决策蕴涵规范基

翟岩慧,李德玉,曲开社
(山西大学计算机与信息技术学院,山西太原 030006)

摘 要: 主要给出一个决策蕴涵下的决策蕴涵规范基.首先给出决策前提的概念,然后生成以决策前提为前提,决策前提相对于决策子背景的闭包为结论的决策蕴涵集——决策蕴涵规范基.证明了该决策决策蕴涵集是完备的,无冗余的,并且是最优的,即在所有完备的决策蕴涵集中,决策蕴涵规范基所含的决策蕴涵数最少.最后给出了该决策蕴涵基的生成算法.实验表明,决策蕴涵规范基可以很好地抑制冗余决策蕴涵的生成,比已有的决策蕴涵集也更为紧凑和有效.

关键词: 决策蕴涵; 规范基; 概念格; 最小生成子
中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112 (2015)01-0018-06
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2015.01.004

Canonical Basis for Decision Implications

ZHAI Yan-hui, LI De-yu, QU Kai-she
(*School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi* 030006, *China*)

Abstract: In this paper we introduced the notion of decision premise, and formed decision implications with decision premises as premises and closures w.r.t. decision subcontext as consequences. It was proven that such decision implications constitute the so-called decision canonical basis, i.e., it is complete, non-redundant and of minimal cardinality among all complete sets of decision implications. We also described an algorithm to generate decision implication canonical basis and analyzed time complexity of this algorithm. Experiments showed that decision canonical basis can greatly reduce redundant decision implications and is more efficient than other decision implication bases.

Key words: decision implication; canonical basis; concept lattice; minimal generator

1 引言

形式概念分析(Formal Concept Analysis, FCA)是 R Wille[1]提出的一种从形式背景建立概念格来进行数据分析和规则提取的强有力工具,已被广泛地研究[2~6],并应用到机器学习[7]、信息检索[8]、软件工程[9]和社会网络分析[10]等领域.

属性逻辑在 FCA 的研究中占有重要的地位[2,3,11,12],该研究通过研究属性子集之间的关系来揭示属性间的依赖.在 FCA 中,属性逻辑通过蕴涵来表示.蕴涵是具有形如 $A \rightarrow B$ 的公式,其含义是我们可以从 A 推导出 B,即任意具有属性集 A 的对象也必然具有属性集 B.给定数据集,蕴涵可以在该数据集中成立,这意味着该数据集中没有与该蕴涵矛盾的数据.相对于数据集,一个蕴涵组成的集合可以是完备的或无冗余的.完备的蕴涵集是数据集一个完整表示,而无冗余集用来表示该蕴涵集中没有冗余蕴涵,即该集合是数据集的一个紧表示.

决策背景和决策蕴涵是形式概念分析在决策情形下的一种扩展[3,11,13].在文献[3]中,我们提出了一种特殊的推理规则——α 推理规则.该推理规则通过增加决策蕴涵的前提或者减小决策蕴涵的结论来导出新的决策蕴涵.通过 α-推理规则,我们可以得到一个相对完备和无冗余的决策蕴涵集.文献[3]还证明了通过最小生成子[14]可以生成该决策蕴涵集,并给出了一个基于最小生成子的规则集生成算法.魏玲等[13]研究了决策背景中的属性约简理论.通过定义决策背景的强协调性与弱协调性,文献[13]研究了两种协调性下决策背景的属性约简.另外,我们还讨论了模糊决策蕴涵的语义和语构特征,给出了模糊决策蕴涵的逻辑理论[11].

