第 31 卷 第 10 期
Vol. 31 No. 10

控 制 与 决 策
Control and Decision

2016 年 10 月
Oct. 2016

文章编号: 1001-0920 (2016) 10-1779-06 **DOI:** 10.13195/j.kzyjc.2015.1305

三支近似概念格中基于对象-概念辨识矩阵的属性约简方法

李美争, 王国胤

(1. 西南交通大学 信息科学与技术学院，成都 610031；2. 重庆邮电大学 计算智能重庆市重点实验室，重庆 400065)

摘 要: 属性约简是概念格理论的一个重要研究内容, 基于辨识矩阵计算约简是一种经典方法, 传统辨识矩阵的计算复杂度为 $O(nl^2)$. 鉴于此, 在三支近似概念格模型中, 构造一种对象-概念辨识矩阵, 其计算复杂度为 $O(mnl)$, 一般情况下, m 远远小于 l, 辨识矩阵的计算复杂度大大降低, 并结合概念格的偏序关系进一步简化对象-概念辨识矩阵. 通过理论分析和实验结果表明了所提出方法的高效性.

关键词: 概念格；三支决策；不完备形式背景；属性约简；辨识矩阵

中图分类号: TP18 **文献标志码:** A

Object-concept discernibility matrix based approach to attribute reduction in three-way approximate concept lattice

LI Mei-zheng, *WANG Guo-yin*

(1. School of Information Science and Technology，Southwest Jiaotong University，Chengdu 610031，China; 2. Chongqing Key Laboratory of Computational Intelligence，Chongqing University of Posts and Telecommunications，Chongqing 400065，China．Correspondent：WANG Guo-yin，E-mail：wanggy@ieee.org)

Abstract: Attribute reduction is a core issue in formal concept analysis(FCA). Of all attribute reduction approaches, the ones based on discernibility matrix and discernibility function are of most importance. However, in the traditional discernibility matrices, the comparisons between every two concepts result in a high computation complexity: $O(nl^2)$. Therefore, an object-concept discernibility matrix is constructed to obtain the reducts of the incomplete contexts, and the computation complexity is reduced to $O(mnl)$. In most cases, m is much smaller than l, so $O(mnl) \ll O(nl^2)$. The partial order of the concept lattice is further used to simplify the object-concept discernibility matrix. Theoretical analysis and experimental results show the effectiveness of the proposed methods.

Keywords: concept lattice；three-way decisions；incomplete context；attribute reduction；discernibility matrix

0 引 言

属性约简是概念格理论中的一个重要研究内容, 基于辨识矩阵计算约简是一种经典方法, 在多个概念格模型中得到了成功应用[1-7]. 由于传统辨识矩阵需要区分任意两个概念, 计算复杂度为 $O(nl^2)$ (其中: n 表示形式背景中的属性个数, l 表示概念格大小), 即使对于一个很小的形式背景而言, l 也可能很大[8], 导致构造辨识矩阵时计算量非常大.

为了解决这一问题, 本文在三支近似概念格模型中构造一种对象-概念辨识矩阵, 仅需比较每个对象和不包含该对象的概念, 计算复杂度为 $O(mnl)$ (其中 m 表示形式背景中的对象个数). 由于 m 往往远小于 l, $O(mnl)$ 远小于 $O(nl^2)$, 即对象-概念辨识矩阵的计算复杂度大大降低. 充分利用概念格的偏序关系, 仅需比较每个对象和不包含该对象的极大概念, 进一步简化了对象-概念辨识矩阵. 通过理论分析和实验结果表明了所提出方法的高效性.

1 三支近似概念格理论基本知识

首先介绍三支近似概念格的相关知识[7]. 称四元组 (U, A, V, T) 是一个不完备形式背景. 其中: U 为对象集合, A 为属性集合, $V = \{1, ?, 0\}$ 为值域, $T \subseteq U \times A \times V$ 为一个三元关系, $(x, a, 1) \in T$ 表示对象 x 具有属性 a, $(x, a, 0) \in T$ 表示对象 x 不具有属性 a, $(x, a, ?) \in T$ 表示对象 x 和属性 a 的关系未知.

收稿日期: 2015-10-24；**修回日期:** 2016-03-14.

基金项目: 国家自然科学基金项目(61272060, 61379114)；重庆市自然科学基金重点项目(cstc2013jjB40003).

作者简介: 李美争(1984−), 女, 博士生, 从事概念格与粗糙集的研究；王国胤(1970−), 男, 教授, 博士生导师, 从事粗糙集、粒计算、认知计算等研究.

注 1 经典形式背景[9]可以看作是一种特殊的不完备形式背景, 这是因为经典形式背景 (U,A,R) 也可记作 (U,A,V,T), 其中 $(x,a,1)\in T$ 当且仅当 $(x,a)\in R$, 且 $(x,a,0)\in T$ 当且仅当 $(x,a)\notin R$.

本文假设所有的不完备形式背景在进行属性约简之前是正则的, 即三元关系 T 满足如下命题.

