



· 形式概念分析中的概念约简 ·

保持规则前件信息的概念约简

李 炎^{1,2}, 赵思雨^{1,2}, 任睿思^{1,2}, 魏 玲^{1,2,3}

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

摘要 该文将形式背景上保持二元关系不变的概念约简扩展到决策形式背景, 在弱协调决策形式背景上, 研究保持规则前件信息的概念约简问题。首先, 根据决策形式背景获得规则, 仅考虑规则前件集合中的二元关系将条件子背景进行压缩, 构造前件背景; 其次, 针对前件背景提出保持规则前件信息的概念约简, 借助代表概念矩阵给出概念协调集的判定定理; 最后, 根据概念在概念约简中所起的作用, 将所有概念分成 3 类, 并从最小代表概念矩阵的角度讨论 3 类概念的概念特征。

关键词 决策形式背景; 规则; 前件背景; 概念约简; 概念特征

中图分类号: O29; TP18 **DOI**: 10.16152/j.cnki.xdxbr.2023-05-012

Concept reduction preserving antecedent information of rules

LI Yan^{1,2}, ZHAO Siyu^{1,2}, REN Ruisi^{1,2}, WEI Ling^{1,2,3}

(1. School of Mathematics, Northwest University, Xi'an 710127, China;

2. Institute of Concepts, Cognition and Intelligence, Northwest University, Xi'an 710127, China;

3. School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China)

Abstract This paper extends the concept reduction in formal context to formal decision context, and studies the concept reduction preserving antecedent information of rules in weakly consistent formal decision context. Firstly, the conditional subcontext is compressed to form the antecedent context by only considering the binary relation in the antecedent of the rules of formal decision context. Then, we propose the concept reduction preserving antecedent information of rules for the antecedent context, and give the judgment theorem of concept consistent set by representative concept matrix. Finally, we classify all formal concepts into three parts according to the role of each formal concept in concept reduction, and discuss the concept characteristics of three types of formal concepts from the perspective of the minimal representative concept matrix.

Keywords formal decision context; rule; antecedent context; concept reduction; concept characteristics

形式概念分析(formal concept analysis, FCA)是由德国数学家 Wille 于 1982 年首次提出^[1], 此后, 由 Ganter 等多位学者逐步发展完善^[2]。形式概念分析已成为数据分析与处理的有效工具, 被

广泛应用于关联分析^[3]、推荐系统^[4]、知识发现^[5]等领域。

约简理论是形式概念分析中重要的研究方向之一。2005 年张文修等首次提出保持概念格结

收稿日期: 2023-06-20

基金项目: 国家自然科学基金(12171392, 62006190); 西北大学研究生创新项目(CX2023065)

第一作者: 李炎, 女, 从事形式概念分析、粒计算、三支决策理论研究, leeyan1026@163.com。

通信作者: 魏玲, 女, 教授, 博士生导师, 从事形式概念分析、粒计算、三支决策理论等研究, wl@nwu.edu.cn。

构的属性约简理论,并给出了属性约简的一般研究框架^[6]。随后,不少学者在此框架下深入研究了保持其他特性的属性约简^[7-9]。虽然在特定语义下进行的属性约简可以简化形式背景,但在一定程度上会损失部分原始信息。鉴于此,魏玲等于 2018 年提出了概念约简这个新的研究方向,已逐步引起学者们的关注^[10-11]。谢小贤等通过布尔矩阵运算研究了概念特征与概念约简方法^[12];李俊余等研究了三元背景上保持所有三元关系不变的三元概念约简^[13];王霞等通过定义概念可辨识矩阵给出了获取全部概念约简的方法^[14];智慧来等研究了面向对象概念的概念约简问题^[15];Zhao 等通过定义代表概念矩阵,给出了一种较为简单、直观的概念约简求解方法^[16];Zhou 等基于形式概念的矩阵理论,对 3 类概念进行概念特征分析^[17]。

规则获取一直是知识发现研究领域的热点问题,众多学者从不同角度研究了决策形式背景的属性约简问题。魏玲等通过定义强弱协调性,探讨了决策形式背景的属性约简^[18];陈雪等在属性三支协调下,探讨了保持非冗余规则信息不丢失的属性约简^[19];Wei 等从三支概念格的角度给出了决策形式背景的规则提取方法^[20];Chen 等研究了大规模决策形式背景的属性约简问题^[21];李金海等对规则提取的相关研究成果进行了梳理、总结与展望^[22]。

在属性约简理论框架下,决策形式背景中的属性约简研究已相对完善,通过挖掘有意义的决策知识,以满足决策者的不同需求。而在概念约简理论框架下,目前还没有对决策形式背景展开概念约简的相关研究。因此,本文基于文献[10-11]中概念约简的思想,在保持规则前件信息不变的条件下,设法用部分概念反映规则前件的全部信息,继而探讨弱协调决策形式背景中的概念约简问题。

1 预备知识

定义 1^[1] 设 (G, M, I) 是形式背景。其中: $G = \{g_1, g_2, \dots, g_p\}$ 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, $I \subseteq G \times M$ 。若 $(g, m) \in I$, 则表示对象 g 具有属性 m , 记为 gIm 。

在形式背景 (G, M, I) 中, 对任意的 $X \subseteq G, B \subseteq M$, Wille^[1] 定义了一对诱导算子为

$$X^* = \{m \mid m \in M, \forall g \in X, gIm\}, \quad (1)$$

$$B^* = \{g \mid g \in G, \forall m \in B, gIm\}。 \quad (2)$$

式中: X^* 表示集合 X 中所有对象共同具有的属性集合; B^* 表示共同拥有集合 B 中所有属性的对象集合。这对诱导算子的性质见文献[2]。

对于任意 $X \subseteq G, B \subseteq M$, 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。其中 X 称为概念 (X, B) 的外延, B 称为其内涵。