收稿日期:2013-09-18;修回日期:2014-03-04;责任编辑:李勇锋
基金项目:国家自然科学基金(No.61303107, No.61272095, No.61175067, No.41101440, No.61202018);山西省回国留学人员科研项目(No.2013-014);山西省自然科学基金(No.2013011066-4)

本文将深入考虑决策背景下决策蕴涵的生成.类似于蕴涵的情形,即使小规模的决策背景也可能产生大量的(甚至可能是指数阶的)决策蕴涵.虽然文献[3]中给出了一个较为紧凑的完备无冗余决策蕴涵集,并在一定程度上抑制了冗余决策蕴涵的生成,然而,通过本文我们将看到,存在一个更为紧凑的决策蕴涵集——决策蕴涵规范基.该规范基基于决策前提(D-前提),即,由决策前提作为该决策蕴涵集的前提,由决策蕴涵相对于决策子背景的闭包作为该决策蕴涵集的结论.本文将证明该决策蕴涵规范基是决策蕴涵基,即该规范基是完备无冗余的决策蕴涵集.进一步,因为一个决策背景可能存在多个决策蕴涵基(如文献[3]中的决策蕴涵基),而这些决策蕴涵基中可能含有不同的决策蕴涵数,因此找到所含决策蕴涵数最小的决策蕴涵基将更有利于理论的研究和实际的应用.基于此,本文还将证明,决策蕴涵规范基是最优的,即,在所有完备的决策蕴涵基中,其所含的决策蕴涵数是最小的.这些结果表明,决策蕴涵规范基是蕴涵规范基[2]在决策背景下的对应概念,并具有蕴涵规范基的所有优点.

2 决策背景和决策蕴涵[3]

形式概念分析的主要记号和结论见文献[2],下面我们直接引入决策背景和决策蕴涵的概念[3].

定义 1 决策背景为一个三元组 $K=(G, C\cup D, I_C\cup I_D)$,其中 C 为条件属性集,D 为决策属性集且 $C\cap D=\varnothing$,$I_C\subseteq G\times C$ 为条件二元关系,$I_D\subseteq G\times D$ 为决策二元关系.

例 1 一个决策背景 K 如表 1[3]所示,其中条件属性集为$\{a_1, a_2, \cdots, a_6\}$,决策属性集为$\{d_1, d_2\}$.

表 1 形式背景

	a_1	a_2	a_3	a_4	a_5	a_6	d_1	d_2
x_1	×	×			×	×	×	
x_2			×					×
x_3	×		×			×	×	
x_4		×						
x_5		×	×	×	×	×	×	×
x_6			×		×			
x_7		×	×	×		×		×
x_8	×	×		×	×	×	×	

定义 2 设 K 为决策背景.对于集合 $A\subseteq G$,$B_1\subseteq C$ 和 $B_2\subseteq D$,记:

$$A^C=\{m\in C\mid (g,m)\in I_C, \forall g\in A\}$$

$$A^D=\{m\in D\mid (g,m)\in I_D, \forall g\in A\}$$

$$B_1^C=\{g\in G\mid (g,m)\in I_C, \forall m\in B_1\}$$

$$B_2^D=\{g\in G\mid (g,m)\in I_D, \forall m\in B_2\}$$

定义 3[3] 集合 C 和 D 上的决策蕴涵(简称决策蕴涵)具有形式 $B_1\to B_2$,并满足 $B_1\subseteq C$ 和 $B_2\subseteq D$,其中 B_1 称为该决策蕴涵的前提,B_2 称为该决策蕴涵的结论.一个集合 $T\subseteq C\cup D$ 满足 $B_1\to B_2$(记作 $T\vDash B_1\to B_2$),若 $B_1\not\subseteq T\cap C$ 或 $B_2\subseteq T\cap D$.T 满足一个决策蕴涵集合 L(记为 $T\vDash L$),若 T 满足该决策蕴涵集合中的所有决策蕴涵.$B_1\to B_2$ 在集合族$\{T_1, T_2, \cdots, T_n\}$中成立,若每个 T_i 都满足 $B_1\to B_2$.

将决策蕴涵与决策背景关联,有下面的定义[3]:

定义 4 设 K 为一决策背景,B_1, B_2 为属性子集.$B_1\to B_2$ 为 K 的决策蕴涵,若对于任意对象 g,若 $B_1\subseteq g^C$,则 $B_2\subseteq g^D$.

