

不协调决策形式背景的属性约简

李仲玲¹ 米据生^{2,3} 解 滨⁴

(河北师范大学汇华学院 石家庄 050091)¹ (河北师范大学数学与信息科学学院 石家庄 050024)²

(河北省计算数学与应用重点实验室 石家庄 050024)³

(数据科学与智能应用福建省高校重点实验室 福建漳州 363000)⁴

摘 要 形式概念分析由德国数学家 Wille 于 1982 年提出,是刻画概念和概念之间层次化结构的数据分析工具,主要用于概念的发现、排序和显示。作为知识发现的有效工具,该理论已被成功地运用到信息检索、数据挖掘、模式识别等领域。而在实际问题中呈现出的形式背景往往具有冗余的属性,使得所生成的概念格结构非常复杂。为了提取更加简洁有效的概念格,需要对形式背景中的属性进行约简。因此,寻找更高效的属性约简方法成为了形式概念分析中的一个重要研究问题。文中把基于粗糙集理论的属性约简思想引入到形式背景中,进一步研究不协调决策形式背景的属性约简问题。一些学者在不协调信息系统中基于等价类提出了分布约简、最大分布约简、分配约简、近似约简,并且讨论了 4 个约简之间的关系。而形式背景是一种特殊的信息系统,文中用粒集代替信息系统中的等价类,提出了 4 种基于包含度的属性约简,即分布约简、最大分布约简、分配约简与上近似和约简,并证明了分布协调集必为最大分布协调集,分布协调集必为分配协调集,分配协调集与上近似和协调集等价等结论。最后以分配约简为例,给出了分配协调集的判定定理,构建了不同粒集之间的可辨识属性集合,得到了求解分配约简的布尔计算方法。

关键词 辨识矩阵,协调集,属性约简,决策形式背景

中图分类号 O29 文献标识码 A DOI 10.11896/jsjx.181102137

Attribute Reduction in Inconsistent Decision Formal Contexts

LI Zhong-ling¹ MI Ju-sheng^{2,3} XIE Bin⁴

(HuiHua College, Hebei Normal University, Shijiazhuang 050091, China)¹

(College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, China)²

(Hebei Key Laboratory of Computational Mathematics and Applications, Shijiazhuang 050024, China)³

(Key Laboratory of Data Science and Intelligence Application, Fujian Province University, Zhangzhou, Fujian 363000, China)⁴

Abstract Formal concept analysis was proposed by Wille R. in 1982. It is a model for the study of formal concepts and conceptual hierarchies. As an effective tool in knowledge discovery, it has been applied in various research areas such as information retrieval, data mining and pattern recognition. In practical applications, there may be a lot of redundant attributes in the formal context. Therefore, it is necessary to study the attribute reduction in formal concept analysis, and finding more concise approaches of attribute reduction is an important aspect in formal contexts. In this paper, inspired by the idea of rough set theory, attribute reduction in inconsistent decision formal contexts was studied. Some scholars proposed four definitions of distribution reduction, maximum distribution reduction, assignment reduction, and approximation reduction based on equivalency class in inconsistent information systems. As a formal context is a special information system, in this paper, substituting the equivalency class by the granular set, four new definitions of distribution reduction, maximum distribution reduction, assignment reduction, and upper approximation reduction based on inclusion degree were proposed. It is proved that the distribution reduction must be the maximum distribution reduction, the distribution reduction must be the assignment reduction, and the assignment reduction is equivalent to the upper approximation reduction. As an example, the judgement theorem for assignment consistent set was proved, and Boolean method for assignment reduction were given.

Keywords Discernibility matrix, Consistent set, Attribute reduction, Decision formal contexts

形式概念分析由 Wille 于 1982 年首先提出,在数据分析和知识获取等方面有着非常重要的应用。形式背景的属性约简是形式概念分析中的一个重要研究课题,因此找到有效的属性约简方法十分必要。目前,许多学者已开始关注形式背景的属性约简研究,并取得了很多成果^[1-15]。Zhang 等^[1]给出了概念格同构意义下属性协调集的判定方法,提出了形式概念的可辨识属性集与区分矩阵的概念,并进一步给出了认知的粒化描述和新的认知模型。Liu 等^[2]针对基于属性的概念格与基于对象的概念格给出了协调集判定定理,并通过概念的区分属性给出了约简的方法,且刻画了概念格约简与信息系统的约简之间的关系。Wu 等^[3]提出粒协调决策形式背景的概念,并通过对象的区分属性矩阵给出了粒协调决策形式背景的约简方法。Shao 等^[4-5]针对模糊形式背景提出了粒协调集与粒约简的概念,给出了粒约简的计算方法。Li 等^[6]提出了概念格外延信息量的概念,给出了信息量计算方法,并提出了一种启发式属性约简计算方法。已有研究中讨论比较多的是无决策或者协调决策形式背景的约简问题,不协调决策形式背景的约简问题讨论得并不多。本文借鉴粗糙集属性约简的思想来研究不协调决策形式背景的约简问题。