用 $L(G, M, I)$ 表示形式背景 (G, M, I) 的全体概念, 记 $(X_1, B_1) \leq (X_2, B_2) \Leftrightarrow X_1 \subseteq X_2 (\Leftrightarrow B_2 \subseteq B_1)$, 则 $\leq$ 是 $L(G, M, I)$ 上的偏序关系, 从而 $(L(G, M, I), \leq)$ 是偏序集。其中, 下确界和上确界分别为

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

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

时, $L(G, M, I)$ 是完备格, 称为 (G, M, I) 的概念格^[2]。

对于任意两个概念 $(X_1, B_1), (X_2, B_2)$, 若 $(X_1, B_1) \leq (X_2, B_2)$, 则称 (X_1, B_1) 为 (X_2, B_2) 的亚概念, (X_2, B_2) 为 (X_1, B_1) 的超概念。

为了得到由概念格形成的规则, 文献[18, 23] 提出了决策形式背景的概念, 并给出了决策形式背景协调性的定义。

定义 2^[23] 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。

设 $L(G, M_1, I_1)$ 和 $L(G, M_2, I_2)$ 是两个概念格。若存在单射 $f: L(G, M_2, I_2) \rightarrow L(G, M_1, I_1)$, 满足: ① $f((G, \emptyset)) = (G, \emptyset), f((\emptyset, M_2)) = (\emptyset, M_1)$; ② 对任意 $(X, B) \in L(G, M_2, I_2)$, $f((X, B))$ 的外延包含于 X , 则称 f 是 $L(G, M_2, I_2)$ 到 $L(G, M_1, I_1)$ 的蕴含映射。若存在 $L(G, M_2, I_2)$ 到 $L(G, M_1, I_1)$ 的一个蕴含映射, 则称 $L(G, M_1, I_1)$ 弱细于 $L(G, M_2, I_2)$, 记作 $L(G, M_1, I_1) \leq L(G, M_2, I_2)$ 。

定义 3^[18] 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) 是弱协调的, 否则为不协调的。

定义 4^[24] 设 (G, M, I, T, J) 为决策形式背景, 对于任意的 $(X, B) \in L(G, M, I), (Y, C) \in$

$L(G, T, J)$, 若 $X \subseteq Y$, 且 $X, B, Y, C \neq \emptyset$, 则称 $(X, B) \rightarrow (Y, C)$ 是规则, 其中 (X, B) 称为 $(X, B) \rightarrow (Y, C)$ 的前件, (Y, C) 称为 $(X, B) \rightarrow (Y, C)$ 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 $\mathcal{R}(M, T) = \{(X, B) \rightarrow (Y, C) \mid X \subseteq Y, (X, B) \in L(G, M, I), (Y, C) \in L(G, T, J)\}$ 。

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

2 保持规则前件信息的概念约简

2.1 条件子背景的压缩

条件子背景的二元关系 I 存储了对象与属性之间的所有信息, 但其中存在一些对获取规则无关的信息, 为简化数据, 首先对条件子背景进行压缩。压缩时为保留规则前件的完整信息, 考虑定义 4 的所有规则, 给出前件背景的定义。

定义 5 设 (G, M, I, T, J) 为弱协调决策形式背景, $R(G, M, I) = \{(X, B) \mid (X, B) \rightarrow (Y, C), (X, B) \in L(G, M, I)\}$ 为所有规则前件的概念集合。令 $G' = \bigcup \{X \mid (X, B) \in R(G, M, I)\}, M' = \bigcup \{B \mid (X, B) \in R(G, M, I)\}, I_R = \bigcup \{X \times B \mid (X, B) \in R(G, M, I)\}$, 称 (G', M', I_R) 为弱协调决策形式背景 (G, M, I, T, J) 的前件背景, 其概念格记为 $L(G', M', I_R)$ 。

由定义 5 知, 前件背景即由决策形式背景的所有规则前件重新构造的背景, 它是条件子背景的一个子背景, 且 $R(G, M, I) \subseteq L(G', M', I_R)$ 。

性质 1 设 $R(G, M, I)$ 是所有规则前件的概念集合, 对于任意的 $(X, B) \in R(G, M, I)$, 令 $S = \{(X_i, B_i) \mid (X_i, B_i) \leq (X, B), (X_i, B_i) \in L(G, M, I), i \in \tau\}$, 则 $S \subseteq R(G, M, I)$ 。

证明 设 $(X, B) \in R(G, M, I)$, 则存在 $(Y, C) \in L(G, T, J)$, 使得 $X \subseteq Y$ 。对于任意 $(X_i, B_i) \in L(G, M, I), (X_i, B_i) \leq (X, B)$, 有 $X_i \subseteq X$, 故 $X_i \subseteq Y$, 则 $(X_i, B_i) \rightarrow (Y, C)$, 即 $(X_i, B_i) \in R(G, M, I)$, 故 $S \subseteq R(G, M, I)$ 。

性质 1 表明对于任意规则前件 $(X, B), (X, B)$ 的所有亚概念都是规则前件。

性质 2 设 (G', M', I_R) 为弱协调决策形式背景 (G, M, I, T, J) 的前件背景, 对于任意 $g \in G'$, 有 $\{g\} \times g^* \in I_R$ 。

证明 对于任意 $g \in G'$, 由定义 5 知, 存在 $(X, B) \in R(G, M, I)$, 使得 $g \in X$, 且 $(X, B) \in L(G, M, I)$, 显然有 $X = B^*, g \in B^*$, 则 $B = B^{**} \subseteq g^*$ 。又因为 $B = X^*$, 故 $X^* \subseteq g^*$, 则有 $g^{**} \subseteq X^{**} = X$, 因此, $(g^{**}, g^*) \leq (X, B)$ 。由性质 1 知, $(g^{**}, g^*) \in R(G, M, I)$, 由 $R(G, M, I) \subseteq L(G', M', I_R)$, 进而 $(g^{**}, g^*) \in L(G', M', I_R)$, 则 $g^{**} \times g^* \in I_R$, 又 $\{g\} \subseteq g^{**}$, 故 $\{g\} \times g^* \in I_R$ 。

