

基于概念约简的强协调决策形式背景的规则提取

王一帆^{1,3}, 金铭^{1,3}, 张琴^{1,3}, 魏玲^{1,2,3*}

(1.西北大学数学学院, 陕西 西安 710127; 2.闽南师范大学数学与统计学院, 福建 漳州 363000; 3.西北大学概念、认知与智能研究中心, 陕西 西安 710127)

摘要:运用形式背景保持二元关系不变的概念约简的思想,研究强协调决策形式背景的规则提取。依据决策子背景的概念约简,证明概念约简诱导的条件属性集为强协调决策形式背景的属性协调集,对条件属性集进行简化,给出概念约简与其他概念的关系,证明概念约简诱导的条件关系集包含所有简化后的规则前件信息,给出利用概念约简获取强协调决策形式背景的非冗余规则的具体步骤。

关键词:决策形式背景;概念约简;规则提取;规则前件;条件属性集

中图分类号: TP181 **文献标志码:** A

引用格式:王一帆,金铭,张琴,等.基于概念约简的强协调决策形式背景的规则提取[J].山东大学学报(理学版),2025,60(1):83-90,100.

Rule acquisition based on concept reduction in strongly consistent formal decision context

WANG Yifan^{1,3}, JIN Ming^{1,3}, ZHANG Qin^{1,3}, WEI Ling^{1,2,3*}

(1. School of Mathematics, Northwest University, Xi'an 710127, Shaanxi, China; 2. School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, Fujian, China; 3. Institute of Concepts, Cognition and Intelligence, Northwest University, Xi'an 710127, Shaanxi, China)

Abstract: Rule acquisition in strongly consistent formal decision context is studied by using the idea of concept reduction of keeping the binary relation unchanged in the formal context. According to the concept reduction of decision subcontext, the conditional attribute set induced by concept reduction is defined, and it is proved that it is the attribute consistent set of the strongly consistent formal decision context, so as to simplify the conditional attribute set. The relationship between concept reduction and other concepts is given, and then it is proved that the conditional relation set induced by concept reduction contains all the information of the simplified rule antecedent. The concrete steps of obtaining non-redundant rules for the strongly consistent formal decision context by concept reduction are given.

Key words: formal decision context; concept reduction; rule acquisition; rule antecedent; conditional attribute set

0 引言

形式概念分析(formal concept analysis, FCA)理论^[1-2]以形式背景作为基础,对哲学中的概念进行形式化描述,形成概念格,以格结构为基础进行知识发现。李金海等^[3]指出了概念格理论与方法的多个研究方向,并对现有的研究成果进行了梳理、总结与展望。作为数据分析与知识挖掘的有效工具,概念格理论已广泛应用到机器学习^[4]、知识发现^[5]、关联分析^[6]等领域。

收稿日期:2024-05-14;网络出版时间:2024-12-03 13:22:10

基金项目:国家自然科学基金资助项目(12171392);陕西数理基础科学研究资助项目(23JSZ008);西北大学研究生科研创新资助项目(CX2024128)

第一作者:王一帆(2000—),女,硕士研究生,研究方向为形式概念分析、三支决策、粒计算. E-mail: wangyifan762@163.com

*通信作者:魏玲(1972—),女,教授,博士生导师,博士,研究方向为形式概念分析、粗糙集、三支决策、粒计算. E-mail: wl@nwu.edu.cn

决策形式背景^[7-8]的规则提取的研究得到很多学者的关注。Li等^[9]研究了决策形式背景上的规则提取问题,提出了一种从决策形式背景导出非冗余决策规则的算法。钱婷等^[10]将粒计算引入决策形式背景,定义了多粒度决策形式背景,给出了粒化前后多粒度决策形式背景上的规则之间的关系。Wei等^[11]基于三支概念格在决策形式背景上研究了规则提取问题,并将结果与经典概念格上的常见决策规则进行了比较。

约简理论是形式概念分析中重要的研究方向之一。张文修等^[12]首次提出了保持概念格结构不变的属性约简理论,并给出研究属性约简的一般框架;随后,魏玲等^[13]扩展了这一理论到决策形式背景,提出了决策形式背景的强、弱协调性,给出了强、弱协调决策形式背景的属性约简方法。属性约简消除冗余的属性,用部分属性来保持全部属性所具有的某种特性不变,但在一定程度上会损失部分原始信息。从FCA的研究角度出发,受布尔因子分析中因子分解的启发,文献[14-15]中提出保持二元关系不变的概念约简。概念约简可以在不损失形式背景所有原始信息的情况下,减少概念的数量,节省存储空间。随后,众多学者从不同角度对概念约简进行了研究。为了求解所有的概念约简,谢小贤等^[16]通过布尔矩阵运算来研究保持二元关系不变的概念约简。王霞等^[17]定义了概念可辨识矩阵,给出了获取全部概念约简的方法。Zhao等^[18]定义了代表概念矩阵,并给出了一种较为简便的求解所有概念约简的方法。概念约简理论也逐渐被扩展到其他形式背景上。魏玲等^[19]定义了对称形式背景,研究了对称形式背景的概念约简;朱朵朵等^[20]研究了不完备背景中3类概念约简;李炎等^[21]首次将概念约简的思想运用到决策形式背景,定义了前件背景,研究保持规则前件信息概念约简。

文献[21]中保持规则前件为前提,依据弱协调决策形式背景的规则考虑保持规则前件信息不变的概念约简,受其启发,本文以保持二元关系不变的概念约简为基础,研究其对强协调决策形式背景中规则提取的作用和意义,研究了决策子背景的概念约简对强协调决策形式背景的非冗余规则前件的作用,继而探讨强协调决策形式背景基于概念约简的规则提取方法。

1 预备知识

定义 1^[1] 设 (G, M, I) 是形式背景, $G = \{g_1, g_2, \dots, g_p\}$ 为对象集, $g_i (i \leq p)$ 称为对象, $M = \{m_1, m_2, \dots, m_q\}$ 为属性集, $m_j (j \leq q)$ 称为属性, I 为 G 与 M 之间的二元关系, $I \subseteq G \times M$,如果 $(g, m) \in I$,则表示对象 g 具有属性 m ,记为 gIm 。