命题 1 $\forall\, x\in U$, $\exists\, a,\ b\in A$, 使得 $(x,a,1)\in T$ 且 $(x,b,0)\in T$.

命题 2 $\forall\, a\in A$, $\exists\, x,\ y\in U$, 使得 $(x,a,1)\in T$ 且 $(y,a,0)\in T$.

设 S 是一个有限非空集合, 令 $\mathcal{P}(S)$ 表示 S 的幂集, $\mathcal{DP}(S)=\mathcal{P}(S)\times\mathcal{P}(S)$. $\mathcal{DP}(S)$ 上的运算交 $\bigcap$, 并 $\bigcup$, 补 c 和包含 $\subseteq$ 分别定义如下[10]: 对于任意 (X,Y), (X_1,Y_1), $(X_2,Y_2)\in\mathcal{DP}(S)$, 有

$$(X_1,Y_1)\bigcap(X_2,Y_2)=(X_1\bigcap X_2,Y_1\bigcap Y_2);$$
$$(X_1,Y_1)\bigcup(X_2,Y_2)=(X_1\bigcup X_2,Y_1\bigcup Y_2);$$
$$(X,Y)^c=(X^c,Y^c)=(S-X,S-Y);$$
$$(X_1,Y_1)\subseteq(X_2,Y_2),\ X_1\subseteq X_2,\ Y_1\subseteq Y_2.$$

定义 1[7] 设 $\mathrm{IC}=(U,A,V,T)$ 是不完备形式背景, $X\subseteq U$, $B,C\subseteq A$, 定义三支近似算子为

$$\lessdot:\mathcal{P}(U)\to\mathcal{DP}(A),\ X^{\lessdot}=(X^{*},X^{\bar{*}}),$$

三支近似算子的对偶算子定义为

$$\gtrdot:\mathcal{DP}(A)\to\mathcal{P}(U),\ (B,C)^{\gtrdot}=B^{*}\bigcap C^{\bar{*}}.$$

其中

$$X^{*}=\{a\in A|(x,a,1)\in T,\ \forall x\in X\},$$
$$X^{\bar{*}}=\{a\in A|(x,a,0)\in T,\ \forall x\in X\},$$
$$B^{*}=\{x\in U|(x,a,1)\in T,\ \forall a\in B\},$$
$$C^{\bar{*}}=\{x\in U|(x,a,0)\in T,\ \forall a\in C\}.$$

定理 1[7] 对于任意 $X,Y\subseteq U$, $\varTheta,\varOmega\in\mathcal{DP}(A)$, 下列性质成立:

1) $X\subseteq Y\Rightarrow X^{\lessdot}\supseteq Y^{\lessdot}$, $\varTheta\subseteq\varOmega\Rightarrow\varTheta^{\gtrdot}\supseteq\varOmega^{\gtrdot}$;

2) $X\subseteq X^{\lessdot\gtrdot}$, $\varTheta\subseteq\varTheta^{\gtrdot\lessdot}$;

3) $X^{\lessdot}=X^{\lessdot\gtrdot\lessdot}$, $\varTheta^{\gtrdot}=\varTheta^{\gtrdot\lessdot\gtrdot}$;

4) $X\subseteq\varTheta^{\gtrdot}\Leftrightarrow\varTheta\subseteq X^{\lessdot}$;

5) $(X\bigcup Y)^{\lessdot}=X^{\lessdot}\bigcap Y^{\lessdot}$, $(\varTheta\bigcup\varOmega)^{\gtrdot}=\varTheta^{\gtrdot}\bigcap\varOmega^{\gtrdot}$.

定义 2[7] 设 $\mathrm{IC}=(U,A,V,T)$ 是不完备形式背景, $X\subseteq U$ 且 $(B,C)\in\mathcal{DP}(A)$. 如果 $X^{\lessdot}=(B,C)$ 且 $(B,C)^{\gtrdot}=X$, 则称 $(X,(B,C))$ 是 IC 的一个对象诱导的三支近似概念, 简称为 OE 近似概念. X 和 (B,C) 分别称为 $(X,(B,C))$ 的外延和内涵.

IC 的所有 OE 近似概念形成一个完备格, 称为 OE 近似概念格, 记作 OEL(IC), 其中下确界与上确界分别定义如下 ($\forall\mathcal{C}_i=(X_i,(B_i,C_i))\in\mathrm{OEL(IC)}$, $i=1,2$):

$$\mathcal{C}_1\bigwedge\mathcal{C}_2=(X_1\bigcap X_2,(B_1\bigcup B_2,C_1\bigcup C_2)^{\gtrdot\lessdot});$$
$$\mathcal{C}_1\bigvee\mathcal{C}_2=((X_1\bigcup X_2)^{\lessdot\gtrdot},(B_1\bigcap B_2,C_1\bigcap C_2)).$$

OEL(IC) 上的偏序关系 $\leqslant$ 定义为

$$\mathcal{C}_1\leqslant\mathcal{C}_2\Leftrightarrow X_1\subseteq X_2\Leftrightarrow(B_1,C_1)\supseteq(B_2,C_2).$$

如果 $\mathcal{C}_1\leqslant\mathcal{C}_2$, $\mathcal{C}_1\neq\mathcal{C}_2$, 则记作 $\mathcal{C}_1<\mathcal{C}_2$. 进一步, 如果不存在 OE 近似概念 $\mathcal{C}$, 使得 $\mathcal{C}_1<\mathcal{C}<\mathcal{C}_2$, 则称 $\mathcal{C}_1$ 是 $\mathcal{C}_2$ 的一个下近邻, 记作 $\mathcal{C}_1\prec\mathcal{C}_2$.