为进一步说明前件背景与条件子背景之间的关系, 给出封闭关系的定义。

定义 6^[2] 如果形式背景 (G, M, J) 的每个概念都是形式背景 (G, M, I) 的概念, 那么, 关系 $J \subseteq I$ 称为形式背景 (G, M, I) 的一个封闭关系。

性质 1 和性质 2 表明, $R(G, M, I)$ 是由一些概念及其所有亚概念构成的集合。利用 $R(G, M, I)$ 构造前件背景, 它会保留条件子背景中对象集为 G' 的整行对象信息。基于此, 前件背景不会生成除条件格外其他的概念, 即前件背景的二元关系是条件子背景的封闭关系。因此, 得出以下性质。

性质 3 设 (G', M', I_R) 为弱协调决策形式背景 (G, M, I, T, J) 的前件背景, 则 I_R 是条件子背景 (G, M, I) 的封闭关系。

例 1^[25] 表 1 为中国港口危险品事故分析的决策形式背景 (G, M, I, T, J) 。 $G = \{1, 2, 3, 4, 5, 6, 7, 8\}$ 表示 8 起港口危险品事故, $M = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$ 表示危险品事故发生因素。其中: a_1 为人员专业性; a_2 为操作规范; a_3 为仓库储存; a_4 为企业经营; a_5 为监督管理; a_6 为设施与设备; a_7 为应急管理; a_8 为货物登记; a_9 为自然因素。 I 为对象和条件属性之间的二元关系。其中: 0 表示对应行事故在对应列因素下合格; 1 表示不合格。 $T = \{d_1, d_2, d_3\}$ 表示事故类型。其中: d_1 为火灾; d_2 为爆炸; d_3 为泄露。 J 为对象和决策属性之间的二元关系。其中: 0 表示未发生该事故; 1 表示发生该事故。

条件格和决策格分别如图 1 和图 2 所示。其中, 条件格 $L(G, M, I)$ 有 18 个概念。为便于讨论, 将每个概念以 $c_i (i = 1, 2, \dots, 18)$ 编号。

表 1 决策形式背景 (G, M, I, T, J)

Tab. 1 Formal decision context (G, M, I, T, J)

G	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	d_1	d_2	d_3
1	1	1	0	0	1	0	0	0	0	1	1	0
2	0	0	0	1	0	0	1	0	0	0	0	1
3	1	1	0	1	0	0	0	0	0	1	1	0
4	1	1	0	1	1	0	1	1	0	1	1	0
5	1	1	0	1	1	1	1	1	0	0	1	1
6	0	0	1	0	1	0	1	1	0	0	0	1
7	1	1	1	1	1	1	1	0	1	1	1	0
8	0	0	0	0	0	0	1	0	0	0	0	1

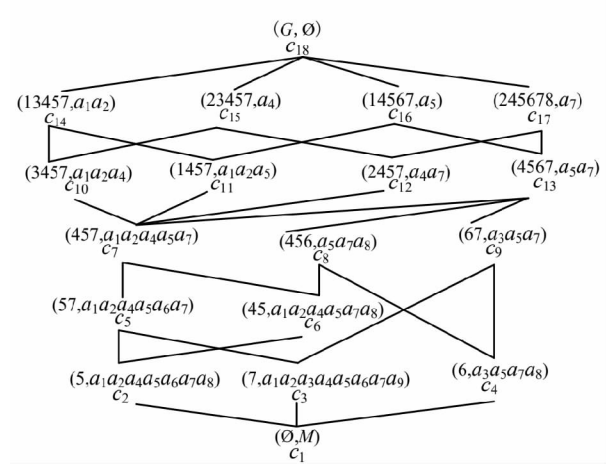


图 1 概念格 $L(G, M, I)$
Fig. 1 Concept lattice $L(G, M, I)$

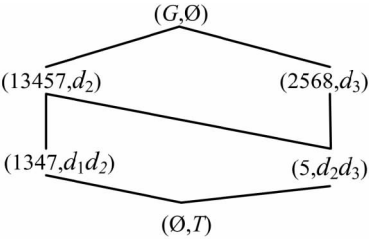


图 2 概念格 $L(G, T, J)$
Fig. 2 Concept lattice $L(G, T, J)$

根据条件格 $L(G, M, I)$ 和决策格 $L(G, T, J)$, 可以构造蕴含映射 $f_1: (G, \emptyset) \rightarrow c_{18}, (13457, d_2) \rightarrow c_{14}, (2568, d_3) \rightarrow c_4, (1347, d_1d_2) \rightarrow c_3, (5, d_2d_3) \rightarrow c_2, (\emptyset, T) \rightarrow c_1$ 。因此, 由定义 3 可知, 决策形式背景 (G, M, I, T, J) 是弱协调的。

由定义 4 可知, 条件子背景与决策子背景之间的所有规则构成的集合为

$$\begin{aligned} \mathcal{R}(M, T) = \{ & (13457, a_1a_2) \rightarrow (13457, d_2), \\ & (3457, a_1a_2a_4) \rightarrow (13457, d_2), \\ & (1457, a_1a_2a_5) \rightarrow (13457, d_2), \\ & (457, a_1a_2a_4a_5a_7) \rightarrow (13457, d_2), \\ & (45, a_1a_2a_4a_5a_7a_8) \rightarrow (13457, d_2), \\ & (57, a_1a_2a_4a_5a_6a_7) \rightarrow (13457, d_2), \\ & (7, a_1a_2a_3a_4a_5a_6a_7a_9) \rightarrow (13457, d_2), \\ & (5, a_1a_2a_4a_5a_6a_7a_8) \rightarrow (13457, d_2), \\ & (6, a_3a_5a_7a_8) \rightarrow (2568, d_3), \\ & (5, a_1a_2a_4a_5a_6a_7a_8) \rightarrow (2568, d_3), \\ & (7, a_1a_2a_3a_4a_5a_6a_7a_9) \rightarrow (1347, d_1d_2), \\ & (5, a_1a_2a_4a_5a_6a_7a_8) \rightarrow (5, d_2d_3) \}。 \end{aligned}$$