在形式背景 (G, M, I) 中,对于任意 $X \subseteq G, B \subseteq M$,定义一对诱导算子如下:

$$X^* = \{m \mid m \in M, \forall g \in X, gIm\},$$

$$B^* = \{g \mid g \in G, \forall m \in B, gIm\},$$

式中, X^* 表示集合 X 中所有对象共同具有的属性集合, B^* 表示共同具有集合 B 中所有属性的对象集合。

性质 1^[2] (G, M, I) 为形式背景,对于任意 $X, X_1, X_2 \subseteq G$ 和 $B, B_1, B_2 \subseteq M$,则

- (1) $X_1 \subseteq X_2 \Rightarrow X_2^* \subseteq X_1^*, B_1 \subseteq B_2 \Rightarrow B_2^* \subseteq B_1^*$;
- (2) $X \subseteq X^{**}, B \subseteq B^{**}$;
- (3) $X^* = X^{***}, B^* = B^{***}$;
- (4) $(X_1 \cup X_2)^* = X_1^* \cap X_2^*, (B_1 \cup B_2)^* = B_1^* \cap B_2^*$;
- (5) $(X_1 \cap X_2)^* \supseteq X_1^* \cup X_2^*, (B_1 \cap B_2)^* \supseteq B_1^* \cup B_2^*$ 。

对任意 $X \subseteq G, B \subseteq M$,如果二元组 (X, B) 满足 $X^* = B$ 且 $B^* = X$,则称 (X, B) 是一个形式概念,简称概念,其中 X 称为概念 (X, B) 的外延, B 称为其内涵。

为了简化书写,本文在表达形式概念时,一般集合用其元素的序列表示,例如,属性子集 $\{a, b, c\}$ 简写为 abc ,对象子集 $\{1, 2, 3\}$ 简写为 123 。

用 $L(G, M, I)$ 表示形式背景 (G, M, I) 的全体概念,对于任意的 $(X_1, B_1), (X_2, B_2) \in L(G, M, I)$,定义偏序关系:

$$(X_1, B_1) \leq (X_2, B_2) \Leftrightarrow X_1 \subseteq X_2 (\Leftrightarrow B_1 \supseteq B_2)。$$

进一步地,给出任意 2 个概念的下确界和上确界:

$$(X_1, B_1) \wedge (X_2, B_2) = (X_1 \cap X_2, (B_1 \cup B_2)^{**}),$$

$$(X_1, B_1) \vee (X_2, B_2) = ((X_1 \cup X_2)^{**}, B_1 \cap B_2),$$

$L(G, M, I)$ 形成了完备格,称为 (G, M, I) 的概念格^[2]。

定义 2^[14] 设 (G, M, I) 为形式背景, $L(G, M, I)$ 为其概念格,对于 $\mathcal{F} \subseteq L(G, M, I)$,如果 $I = \bigcup_{(X_i, B_i) \in \mathcal{F}} X_i \times B_i$,称 \mathcal{F} 为保持二元关系不变的概念协调集。进一步地,若对任意 $(X, B) \in \mathcal{F}$, $\mathcal{F}' = \mathcal{F} \setminus \{(X, B)\}$,有 $I = \bigcup_{(X_i, B_i) \in \mathcal{F}'} X_i \times B_i$,称 \mathcal{F} 为保持二元关系不变的概念约简。

为了方便起见,下文统称为概念约简。

定义 3^[7] 设 (G, M, I) 和 (G, T, J) 为形式背景且 $M \cap T = \emptyset$,称五元组 (G, M, I, T, J) 为决策形式背景,其中, M 为条件属性集, T 为决策属性集,称 (G, M, I) 为条件子背景, (G, T, J) 为决策子背景,相应的概念格分别称为条件格和决策格。

定义 4^[13] 设 $L(G, M_1, I_1)$ 和 $L(G, M_2, I_2)$ 是 2 个概念格,如果对任意 $(X, B) \in L(G, M_2, I_2)$,总存在 $(X', B') \in L(G, M_1, I_1)$,使得 $X' = X$,则称 $L(G, M_1, I_1)$ 细于 $L(G, M_2, I_2)$,记作 $L(G, M_1, I_1) \leq L(G, M_2, I_2)$ 。称概念 (X', B') 与概念 (X, B) 等价,记作 $(X', B') \leftrightarrow (X, B)$ 。

定义 5^[13] 设 (G, M, I, T, J) 为决策形式背景,如果 $L(G, M, I) \leq L(G, T, J)$,那么称 (G, M, I, T, J) 是强协调的。如果存在属性集 $D \subseteq M$,使得决策形式背景 (G, D, I_D, T, J) 也是强协调的,则称 D 是 (G, M, I, T, J) 的属性协调集。进一步地,若对于任意 $d \in D$,决策形式背景 $(G, D - \{d\}, I_{D - \{d\}}, T, J)$ 不是强协调的,则称 D 是 (G, M, I, T, J) 的属性约简。

定义 6^[7] 设 (G, M, I, T, J) 为决策形式背景,对于任意 $(X, B) \in L(G, M, I)$, $(Y, C) \in L(G, T, J)$,若 $X \subseteq Y$,且 $X, B, Y, C \neq \emptyset$,则称 $B \rightarrow C$ 是一个规则,其中 B 称为 $B \rightarrow C$ 的前件, C 称为 $B \rightarrow C$ 的结论。

记条件格 $L(G, M, I)$ 到决策格 $L(G, T, J)$ 的所有规则集合为 $\mathcal{R}(M, T)$,即 $\mathcal{R}(M, T) = \{B \rightarrow C \mid X \subseteq Y, (X, B) \in L(G, M, I), (Y, C) \in L(G, T, J)\}$ 。

设 $B \rightarrow C, B' \rightarrow C' \in \mathcal{R}(M, T)$,若 $B \subseteq B'$ 且 $C' \subseteq C$,则称 $B \rightarrow C$ 蕴含 $B' \rightarrow C'$,并称规则 $B' \rightarrow C'$ 是冗余的,记所有非冗余规则构成的集合为 $\mathcal{R}^\circ(M, T)$ 。