粗糙集理论^[16-18]是由波兰数学家 Pawlak 于 1982 年提出的用于数据分析的理论,它是一种刻画不完整性和不确定性知识的数学工具。粗糙集理论能够有效分析和处理不精确、不完备和不一致等信息,从中发现隐含的知识并揭示潜在规律,因此它作为一种具有极大潜力的有效知识获取工具受到了广大研究者的关注。粗糙集理论已被成功应用在机器学习与知识发现、数据挖掘、决策支持和分析、模式识别等领域。

本文基于粗糙集属性约简理论进一步研究不协调决策形式背景的属性约简问题,提出了 4 种新的基于包含度的属性约简,即分布约简、最大分布约简、分配约简与上近似和约简,并讨论了这 4 种约简之间的关系;以分配协调集为例,给出了分配协调集的判定定理,及求解分配约简的辨识矩阵法。

1 基本概念

首先给出一些基本概念和结论,具体可参考文献^[3]。

定义 1 一个形式背景 $K=(U, A, I)$ 由两个集合 U 和 A 以及 U 与 A 之间的关系 I 组成。其中, $U=\{x_1, x_2, x_3, \dots, x_m\}$ 是有限对象集; $A=\{a_1, a_2, a_3, \dots, a_n\}$ 是有限属性集; $(x_i, a_j) \in I$ 或者 $x_i I a_j$ 表示对象 x_i 具有属性 a_j , 这里假设每一个对象至少具有一个属性。

设 (U, A, I) 和 (U, C, J) 是两个形式背景, 则称 (U, A, I, C, J) 是一个决策形式背景, 其中 $U=\{x_1, x_2, x_3, \dots, x_m\}$ 是对象集, $A=\{a_1, a_2, a_3, \dots, a_n\}$ 是条件属性集, $C=\{c_1, c_2, c_3, \dots, c_k\}$ 是决策属性集, 并且 $A \cap C = \emptyset$ 。

用 $*$ 来表示 $(x_i, a_j) \in I$, 用空格来表示 $(x_i, a_j) \notin I$, 这样, 形式背景就可以表示成一个表格的形式。

例 1 表 1 是一个决策形式背景 $S=(U, A, I, C, J)$, 其中 $U=\{x_1, x_2, x_3, x_4\}$, $A=\{a, b, c\}$, $D=\{d, e\}$ 。

表 1 形式背景

Table 1 Formal context

	a	b	c	d	e
x_1	*			*	*
x_2		*		*	
x_3			*		
x_4		*	*	*	

定义 2 设 $K=(U, A, I)$ 是一个形式背景, $X \subseteq U, B \subseteq A$,

定义:

$$X^{*B} = \{a \in B \mid (x, a) \in I, \forall x \in X\}$$

$$B^{\downarrow I} = \{x \in U \mid (x, a) \in I, \forall a \in B\}$$

特别地, 记 $x_i^{*B} = \{x_i\}^{*B}, a^{\downarrow I} = \{a^{\downarrow I}\}$ 。

这样, $x_i^{*B \downarrow I} = \{x \in U \mid (x, a) \in I, \forall a \in x_i^{*B}\}$, 即 $x_i^{*B \downarrow I}$ 是包含 x_i 的一个对象子集, 称 $x_i^{*B \downarrow I}$ 为 x_i 关于集合 B 的粒集。由此, $U_B^* = \{x_1^{*B \downarrow I}, \dots, x_m^{*B \downarrow I}\}$ 构成 U 的一个覆盖。

同理, 在决策形式背景 $S=(U, A, I, C, J)$ 中, $U_C^* = \{x_1^{*C \downarrow J}, x_2^{*C \downarrow J}, \dots, x_m^{*C \downarrow J}\}$ 也构成 U 的一个覆盖。