一般来说,决策背景中可以有很多决策蕴涵,这些决策蕴涵中有许多的冗余决策蕴涵,这些冗余的决策蕴涵可以从其他的决策蕴涵中导出.

定义 5[3] 决策蕴涵 $B_1\to B_2$ 可以从决策蕴涵集 L 中语义导出,若对任意的 $T\subseteq C\cup D$,$T\vDash L$ 蕴含 $T\vDash B_1\to B_2$,记为 $L\vdash B_1\to B_2$.决策蕴涵集 L 是无冗余的,若对任意的 $B_1\to B_2\in L$,$L\setminus\{B_1\to B_2\}\nvdash B_1\to B_2$ 成立.L 是封闭的,若任意可以从 L 中语义导出的决策蕴涵都包含在 L 中.对于封闭的决策蕴涵集 L,说 $O\subseteq L$ 是完备的,若 L 可以从 O 语义导出.

应用定义 5 到决策背景,我们有下面的定义[3].

定义 6 决策背景 K 的决策蕴涵集 L 是完备的,若 K 中成立的任何决策蕴涵都可以从 L 导出.

推论 1 若 $T\vDash B_1\to B_1^{CD}$ 且 $B_1\to B_2$ 是 K 的一个决策蕴涵,则 $T\vDash B_1\to B_2$.

证明 若 $B_1\to B_2$ 是 K 的一个决策蕴涵,由文献[3]中的定理 2 可知 $B_2\subseteq B_1^{CD}$.若 $B_1\subseteq T\cap C$,则由 $T\vDash B_1\to B_1^{CD}$ 可知 $B_1^{CD}\subseteq T\cap D$,从而 $B_2\subseteq B_1^{CD}\subseteq T\cap D$,即 $T\vDash B_1\to B_2$.

3 决策前提和决策蕴涵规范基

这节我们引入决策前提,并生成一个基于决策前提的决策蕴涵基.我们将证明该决策蕴涵基是完备的,无冗余的,而且是最优的,即在所有完备集中,决策蕴涵基所含有的决策蕴涵数是最少的.

定义 7 设 K 为决策背景.一个相对于 C 和 D 的决策前提(或 D-前提)是一个子集 $P\subseteq C$,满足

(1)P 相对于 P^{CD}是最小的,即若 $Q\subset P$,则 $Q^{CD}\subset P^{CD}$;

(2)P 是恰当的,即 $P^{CD}\neq\bigcup\{B_1^{CD}\mid B_1$ 是决策前提,$B_1\subset P\}$

例 2 表 1 的所有决策前提如表 2 所示.

对于任意的子集 $P\subseteq C$,特别是决策前提,我们可以得到一个决策蕴涵 $P\to P^{CD}$,其中 P^{CD}为 P 相对于决

策子背景的闭包.根据文献[3]中定理 2,该决策蕴涵必然为 K 的决策蕴涵.进一步,我们可以证明由决策前提组成的决策蕴涵集合是完备的和无冗余的.

表 2 表 1 中的决策前提

$\varnothing$	$\{a_2,a_3\}$	$\{a_4,a_5\}$	$\{a_1,a_3,a_4\}$	$\{a_1,a_3,a_5\}$	$\{a_3,a_4,a_5\}$
$\{a_1\}$	$\{a_2,a_4\}$	$\{a_5,a_6\}$	$\{a_3,a_5,a_6\}$	$\{a_2,a_3,a_5\}$	

定理 1 集合 $O=\{P\to P^{CD}\mid P$ 是决策前提$\}$是完备的和无冗余的.称为 K 的决策蕴涵规范基.

证明 设 $T\vDash O$,$B_1\to B_2$ 是 K 的一个决策蕴涵.为了证明 O 的完备性,我们需要证明 $T\vDash B_1\to B_2$,而由推论 1,我们只需要证明 $T\vDash B_1\to B_1^{CD}$.