由定义 4,6 可知,强协调决策形式背景 (G, M, I, T, J) 的所有非冗余规则集合为 $\mathcal{R}^\circ(M, T) = \{B \rightarrow C \mid X = Y, (X, B) \in L(G, M, I), (Y, C) \in L(G, T, J)\}$ 。

例 1 表 1 为决策形式背景 $S = (G, M, I, T, J)$,其中对象集为 $G = \{1, 2, 3, 4, 5, 6\}$,条件属性集为 $M = \{a, b, c, d, e, f, g\}$,决策属性集为 $T = \{k, l, m, n\}$ 。

表 1 决策形式背景 (G, M, I, T, J)
Table 1 Formal decision context (G, M, I, T, J)

| G | a | b | c | d | e | f | g | k | l | m | n |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 3 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 4 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| 5 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| 6 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |

决策形式背景 S 的条件格和决策格分别如图 1、2 所示,其中,决策格 $L(G, T, J)$ 有 10 个概念,以 $c_i (i = 1, 2, \dots, 10)$ 标记于图 2 中相应概念的左侧。

根据条件格 $L(G, M, I)$ 和决策格 $L(G, T, J)$ 可知, $L(G, T, J)$ 的每个概念在 $L(G, M, I)$ 中都有一个等价概念,故该决策形式背景是强协调的。强协调决策形式背景 S 的所有非冗余规则集 $\mathcal{R}^\circ(M, T)$ 为 $\{cg \rightarrow l, f \rightarrow m, a \rightarrow n, adg \rightarrow kn, df \rightarrow km, d \rightarrow k, abcfg \rightarrow lmn, acf \rightarrow mn\}$ 。

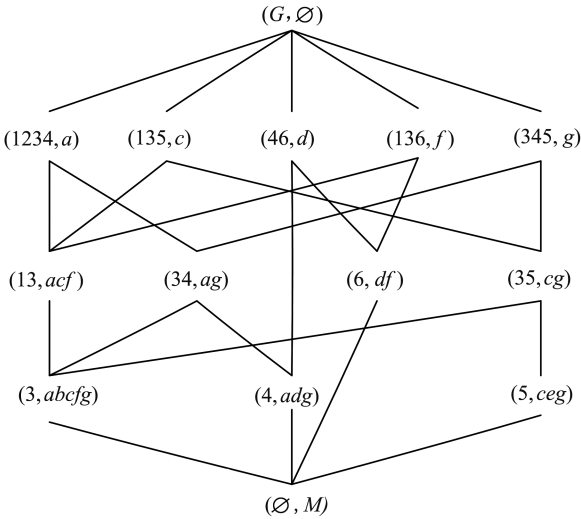


图 1 条件格 $L(G, M, I)$
Fig.1 Condition lattice $L(G, M, I)$

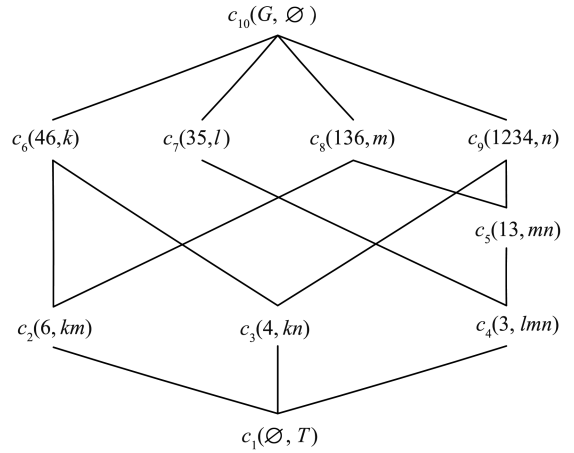


图 2 决策格 $L(G, T, J)$
Fig.2 Decision lattice $L(G, T, J)$

2 基于概念约简的决策形式背景的规则提取

2.1 基于概念约简的条件属性集简化

从决策子背景的概念约简出发,定义概念约简诱导的条件属性集,并探究条件属性集与决策形式背景的属性协调集的关系,简化条件属性集,使得非冗余规则的前件更简洁。

定义 7 设 $S = (G, M, I, T, J)$ 为强协调决策形式背景, \mathcal{F} 是 (G, T, J) 的一个概念约简, $Q = \{(X_i, C_i) \in L(G, M, I) \mid (X_i, B_i) \in \mathcal{F}, i \in \tau\}$, $D = \bigcup_{i \in \tau} C_i \subseteq M$, 称 D 为概念约简 \mathcal{F} 诱导的条件属性集。

引理 1 设 $S = (G, M, I, T, J)$ 为强协调决策形式背景, 则决策格 $L(G, T, J)$ 与条件格 $L(G, M, I)$ 之间存在保并保交的映射 $\varphi: L(G, T, J) \rightarrow L(G, M, I)$ 。

证明 定义映射 $\varphi: L(G, T, J) \rightarrow L(G, M, I)$ 为 $\varphi((X, B)) = (X, C)$, 对于任意 $(X_s, B_s), (X_t, B_t) \in L(G, T, J)$, 有 $\varphi((X_s, B_s) \vee (X_t, B_t)) = \varphi(((X_s \cup X_t)^{**}, B_s \cap B_t))$ 。又因为 $\varphi((X_s, B_s)) \vee \varphi((X_t, B_t)) = (X_s, C_s) \vee (X_t, C_t) = ((X_s \cup X_t)^{**}, C_s \cap C_t)$, 所以 $\varphi((X_s, B_s) \vee (X_t, B_t)) = \varphi((X_s, B_s) \vee (X_t, B_t))$, 因此 φ 是一个保并映射。同理可证, φ 也是一个保交映射。