例 1 一个不完备形式背景 IC 见表 1, OEL(IC) 的 Hasse 图见图 1.

表 1 不完备形式背景 IC

U	a	b	c	d	e	h	j
1	1	1	0	1	1	0	1
2	?	?	1	0	0	?	?
3	0	0	0	1	0	0	?
4	1	1	1	0	0	1	0

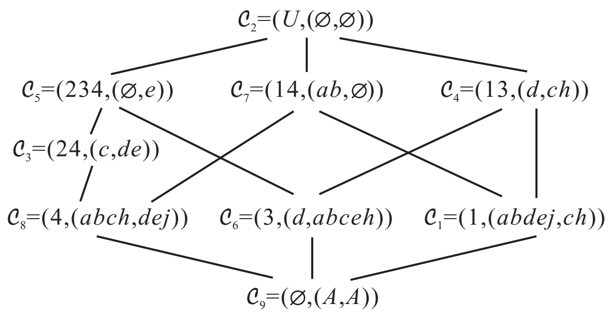

图 1 OEL(IC) 的 Hasse 图

定义 3[7] 设 $\mathrm{IC}_i=(U,A_i,V,T_i)\,(i=1,2)$ 是不完备形式背景. 如果 $\forall(X_2,(B_2,C_2))\in\mathrm{OEL(IC_2)}$, 则存在 $(X_1,(B_1,C_1))\in\mathrm{OEL(IC_1)}$, 使得 $X_1=X_2$, 记作 $\mathrm{OEL(IC_1)}\leqslant\mathrm{OEL(IC_2)}$. 若 $\mathrm{OEL(IC_1)}\leqslant\mathrm{OEL(IC_2)}$ 和 $\mathrm{OEL(IC_2)}\leqslant\mathrm{OEL(IC_1)}$ 同时成立, 则记作 $\mathrm{OEL(IC_1)}\cong\mathrm{OEL(IC_2)}$.

设 $\mathrm{IC}=(U,A,V,T)$ 是不完备形式背景, $D\subseteq A$. 令 $T_D=T\bigcap(U\times D\times V)$, 那么 $\mathrm{IC}_D=(U,D,V,T_D)$ 也是不完备形式背景. $\forall X\subseteq U$, $\forall(B,C)\in\mathcal{DP}(D)$, $X^{\lessdot}$ 在 IC 和 IC_D 中分别用 $X^{\lessdot A}$ 和 $X^{\lessdot D}$ 表示. $(B,C)^{\gtrdot A}$ 与 $(B,C)^{\gtrdot D}$ 的含义类似. 显然 $T_A=T$, $X^{\lessdot A}=X^{\lessdot}$, $X^{\lessdot D}=X^{\lessdot}\bigcap(D,D)$, $X^{\lessdot D}\subseteq X^{\lessdot}$, 且 $(B,C)^{\gtrdot A}=(B,C)^{\gtrdot}=(B,C)^{\gtrdot D}$.

定义 4[7] 设 $\mathrm{IC}=(U,A,V,T)$ 是不完备形式背景, $D\subseteq A$. 如果 $\mathrm{OEL(IC}_D)\cong\mathrm{OEL(IC)}$, 则称 D 是 IC 的一个 OE 协调集. 进一步, 如果 $\forall d\in D$, 有

$$\mathrm{OEL(IC}_{D-\{d\}})\ncong\mathrm{OEL(IC)},$$

则称 D 是 IC 的一个 OE 约简.

2　基于传统辨识矩阵的属性约简方法

参考文献 [7] 给出了一种基于传统辨识矩阵计算 OE 约简的方法.

定义 5[7]　OE 近似概念 $\mathcal{C}_i = (X_i, (B_i, C_i))$ 和 $\mathcal{C}_j = (X_j, (B_j, C_j))$ 的辨识属性集对定义为

$$\mathfrak{D}(\mathcal{C}_i, \mathcal{C}_j) =$$
$$(B_i \bigcup B_j - B_i \bigcap B_j, C_i \bigcup C_j - C_i \bigcap C_j),$$

不完备形式背景 IC 的辨识矩阵定义为

$$\varLambda = \{\mathfrak{D}(\mathcal{C}_i, \mathcal{C}_j) | \mathcal{C}_i,\ \mathcal{C}_j \in \mathrm{OEL(IC)}\},$$

IC 的辨识函数定义为

$$f(\varLambda) = \bigwedge_{(H,F)\in\varLambda, H\cup F\neq\varnothing} \Big(\bigvee_{h\in H\cup F} h\Big).$$

D 是 IC 的一个 OE 约简, 当且仅当 $\wedge D$ 是 $f(\varLambda)$ 的主蕴涵.