定义 3 设 $K=(U, A, I)$ 是一个形式背景, $B \subseteq A, X \subseteq U$,

定义:

$$R_B^-(X) = \{x \in U \mid x^{*B \downarrow I} \subseteq X\}$$

$$R_B^+(X) = \{x \in U \mid x^{*B \downarrow I} \cap X \neq \emptyset\}$$

则 $R_B^-(X)$ 和 $R_B^+(X)$ 分别称为 X 关于 B 的粒下近似集和粒上近似集。

定义 4 设 $S=(U, A, I, C, J)$ 是一个决策形式背景, 如果对于任意的 $x \in U$ 都有 $x^{*A \downarrow I} \subseteq x^{*C \downarrow J}$, 则称 S 是协调的决策形式背景, 否则称为不协调的。

例 2 (续例 1) $x_1^{*A \downarrow I} = \{x_1\}, x_4^{*A \downarrow I} = \{x_4\}, x_2^{*A \downarrow I} = \{x_2, x_4\}, x_3^{*A \downarrow I} = \{x_3, x_4\}$ 。 $x_1^{*C \downarrow J} = \{x_1\}, x_2^{*C \downarrow J} = \{x_1, x_2, x_4\}, x_3^{*C \downarrow J} = \{x_1, x_3\}, x_4^{*C \downarrow J} = \{x_1, x_2, x_4\}$ 。

由于 $x_3^{*A \downarrow I} \not\subseteq x_3^{*C \downarrow J}$, 因此 $S=(U, A, I, C, J)$ 是不协调的决策形式背景。

性质 1 设 $K=(U, A, I)$ 是一个形式背景, $B \subseteq A$, 对于任意的 $x \in U$ 都有 $x^{*A \downarrow I} \subseteq x^{*B \downarrow I}$; 同理, 在决策形式背景 $S=(U, A, I, C, J)$ 中, 如果 $B \subseteq C$, 则对于任意的 $x \in U$ 都有 $x^{*C \downarrow J} \subseteq x^{*B \downarrow J}$ 。

2 不协调决策形式背景基于粒集的约简

下面给出不协调决策形式背景基于粒集的分布函数、最大分布函数、分配函数、上近似和函数的定义。

定义 5 设 $S=(U, A, I, C, J)$ 是一个决策形式背景, 对于 $B \subseteq A, \forall x \in U$, 记:

$$U_B^* = \{x_1^{*B \downarrow I}, x_2^{*B \downarrow I}, \dots, x_m^{*B \downarrow I}\}$$

$$U_C^* = \{x_1^{*C \downarrow J}, x_2^{*C \downarrow J}, \dots, x_m^{*C \downarrow J}\}$$

$$\mu_B^*(x) = \left(\frac{|x^{*B \downarrow I} \cap x_1^{*C \downarrow J}|}{|U|}, \frac{|x^{*B \downarrow I} \cap x_2^{*C \downarrow J}|}{|U|}, \dots, \right.$$

$$\left. \frac{|x^{*B \downarrow I} \cap x_m^{*C \downarrow J}|}{|U|} \right)$$

$$\gamma_B^*(x) = \max \left(\frac{|x^{*B \downarrow I} \cap x_1^{*C \downarrow J}|}{|U|}, \frac{|x^{*B \downarrow I} \cap x_2^{*C \downarrow J}|}{|U|}, \dots, \right.$$

$$\left. \frac{|x^{*B \downarrow I} \cap x_m^{*C \downarrow J}|}{|U|} \right)$$

$$\sigma_B^*(x) = \{x_i^{*C \downarrow J} \mid x^{*B \downarrow I} \cap x_i^{*C \downarrow J} \neq \emptyset, x_i \in U\}$$

$$\eta_B^* = \frac{\sum_{i=1}^m |R_B^-(x_i^{*C \Delta J})|}{|U|}$$

称 μ_B^* 为论域 U 上的关于属性子集 B 的分布函数, γ_B^* 为论域 U 上的关于属性子集 B 的最大分布函数, σ_B^* 为论域 U 上的关于属性子集 B 的分配函数, η_B^* 为论域 U 上的关于属性子集 B 的上近似和函数。

定义 6 设 $S=(U, A, I, C, J)$ 是一个决策形式背景, $B \subseteq A$ 。

(1) 若 $\forall x \in U$, 有 $\mu_B^*(x) = \mu_A^*(x)$, 则称 B 是分布协调集; 若 B 是分布协调集, 且 B 的任何真子集不是分布协调集, 则称 B 是分布约简。

(2) 若 $\forall x \in U$, 有 $\gamma_B^*(x) = \gamma_A^*(x)$, 则称 B 是最大分布协调集; 若 B 是最大分布协调集, 且 B 的任何真子集不是最大分布协调集, 则称 B 是最大分布约简。