引理 2 设 $S = (G, M, I, T, J)$ 为强协调决策形式背景, \mathcal{F} 是 (G, T, J) 的一个概念约简, D 为概念约简 \mathcal{F} 诱导的条件属性集, 则对任意 $(X_s, B_s), (X_t, B_t) \in L(G, T, J)$, 存在 $(X_s, C_s), (X_t, C_t) \in L(G, M, I)$, 使得 $(X_s, C_s) \vee (X_t, C_t), (X_s, C_s) \wedge (X_t, C_t) \in L(G, D, I_D)$ 。

证明 S 是强协调决策形式背景, 由引理 1 可知, 存在保并保交的映射 $\varphi: L(G, T, J) \rightarrow L(G, M, I)$, 满足 $\varphi((X, B)) = (X, C)$, 且对于任意 $(X_s, B_s), (X_t, B_t) \in L(G, T, J)$, 有

$$\varphi((X_s, B_s) \vee (X_t, B_t)) = \varphi((X_s, B_s)) \vee \varphi((X_t, B_t)) = (X_s, C_s) \vee (X_t, C_t) = ((X_s \cup X_t)^{**}, C_s \cap C_t)。$$

下证 $(X_s, C_s) \vee (X_t, C_t) \in L(G, D, I_D)$ 。由于 $C_s, C_t \subseteq D$, 因此 $C_s \cap C_t \subseteq D$, 则 $((C_s \cap C_t) \cap D)^* = (C_s \cap C_t)^* = (X_s^* \cap X_t^*)^* = (X_s \cup X_t)^{**}$, 故 $(X_s, C_s) \vee (X_t, C_t) \in L(G, D, I_D)$ 。

由于 $\varphi((X_s, B_s) \wedge (X_t, B_t)) = \varphi((X_s, B_s)) \wedge \varphi((X_t, B_t)) = (X_s, C_s) \wedge (X_t, C_t) = (X_s \cap X_t, (C_s \cup C_t)^{**})$, 因此要证 $(X_s, C_s) \wedge (X_t, C_t) \in L(G, D, I_D)$ 只须要证 $((C_s \cup C_t)^{**} \cap D)^* = X_s \cap X_t$ 。

一方面, 由于 $(C_s \cup C_t) \subseteq (C_s \cup C_t)^{**}$, 因此 $((C_s \cup C_t)^{**} \cap D)^* \subseteq ((C_s \cup C_t) \cap D)^*$, 又因为 $((C_s \cup C_t) \cap D)^* = ((C_s \cap D) \cup (C_t \cap D))^* = (C_s \cap D)^* \cap (C_t \cap D)^* = X_s \cap X_t$, 所以 $((C_s \cup C_t)^{**} \cap D)^* \subseteq X_s \cap X_t$; 另一方面, 因为

$$\begin{aligned} ((C_s \cup C_t)^{**})^* \cup D^* &= (C_s \cup C_t)^* \cup (C_s^* \cap C_t^*) \cup D^* = (X_s \cap X_t) \cup D^*, \\ ((C_s \cup C_t)^{**})^* \cup D^* &\subseteq ((C_s \cup C_t)^{**} \cap D)^*, \end{aligned}$$

所以 $X_s \cap X_t \subseteq (X_s \cap X_t) \cup D^* \subseteq ((C_s \cup C_t) ** \cap D)^*$ 。综上所述, $((C_s \cup C_t) ** \cap D)^* = X_s \cap X_t$, 故 $(X_s, C_s) \wedge (X_t, C_t) \in L(G, D, I_D)$ 。

引理 2 表明, S 的决策格 $L(G, T, J)$ 与由概念约简 \mathcal{F} 诱导的条件属性集 D 构成的概念格 $L(G, D, I_D)$ 的概念存在对应关系, 且概念的上确界和下确界也存在对应关系。

例 2 (续例 1) 决策格 $L(G, T, J)$ 的概念约简为 $\mathcal{F}_1 = \{c_6, c_7, c_8, c_9\}$, $\mathcal{F}_2 = \{c_2, c_5, c_6, c_7, c_9\}$, $\mathcal{F}_3 = \{c_2, c_3, c_7, c_8, c_9\}$, $\mathcal{F}_4 = \{c_2, c_3, c_5, c_7, c_9\}$, 以 \mathcal{F}_3 为例, 找到 $L(G, M, I)$ 中与 \mathcal{F}_3 中概念等价的概念集 $Q = \{(4, adg), (6, df), (35, cg), (136, f), (1234, a)\} \subseteq L(G, M, I)$, 故概念约简 \mathcal{F}_3 诱导的条件属性集 $D = \{a, d, g\} \cup \{d, f\} \cup \{c, g\} \cup \{f\} \cup \{a\} = \{a, c, d, f, g\}$, 进一步地, 得到由 D 构成的背景 (G, D, I_D) , 如表 2 所示, (G, D, I_D) 的概念格 $L(G, D, I_D)$ 如图 3 所示。

表 2 形式背景 (G, D, I_D)
Table 2 Formal context (G, D, I_D)

| G | a | c | d | f | g |
|-----|-----|-----|-----|-----|-----|
| 1 | 1 | 1 | 0 | 1 | 0 |
| 2 | 1 | 0 | 0 | 0 | 0 |
| 3 | 1 | 1 | 0 | 1 | 1 |
| 4 | 1 | 0 | 1 | 0 | 1 |
| 5 | 0 | 1 | 0 | 0 | 1 |
| 6 | 0 | 0 | 1 | 1 | 0 |

在 $L(G, T, J)$ 中, 对于 $c_2 = (6, km)$, $c_3 = (13, mn) \in L(G, T, J)$, 存在 $(6, df), (13, acf) \in L(G, M, I)$, 使得 $(6, df) \vee (13, acf) = (136, f) \in L(G, D, I_D)$ 。对于 $c_8 = (136, m)$, $c_9 = (1234, n) \in L(G, T, J)$, 有 $(136, f), (1234, a) \in L(G, M, I)$, 使得 $(136, f) \wedge (1234, a) = (13, acf) \in L(G, D, I_D)$ 。