分三种情况讨论:

(1)B_1 相对于 B_1^{CD} 是最小的,并且是恰当的.此时,B_1 是决策前提,并且 $B_1\to B_1^{CD}\in O$,由 $T\vDash O$ 可知断言成立;

(2)B_1 相对于 B_1^{CD} 是最小的,但不是恰当的.此时有 $B_1^{CD}=\bigcup\{B_3^{CD}\mid B_3$ 是决策前提,$B_3\subset B_1\}$.假设 $B_1\subseteq T$.因为 $T\vDash O$,特别地,对于满足 $B_3\subset B_1$ 的所有决策前提 B_3,我们有 $T\vDash B_3\to B_3^{CD}$.再由 $B_3\subset B_1\subseteq T$,我们有 $B_3^{CD}\subset T$.因此 $B_1^{CD}=\bigcup B_3^{CD}\subseteq T$,从而 $T\vDash B_1\to B_1^{CD}$;

(3)B_1 相对于 B_1^{CD} 不是最小的.此时,我们可以找到一个满足 $B_3^{CD}=B_1^{CD}$ 的最小集合 $B_3\subset B_1$.接下来,我们用 B_3 代替 B_1,并且按照情形(2)进行处理,此时我们可以证明 $T\vDash B_3\to B_1^{CD}$.易证明 $T\vDash B_1\to B_1^{CD}$ 也成立.

下面我们证明 O 是无冗余的,即对任意的 $B_1\to B_1^{CD}\in O$,我们有 $O\setminus\{B_1\to B_1^{CD}\}\nvdash B_1\to B_1^{CD}$.

我们只需找到一个集合 T 满足 $T\vDash O\setminus\{B_1\to B_1^{CD}\}$ 且 $T\nvDash B_1\to B_1^{CD}$ 即可.

记 $T=B_1\cup\bigcup\{B_3^{CD}\mid B_3$ 是决策前提,$B_3\subset B_1\}$

我们断言:(1)$T\vDash O\setminus\{B_1\to B_1^{CD}\}$;(2)$T\nvDash B_1\to B_1^{CD}$.意味着 $O\setminus\{B_1\to B_1^{CD}\}\nvdash B_1\to B_1^{CD}$.

对于第一条断言,设 $B_3\to B_3^{CD}\in O\setminus\{B_1\to B_1^{CD}\}$.若 $B_3\subseteq T\cap C=B_1$,那么 $B_3\subset B_1$(显然 $B_3\neq B_1$),因此 $B_3^{CD}\subseteq\bigcup_{B\subset B_1}B^{CD}=T\cap D$.对于第二条断言,因为 B_1 为决策前提,我们有 $B_1^{CD}\neq\bigcup\{B_3^{CD}\mid B_3$ 是决策前提,$B_3\subset B_1\}=V$,因此存在 $m\in B_1^{CD}\setminus V$.因为 $B_1\subseteq T\cap C=B_1$,但 $m\in B_1^{CD}\nsubseteq T\cap D=V$,因此 $T\nvDash B_1\to B_1^{CD}$.

例 3 表 1 的决策蕴涵规范基列于表 3.由文献[3]可知$\{a_1,a_2,a_3\}\to\{d_1,d_2\}$为 α 极大决策蕴涵,即并非冗余蕴涵,但本文其为冗余决策蕴涵,因此可以删除.这也说明决策蕴涵规范基比 α 极大决策蕴涵集更为紧凑.

对于决策蕴涵来说,决策前提在决策蕴涵上的作用正如伪内涵在蕴涵集上的作用一样.定理 1 已经证明了 O 是完备的和无冗余的.下面我们证明 O 是最优的,即在所有完备集中 O 是最小的.首先我们给出下面的结论.