算法 1 给出了计算传统辨识矩阵的方法. 由于传统辨识矩阵具有对称性, 算法 1 仅计算了 $\mathfrak{D}(\mathcal{C}_i, \mathcal{C}_j)$ $(i < j)$. 令 $|A| = n$, $|\mathrm{OEL(IC)}| = l$, 第 2 行和第 3 行的计算复杂度都是 $O(l)$, 第 4 行的计算复杂度是 $O(n)$, 所以算法 1 的计算复杂度为 $O(nl^2)$.

算法 1　计算传统辨识矩阵.

输入: OE 近似概念格 OEL(IC);

输出: IC 的传统辨识矩阵 $\varLambda$.

$\varLambda = \varnothing$

for $i = 1, 2, \cdots, |\mathrm{OEL(IC)}|$ do

　for $j = i+1, i+2, \cdots, |\mathrm{OEL(IC)}|$ do

　　根据定义 5 计算 $\mathfrak{D}(\mathcal{C}_i, \mathcal{C}_j)$

　　$\varLambda = \varLambda \bigcup \{\mathfrak{D}(\mathcal{C}_i, \mathcal{C}_j)\}$

　end for

end for

输出 $\varLambda$.

在经典概念格模型中, Qi 等[2] 充分利用概念格的偏序关系, 提出了一种简化辨识矩阵, 仅需比较当前概念和其下近邻. 本文将这种辨识矩阵推广到 OE 近似概念格模型中.

定义 6　OE 近似概念 $\mathcal{C}_i = (X_i, (B_i, C_i))$ 和 $\mathcal{C}_j = (X_j, (B_j, C_j))$ 的辨识属性集对定义为

$$\mathfrak{D}_{\mathrm{QI}}(\mathcal{C}_i, \mathcal{C}_j) = \begin{cases} (B_i - B_j, C_i - C_j),\ \mathcal{C}_i \prec \mathcal{C}_j; \\ (\varnothing, \varnothing),\ \text{otherwise}. \end{cases}$$

IC 的辨识矩阵 (本文称为 QI 辨识矩阵) 定义为

$$\varLambda_{\mathrm{QI}} = \{\mathfrak{D}_{\mathrm{QI}}(\mathcal{C}_i, \mathcal{C}_j) | \mathcal{C}_i, \mathcal{C}_j \in \mathrm{OEL(IC)}\}.$$

通过 QI 辨识矩阵计算 OE 约简的方法类似于经典概念格的情况[2], 这里不再赘述. 算法 2 给出了计算 QI 辨识矩阵的方法. 令 $|U| = m$, OE 近似概念的最大下近邻个数为 k, 则第 2 ∼ 第 4 行的计算复杂度分别为 $O(l)$、$O(k)$ 和 $O(n)$, 所以算法 2 的计算复杂度为 $O(nkl)$. 一般情况下, $k \ll l$, 算法 2 的计算复杂度小于算法 1.

算法 2　计算 QI 辨识矩阵.

输入: OEL(IC);

输出: IC 的 QI 辨识矩阵 $\varLambda_{\mathrm{QI}}$.

$\varLambda_{\mathrm{QI}} = \varnothing$

for each $\mathcal{C}_i \in \mathrm{OEL(IC)}$ do

　for $\mathcal{C}_i$ 的每个下近邻 $\mathcal{C}_j$ do

　　根据定义 6 计算 $\mathfrak{D}_{\mathrm{QI}}(\mathcal{C}_i, \mathcal{C}_j)$

　　$\varLambda_{\mathrm{QI}} = \varLambda_{\mathrm{QI}} \bigcup \{\mathfrak{D}_{\mathrm{QI}}(\mathcal{C}_i, \mathcal{C}_j)\}$

　end for

end for

输出 $\varLambda_{\mathrm{QI}}$

例 2 (续例 1)　表 2 给出了 IC 的传统辨识矩阵, 其非空项有 36 个. 表 3 给出了 IC 的 QI 辨识矩阵 (仅列出非空项), 其非空项只有 13 个.

3　基于 OC 辨识矩阵的属性约简方法

上述两种辨识矩阵区分的都是 OE 近似概念. Pang 等[4] 在直觉模糊概念格模型中构造了一种辨识矩阵, 不仅区分了任意两个概念, 还区分了对象、概念和任意两个对象. 受其启发, 本文提出一种仅区分对象和概念的辨识矩阵 —— 对象-概念辨识矩阵, 简称为 OC 辨识矩阵.