根据 $\mathcal{R}(M, T)$ 可得, 所有规则前件的概念集合为 $R(G, M, I) = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_{10}, c_{11}, c_{14}\}$ 。从图 1 可看出, $c_{14} \in R(G, M, I)$, c_{14} 的所有亚概念为 $c_1, c_2, c_3, c_5, c_6, c_7, c_{10}, c_{11}$, 显然这些亚概念都属于 $R(G, M, I)$, 即满足性质 1。

根据定义 5, 利用 $R(G, M, I)$ 构造前件背景 (G', M', I_R) 如表 2 所示, 其概念格 $L(G', M', I_R)$ 如图 3 所示。其中, $L(G', M', I_R)$ 有 15 个概念。为便于讨论, 将每个概念以 $c_i (i = 1, 2, \dots, 15)$ 编号。

表 2 形式背景 (G', M', I_R)
Tab. 2 Formal context (G', M', I_R)

G'	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
1	1	1	0	0	1	0	0	0	0
3	1	1	0	1	0	0	0	0	0
4	1	1	0	1	1	0	1	1	0
5	1	1	0	1	1	1	1	1	0
6	0	0	1	0	1	0	1	1	0
7	1	1	1	1	1	1	1	0	1

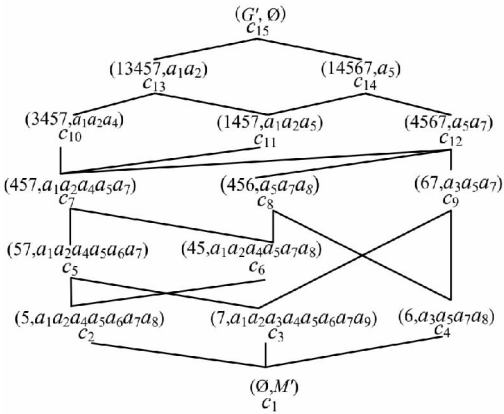


图 3 概念格 $L(G', M', I_R)$
Fig. 3 Concept lattice $L(G', M', I_R)$

对比表1和表2发现, (G, M, I) 中危险品事故1、事故3、事故4、事故5、事故6、事故7的整行信息, 完全保留在前件背景 (G', M', I_R) , 即满足性质2。

从图1和图3可以看出, $L(G', M', I_R)$ 中的任意概念都是 $L(G, M, I)$ 中的概念, 即 I_R 是条件子背景的封闭关系, 则通过前件背景和决策子背景即可获取全部规则。

为便于后续讨论, 由图2和图3找出该决策形式背景的所有非冗余规则集合为

$$\begin{aligned} \mathcal{R}^o(M, T) = & \{ (13457, a_1a_2) \rightarrow (13457, d_2), \\ & (6, a_3a_5a_7a_8) \rightarrow (2568, d_3), \\ & (7, a_1a_2a_3a_4a_5a_6a_7a_9) \rightarrow (1347, d_1d_2), \\ & (5, a_1a_2a_4a_5a_6a_7a_8) \rightarrow (5, d_2d_3) \}。 \end{aligned}$$

依次记上述4条规则分别为 r_1, r_2, r_3, r_4 。规则 r_1 表明如果人员专业性、操作规范不合格时, 则会发生爆炸; 规则 r_2 表明如果仓库存储、监督管理、应急管理、货物登记不合格时, 则会发生泄露; 规则 r_3 表明如果人员专业性、操作规范、仓库存储、企业经营、监督管理、设施与设备、应急管理、自然因素不合格时, 则会发生火灾和爆炸; 规则 r_4 表明如果人员专业性、操作规范、企业经营、监督管理、设施与设备、应急管理、货物登记不合格时, 则会发生爆炸和泄露。

2.2 概念约简的定义及存在性

给定前件背景 (G', M', I_R) , I_R 表示对象集 G' 与属性集 M' 之间的二元关系, 同时 I_R 也存储了决策形式背景所有规则前件的信息, 为保留这些信息并简化概念格, 给出保持规则前件信息概念约简的定义。

定义7 设 (G, M, I, T, J) 为弱协调决策形式背景, (G', M', I_R) 为其前件背景, $L(G', M', I_R)$ 是其概念格。对于 $\mathcal{F} \subseteq L(G', M', I_R)$, 如果 $I_R = \bigcup_{(X_i, B_i) \in \mathcal{F}} X_i \times B_i$, 称 $\mathcal{F}$ 为保持规则前件信息概念协调集, 本文简称概念协调集; 若进一步, 对任意 $(X, B) \in \mathcal{F}, \mathcal{F}' = \mathcal{F} \setminus \{(X, B)\}$, 有 $I_R \neq \bigcup_{(X_i, B_i) \in \mathcal{F}'} X_i \times B_i$, 称 $\mathcal{F}$ 为保持规则前件信息概念约简, 本文简称概念约简。

概念协调集是保持前件背景中所有规则前件信息不丢失的概念格子集, 概念约简是极小概念协调集。

定理1 弱协调决策形式背景的保持规则前件信息概念约简必存在。

证明 设 (G, M, I, T, J) 为弱协调决策形式背景, (G', M', I_R) 为其前件背景, $L(G', M', I_R)$ 是概念格。若对于任意 $(X, B) \in L(G', M', I_R)$, 有 $\bigcup_{(X_i, B_i) \in L(G', M', I_R)} X_i \times B_i \neq \bigcup_{(X_i, B_i) \in L(G', M', I_R) \setminus \{(X, B)\}} X_i \times B_i$, 则 $L(G', M', I_R)$ 本身即为概念约简。若存在 $(X_1, B_1) \in L(G', M', I_R)$, 使得 $\bigcup_{(X_i, B_i) \in L(G', M', I_R)} X_i \times B_i = \bigcup_{(X_i, B_i) \in L(G', M', I_R) \setminus \{(X_1, B_1)\}} X_i \times B_i$, 则考虑 $\mathcal{F} = L(G', M', I_R) \setminus \{(X_1, B_1)\}$, 若对于任意的 $(X_j, B_j) \in \mathcal{F}$, 有 $\bigcup_{(X_i, B_i) \in \mathcal{F} \setminus \{(X_j, B_j)\}} X_i \times B_i \neq \bigcup_{(X_i, B_i) \in \mathcal{F}} X_i \times B_i$, 则 $\mathcal{F}$ 为概念约简。否则, 进一步考虑 $\mathcal{F}_1 = \mathcal{F} \setminus \{(X_2, B_2)\}$ 。重复上述过程, 由于 $L(G', M', I_R)$ 为有限集, 则至少可以找到一个约简。因此, 概念约简必存在。