(3) 若 $\forall x \in U$, 有 $\sigma_B^*(x) = \sigma_A^*(x)$, 则称 B 是分配协调集; 若 B 是分配协调集, 且 B 的任何真子集不是分配协调集, 则称 B 是分配约简。

(4) 若 $\eta_B^* = \eta_A^*$, 则称 B 是上近似和协调集; 若 B 是上近似和协调集, 且 B 的任何真子集不是上近似和协调集, 则称 B 是上近似和约简。

显然, 对于协调决策形式背景, 分布约简、最大分布约简、分配约简与上近似和约简都是一致的。

接下来讨论在不协调的决策形式背景中 4 种约简之间的关系。

性质 2 设 $S=(U, A, I, C, J)$ 是决策形式背景, $B \subseteq A$, 则对于 $\forall x \in U$, 有以下结论成立:

1) $|\mu_A^*(x)| \leq |\mu_B^*(x)|$; 2) $\gamma_A^*(x) \leq \gamma_B^*(x)$; 3) $\sigma_A^*(x) \subseteq \sigma_B^*(x)$ 。

其中, $|\mu_A^*(x)| \leq |\mu_B^*(x)|$ 表示两个向量对应的每个分量之间具有相应的大小关系。

证明: 由性质 1 很容易得证。

引理 1 设 $S=(U, A, I, C, J)$ 是不协调决策形式背景, 则 $B \subseteq A$ 是分配协调集的充分必要条件是: 对于 $\forall x_i, x_j \in U$, 当 $\sigma_A^*(x_j) \not\subseteq \sigma_A^*(x_i)$ 时, 有 $x_j^{*B \Delta I} \not\subseteq x_i^{*B \Delta I}$ 。

证明: (必要性) 假设 $x_j^{*B \Delta I} \subseteq x_i^{*B \Delta I}$, 从而有 $\sigma_B^*(x_j) \subseteq \sigma_B^*(x_i)$, 而 B 是分配协调集, 所以有 $\sigma_A^*(x_j) = \sigma_B^*(x_j)$ 且 $\sigma_A^*(x_i) = \sigma_B^*(x_i)$, 进而有 $\sigma_A^*(x_j) \subseteq \sigma_A^*(x_i)$, 这与 $\sigma_A^*(x_j) \not\subseteq \sigma_A^*(x_i)$ 相矛盾。必要性得证。

(充分性) 要证 B 是分配协调集, 只要证明对于 $\forall x \in U$ 都有 $\sigma_A^*(x) = \sigma_B^*(x)$, 而由性质 2 知 $\sigma_A^*(x) \subseteq \sigma_B^*(x) (\forall x \in U)$ 恒成立, 所以只须证明 $\sigma_B^*(x) \subseteq \sigma_A^*(x) (\forall x \in U)$ 。任取 $x_j^{*C \Delta J} \in \sigma_B^*(x)$, 则 $x_j^{*C \Delta J} \cap x^{*B \Delta I} \neq \emptyset$ 。不妨设 $x_i \in x_j^{*C \Delta J} \cap x^{*B \Delta I}$, 于是有 $x_i \in x_j^{*C \Delta J}$ 且 $x_i \in x^{*B \Delta I}$ 。由 $x_i \in x^{*B \Delta I}$ 得 $x_i^{*B \Delta I} \subseteq x^{*B \Delta I}$, 所以有 $\sigma_A^*(x_i) \subseteq \sigma_A^*(x)$ 。而由 $x_i \in x_j^{*C \Delta J}$ 且 $x_i \in x_i^{*B \Delta I}$ 得 $x_i \in x_i^{*B \Delta I} \cap x_j^{*C \Delta J}$, 因此 $x_j^{*C \Delta J} \in \sigma_A^*(x_i) \subseteq \sigma_A^*(x)$ 。由 $x_j^{*C \Delta J}$ 的任意性得 $\sigma_B^*(x) \subseteq \sigma_A^*(x) (\forall x \in U)$, 因此得出 B 是分配协调集。充分性得证。

定理 1 分布协调集必为最大分布协调集。

证明: 设 B 是分布协调集, 则 $\forall x \in U$, 有 $\mu_B^*(x) = \mu_A^*(x)$, 即 $\forall x_i \in U$, 有 $\frac{|x^{*B \Delta I} \cap x_i^{*C \Delta J}|}{|U|} = \frac{|x^{*A \Delta I} \cap x_i^{*C \Delta J}|}{|U|}$, 从