表 3 表 1 的决策规范基

$\varnothing\to\varnothing$	$\{a_2,a_3\}\to\{d_2\}$	$\{a_2,a_3,a_5\}\to\{d_1,d_2\}$
$\{a_1\}\to\{d_1\}$	$\{a_5,a_6\}\to\{d_1\}$	$\{a_3,a_4,a_5\}\to\{d_1,d_2\}$
$\{a_4,a_5\}\to\{d_1\}$	$\{a_1,a_3,a_4\}\to\{d_1,d_2\}$	$\{a_1,a_3,a_5\}\to\{d_1,d_2\}$
$\{a_2,a_4\}\to\{d_1\}$	$\{a_3,a_5,a_6\}\to\{d_1,d_2\}$	

定理 2 设 L 为 K 上的一决策蕴涵集,P 为决策前提.那么我们有 $P\to P^{CD}\in L$ 或者 $L\nvdash P\to P^{CD}$.

证明 首先,我们将 L 中的任意蕴涵 $B_1\to B_2$ 改为形式 $B_1\to B_1^{CD}$,并记此时的蕴涵集为 $\bar{L}$.容易看出若 $\bar{L}\nvdash P\to P^{CD}$,则 $L\nvdash P\to P^{CD}$.因此只需证明,若 $T\vDash\bar{L}$ 则 $T\nvDash P\to P^{CD}$即可.

设 $P\to P^{CD}\notin L$ 且 $T\vDash\bar{L}$.假设 $P\subseteq T$.因为 P 为决策前提,所以 $P^{CD}\supset\bigcup\{B_1^{CD}\mid B_1$ 是决策前提,$B_1\subset P\}$.

现在我们断言:

$$W=\bigcup\{B_1^{CD}\mid B_1\text{ 是决策前提},B_1\subset P\}\supseteq\bigcup\{B_1^{CD}\mid B_1\to B_1^{CD}\in\bar{L},B_1\subset P\}=V \tag{1}$$

如果式(1)成立,那么 $P^{CD}\supset V$,此时存在 $m\in P^{CD}\setminus V$.记 $T=P\cup\bigcup\{B_1^{CD}\mid B_1$ 是决策前提,$B_1\subset P\}$,类似于定理 2 的证明,我们可以证明 $T\vDash\bar{L}$ 但 $T\nvDash P\to P^{CD}$,此时有 $\bar{L}\nvdash P\to P^{CD}$,因此 $L\nvdash P\to P^{CD}$.

现在我们证明(1).对任意的 $B_1\to B_1^{CD}\in\bar{L}$,有三种情况:(1)$B_1$ 是决策前提.此时 $B_1^{CD}\subseteq W$ 并且断言成立;(2)B_1 相对于 B_1^{CD} 是最小的,但不是恰当的.此时我们有 $B_1^{CD}=\bigcup\{B_3^{CD}\mid B_3$ 是决策前提,$B_3\subset B_1\}$,这意味着 B_1^{CD} 可以由 W 的子集合并得到,因此断言成立;(3)B_1 相对于 B_1^{CD} 不是最小的,此时我们可找到一个最小集 $B_3\subset B_1$ 满足 $B_3^{CD}=B_1^{CD}$,类似于情形 2,我们可以证明断言.

从上面的定理,我们可以得到:

推论 2 O 在所有完备集中,决策蕴涵数目是最小的.

证明 由定理 2,对任意的完备集 L 和决策前提 P,若 $P\to P^{CD}$不属于 L,那么 $L\nvdash P\to P^{CD}$.因为 L 是完备的,于是 $P\to P^{CD}$将属于 L.由 P 的任意性,我们知道 O 中任意决策蕴涵包含于 L.因此 O 的决策蕴涵的数目是最小的.

定理 1 和推论 2 事实上说明决策蕴涵规范基是一个自然决策蕴涵基,即该蕴涵基不仅是完备无冗余的,而且所含的决策蕴涵数目是最小的.

4　决策蕴涵规范基生成及实验比较

4.1　生成决策蕴涵规范基

为了生成决策蕴涵规范基,直观的方法是对 C 的所有子集检查其是否满足定义 7,然后生成决策前提,进而生成决策蕴涵规范基.然而,对于大规模的决策背景来说,该方法是不现实的.

另外一种方法来源于文献[3],该文献中我们提出一种算法来获取 α 极大决策蕴涵集.该方法基于一个 α 极大决策蕴涵的充要条件[3].