2.3 概念协调集判定定理

定义8 设 (G, M, I, T, J) 为弱协调决策形式背景, (G', M', I_R) 为其前件背景, $L(G', M', I_R)$ 是其概念格。对于任意 $g \in G', m \in M', (X, B) \in L(G', M', I_R)$ 。若 $(g, m) \in X \times B$, 则称 (X, B) 为 (g, m) 的代表概念。进一步, 称 $\text{REP}((g, m)) = \{(X, B) \in L(G', M', I_R) \mid (g, m) \in X \times B\}$ 为 (g, m) 的代表概念集, $\mathbf{A} = (\text{REP}((g, m)), g \in G', m \in M')_{|G'| \times |M'|}$ 为 (G', M', I_R) 的代表概念矩阵。

代表概念矩阵中, 部分代表概念集之间存在包含关系, 记 $\mathbf{A}_{\min}$ 是由包含关系下所有代表概念极小集组成的矩阵, 称 $\mathbf{A}_{\min}$ 为 (G', M', I_R) 的最小代表概念矩阵。

下面从代表概念矩阵角度给出概念协调集的判定定理。

定理2 设 (G, M, I, T, J) 为弱协调决策形式背景, (G', M', I_R) 为其前件背景, $L(G', M', I_R)$ 是其概念格。对于 $\mathcal{F} \subseteq L(G', M', I_R), \mathcal{F} \neq \emptyset$, 以下命题等价:

- 1) $\mathcal{F}$ 是概念协调集;
- 2) 对任意 $(g, m) \in I_R, \mathcal{F} \cap \text{REP}((g, m)) \neq \emptyset$;
- 3) 对任意 $D \subseteq L(G', M', I_R)$, 若 $\mathcal{F} \cap D = \emptyset$, 则 $D \notin \mathbf{A}$ 。

证明 1) $\Rightarrow$ 2)。假设 $\mathcal{F}$ 是概念协调集, 则有 $I_R = \bigcup_{(X, B) \in \mathcal{F}} X \times B$ 。那么, 对于任意 $(g, m) \in I_R$, 存在 $(X, B) \in \mathcal{F}$, 使得 $(g, m) \in X \times B$ 成立。因此, $(X, B) \in \text{REP}((g, m))$ 。故 $\mathcal{F} \cap \text{REP}((g, m)) \neq \emptyset$ 。

2) $\Rightarrow$ 1)。若对于任意 $(g, m) \in I_R, \mathcal{F} \cap \text{REP}((g, m)) \neq \emptyset$, 则存在 $(X, B) \in \mathcal{F}$, 满足 $(X, B) \in \text{REP}((g, m))$, 使得 $(X, B) \in \mathcal{F}$ 和 $(g, m) \in X \times B$ 成立。因此, $\bigcup_{(X, B) \in \mathcal{F}} X \times B = \bigcup_{(g, m) \in I_R, (X, B) \in \mathcal{F}} \{(g, m)\} = I_R$, 故 $\mathcal{F}$ 是概念协调集。

2) $\Rightarrow$ 3)。若对于任意 $(g,m) \in I_R, D \subseteq L(G', M', I_R)$, 有 $\mathcal{F} \cap \text{REP}((g,m)) \neq \emptyset$ 和 $\mathcal{F} \cap D = \emptyset$ 成立。假设 $D \in \mathbf{A}$, 则存在 $(g_0, m_0) \in I_R$, 使得 $\text{REP}((g_0, m_0)) = D$ 。进一步地, 有 $\mathcal{F} \cap D = \mathcal{F} \cap \text{REP}((g_0, m_0)) \neq \emptyset$, 与条件 $\mathcal{F} \cap D = \emptyset$ 矛盾。故 $D \notin \mathbf{A}$ 。

3) $\Rightarrow$ 2)。对于任意 $D \subseteq L(G', M', I_R)$, 若 $\mathcal{F} \cap D = \emptyset$, 有 $D \notin \mathbf{A}$ 。反之, 若 $D \in \mathbf{A}$, 则有 $\mathcal{F} \cap D \neq \emptyset$ 。又因为 $(g,m) \in I_R$, 那么, $\text{REP}((g,m)) \in \mathbf{A}$, 故 $\mathcal{F} \cap \text{REP}((g,m)) \neq \emptyset$ 。

结合定义 7 和定理 2 可得概念约简的判定定理。

定理 3 设 (G,M,I,T,J) 为弱协调决策形式

背景, (G', M', I_R) 为其前件背景, $L(G', M', I_R)$ 是其概念格。对于 $\mathcal{F} \subseteq (G', M', I_R)$, 如果满足条件:

- 1) 对任意 $(g,m) \in I_R, \mathcal{F} \cap \text{REP}((g,m)) \neq \emptyset$;
- 2) 对任意 $(X_i, B_i) \in \mathcal{F}$, 存在 $(g_0, m_0) \in I_R$, 使得 $(\mathcal{F} \setminus \{(X_i, B_i)\}) \cap \text{REP}((g_0, m_0)) = \emptyset$, 则 $\mathcal{F}$ 是概念约简。