而有 $\max_{1 \leq i \leq m} \frac{|x^{*B \Delta I} \cap x_i^{*C \Delta J}|}{|U|} = \max_{1 \leq i \leq m} \frac{|x^{*A \Delta I} \cap x_i^{*C \Delta J}|}{|U|}$ 。因此 $\gamma_B^*(x) = \gamma_A^*(x)$, 即 B 是最大分布协调集。

定理 2 分布协调集必为分配协调集。

证明: 设 B 是分布协调集, 则 $\forall x \in U$, 有 $\mu_B^*(x) = \mu_A^*(x)$, 即 $\forall x_i^{*C \Delta J} (1 \leq i \leq m)$ 都有 $|x^{*B \Delta I} \cap x_i^{*C \Delta J}| = |x^{*A \Delta I} \cap x_i^{*C \Delta J}|$ 。

一方面, 若 $x_i^{*C \Delta J} \in \sigma_B^*(x)$, 则 $x^{*B \Delta I} \cap x_i^{*C \Delta J} \neq \emptyset$, 即 $|x^{*B \Delta I} \cap x_i^{*C \Delta J}| \neq 0$, 从而 $|x^{*A \Delta I} \cap x_i^{*C \Delta J}| \neq 0$, 即 $x^{*A \Delta I} \cap x_i^{*C \Delta J} \neq \emptyset$, 所以 $x_i^{*C \Delta J} \in \sigma_A^*(x)$, 即 $\sigma_B^*(x) \subseteq \sigma_A^*(x)$ 。另一方面, 若 $x_i^{*C \Delta J} \in \sigma_A^*(x)$, 则 $x^{*A \Delta I} \cap x_i^{*C \Delta J} \neq \emptyset$, 又因为 $x^{*A \Delta I} \subseteq x^{*B \Delta I}$, 所以 $x^{*B \Delta I} \cap x_i^{*C \Delta J} \neq \emptyset$, 从而 $x_i^{*C \Delta J} \in \sigma_B^*(x)$, 因此 $\sigma_A^*(x) \subseteq \sigma_B^*(x)$ 。

综上所述, $\sigma_A^*(x) = \sigma_B^*(x)$, 因此 B 是分配协调集。

定理 3 B 是分配协调集的充要条件为 B 是上近似和协调集。

证明: 首先证明分配协调集一定是上近似和协调集。

因为 B 是分配协调集, 所以对于 $\forall x \in U$, 都有 $\sigma_A^*(x) = \sigma_B^*(x)$, 即对于 $\forall x \in U$ 有 $\{x_i^{*C \Delta J} | x^{*B \Delta I} \cap x_i^{*C \Delta J} \neq \emptyset, 1 \leq i \leq m\} = \{x_j^{*C \Delta J} | x^{*A \Delta I} \cap x_j^{*C \Delta J} \neq \emptyset, 1 \leq j \leq m\}$ 。而 $R_B^-(x_i^{*C \Delta J}) = \{x \in U | x^{*B \Delta I} \cap x_i^{*C \Delta J} \neq \emptyset\}$, 即对于 $\forall x_i^{*C \Delta J}$, 有 $R_B^-(x_i^{*C \Delta J}) = R_A^-(x_i^{*C \Delta J})$ 。否则, 若存在 $x_i^{*C \Delta J}$ 使得 $R_B^-(x_i^{*C \Delta J}) \neq R_A^-(x_i^{*C \Delta J})$, 则必存在 $x \in U$ 使得 $x^{*A \Delta I} \cap x_i^{*C \Delta J} \neq \emptyset$, 但 $x^{*B \Delta I} \cap x_i^{*C \Delta J} = \emptyset$, 这是不可能的; 或者肯定存在 $x \in U$ 使得 $x^{*A \Delta I} \cap x_i^{*C \Delta J} = \emptyset$, 但 $x^{*B \Delta I} \cap x_i^{*C \Delta J} \neq \emptyset$, 这与 $\sigma_A^*(x) = \sigma_B^*(x)$ 相矛盾, 所以 $R_B^-(x_i^{*C \Delta J}) = R_A^-(x_i^{*C \Delta J})$, 因此 $\eta_B^* = \eta_A^*$, 故 B 是上近似和协调集。

