3	6	5.8	5,6	10.2	
4	House	22.2	0	3.1	0,2	6.7	
5	Connectionist	90.8	0	8.3	0,2	16.6	
6	Yeast	53.5	5	4.3	5,8	7.1	

3.3 约简效率对比

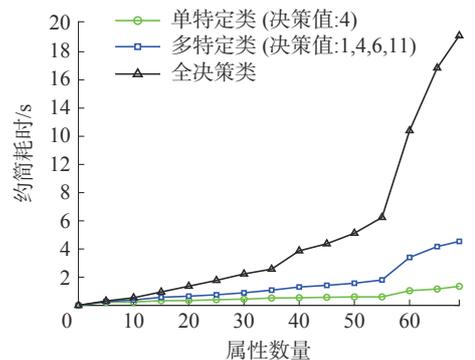
本节将 6 组数据集, 依据随属性数目递增的时间消耗进行经典不完备决策系统广义决策约简与 MGDRDM 算法之间约简效率对比。约简效率如图 1 所示, 其中“全决策类”表示经典全类不完备决策系统广义决策约简算法随属性数目增加的约简耗时变化曲线, “单特定类”和“多特定类”则表示 MGDRDM 算法分别选定一个和多个特定类随属性数目增加的约简耗时变化曲线, “决策值”表示选定特定类所表现的决策值。

由图 1 可知, 选取特定类数量相对全部决策类数量较少时, 约简效率相对对比算法会有明显的提升, 这是由于多特定类广义决策约简的差别矩阵中非空属性集的数目小于全决策类广义决策约简的差别矩阵中非空属性集的数目, 所以 MGDRDM 算法在构造差别矩阵时耗时会有所减少。另外, 因多特定类广义决策约简的差别矩阵中非空属性集数目偏少, 所以构造区分函数以及求取极小析取范式耗时相对较少, 因此约简效率更高。除此之外, 随着属性数目的增加, 算法之间耗时差距越发明显, 这是由于随着属性数目的增加, 差别矩阵也愈加复杂, 使得构造区分函数以及计算极小析取范式时计算量增大, 而且 MGDRDM 算法构造差别矩阵中非空属性集数目较少, 所以耗时增长缓慢, 两个算法之间耗时差距越明显。当多特定类选择为所有决策类时, 算法退化为全类约简, 此时约简效率并无明显提升。

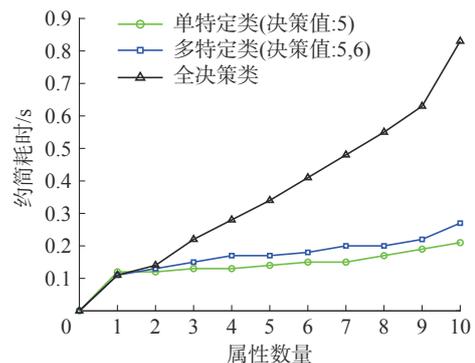
某些情况下, 添加属性求解约简反而会使耗时减少, 这可能是因为添加属性之后构造差别矩阵的过程中产生了较添加属性之前更短的非空属性集, 这使得构造区分函数以及求取极小析取范式的过程耗时减少, 所以在某些情况添加属性反而求解更快。



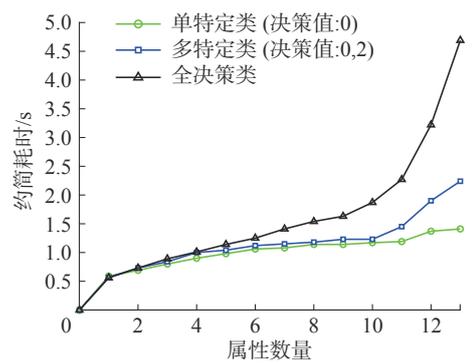
(a) Zoo



(b) Audiology



(c) Glass



(d) House

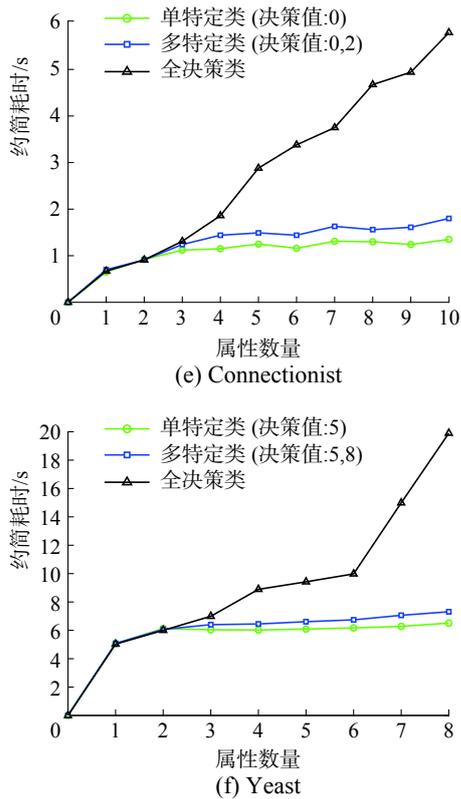


图 1 约简效率

Fig. 1 Reduction efficiency

4 结束语

属性约简作为数据分析预处理的重要手段之一,能够提取更加泛化的规则,压缩数据规模,使得数据分析更加准确、高效,具有重要意义。

本文提出了基于多特定类的不完备决策系统广义决策约简基本理论框架,定义了多特定类的不完备决策系统广义决策约简基本定义,并构造相应差别矩阵以及区分函数,提出基于差别矩阵的约简算法并通过 6 组 UCI 数据集验证了算法的有效性。本文提出约简算法是基于差别矩阵求取所有约简,为提升算法效率,差别矩阵优化是将来研究的重要问题。