定理 3 实际上给出了一种获取所有概念约简的方法, 要找概念约简, 就是对任意的 $(g,m) \in I_R$, 找满足 $\mathcal{F} \cap \text{REP}((g,m)) \neq \emptyset$ 的极小概念集合 $\mathcal{F}$ 。

例 2 (续例 1) 根据定义 8 可得到前件背景 (G', M', I_R) 的代表概念矩阵 $\mathbf{A}$ 为

$$\mathbf{A} = \begin{pmatrix} \{c_1, c_{13}\} & \{c_1, c_{13}\} & \emptyset & \emptyset & \{c_{11}, c_{14}\} & \emptyset & \emptyset & \emptyset & \emptyset \\ \{c_{10}, c_{13}\} & \{c_{10}, c_{13}\} & \emptyset & \{c_{10}\} & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \\ \{c_6, c_7, c_{10}, c_{11}, c_{13}\} & \{c_6, c_7, c_{10}, c_{11}, c_{13}\} & \emptyset & \{c_6, c_7, c_{10}\} & \{c_6, c_7, c_8, c_{11}, c_{12}, c_{14}\} & \emptyset & \{c_6, c_7, c_8, c_{12}\} & \{c_6, c_8\} & \emptyset \\ \{c_2, c_5, c_6, c_7, c_{10}, c_{11}, c_{13}\} & \{c_2, c_5, c_6, c_7, c_{10}, c_{11}, c_{13}\} & \emptyset & \{c_2, c_5, c_6, c_7, c_{10}\} & \{c_2, c_5, c_6, c_7, c_8, c_{11}, c_{12}, c_{14}\} & \{c_2, c_5\} & \{c_2, c_5, c_6, c_7, c_8, c_{12}\} & \{c_2, c_6, c_8\} & \emptyset \\ \emptyset & \emptyset & \{c_4, c_9\} & \emptyset & \{c_4, c_6, c_9, c_{12}, c_{14}\} & \emptyset & \{c_4, c_6, c_9, c_{12}\} & \{c_4, c_8\} & \emptyset \\ \{c_3, c_5, c_7, c_{10}, c_{11}, c_{13}\} & \{c_3, c_5, c_7, c_{10}, c_{11}, c_{13}\} & \{c_3, c_9\} & \{c_3, c_5, c_7, c_{10}\} & \{c_3, c_5, c_7, c_9, c_{11}, c_{12}, c_{14}\} & \{c_3, c_5\} & \{c_3, c_5, c_7, c_9, c_{12}\} & \emptyset & \{c_3\} \end{pmatrix}$$

最小代表概念矩阵 $\mathbf{A}_{\min}$ 为

$$\mathbf{A}_{\min} = \begin{pmatrix} \{c_{11}, c_{13}\} & \emptyset & \emptyset & \emptyset & \{c_{11}, c_{14}\} & \emptyset & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \{c_{10}\} & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \{c_6, c_8\} & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \{c_2, c_5\} & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \{c_4, c_9\} & \emptyset & \emptyset & \emptyset & \emptyset & \{c_4, c_8\} & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \{c_3\} \end{pmatrix}$$

通过代表概念矩阵可知, 该前件背景有 12 个概念约简:

- $\mathcal{F}_1 = \{c_2, c_3, c_4, c_6, c_{10}, c_{11}\},$
- $\mathcal{F}_2 = \{c_3, c_4, c_5, c_6, c_{10}, c_{11}\},$
- $\mathcal{F}_3 = \{c_2, c_3, c_4, c_8, c_{10}, c_{11}\},$
- $\mathcal{F}_4 = \{c_3, c_4, c_5, c_8, c_{10}, c_{11}\},$
- $\mathcal{F}_5 = \{c_2, c_3, c_8, c_9, c_{10}, c_{11}\},$
- $\mathcal{F}_6 = \{c_3, c_5, c_8, c_9, c_{10}, c_{11}\},$
- $\mathcal{F}_7 = \{c_2, c_3, c_4, c_6, c_{10}, c_{13}, c_{14}\},$
- $\mathcal{F}_8 = \{c_3, c_4, c_5, c_6, c_{10}, c_{13}, c_{14}\},$
- $\mathcal{F}_9 = \{c_2, c_3, c_4, c_8, c_{10}, c_{13}, c_{14}\},$
- $\mathcal{F}_{10} = \{c_3, c_4, c_5, c_8, c_{10}, c_{13}, c_{14}\},$
- $\mathcal{F}_{11} = \{c_2, c_3, c_8, c_9, c_{10}, c_{13}, c_{14}\},$
- $\mathcal{F}_{12} = \{c_3, c_5, c_8, c_9, c_{10}, c_{13}, c_{14}\}。$

以概念约简 $\mathcal{F}_1$ 为例, 如果需要知道所有规则前件拥有的信息, 不需要图 1 中的所有概念, 仅通过 $\mathcal{F}_1$ 中的 6 个概念就能得到所有规则前件的信息。概念约简保持所有规则前件的信息, 进而也会保持非冗余规则前件的信息。由例 1 可知, 所有非冗余规则前件的概念集合为 $\{c_2, c_3, c_4, c_{13}\}$, 而 $\mathcal{F}_1$ 含有规则前件 c_2, c_3, c_4 , 相应就会直接保留与之对应的非冗余规则 $\mathcal{R}_1 \circ (M, T) = \{r_2, r_3, r_4\}。$

3 概念特征分析

定义 9 设 (G,M,I,T,J) 为弱协调决策形式背景, (G', M', I_R) 为其前件背景, $L(G', M', I_R)$ 是其概念格, 记 $W = \{\mathcal{F}_i | i \in \tau, \tau$ 为指标集 $\}$ 为 $(G',$

M', I_R 的所有概念约简的集合, 可将 $L(G', M', I_R)$ 中的概念分为3类:

- 1) 核心概念集 $\mathcal{C} = \bigcap_{i \in \tau} \mathcal{F}_i$;
- 2) 相对必要概念集 $\mathcal{R} = \bigcup_{i \in \tau} \mathcal{F}_i \setminus \bigcap_{i \in \tau} \mathcal{F}_i$;
- 3) 绝对不必要概念集 $\mathcal{U} = L(G', M', I_R) \setminus \bigcup_{i \in \tau} \mathcal{F}_i$ 。

例3 (续例2) 由例2的概念约简结果, 可将 $L(G', M', I_R)$ 中的概念分为3类。核心概念集为 $\mathcal{C} = \{c_3, c_{10}\}$, 相对必要概念集为 $\mathcal{R} = \{c_2, c_4, c_5, c_6, c_8, c_9, c_{11}, c_{13}, c_{14}\}$, 绝对不必要概念集为 $\mathcal{U} = \{c_1, c_7, c_{12}, c_{15}\}$ 。

下面给出3类概念特征的判定定理。

定理4 设 (G, M, I, T, J) 为弱协调决策形式背景, (G', M', I_R) 是其前件背景, $L(G', M', I_R)$ 是其概念格。对于任意 $(X, B) \in L(G', M', I_R)$, (X, B) 是核心概念当且仅当存在 $(g_0, m_0) \in I_R$, 有 $\text{REP}((g_0, m_0)) = \{(X, B)\}$ 。

证明 若 (X, B) 是核心概念 $\Leftrightarrow L(G', M', I_R) \setminus \{(X, B)\}$ 不是概念协调集 $\Leftrightarrow$ 存在 $(g_0, m_0) \in I_R$, 使 $(L(G', M', I_R) \setminus \{(X, B)\}) \cap \text{REP}((g_0, m_0)) = \emptyset$ 成立 $\Leftrightarrow \text{REP}((g_0, m_0)) = \{(X, B)\}$ 。

例4 (续例2) 由于 $\text{REP}((7, a_9)) = \{c_3\}$, 根据定理4知 c_3 是核心概念, 则 c_3 存在于每个概念约简。同时, c_3 是非冗余规则 $r_3: (7, a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_9) \rightarrow (1347, d_1 d_2)$ 的前件, 因此, 每个概念约简都会得到与 c_3 对应的非冗余规则 r_3 。

下面给出核心概念的判定定理。

定理5 设 (G, M, I, T, J) 为弱协调决策形式背景, (G', M', I_R) 是其前件背景, $L(G', M', I_R)$ 是其概念格。 $\mathcal{F}_i$ 是 (G', M', I_R) 的概念约简, 每个概念约简对应保留的非冗余规则集合记为 $\mathcal{R}_i^\circ(M, T)$ 。若 $\bigcap \mathcal{R}_i^\circ(M, T) = \{(X_0, B_0) \rightarrow (Y_0, C_0) \mid (X_0, B_0) \in L(G', M', I_R), (Y_0, C_0) \rightarrow L(G, T, J)\}$, 则 (X_0, B_0) 是 $L(G', M', I_R)$ 的核心概念。其中, $i \in \tau, \tau$ 指标集。

证明 $\mathcal{R}_i^\circ(M, T)$ 是与概念约简 $\mathcal{F}_i$ 对应保留的非冗余规则集合。若 $\bigcap \mathcal{R}_i^\circ(M, T) = \{(X_0, B_0) \rightarrow (Y_0, C_0) \mid (X_0, B_0) \in L(G', M', I_R), (Y_0, C_0) \rightarrow L(G, T, J)\}$, 则 $(X_0, B_0) \in \bigcap \mathcal{F}_i$, 即 $(X_0, B_0) \in \mathcal{C}$, 故 (X_0, B_0) 是 $L(G', M', I_R)$ 的核心概念。

一个绝对不必要概念是不包含在任何概念约简中的概念, 下面给出绝对不必要概念的判定定理。

定理6 设 (G, M, I, T, J) 为弱协调决策形式背景, (G', M', I_R) 是其前件背景, $L(G', M', I_R)$ 是其概念格。对于任意 $(X, B) \in L(G', M', I_R)$, (X, B) 是绝对不必要概念当且仅当对于任意 $\text{REP}((g, m)) \in \mathbf{A}_{\min}$, 有 $(X, B) \notin \text{REP}((g, m))$ 。

证明 下面证明定理6的逆否命题。

由定理3知, 若存在 $\text{REP}((g_0, m_0)) \in \mathbf{A}_{\min}$, 满足 $(X, B) \in \text{REP}((g_0, m_0)) \Leftrightarrow$ 存在概念约简 $\mathcal{F}_0$, 使得 $(X, B) \in \mathcal{F}_0 \Leftrightarrow (X, B) \notin \mathcal{U}$, 即 (X, B) 不是绝对不必要概念。

结合定理5和定理6, 容易得到相对必要概念的概念特征。

定理7 设 (G, M, I, T, J) 为弱协调决策形式背景, (G', M', I_R) 是其前件背景, $L(G', M', I_R)$ 是其概念格。对于任意 $(X, B) \in L(G', M', I_R)$, (X, B) 是相对必要概念当且仅当存在 $\text{REP}((g, m)) \in \mathbf{A}_{\min}$, 有 $(X, B) \in \text{REP}((g, m))$ 成立, 且对于任意 $(g_0, m_0) \in I_R$, 有 $\text{REP}((g_0, m_0)) \neq \{(X, B)\}$ 。

例5 (续例2) 对任意的 $\text{REP}((g, m)) \in \mathbf{A}_{\min}$, $c_7 \notin \text{REP}((g, m))$, 根据定理6可知 c_7 是绝对不必要概念; 存在 $\text{REP}((4, a_8)) \in \mathbf{A}_{\min}$, 使 $c_8 \in \text{REP}((4, a_8))$, 且对于任意 $(g_0, m_0) \in I_R$, $\text{REP}((g_0, m_0)) \neq \{c_8\}$, 根据定理7可知 c_8 是相对必要概念。