下面证明上近似和协调集一定是分配协调集。

因为 B 是上近似和协调集, 所以 $\eta_B^* = \eta_A^*$, 即 $\frac{\sum_{i=1}^m |R_B^-(x_i^{*C \Delta J})|}{|U|} = \frac{\sum_{i=1}^m |R_A^-(x_i^{*C \Delta J})|}{|U|}$ 。假设 $\sigma_A^*(x) \neq \sigma_B^*(x)$, 则 $\{x_i^{*C \Delta J} | x^{*B \Delta I} \cap x_i^{*C \Delta J} \neq \emptyset, 1 \leq i \leq m\} \neq \{x_j^{*C \Delta J} | x^{*A \Delta I} \cap x_j^{*C \Delta J} \neq \emptyset, 1 \leq j \leq m\}$, 即存在 $x_r^{*C \Delta J}$ 使得 $x^{*B \Delta I} \cap x_r^{*C \Delta J} = \emptyset$, 但 $x^{*A \Delta I} \cap x_r^{*C \Delta J} \neq \emptyset$, 这是不可能的; 或者存在 $x_r^{*C \Delta J}$ 使得 $x^{*A \Delta I} \cap x_r^{*C \Delta J} = \emptyset$, 但 $x^{*B \Delta I} \cap x_r^{*C \Delta J} \neq \emptyset$, 于是有 $|\{x \in U | x^{*A \Delta I} \cap x_r^{*C \Delta J} = \emptyset\}| < |\{x \in U | x^{*B \Delta I} \cap x_r^{*C \Delta J} = \emptyset\}|$, 从而得到 $\eta_A^* < \eta_B^*$, 这与 $\eta_B^* = \eta_A^*$ 矛盾, 所以 $\sigma_A^*(x) = \sigma_B^*(x)$, 即 B 是分配协调集。

3 不协调决策形式背景分配约简的判定

针对分配约简, 下文给出它的可辨识属性集的定义。

定义 7 设 $S=(U, A, I, C, J)$ 是一个不协调的决策形式背景, 定义:

$$D^*(x_i, x_j) = \begin{cases} x_i^A - x_j^A, & \sigma_A^*(x_j) \not\subseteq \sigma_A^*(x_i) \\ \emptyset, & \text{其他} \end{cases}$$

称 $D^*(x_i, x_j)$ 为 x_i 与 x_j 的可辨识属性集。

记 $\Omega = \{D^*(x_i, x_j) | 1 \leq i, j \leq m\}$, 称之为不协调决策形式背景 S 的分配可辨识矩阵。记 $G = \bigwedge \{V a | a \in D^*(x_i, x_j), 1 \leq i, j \leq m\}$, 称之为分配辨识公式。

定理 4 设 $S=(U, A, I, C, J)$ 是一个不协调的决策形式背景, $B \subseteq A$, 则 B 是分配协调集, 当且仅当对于非空的 $D^*(x_i, x_j)$, $B \cap D^*(x_i, x_j) \neq \emptyset$ 。

证明: (必要性) 因为 $D^*(x_i, x_j) \neq \emptyset$, 由 $D^*(x_i, x_j)$ 的定

义知, $\sigma_A^*(x_j) \not\subseteq \sigma_A^*(x_i)$, 又因为 B 是分配协调集, 所以由引理 1 得 $x_j^{*B \triangleleft I} \not\subseteq x_i^{*B \triangleleft I}$. 因此, 可分为以下 3 种情况进行讨论:

1) $x_i^{*B \triangleleft I} \subset x_j^{*B \triangleleft I}$; 2) $x_i^{*B \triangleleft I} \cap x_j^{*B \triangleleft I} = \emptyset$; 3) $x_i^{*B \triangleleft I} \cap x_j^{*B \triangleleft I} \subset x_i^{*B \triangleleft I}$ 且 $x_i^{*B \triangleleft I} \cap x_j^{*B \triangleleft I} \subset x_j^{*B \triangleleft I}$.

情况 1) 因为 $x_i^{*B \triangleleft I} \subset x_j^{*B \triangleleft I}$, 所以至少存在一个 x_m 满足 $x_m \in x_j^{*B \triangleleft I}$, 但 $x_m \notin x_i^{*B \triangleleft I}$, 从而有 $x_j^{*B} \subseteq x_m^{*B}$ 且 $x_i^{*B} \not\subseteq x_m^{*B}$, 于是存在 $a \in B$, 满足 $a \in x_i^{*B}$, 但是 $a \notin x_m^{*B}$, 进一步得到 $a \notin x_j^{*B}$, 即 $a \in x_i^{*B} - x_j^{*B}$, 从而有 $a \in (x_i^{*B} - x_j^{*B}) \cap B$, 即 $B \cap D^*(x_i, x_j) \neq \emptyset$.