引理 1　设 $K=(G,C\cup D,I_C\cup I_D)$为决策背景.那么 $B_{C1}\rightarrow B_{D1}$是 K 的α-极大决策蕴涵当且仅当下面三个条件成立:

(1) B_{C1}是条件子背景 $K_C=(G,C,I_C)$的最小生成子;

(2) $B_{D1}=B_{C1}^{CD}$;

(3)若 $B_{C2}\subset B_{C1}$,那么 $B_{C1}^{CD}\neq B_{C2}^{CD}$.

事实上,定义 7 的条件(1)和引理 1 的条件(3)相同.进一步,该条件事实上蕴含引理 1 的条件(1).

定理 3　若 B_{C1}是 B_{C1}^{CD}的最小生成子(即引理1 的条件(3)),那么 B_{C1}是条件子背景 K_C 的最小生成子(即不存在 $B_{C2}\subset B_{C1}$满足 $B_{C2}^{CC}=B_{C1}^{CC}$).

证明　若 B_{C1}不是(B_{C1}^{C},B_{C1}^{CC})的最小生成子,那么存在最小集合 $B_{C2}\subset B_{C1}$满足 $B_{C2}^{CC}=B_{C1}^{CC}$,于是 $B_{C2}^{C}=B_{C2}^{CCC}=B_{C1}^{CCC}=B_{C1}^{C}$.因此 $B_{C2}^{CD}=B_{C1}^{CD}$,与 B_{C1}相对于 B_{C1}^{CD}是最小的矛盾.

因为每个基于决策前提的决策蕴涵都满足引理 1 的 3 个条件,所以是 α 极大决策蕴涵,因此决策蕴涵规范基包含在 α 极大决策蕴涵集 Σ 中.为了生成决策蕴涵规范基,我们只需要检查 Σ 中的所有决策蕴涵,并且移去不满足定义 7 条件(2)的所有决策蕴涵即可,如算法 1 所示.

算法 1　决策蕴涵基生成算法

输入:决策背景 K

输出:决策蕴涵规范基 O;

1　$O=\varnothing$//决策蕴涵规范基

2　$\Pi=\varnothing$//最小生成子集合

3　使用 Titanic 算法[14]生成(G,C,I_C)的所有最小生成子集合 Π 并以字典序对 Π 排序)

4　for all $B_{C1}\in\Pi$ do

5　$T=\varnothing$//用于检验定义 12 的条件(2)

6　for all 满足 $B_{C2}\subset B_{C1}$的 $B_{C2}\rightarrow B_{C2}^{CD}\in O$ do

7　　if $B_{C1}^{CD}\neq B_{C2}^{CD}$ then

8　　　$T=T\cup\{B_{C2}^{CD}\}$

9　　else

10　　　$T=B_{C2}^{CD}$;break

11　　end if

12　end for

13　从 Π 移除 B_{C1}

14　if $B_{C1}^{CD}\neq T$ then

15　　生成决策蕴涵 $B_{C1}\rightarrow B_{C1}^{CD}$,并添加到 O

16　　end if

17　end for

18　return O

下面我们来分析算法 1 的时间复杂度.算法 1 的第 3 行,提取并排序最小生成子需要时间:

$$O\left(|C|\cdot\left(db+\binom{|C|}{\lfloor |C|/2\rfloor}\cdot|G|\cdot|C|\right)\right)$$

其中 db 为形式背景的访问时间.对算法 4－17 行程序的处理和最小生成子数目是直接相关的,最坏情形下,需要时间 $O(|\Pi|^2)$,其中 Π 是最小生成子的数目.总之,该算法的时间复杂度为

$$O\left(|C|\cdot\left(db+\binom{|C|}{\lfloor |C|/2\rfloor}\cdot|G|\cdot|C|\right)+|\Pi|^2\right)$$

4.2　实验比较

我们在 matlab 中实现了算法 1.由于没有类似的算法生成规范基,我们比较了规范基与 α 极大决策蕴涵集在抑制冗余决策蕴涵方面的差异.