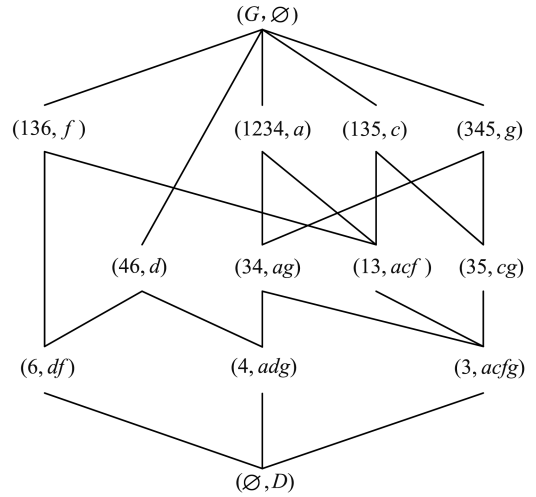


图 3 概念格 $L(G, D, I_D)$
Fig.3 Concept lattice $L(G, D, I_D)$

将引理 2 的结论推广到多个概念中可得到如下推论 1。

推论 1 设 $S = (G, M, I, T, J)$ 为强协调决策形式背景, \mathcal{F} 是 (G, T, J) 的一个概念约简, D 为概念约简 \mathcal{F} 诱导的条件属性集, 则对于任意 $\mathcal{H} = \{(X_j, B_j) \in L(G, T, J) \mid j \in \tau\}$, 存在 $\mathcal{H}' = \{(X_j, C_j) \in L(G, M, I) \mid j \in \tau\}$, 使得 $\bigvee \mathcal{H}', \bigwedge \mathcal{H}' \in L(G, D, I_D)$ 。

设 (G, M, I) 是形式背景, $L(G, M, I)$ 为其概念格, \mathcal{F} 是任一概念约简, 记 $\mathcal{E} = \{\bigwedge \mathcal{H} \mid \mathcal{H} \subseteq \mathcal{F}\}$, 则 \mathcal{E} 在 $L(G, M, I)$ 中并稠密。进一步地, 若 $\mathcal{E} = \{\bigwedge \mathcal{H} \mid \mathcal{H} \subseteq \mathcal{F}\} = \{(X_i, B_i) \mid i \in \tau\}$, 则由并稠密的定义可知, 对于任意 $(X, B) \in L(G, M, I)$, 存在 $\omega \subseteq \tau$, 有 $(X, B) = \bigvee_{i \in \omega} (X_i, B_i)$ 。

例 3 (续例 2) 在 $L(G, T, J)$ 中, 以 $\mathcal{F}_3 = \{c_2, c_3, c_7, c_8, c_9\}$ 为例, $\mathcal{E} = \{\bigwedge \mathcal{H} \mid \mathcal{H} \subseteq \mathcal{F}_3\}$, 对于 $c_4 \in L(G, T, J)$, 存在 $c_7 \wedge c_8 \wedge c_9 \in \mathcal{E}$, 有 $c_4 = c_7 \wedge c_8 \wedge c_9 = (35, l) \wedge (136, m) \wedge (1234, n) = (3, lmn)$; 对于 $c_6 \in L(G, T, J)$, 存在 $c_2, c_3 \in \mathcal{E}$, 使得 $c_6 = c_2 \vee c_3 = (6, km) \vee (4, kn) = (46, k)$ 。

定理 1 设 $S = (G, M, I, T, J)$ 为强协调决策形式背景, \mathcal{F} 是 (G, T, J) 的一个概念约简, D 为概念约简 \mathcal{F} 诱导的条件属性集, 则 $L(G, D, I_D) \leq L(G, T, J)$, 即 D 为 S 的属性协调集。

证明 要证 $L(G, D, I_D) \leq L(G, T, J)$ 等价于证明对任意 $(X, B) \in L(G, T, J)$, 存在 $(X, C) \in L(G, D, I_D)$ 。若 $(X, B) \in \mathcal{F}$, 则存在 $(X, C) \in Q \subseteq L(G, D, I_D)$ 。若 $(X, B) \in L(G, T, J) \setminus \mathcal{F}$, 记 $\mathcal{E} = \{\bigwedge \mathcal{H} \mid \mathcal{H} \subseteq \mathcal{F}\} = \{(X_i, B_i) \mid i \in \tau\}$, 有 $(X, B) = \bigvee_{i \in \omega} (X_i, B_i)$, $\omega \subseteq \tau$, 则由推论 1, 存在 $\eta = \{\bigwedge \mathcal{H}' \mid \mathcal{H}' \subseteq L(G, M, I)\} = \{(X_i, C_i) \mid i \in \tau\} \in L(G, D, I_D)$ 。进一步地, $(X, C) = \bigvee_{i \in \omega} (X_i, C_i) \in L(G, D, I_D)$, $\omega \subseteq \tau$, 故 $L(G, D, I_D) \leq L(G, T, J)$, 即 D 为 S 的属性协调集。

定理 1 表明, 概念约简 \mathcal{F} 诱导的条件属性集 D 是 S 的属性协调集。进一步地, 依据文献 [13] 可得到 S 的属性约简, 表明利用决策子背景的概念约简可以得到决策形式背景的属性协调集, 故通过决策子背景的概

念约简来简化条件属性集是可行的。

例 4 (续例 2) 由图 2、3 可知,对于 $L(G, T, J)$ 的任一概念,在 $L(G, D, I_D)$ 中可以找到等价概念,故 $L(G, D, I_D) \leq L(G, T, J)$, 即 D 为 S 的属性协调集。进一步地,可得到 S 的属性约简。

另外,注意到 $Q = \{(4, adg), (6, df), (35, cg), (136, f), (1234, a)\}$ 中所有概念包含的二元关系不包括 $(1, c) \in (G, D, I_D)$, 所以 Q 不能恢复 $L(G, D, I_D)$ 。借鉴概念约简中局部替代全局的思想,将进一步地讨论通过 Q 能否得到简化后的所有非冗余规则前件。

2.2 基于概念约简的规则前件信息获取

利用概念约简对条件属性集进行简化后,定义概念约简诱导的条件关系集,通过给出概念约简与其他概念的关系来探究其是否包含所有简化后的非冗余规则前件信息。

引理 4 设 (G, M, I) 是形式背景, $L(G, M, I)$ 为其概念格, \mathcal{F} 是任一概念约简,对于任意 $(X, B) \in L(G, M, I)$, 则:

- (1) 若存在 $(X_s, B_s), (X_t, B_t) \in \mathcal{F}$, 使得 $(X, B) = (X_s, B_s) \vee (X_t, B_t)$, 则 $(X, B) = (X_s \cup X_t, B_s \cap B_t)$;
- (2) 若存在 $(X_s, B_s), (X_t, B_t) \in \mathcal{F}$, 使得 $(X, B) = (X_s, B_s) \wedge (X_t, B_t)$, 则 $(X, B) = (X_s \cap X_t, B_s \cup B_t)$ 。

证明 (1) 因为 $(X, B) = (X_s, B_s) \vee (X_t, B_t) = ((X_s \cup X_t)^{**}, B_s \cap B_t)$, 所以要证 $(X, B) = (X_s \cup X_t, B_s \cap B_t)$ 等价于证 $(X_s \cup X_t)^{**} = X_s \cup X_t$ 。一方面,对于任意 $g \in (X_s \cup X_t)^{**} \setminus X_s \cup X_t$, 有 $g \in (X_s \cup X_t)^{**} \setminus X_s$ 且 $g \in (X_s \cup X_t)^{**} \setminus X_t$, 则由概念约简的定义,存在 $(X_p, B_p) \in \mathcal{F}$, 使得 $g \in X_p$, 且 $(X, B) \leq (X_p, B_p)$, 故 $B_p \subseteq B_s \cap B_t$, 则存在 $m \in (B_s \cap B_t) \setminus B_p$, 使得

$$(g, m) \in X \times B \text{ 且 } (g, m) \notin X_s \times B_s, X_t \times B_t, X_p \times B_p;$$

另一方面,

$$g \times (B_s \cap B_t) \subseteq X_p \times (B_s \cap B_t) = (X_p \times B_p) \cup (X_p \times (B_s \cap B_t) \setminus B_p)$$

$$\subseteq (X_p \times B_p) \cup (X_p \times B_s) \cup (X_p \times B_t) = (X_p \times B_p) \cup (X_s \times B_s) \cup (X_t \times B_t),$$

因此 $X \times B \subseteq (X_p \times B_p) \cup (X_s \times B_s) \cup (X_t \times B_t)$, 与存在 $(g, m) \in X \times B$ 且 $(g, m) \notin X_s \times B_s, X_t \times B_t, X_p \times B_p$ 矛盾,故 $(X_s \cup X_t)^{**} = X_s \cup X_t$ 。

(2) 由(1)同理可证。

引理 4 说明,若任一概念可通过概念约简中的 2 个概念求上确界或下确界得到,那么可将计算过程中的 * 运算省略,只须要进行集合的交并运算。

例 5 (续例 3) 对于 $c_5 \in L(G, T, J)$, 存在 $c_8, c_9 \in \mathcal{F}_3$, 使得 $c_5 = c_8 \wedge c_9 = (136, m) \wedge (1234, n) = (13, mn)$; 对于 $c_6 \in L(G, T, J)$, 存在 $c_2, c_3 \in \mathcal{F}_3$, 使得 $c_6 = c_2 \vee c_3 = (6, km) \vee (4, kn) = (46, k)$ 。

将引理 4 推广到概念约简中的多个概念求上确界或下确界,可得推论 2。

推论 2 设 (G, M, I) 是形式背景, $L(G, M, I)$ 为其概念格, \mathcal{F} 是任一概念约简,对于任意 $(X, B) \in L(G, M, I)$,

- (1) 若存在 $\mathcal{N} = \{(X_i, B_i) \mid i \in \tau\} \subseteq \mathcal{F}$, 使得 $(X, B) = \bigvee \mathcal{N}$, 则 $(X, B) = (\bigcup_{i \in \tau} X_i, \bigcap_{i \in \tau} B_i)$;
- (2) 若存在 $\mathcal{N} = \{(X_i, B_i) \mid i \in \tau\} \subseteq \mathcal{F}$, 使得 $(X, B) = \bigwedge \mathcal{N}$, 则 $(X, B) = (\bigcap_{i \in \tau} X_i, \bigcup_{i \in \tau} B_i)$ 。

定理 2 设 (G, M, I) 是形式背景, $L(G, M, I)$ 为其概念格, \mathcal{F} 是任一概念约简,记 $\mathcal{E} = \{\bigwedge \mathcal{N} \mid \mathcal{N} \subseteq \mathcal{F}\} = \{(X_i, B_i) \mid i \in \tau\}$, 则对任意 $(X, B) \in L(G, M, I)$, 存在 $\omega \subseteq \tau$, 使得 $(X, B) = \bigvee_{i \in \omega} (X_i, B_i) = (\bigcup_{i \in \omega} X_i, \bigcap_{i \in \omega} B_i)$ 。

定理 2 表明,已知形式背景的任一概念约简,便可通过集合的交并运算得到其他概念。

定义 8 设 $S = (G, M, I, T, J)$ 为强协调决策形式背景, \mathcal{F} 是 (G, T, J) 的一个概念约简, $Q = \{(X_i, C_i) \in L(G, M, I) \mid (X_i, B_i) \in \mathcal{F}, i \in \tau\}$, $K = \bigcup_{i \in \tau} (X_i \times C_i)$, 称 K 为概念约简 \mathcal{F} 诱导的条件关系集。

定理 3 设 $S = (G, M, I, T, J)$ 为强协调决策形式背景, \mathcal{F} 是 (G, T, J) 的一个概念约简, D 为概念约简 \mathcal{F} 诱导的条件属性集, K 为概念约简 \mathcal{F} 诱导的条件关系集, 则 $L(G, D, K) \leq L(G, T, J)$ 。

证明 若 $(X, B) \in \mathcal{F}$, 则存在 $(X, C) \in Q \subseteq L(G, D, K)$; 若 $(X, B) \in L(G, T, J) \setminus \mathcal{F}$, 记 $\mathcal{E} = \{\bigwedge \mathcal{N} \mid \mathcal{N} \subseteq \mathcal{F}\} =$