情况 2) 因为 $x_i^{*B \triangleleft I} \cap x_j^{*B \triangleleft I} = \emptyset$, 显然存在 $a \in B$, 使得 $a \in x_i^{*B} - x_j^{*B}$. 否则, 若对于 $\forall a \in B$, 都有 $a \notin x_i^{*B} - x_j^{*B}$, 而 $x_i^{*B} - x_j^{*B} \subseteq B$, 于是有 $x_i^{*B} - x_j^{*B} = \emptyset$, 从而有 $x_i^{*B} \subseteq x_j^{*B}$, 于是 $x_j^{*B \triangleleft I} \subseteq x_i^{*B \triangleleft I}$, 因此得到 $x_j^{*B \triangleleft I} \cap x_i^{*B \triangleleft I} = x_j^{*B \triangleleft I} \neq \emptyset$, 这与 $x_i^{*B \triangleleft I} \cap x_j^{*B \triangleleft I} = \emptyset$ 矛盾, 因此 $B \cap D^*(x_i, x_j) \neq \emptyset$.

情况 3) 证明类似于情况 1)。

(充分性) 因为 $D^*(x_i, x_j) \neq \emptyset$, 于是有 $\sigma_A^*(x_j) \not\subseteq \sigma_A^*(x_i)$, 又因为 $B \cap D^*(x_i, x_j) \neq \emptyset$, 于是必存在 $a \in B$ 且 $a \in D^*(x_i, x_j)$, 即 $a \in (x_i^{*A} - x_j^{*A}) \cap B$, 也就是说 $a \in x_i^{*B} - x_j^{*B}$, 即 $a \in x_i^{*B}$ 但 $a \notin x_j^{*B}$, 继而得到 $x_i^{*B} \not\subseteq x_j^{*B}$, 所以 $x_j^{*B \triangleleft I} \not\subseteq x_i^{*B \triangleleft I}$, 因此由引理 1 得 B 是分配协调集。

定理 5 设 $S = (U, A, I, C, J)$ 是一个不协调的决策形式背景, 辨识公式 G 的极小析取范式为 $G = \bigvee_m^p (\bigvee_{n=1}^q a_n)$, 其中 $a_n \in \Omega$, 记 $B_m = \{a_n \mid n = 1, 2, 3, \dots, q\}$, 并且 Φ 是 S 的所有分配约简所构成的集合, 则 $\Theta = \{B_m \mid m = 1, 2, 3, \dots, p\} = \Phi$.

证明: 首先证 $\Phi \subseteq \Theta$. 对于 $\forall F \in \Phi$, 由定理 4 可以得出 $F \cap D^*(x, y) \neq \emptyset (\forall D^*(x, y) \in \Omega)$, 即 $F \in \Theta$, 所以 $\Phi \subseteq \Theta$; 反之, 对于 $\forall D^*(x, y) \in \Omega$, 有 $B \cap D^*(x, y) \neq \emptyset (\forall B \in \Theta)$, 由定理 4 知 B 是分配协调集. 从 B 中任意去掉一个属性, 将此时的集合记为 B_1 , 因此得到存在 $x, y \in U$, 使得 $B_1 \cap D^*(x, y) = \emptyset$, 即 B_1 不是分配协调集, 又由 B_1 的任意性知, B 是分配约简, 也就是说 $B \in \Phi$, 因此 $\Theta \subseteq \Phi$. 综上所述, $\Theta = \Phi$, 即 $\{B_m \mid m = 1, 2, 3, \dots, p\}$ 是所有分配约简所构成的集合。

例 3 (续例 2) 由定义得: $\sigma_A^*(x_1) = \{x_1^{*C \triangleleft J}, x_2^{*C \triangleleft J}, x_3^{*C \triangleleft J}, x_4^{*C \triangleleft J}\}$, $\sigma_A^*(x_2) = \sigma_A^*(x_4) = \{x_2^{*C \triangleleft J}, x_4^{*C \triangleleft J}\}$, $\sigma_A^*(x_3) = \{x_2^{*C \triangleleft J}, x_3^{*C \triangleleft J}, x_4^{*C \triangleleft J}\}$, 所以辨识矩阵如表 2 所列。