表 2　IC 的传统辨识矩阵

	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$	$\mathcal{C}_6$	$\mathcal{C}_7$	$\mathcal{C}_8$	$\mathcal{C}_9$
$\mathcal{C}_1$	$(abdej, ch)$	$(abcdej, cdeh)$	$(abej, \varnothing)$	$(abdej, ceh)$	$(abej, abe)$	(dej, ch)	$(cdehj, chdej)$	$(ch, abdej)$
$\mathcal{C}_2$		(c, de)	(d, ch)	$(\varnothing, e)$	$(d, abceh)$	$(ab, \varnothing)$	$(abch, dej)$	(A, A)
$\mathcal{C}_3$			$(cd, cdeh)$	(c, d)	$(cd, abcdh)$	(abc, de)	(abh, j)	$(abdehj, abchj)$
$\mathcal{C}_4$				(d, ceh)	$(\varnothing, abe)$	(abd, ch)	$(abcdh, cdehj)$	$(abcehj, abdej)$
$\mathcal{C}_5$					$(d, abch)$	(ab, e)	$(abch, dj)$	$(A, abcdhj)$
$\mathcal{C}_6$						$(abd, abceh)$	$(abdch, abcdhj)$	$(abcehj, dj)$
$\mathcal{C}_7$							(ch, dej)	$(cdehj, A)$
$\mathcal{C}_8$								$(dej, abch)$

表 3 IC 的 QI 辨识矩阵

	OC_1	OC_2	OC_3	OC_4	OC_5	OC_6	OC_7	OC_8	OC_9
OC_1									$(ch, abdej)$
OC_2				(d, ch)	$(\varnothing, e)$		$(ab, \varnothing)$		
OC_3								(abh, j)	
OC_4	$(abej, \varnothing)$					$(\varnothing, abe)$			
OC_5			(c, d)			$(d, abch)$			
OC_6									$(abcehj, dj)$
OC_7	(dej, ch)							(ch, dej)	
OC_8									$(dej, abch)$

3.1 OC 辨识矩阵

定义 7 对象 x 和 OE 近似概念 $\mathcal{C} = (X, (B, C))$ 的辨识属性集对为

$$\mathfrak{D}_{\mathrm{OC}}(x, \mathcal{C}) = \begin{cases} (B - x^*, C - x^{\bar{*}}),\ x \notin X; \\ (\varnothing, \varnothing),\ \text{otherwise}. \end{cases}$$

IC 的 OC 辨识矩阵定义为

$$\Lambda_{\mathrm{OC}} = \{\mathfrak{D}_{\mathrm{OC}}(x, \mathcal{C}) | x \in U,\ \mathcal{C} \in \mathrm{OEL(IC)}\}.$$

定理 2 设 IC 是不完备形式背景, $D \subseteq A$, $D \neq \varnothing$, $\forall x \in U$, $\forall \mathcal{C} = (X, (B, C)) \in \mathrm{OEL(IC)}$ 且 $x \notin X$. 下列命题等价:

1) D 是 IC的一个协调集;

2) $(D, D) \bigcap \mathfrak{D}_{\mathrm{OC}}(x, \mathcal{C}) \neq (\varnothing, \varnothing)$;

3) $\forall (F, H) \in D\mathcal{P}(A) - \{(\varnothing, \varnothing)\}$, 如果 $(D, D) \bigcap (F, H) = (\varnothing, \varnothing)$, 则 $(F, H) \notin \Lambda_{\mathrm{OC}}$.

证明 1) $\Leftrightarrow$ 2). 由参考文献 [7] 的定理 4, D 是 IC 的一个协调集 $\Leftrightarrow$ $\mathrm{OEL(IC}_D) \leqslant \mathrm{OEL(IC)}$ $\Leftrightarrow$ 任意 $(X, (B, C)) \in \mathrm{OEL(IC)}$, 总有 $X = X^{\lessdot D \gtrdot D}$ 成立 $\Leftrightarrow$ $X \supseteq X^{\lessdot D \gtrdot D}$ (由定理 1, $X \subseteq X^{\lessdot D \gtrdot D}$ 自然成立) $\Leftrightarrow$ $\forall x \notin X$, $x \notin X^{\lessdot D \gtrdot D}$ $\Leftrightarrow$ $\forall x \notin X$, $x^{\lessdot D} \not\supseteq X^{\lessdot D \gtrdot D \lessdot D} = X^{\lessdot D}$ (定理 1), 即

$$X^{\lessdot D} - x^{\lessdot D} = (X^{\lessdot} - x^{\lessdot}) \bigcap (D, D) =$$
$$(B - x^*, C - x^{\bar{*}}) \bigcap (D, D) =$$
$$\mathfrak{D}_{\mathrm{OC}}(x, \mathcal{C}) \bigcap (D, D) \neq (\varnothing, \varnothing).$$

2) $\Leftrightarrow$ 3) 显然成立. □

不完备形式背景 IC 的 OC 辨识公式为

$$f(\Lambda_{\mathrm{OC}}) = \bigwedge_{(F,H) \in \Lambda_{\mathrm{OC}}, (F \cup H) \neq \varnothing} \Big(\bigvee_{h \in (F \cup H)} h \Big).$$

由布尔公式中主蕴涵的定义[11]可知, D 是 IC 的一个约简当且仅当 $\wedge D$ 是 $f(\Lambda_{\mathrm{OC}})$ 的一个主蕴涵.