4 结语

概念约简是形式概念分析领域中一种新的约简理论, 它既能保留形式背景的信息, 又可以简化概念格。本文将概念约简思想运用到决策形式背景, 讨论决策形式背景上保持规则前件信息概念约简问题。首先, 为简化数据节省存储空间, 对条件子背景压缩构造前件背景; 其次, 提出了保持规则前件信息概念约简, 并得到了获取所有概念约简的方法; 最后, 从最小代表概念矩阵角度讨论了3类形式概念的概念特征。

决策形式背景上的概念约简还有许多值得研究探讨的问题, 如现实生活中涉及的非协调决策形式背景较多, 未来可以考虑一般的决策形式背景上的概念约简问题。此外, 还可以考虑条件格与决策格之间的联系, 针对决策格讨论约简问题, 这些都将会丰富概念约简理论。

参考文献

- [1] WILLE R. Restructuring lattice theory: An approach based on hierarchies of concepts [C]// Ordered Sets. Dordrecht-Boston: Springer, 1982: 445-470.
- [2] GANTER B, WILLE R. Formal concept analysis: Mathematical foundations [M]. New York: Springer-Verlag, 1999.
- [3] TU X D, WANG Y L, ZHANG M L, 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.
- [4] ZOU C F, ZHANG D Q, WAN J F, et al. Using concept lattice for personalized recommendation system design [J]. IEEE Systems Journal, 2017, 11(1): 305-314.
- [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] 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 W X, WEI L, QI J J. Attribute reduction theory and approach to concept lattice [J]. Science in China Series F-Information Sciences, 2005, 48(6): 713-726.
- [7] WANG X, MA J M. A novel approach to attribute reduction in concept lattices [C]// Proceedings of the International Conference on Rough Sets and Knowledge Technology. Berlin, Heidelberg: Springer, 2006: 522-529.
- [8] LI T J, LI M Z, GAO Y. Attribute reduction of concept lattice based on irreducible elements [J]. International Journal of Wavelets Multiresolution and Information Processing, 2013, 11(6): 2792-2813.
- [9] 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.
- [10] 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 L, WEI L, QI J J. Concept reduction preserving binary relations [J]. Pattern Recognition and Artificial Intelligence, 2018, 31(6): 516-524.
- [11] 魏玲, 曹丽, 祁建军, 等. 形式概念分析中的概念约简与概念特征[J]. 中国科学: 信息科学, 2020, 50(12): 1817-1833.
WEI L, CAO L, QI J J, et al. Concept reduction and concept characteristics in formal concept analysis [J]. Scientia Sinica Informationis, 2020, 50(12): 1817-1833.
- [12] 谢小贤, 李进金, 陈东晓, 等. 基于布尔矩阵的保持二元关系不变的概念约简[J]. 山东大学学报(理学版), 2020, 55(5): 32-45.
XIE X X, LI J J, CHEN D X, et al. Concept reduction of preserving binary relations based on boolean matrix [J]. Journal of Shandong University (Natural Science Edition), 2020, 55(5): 32-45.
- [13] 李俊余, 李星璇, 王霞, 等. 基于三元因子分析的三元概念约简[J]. 南京大学学报(自然科学), 2020, 56(4): 480-493.
LI J Y, LI X X, WANG X, et al. Reduction of triadic concepts based on triadic factor analysis [J]. Journal of Nanjing University (Natural Science Edition), 2020, 56(4): 480-493.
- [14] 王霞, 彭致华, 李俊余, 等. 一种基于概念可辨识矩阵的概念约简方法[J]. 计算机科学, 2021, 48(1): 125-130.
WANG X, PENG Z H, LI J Y, et al. Method of concept reduction based on concept discernibility matrix [J]. Computer Science, 2021, 48(1): 125-130.
- [15] 智慧来, 李逸楠. 形式概念分析中的面向对象概念约简[J]. 海南热带海洋学院学报, 2021, 28(5): 66-71.
ZHI H L, LI Y N. Object-oriented concept reduction in formal concept analysis [J]. Journal of Hainan Tropical Ocean University, 2021, 28(5): 66-71.
- [16] ZHAO S Y, QI J J, LI J A, et al. Concept reduction in formal concept analysis based on representative concept matrix [J]. International Journal of Machine Learning and Cybernetics, 2023, 14(4): 1147-1160.
- [17] ZHOU J Q, YANG S C, WANG X F. Concept and attribute reduction based on rectangle theory of formal concept [J]. Mathematical Foundations of Computing, 2023, 6(2): 178-189.
- [18] 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 L, QI J J, ZHANG W X. Attribute reduction theory of concept lattice based on decision formal contexts [J]. Science in China Series F-Information Sciences, 2008, 51(7): 910-923.
- [19] 陈雪, 魏玲, 钱婷. 基于 AE 概念格的决策形式背

景属性约简[J]. 山东大学学报(理学版), 2017, 52(12): 95-103.

CHEN X, WEI L, QIAN T. Attribute reduction in formal decision contexts based on AE-concept lattices [J]. Journal of Shandong University (Natural Science Edition), 2017, 52(12): 95-103.

[20] WEI L, LIU L, QI J J, et al. Rules acquisition of formal decision contexts based on three-way concept lattices [J]. Information Sciences, 2020, 516: 529-544.

[21] CHEN J K, MI J S, XIE B, et al. Attribute reduction informal decision contexts and its application to finite topological spaces [J]. International Journal of Machine Learning and Cybernetics, 2021, 12(1): 39-52.

[22] 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 J H, WEI L, ZHANG Z, et al. Concept lattice theory and method and their research prospect [J]. Pattern Recognition and Artificial Intelligence, 2020, 33(7): 619-642.

[23] 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.

[24] LI J H, MEI C L, LYU Y J. Knowledge reduction in decision formal contexts [J]. Knowledge-Based Systems, 2011, 24(5): 709-715.

[25] CHEN J H, ZHENG H Y, WEI L, et al. Factor diagnosis and future governance of dangerous goods accidents in China's ports [J]. Environmental Pollution, 2020, 257: 113582.

(编辑 张 欢)