表 2 辨识矩阵

Table 2 Discernibility matrix

	x_1	x_2	x_3	x_4
x_1	\emptyset	$\{b\}$	$\{c\}$	$\{b, c\}$
x_2	\emptyset	\emptyset	$\{b\}$	\emptyset
x_3	\emptyset	\emptyset	\emptyset	\emptyset
x_4	\emptyset	\emptyset	$\{b\}$	\emptyset

由定理 5 得, $S = (U, A, I, C, J)$ 的约简为 $\{b, c\}$. 经验证, 对于任意的 $x \in U$, 都有 $\sigma_{\{b, c\}}^*(x) = \sigma_A^*(x)$, 且存在 $x \in U$, 使得 $\sigma_{\{b\}}^*(x) \neq \sigma_A^*(x)$, $\sigma_{\{c\}}^*(x) \neq \sigma_A^*(x)$, 所以根据定义得 $\{b, c\}$ 是 $S = (U, A, I, C, J)$ 的分配约简。

结束语 本文基于粗糙集理论进一步研究了不协调决策形式背景的属性约简问题, 提出了 4 种新的基于包含度的分布约简、最大分布约简、分配约简及上近似和约简的定义, 并讨论了这 4 种约简之间的关系, 给出了分配协调集的判定定理, 及求解分配约简的辨识矩阵法。未来, 我们将进一步研究约简的算法。

参考文献

- [1] ZHANG W X, WEI L, QI J J. Attribute reduction theory and approach to concept lattice[J]. Science in China Ser. F Information Sciences, 2005, 48(6): 713-726.
- [2] LIU M, SHAO M W, ZHANG W X, et al. Reduction method for concept lattices based on rough set theory and its application [J]. Computers and Mathematics with Applications, 2007, 53(9): 1390-1410.
- [3] WU W Z, LEUNG Y, MI J S. Granular computing and knowledge reduction in formal contexts[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(10): 1461-1474.
- [4] SHAO M W, YANG H Z, WU W Z. Knowledge reduction in formal fuzzy contexts[J]. Knowledge-Based Systems, 2015, 73(1): 265-275.
- [5] SHAO M W, LEUNG Y, WU W Z. Rule acquisition and complexity reduction in formal decision contexts[J]. International Journal of Approximate Reasoning, 2014, 55(1): 259-274.
- [6] LI J H, MEI C L, LV Y J. Knowledge reduction in decision formal contexts[J]. Knowledge-Based Systems, 2011, 24(5): 709-715.
- [7] NIE C P, MI J S, ZHENG F C. Covering reduction of extents in concept lattices[J]. Chinese Journal of Engineering Mathematics, 2009, 26(1): 8-16.
- [8] LI J H, LV Y J, LIANG B M. Algorithm for attribute reduction based on information quantity of concept lattice extension[J]. Computer Engineering and Applications, 2009, 45(10): 144-146.
- [9] XU K, ZHANG Q H, XUE Y B, et al. Attribute reduction based on fuzziness of approximation set in multi-granulation spaces [J]. The Journal of China Universities of Posts and Telecommunications, 2016, 23(6): 16-23.
- [10] FAN X D, ZHAO W D, WANG C Z, et al. Attribute reduction based on max-decision neighborhood rough set model [J]. Knowledge-Based Systems, 2018, 151: 16-23.
- [11] QIAN W B, SHU W H. Attribute reduction in incomplete ordered information systems with fuzzy decision[J]. Applied Soft Computing Journal, 2018, 73: 242-253.
- [12] WANG C Z, HUANG Y, SHAO M W, et al. Fuzzy rough set-based attribute reduction using distance measures[J]. Knowledge-Based Systems, 2018, 164: 205-212.
- [13] CHEN J K, MI J S, XIE B, et al. A fast attribute reduction method for large formal decision contexts [J]. International Journal of Approximate Reasoning, 2018, 106: 1-17.
- [14] LI J Y, WANG X, WU W Z, et al. Attribute reduction in inconsistent formal decision contexts based on congruence relations [J]. International Journal of Machine Learning and Cybernetics, 2017, 8(1): 81-94.
- [15] LI M Z, LI L J, MI J S, et al. Rough entropy based algorithm for attribute reduction in concept lattice [J]. Computer Science, 2018, 45(1): 84-89.
- [16] PAWLAK Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356.
- [17] PAWLAK Z. Rough sets; some extensions [J]. Information Sciences, 2007, 177(1): 3-27.
- [18] QIAN Y H, LIANG J Y. Rough set method based on multi-granulations[C] // Proceedings of 5th IEEE International Conference on Cognitive Informatics, Beijing, China, 2006: 297-304.