$\{(X_i, B_i) \mid i \in \tau\}$, 由定理 2 知, 有 $(X, B) = \bigvee_{i \in \omega} (X_i, B_i) = (\bigcup_{i \in \omega} X_i, \bigcap_{i \in \omega} B_i)$, $\omega \subseteq \tau$ 。由于 $K = \bigcup_{i \in \tau} X_i \times C_i$, 因此 K 是 (G, D, K) 的概念协调集, 记 $\eta = \{\bigwedge \mathcal{H}' \mid \mathcal{H}' \subseteq Q\} = \{(X_i, C_i) \mid i \in \tau\}$, 则有 $(X, C) = (\bigcup_{i \in \omega} X_i, \bigcap_{i \in \omega} C_i)$, 故 $L(G, D, K) \leq L(G, T, J)$ 。

定理 3 表明, 由概念约简 \mathcal{F} 诱导的条件关系集 K 形成的概念格 $L(G, D, K)$ 细于 $L(G, T, J)$, 所以 K 包含所有简化后的非冗余规则的前件信息。又由于 Q 为 (G, D, K) 的概念协调集, 因此结合定理 2 知, 通过集合的交并运算, 便可由 Q 中概念得到所有简化后的非冗余规则的前件。

例 6 (续例 2) 由例 2 可知, $L(G, M, I)$ 中所有与 (G, T, J) 的概念约简 \mathcal{F}_3 中的概念等价的概念为 $Q = \{(4, adg), (6, df), (35, cg), (136, f), (1234, a)\}$, 进一步地, 得到由概念约简 \mathcal{F}_3 诱导的条件关系集 K 形成的形式背景 (G, D, K) , 如表 3 所示, 概念格 $L(G, D, K)$, 如图 4 所示。

表 3 形式背景 (G, D, K)
Table 3 Formal context (G, D, K)

| G | a | c | d | f | g |
|-----|-----|-----|-----|-----|-----|
| 1 | 1 | 0 | 0 | 1 | 0 |
| 2 | 1 | 0 | 0 | 0 | 0 |
| 3 | 1 | 1 | 0 | 1 | 1 |
| 4 | 1 | 0 | 1 | 0 | 1 |
| 5 | 0 | 1 | 0 | 0 | 1 |
| 6 | 0 | 0 | 1 | 1 | 0 |

从图 2、4 中可以看出, $L(G, T, J)$ 的任一概念在 $L(G, D, K)$ 中都可以找到一个等价概念, 即 $L(G, D, K) \leq L(G, T, J)$, 故概念约简 \mathcal{F}_3 诱导的条件关系集 K 包含所有简化后的非冗余规则的前件信息。由于 Q 为 $L(G, D, K)$ 的概念协调集, 因此可由 Q 得到所有简化后的非冗余规则的前件。

2.3 基于概念约简的规则提取方法

对于强协调决策形式背景 $S = (G, M, I, T, J)$, 已知决策子背景 $L(G, T, J)$ 的一个概念约简 \mathcal{F} , 获取 S 的非冗余规则的步骤为:

- (1) 找到条件格 $L(G, M, I)$ 中与 \mathcal{F} 中所有概念等价的概念集 Q , 由 \mathcal{F} 和 Q 得到部分非冗余规则集 $\mathcal{R}^\circ(\mathcal{F}_3, Q)$;
- (2) 计算概念约简 \mathcal{F} 诱导的条件属性集 D 和条件关系集 K , 依据定理 2、3, 利用 \mathcal{F} 和 Q 恢复 $L(G, T, J)$ 和 $L(G, D, K)$;
- (3) 由 $L(G, D, K)$ 和 $L(G, T, J)$ 形成非冗余规则集 $\mathcal{R}^\circ(D, T)$, $\mathcal{R}^\circ(D, T)$ 即为强协调决策形式背景 S 在概念约简 \mathcal{F} 下获取的全部非冗余规则。

例 7 (续例 6) 由 $L(G, T, J)$ 的概念约简 \mathcal{F}_3 和 Q 形成的 S 的部分非冗余规则集 $\mathcal{R}^\circ(\mathcal{F}_3, Q)$ 为 $\{cg \rightarrow l, f \rightarrow m, a \rightarrow n, adg \rightarrow kn, df \rightarrow km\}$ 。

依据定理 2、3, 由 $\mathcal{R}^\circ(\mathcal{F}_3, Q)$ 得到决策形式背景 S 在概念约简 \mathcal{F}_3 下的其他非冗余规则 $d \rightarrow k, acfg \rightarrow lmn, af \rightarrow mn$, 得到非冗余规则集 $\mathcal{R}^\circ(D, T)$ 。表 4 给出了非冗余规则集 $\mathcal{R}^\circ(M, T)$ 和 $\mathcal{R}^\circ(D, T)$, 并将 $\mathcal{R}^\circ(D, T)$ 中不同于 $\mathcal{R}^\circ(M, T)$ 的规则做加粗处理。

由表 4 知, $\mathcal{R}^\circ(M, T)$ 和 $\mathcal{R}^\circ(D, T)$ 中都包含 $\mathcal{R}^\circ(\mathcal{F}_3, Q)$, 本文给出的获取非冗余规则的方式是保持 $\mathcal{R}^\circ(\mathcal{F}_3, Q)$ 不变进行的, 进一步地, 通过集合的交并运算生成所有的非冗余规则 $\mathcal{R}^\circ(D, T)$ 。 $\mathcal{R}^\circ(D, T)$ 中的规则 $acfg \rightarrow lmn$ 和 $af \rightarrow mn$ 相较于