算法 3 给出了计算 IC 的 OC 辨识矩阵的方法. 该算法中, 第 2 ∼ 第 4 行的计算复杂度分别为 $O(l)$、$O(m)$ 和 $O(n)$, 所以算法 3 的最大计算复杂度为 $O(nml)$. 一般情况下, $m \ll l$, $O(nml) \ll O(nl^2)$, 即算法 3 的最大计算复杂度远远小于算法 1.

算法 3 计算 OC 辨识矩阵.

输入: IC, OEL(IC);

输出: IC 的 OC 辨识矩阵 Λ_{OC}.

$\Lambda_{\mathrm{OC}} = \varnothing$

for each $\mathcal{C} = (X, (B, C)) \in \mathrm{OEL(IC)}$ do

  for each $x \notin X$ do

    根据定义7计算$\mathfrak{D}_{\mathrm{OC}}(x, \mathcal{C})$

    $\Lambda_{\mathrm{OC}} = \Lambda_{\mathrm{OC}} \bigcup \{\mathfrak{D}_{\mathrm{OC}}(x, \mathcal{C})\}$

  end for

end for

输出 Λ_{OC}

例 3 (续 例 2) 表 4 给 出 了 IC 的 OC 辨 识 矩 阵 (仅列出非空项), 其非空项有 20 个. 与传统辨识矩阵相比, OC 辨识矩阵更加简单.

表 4 IC 的 OC 辨识矩阵

	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$	$\mathcal{C}_6$	$\mathcal{C}_7$	$\mathcal{C}_8$	$\mathcal{C}_9$
x_1			(c, de)		$(\varnothing, e)$	$(\varnothing, abe)$		(ch, dej)	$(ch, abdej)$
x_2	$(abdej, ch)$			(d, ch)		$(d, abch)$	$(ab, \varnothing)$	(abh, j)	$(abdehj, abchj)$
x_3	$(abej, \varnothing)$		(c, d)				$(ab, \varnothing)$	$(abch, dj)$	$(abcehj, dj)$
x_4	(dej, ch)			(d, ch)		$(d, abch)$			$(dej, abch)$

3.2 简化 OC 辨识矩阵

受 QI 辨识矩阵充分利用概念格的偏序关系这一点启发, 可以对 OC 辨识矩阵进行简化.

定理 3 设 IC 是不完备形式背景, $\forall x \in U$, $\mathcal{C}_i = (X_i, (B_i, C_i))(i = 1, 2) \in \mathrm{OEL(IC)}$. 如果 $X_1 \subseteq X_2$ 且 $x \notin X_2$, 则 $\mathfrak{D}_{\mathrm{OC}}(x, \mathcal{C}_1) \supseteq \mathfrak{D}_{\mathrm{OC}}(x, \mathcal{C}_2)$.

证明 如果 $X_1 \subseteq X_2$, 则 $B_2 \subseteq B_1$, $C_2 \subseteq C_1$, $B_2 - x^* \subseteq B_1 - x^*$, $C_2 - x^{\bar{*}} \subseteq C_1 - x^{\bar{*}}$, 即 $\mathfrak{D}_{\mathrm{OC}}(x, \mathcal{C}_1) \supseteq$

$\mathfrak{D}_{\mathrm{OC}}(x,\mathcal{C}_2)$. □

通过OC辨识公式计算OE约简时, 首先要利用吸收率化简 $f(\Lambda_{\mathrm{OC}})$. 根据定理3, 构造 Λ_{OC} 时仅需比较每个对象 x 和不包含 x 的极大概念.

定义 8 对象 x 和OE近似概念 $\mathcal{C}=(X,(B,C))$ 的简化辨识属性集对为

$$\mathfrak{D}_{\mathrm{SOC}}(x,\mathcal{C})=\begin{cases}(B-x^*,C-x^{\bar{*}}),\\ \qquad X\text{ 是不包含 }x\text{ 的极大外延;}\\ (\varnothing,\varnothing),\ \text{otherwise.}\end{cases}$$

IC的简化OC辨识矩阵定义为

$$\Lambda_{\mathrm{SOC}}=\{\mathfrak{D}_{\mathrm{SOC}}(x,\mathcal{C})|x\in U,\ \mathcal{C}\in \mathrm{OEL(IC)}\}.$$

定理 4 设IC是不完备形式背景, $D\subseteq A$, $D\neq\varnothing$, $\forall x\in U$, $\forall\mathcal{C}=(X,(B,C))\in\mathrm{OEL(IC)}$ 且 X 是不包含 x 的极大外延. 下列命题等价:

1) D 是IC的一个OE协调集;

2) $(D,D)\bigcap\mathfrak{D}_{\mathrm{SOC}}(x,\mathcal{C})\neq(\varnothing,\varnothing)$;

3) $\forall(F,H)\in\mathcal{DP}(A)-\{(\varnothing,\varnothing)\}$, 如果 $(D,D)\bigcap(F,H)=(\varnothing,\varnothing)$, 则 $(F,H)\notin\Lambda_{\mathrm{SOC}}$.