参考文献:

- [1] PAWLAK Z. Rough sets[J]. *International journal of computer and information sciences*, 1982, 11(5): 341–356.
- [2] PAWLAK Z. *Rough sets: theoretical aspects of reasoning about data*[M]. Dordrecht: Kluwer Academic Publishers, 1991.
- [3] ZHANG Qinghua, ZHANG Pei, WANG Guoyin. Research on approximation set of rough set based on fuzzy similarity[J]. *Journal of intelligent and fuzzy systems*, 2017, 32(3): 2549–2562.
- [4] 王国胤, 姚一豫, 于洪. 粗糙集理论与应用研究综述 [J]. *计算机学报*, 2009, 32(7): 1229–1246.
- [5] WANG Guoyin, YAO Yiyu, YU Hong. A survey on rough set theory and applications[J]. *Chinese journal of computers*, 2009, 32(7): 1229–1246.
- [6] 于洪, 王国胤, 姚一豫. 决策粗糙集理论研究现状与展望 [J]. *计算机学报*, 2015, 38(8): 1628–1639.
- [7] YU Hong, WANG Guoyin, YAO Yiyu. Current research and future perspectives on decision-theoretic rough sets[J]. *Chinese journal of computers*, 2015, 38(8): 1628–1639.
- [8] 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.
- [9] ZHANG Wenxiu, MI Jusheng, WU Weizhi. Knowledge reductions in inconsistent information systems[J]. *Chinese journal of computers*, 2003, 26(1): 12–18.
- [10] MIAO D Q, ZHAO Y, YAO Y Y, et al. Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model[J]. *Information sciences*, 2009, 179(24): 4140–4150.
- [11] HU Qinghua, ZHANG Lingjun, ZHOU Yucan, et al. Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets[J]. *IEEE transactions on fuzzy systems*, 2018, 26(1): 226–238.
- [12] QIAN Yuhua, LIANG Jiye, PEDRYCZ W, et al. Positive approximation: an accelerator for attribute reduction in rough set theory[J]. *Artificial intelligence*, 2010, 174(9/10): 597–618.
- [13] WU Weizhi, QIAN Yuhua, LI Tongjun, et al. On rule acquisition in incomplete multi-scale decision tables[J]. *Information sciences*, 2017, 378: 282–302.
- [14] YAO Yiyu, ZHANG Xianyong. Class-specific attribute reducts in rough set theory[J]. *Information sciences*, 2017, 418/419: 601–618.
- [15] LI Jingzheng, YANG Xibei, SONG Xiaoning, et al. Neighborhood attribute reduction: a multi-criterion approach[J]. *International journal of machine learning and cybernetics*, 2019, 10(4): 731–742.
- [16] LIANG Jiye, XU Zongben. The algorithm on knowledge reduction in incomplete information systems[J]. *International journal of uncertainty, fuzziness and knowledge-based systems*, 2002, 10(1): 95–103.
- [17] 周献中, 黄兵, 李华雄, 等. 不完备信息系统知识获取的粗糙集理论与方法 [M]. 南京: 南京大学出版社, 2010.
- [18] QIAN Yuhua, LIANG Jiye, LI Deyu, et al. Approximation reduction in inconsistent incomplete decision tables[J]. *Knowledge-based systems*, 2010, 23(5): 427–433.
- [19] SHU Wenhao, QIAN Wenbin. A fast approach to attribute reduction from perspective of attribute measures in in-

- complete decision systems[J]. *Knowledge-based systems*, 2014, 72: 60–71.
- [17] QIAN Wenbin, SHU Wenhao, XIE Yonghong, et al. Feature selection using compact discernibility matrix-based approach in dynamic incomplete decision system[J]. *Journal of information science and engineering*, 2015, 31(2): 509–527.
- [18] XIE Xiaojun, QIN Xiaolin. A novel incremental attribute reduction approach for dynamic incomplete decision systems[J]. *International journal of approximate reasoning*, 2018, 93: 443–462.
- [19] SKOWRON A. Boolean reasoning for decision rules generation[C]//Proceedings of the 7th International Symposium on Methodologies for Intelligent Systems. Trondheim, Norway, 1993: 295–305.
- [20] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. *Information sciences*, 1998, 112(1/2/3/4): 39–49.
- [21] SKOWRON A, RAUSZER C. The discernibility matrices and functions in information systems[M]//SŁOWIŃSKI R. *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*. Dordrecht, Netherlands: Springer, 1992: 331–362.
- [22] 尹继亮, 张楠, 赵立威, 等. 区间值决策系统的局部属性约简[J]. *计算机科学*, 2018, 45(7): 178–185.
YIN Jiliang, ZHANG Nan, ZHAO Liwei, et al. Local attribute reduction in interval-valued decision systems[J]. *Computer science*, 2018, 45(7): 178–185.
- [23] QIAN Yuhua, LIANG Xinyan, WANG Qi, et al. Local rough set: a solution to rough data analysis in big data[J]. *International journal of approximate reasoning*, 2018, 97: 38–63.
- [24] LIU Guilong, HUA Zheng, ZOU Jiyang. Local attribute reductions for decision tables[J]. *Information sciences*, 2018, 422: 204–217.

作者简介:



唐玉凯, 男, 1996 年生, 硕士研究生, 主要研究方向为粗糙集、数据挖掘与机器学习。



张楠, 男, 1979 年生, 讲师, 主要研究方向为粗糙集、认知信息学与人工智能。



童向荣, 男, 1975 年生, 教授, 主要研究方向为多 Agent 系统、分布式人工智能与数据挖掘技术。主持国家自然科学基金面上项目 2 项, 发表学术论文 50 余篇。