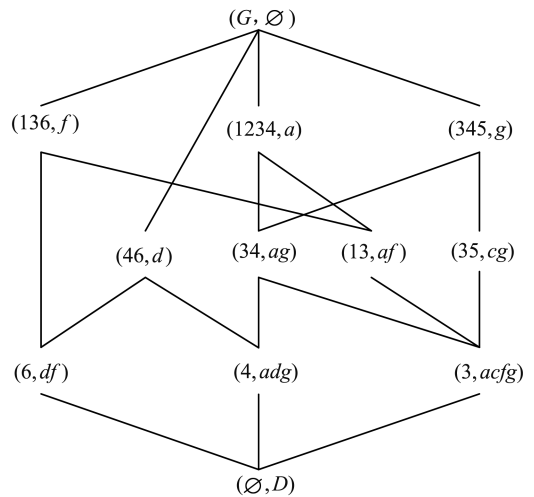


图 4 概念格 $L(G, D, K)$
Fig.4 Concept lattice $L(G, D, K)$

表 4 非冗余规则集 $\mathcal{R}^\circ(M, T)$ 和 $\mathcal{R}^\circ(D, T)$
Table 4 Non-redundant decision rules set $\mathcal{R}^\circ(M, T)$ and $\mathcal{R}^\circ(D, T)$

| $\mathcal{R}^\circ(M, T)$ | $\mathcal{R}^\circ(D, T)$ |
|---------------------------|--|
| $cg \rightarrow l$ | $cg \rightarrow l$ |
| $f \rightarrow m$ | $f \rightarrow m$ |
| $a \rightarrow n$ | $a \rightarrow n$ |
| $adg \rightarrow kn$ | $adg \rightarrow kn$ |
| $df \rightarrow km$ | $df \rightarrow km$ |
| $d \rightarrow k$ | $d \rightarrow k$ |
| $abcfg \rightarrow lmn$ | $acfg \rightarrow lmn$ |
| $acf \rightarrow mn$ | $af \rightarrow mn$ |

$\mathcal{R}^\circ(M, T)$ 中的 $abcf g \rightarrow lmn$ 和 $acf \rightarrow mn$ 有更小的前件,能得到相同的后件,故 $\mathcal{R}^\circ(D, T)$ 中规则的泛化能力更强。

3 结语

本文运用概念约简的思想,考虑强协调决策形式背景的非冗余规则提取问题。基于概念约简诱导的条件属性集是强协调决策形式背景的属性协调集,将条件属性集进行简化。证明了概念约简诱导的条件关系集包含所有简化后的规则前件信息,给出了利用概念约简获取非冗余规则的具体步骤。本文利用概念约简解决强协调决策形式背景上的问题,但是通常情况下,弱协调决策形式背景较多,后续工作考虑弱协调决策形式背景中概念约简对规则提取的作用及意义,此外还可以考虑借助面向属性概念约简和面向对象概念约简来解决决策形式背景的相关问题。

参考文献:

- [1] WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts[M]. Ordered Sets. Dordrecht-Boston: Reidel, 1982:445-470.
- [2] GANTER B, WILLE R. Formal concept analysis: mathematical foundations[M]. Berlin: Springer, 1999.
- [3] 李金海,魏玲,张卓,等. 概念格理论与方法及其研究展望[J]. 模式识别与人工智能, 2020, 33(7):619-642.
LI Jinhai, WEI Ling, ZHANG Zhuo, et al. Concept lattice theory and method and their research prospect[J]. Pattern Recognition and Artificial Intelligence, 2020, 33(7):619-642.
- [4] 胡可云,陆玉昌,石纯一. 概念格及其应用进展[J]. 清华大学学报(自然科学版), 2000, 40(9):77-81.
HU Keyun, LU Yuchang, SHI Chunyi. Advances in concept lattice and its applications[J]. Journal of Tsinghua University (Natural Science Edition), 2000, 40(9):77-81.
- [5] NGUYEN P H P, CORBETT D. A basic mathematical framework for conceptual graphs[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(2):261-271.
- [6] TU Xudong, WANG Yuanliang, ZHANG Maolan, et al. Using formal concept analysis to identify negative correlations in gene expression data[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2016, 13(2):380-391.
- [7] 张文修,仇国芳. 基于粗糙集的不确定决策[M]. 北京:清华大学出版社, 2005.
ZHANG Wenxiu, QIU Guofang. Uncertain decision making based on rough sets[M]. Beijing: Tsinghua University Press, 2005.
- [8] 魏玲. 粗糙集与概念格约简理论与方法[D]. 西安:西安交通大学, 2005.
WEI Ling. Reduction theory and approach to rough set and concept lattice[D]. Xi'an: Xi'an Jiaotong University, 2005.
- [9] LI Jinhai, MEI Changlin, LV Yuejin. Knowledge reduction in decision formal contexts[J]. Knowledge-based Systems, 2011, 24(5):709-715.
- [10] 钱婷,赵思雨,贺晓丽. 基于属性粒度研究决策形式背景的规则提取理论[J]. 山东大学学报(理学版), 2019, 54(10):113-120.
QIAN Ting, ZHAO Siyu, HE Xiaoli. Rules acquisition of decision formal contexts based on attribute granular[J]. Journal of Shandong University(Natural Science), 2019, 54(10):113-120.
- [11] WEI Ling, LIU Lin, QI Jianjun, et al. Rules acquisition of formal decision contexts based on three-way concept lattices[J]. Information Sciences, 2020, 516:529-544.
- [12] 张文修,魏玲,祁建军. 概念格的属性约简理论与方法[J]. 中国科学(E辑:信息科学), 2005, 35(6):628-639.
ZHANG Wenxiu, WEI Ling, QI Jianjun. Attribute reduction theory and approach to concept lattice[J]. Science in China (Series E: Information Sciences), 2005, 35(6):628-639.
- [13] 魏玲,祁建军,张文修. 决策形式背景的概念格属性约简[J]. 中国科学(E辑:信息科学), 2008, 38(2):195-208.
WEI Ling, QI Jianjun, ZHANG Wenxiu. Attribute reduction theory of concept lattice based on decision formal contexts[J]. Science in China (Series F: Information Sciences), 2008, 38(2):195-208.
- [14] 曹丽,魏玲,祁建军. 保持二元关系不变的概念约简[J]. 模式识别与人工智能, 2018, 31(6):516-524.
CAO Li, WEI Ling, QI Jianjun. Concept reduction preserving binary relations[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(6):516-524.