由定理2和定理3可证定理4成立, 此略.

算法4给出了计算IC的简化OC辨识矩阵的方法. 该算法中, 第2行和第4行的计算复杂度分别为 $O(m)$ 和 $O(l)$, 第6～第12行的计算复杂度为 $O(n)$, 所以算法4的最大复杂度为 $O(nml)$. 由于内层循环中队列长度往往远远小于 l, 与算法3相比, 算法4更简单.

算法 4 计算简化辨识矩阵.

输入: IC, OEL(IC);

输出: IC的简化OC辨识矩阵 Λ_{SOC}.

$\Lambda_{\mathrm{SOC}}=\varnothing$

for each $x\in U$ do

  生成队列 $Qx=\{(U,U^{\lessdot})$ 的下近邻$\}$

  while $Qx\neq\varnothing$ do

    $\mathcal{C}=(X,(B,C))=Qx(1)$

    if $x\notin X$

      根据定义8计算 $\mathfrak{D}_{\mathrm{SOC}}(x,\mathcal{C})$

      $\Lambda_{\mathrm{SOC}}=\Lambda_{\mathrm{SOC}}\bigcup\{\mathfrak{D}_{\mathrm{SOC}}(x,\mathcal{C})\}$

      将队列中小于等于OC的OE近似概念从队列中删除, 并标记所有小于OC的OE近似概念

    else

      将 $\mathcal{C}$ 从队列中删除, 并将其从未进入队列且未被标记的下近邻加入队尾

    end if

  end while

end for

输出 Λ_{SOC}.

例 4 (续例3) 以 x_1 为例, 利用算法4计算与 x_1 相关的非空辨识属性集对.

为对象 x_1 生成队列 $Qx_1=\{\mathcal{C}_4,\mathcal{C}_5,\mathcal{C}_7\}$.

令 $\mathcal{C}=\mathcal{C}_4$, 更新 Qx_1: $Qx_1=\{\mathcal{C}_5,\mathcal{C}_7,\mathcal{C}_1,\mathcal{C}_6\}$;

令 $\mathcal{C}=\mathcal{C}_5$, $\mathfrak{D}_{\mathrm{SOC}}(x_1,\ \mathcal{C}_5)=\{\varnothing,e\}$, 更新 Qx_1: $Qx_1=\{\mathcal{C}_7,\mathcal{C}_1\}$;

令 $\mathcal{C}=\mathcal{C}_7$, 更新 Qx_1: $Qx_1=\{\mathcal{C}_1\}$;

令 $\mathcal{C}=\mathcal{C}_1$, 更新 Qx_1: $Qx_1=\varnothing$.

类似地可以计算与对象 x_2、x_3、x_4 相关的非空辨识属性集对. 表5给出了IC的简化OC辨识矩阵(仅列出非空项, $\mathcal{C}_1$、$\mathcal{C}_2$、$\mathcal{C}_6$、$\mathcal{C}_8$ 和 $\mathcal{C}_9$ 对应的列为空, 故省略), 共有6个非空项. 与OC辨识矩阵相比, 简化OC辨识矩阵更加简单.

表 5 IC的简化OC辨识矩阵

	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$	$\mathcal{C}_7$
x_1			$(\varnothing,e)$	
x_2		(d,ch)		$(ab,\varnothing)$
x_3	(c,d)			$(ab,\varnothing)$
x_4		(d,ch)		

4 仿真实验

为了表明本文方法的有效性, 选取一些数据集(见表6)进行实验. 首先利用FcaBedrock[12]将数据集3和数据集4的离散属性转化为布尔属性. 假设不完备属性 a 转化的布尔属性为 $a\_1,\cdots,a\_k,a\_?$, 那么删除 $a\_?$, 并修改 $a\_1,\cdots,a\_k$. 如果原数据集中 $a(x)=?$, 则 $(x,a\_t,?)\in T(1\leqslant t\leqslant k)$.

表 6 数据集

No.	数据集	$\lvert U\rvert$	$\lvert A\rvert$	数据类型	备注
1	本文表1	4	7	布尔型	不完备
2	表2[13]	9	15	布尔型	不完备
3	Zoo[14]	67	18	布尔型, 离散型	完备
4	Post-operative patient[14]	90	9	离散型	不完备

分别比较数据集1～数据4的传统辨识矩阵、OC辨识矩阵、QI辨识矩阵和简化OC辨识矩阵中的非空项个数及其对比值, 实验结果见表7和表8.

表 7 OE近似概念格的4类辨识矩阵非空项个数

No.	l	Λ	Λ_{OC}	Λ_{QI}	Λ_{SOC}
1	9	36	20	13	6
2	134	8 911	741	405	30
3	10 573	5 588 878	503 664	51 066	411
4	31 992	511 728 036	2 398 060	175 852	767

表 8 4类辨识矩阵非空项个数对比值 %

No.	$R_{传统}$	R_{OC}	R_{QI}	R_{SOC}
1	100.0000	55.5556	36.1111	16.6667
2	100.0000	8.3156	4.5449	0.3367
3	100.0000	9.0119	0.9137	0.0074
4	100.0000	0.4686	0.0344	0.0001
均值	100.0000	18.3379	10.4010	4.2527

由表7可见, OC辨识矩阵、QI辨识矩阵的非空项个数都小于传统辨识矩阵; 简化OC辨识矩阵的非空项个数小于OC辨识矩阵, 这与理论分析结果一致. 在所选择的数据集上, 非空项个数从小到大依次为简化OC辨识矩阵、QI辨识矩阵、OC辨识矩阵、传统辨识矩阵.

表8中

$$R_i = \frac{\varLambda_i\text{中非空项个数}}{\varLambda\text{中非空项个数}},\ i \in \{\text{传统, OC, QI, SOC}\}.$$

可以看出, 与传统辨识矩阵相比, OC辨识矩阵的平均非空项个数减少了82%, QI辨识矩阵的平均非空项个数减少了90%, 简化OC辨识矩阵的平均非空项个数减少了96%. 本组实验验证了在OE近似概念格模型中, OC辨识矩阵(不利用概念格的偏序关系时)和简化OC辨识矩阵(利用概念格的偏序关系时)的高效性.

5 结 论

基于辨识矩阵计算概念格属性约简是一种经典的方法, 广泛应用于各种概念格模型的属性约简. 由于传统辨识矩阵计算复杂度非常高, 限制了其在大数据背景下的应用. 本文构造了一种新的辨识矩阵: OC辨识矩阵, 仅需比较对象和OE近似概念, 大大降低了辨识矩阵的计算复杂度. 在此基础上, 通过充分利用概念格的偏序关系, 提出了简化OC辨识矩阵, 进一步简化了辨识矩阵. 理论分析和实验结果都表明了OC辨识矩阵远远优于传统辨识矩阵, 简化OC辨识矩阵远远优于OC辨识矩阵. 此外, 在所选择数据集上, 上述4类辨识矩阵与QI辨识矩阵排名为: 传统辨识矩阵 $\prec$ OC辨识矩阵 $\prec$ QI辨识矩阵 $\prec$ 简化OC辨识矩阵.

参考文献(References)

[1] 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, 2005, 35(6): 628-639.
(Zhang W X, Wei L, Qi J J. Attribute reduction theory and approach to concept lattice[J]. Science China, 2005, 35(6): 628-639.)

[2] Qi J J. Attribute reduction in formal contexts based on a new discernibility matrix[J]. J of Applied Mathematics and Computing, 2009, 30(1/2): 305-314.

[3] 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 & Mathematics with Applications, 2007, 53(9): 1390-1410.

[4] Pang J Z, Zhang X Y, Xu W H. Attribute reduction in intuitionistic fuzzy concept lattices[C]. Abstract and Applied Analysis. 2013: 1-12.

[5] Mi J S, Leung Y, Wu W Z. Approaches to attribute reduction in concept lattices induced by axialities[J]. Knowledge-Based Systems, 2010, 23(6): 504-511.

[6] 崔聘芝. 区间值决策形式背景分析[D]. 临汾: 山西师范大学数学与计算机科学学院, 2014.
(Cui P Z. Analysis of interval-valued decision formal contexts[D]. Linfen: School of Mathematics and Computer Sciences, Shanxi Normal University, 2014.)

[7] Li M Z, Wang G Y. Approximate concept construction with three-way decisions and attribute reduction in incomplete contexts[J]. Knowledge-Based Systems, 2016, 91(1): 165-178.

[8] Bělohlávek R, Macko J. Selecting important concepts using weights[C]. Proc of 2011 Int Conf on Formal Concept Analysis. Berlin: Springer, 2011: 65-80.

[9] Qi J J, Wei L, Yao Y Y. Three-way formal concept analysis[C]. Proc of Rough Sets and Knowledge Technology. Switzerland: Springer, 2014: 732-741.

[10] Ganter B, Wille R. Formal concept analysis: Mathematical foundations[M]. Beilin: Springer, 1999.

[11] Skowron A, Rauszer C. The discernibility matrices and functions in information systems[C]. Proc of Intelligent Decision Support. Netherlands: Springer, 1992: 331-362.

[12] Andrews S, Orphanides C. Fcabedrock, a formal context creator[C]. Proc of Int Conf on Conceptual Structures. Springer, 2010: 181-184.

[13] Li J H, Mei C L, Lv Y J. Incomplete decision contexts: Approximate concept construction, rule acquisition and knowledge reduction[J]. Int J of Approximate Reasoning, 2013, 54(1): 149-165.

[14] Lichman M. UCI machine learning repository[EB/OL]. (2013-01-15)[2013-03-11]. http://archive.ics.uci.edu/ml.

[15] Ganter B. Two basic algorithms in concept analysis[M]. Berlin: Springer-Heidelberg, 2010: 312-340.

[16] Norris E M. An algorithm for computing the maximal rectangles in a binary relation[J]. Revue Roumaine de Mathématiques Pures et Appliquées, 1978, 23(2): 243-250.

(责任编辑：郑晓蕾)