因为即使小型的形式背景也可能生成大量的形式概念,而每个概念至少会有 1 个以上的最小生成子,因此我们选择了 4 个小规模的数据集进行验证性实验.所选择的数据集全部来源于 UCI※.我们对这些数据进行了预处理,包括移除缺失值,对连续值进行归一化,并根据阈值(0.5)生成相应的形式背景,其摘要信息如表 4 所示.

表 4　数据

数据	对象数	属性数
wine	178	14
cleverland	297	14
housing	504	14
ionosphere	351	17

对于每个数据集,我们生成 7 个决策背景.生成的方法是:根据属性数依次选择条件属性,并将剩余属性作为决策属性,例如,wine 有 14 个属性,我们分别依次选取了前 1,3,5,7,9,11,13 个条件属性,并将剩余属性作为决策属性.对于每个决策背景,我们生成了条件子背景的形式概念和最小生成子,并统计其个数,同时还生成决策背景的 α 极大决策蕴涵集和决策蕴涵规范基,统计其所含决策蕴涵数,以及规范基相比 α 极大决

※　http://archive.ics.uci.edu/ml/datasets.html

策蕴涵集减少的冗余决策蕴涵比例. 各个数据集的计算结果如表 5 – 8 所示,其中 # C 表示条件属性数, # Con 表示概念数, # M 表示最小生成子数, # A 表示 α 极大决策蕴涵数, # B 表示规范基所含决策蕴涵数.

表 5 wine 中规范基与 α 极大决策蕴涵的比较

数据	# C	# Con	# M	# A	# B	约简率
wine	1	2	2	2	1	50%
wine	3	8	8	8	1	87.5%
wine	5	32	32	32	2	93.8%
wine	7	68	74	70	18	74.3%
wine	9	117	135	124	34	72.6%
wine	11	168	223	197	42	78.7%
wine	13	260	375	297	58	80.5%

表 6 cleverland 中规范基与 α 极大决策蕴涵的比较

数据	# C	# Con	# M	# A	# B	约简率
cleverland	1	2	2	2	1	50%
cleverland	3	8	8	8	1	87.5%
cleverland	5	19	20	20	4	80%
cleverland	7	65	69	66	11	83.3%
cleverland	9	195	204	198	17	91.4%
cleverland	11	341	392	364	47	87.1%
cleverland	13	761	909	828	35	95.8%

表 7 housing 中规范基与 α 极大决策蕴涵的比较

数据	# C	# Con	# M	# A	# B	约简率
housing	1	2	2	2	2	0
housing	3	4	6	5	5	0
housing	5	10	14	11	10	9.1%
housing	7	29	34	30	12	60%
housing	9	40	52	41	22	46.3%
housing	11	73	98	81	27	66.7%
housing	13	147	185	160	24	85%

表 8 ionosphere 中规范基与 α 极大决策蕴涵的比较

数据	# C	# Con	# M	# A	# B	约简率
ionosphere	1	2	2	2	1	50%
ionosphere	4	4	5	5	2	60%
ionosphere	6	30	31	31	2	93.6%
ionosphere	9	114	115	115	6	94.8%
ionosphere	11	771	797	797	10	98.8%
ionosphere	14	1567	1726	1726	18	99%
ionosphere	16	6023	7156	7156	2	99.8%

从表 5 ~ 8 可以看出:(1) 对于所选的 4 个数据集来说,生成的最小生成子要多于形式概念,但差距不是很大;但也可以看出,随着条件属性的增加,二者之间的差距有加大的趋势;(2) α 极大决策蕴涵的个数与最小生成子的个数相差不大,这说明 α 极大决策蕴涵在实践中并不能很好地抑制冗余决策蕴涵的生成. 尤其对数据集 ionosphere 来说,α 极大决策蕴涵几乎没有效果;(3) 规范基对冗余决策蕴涵的抑制效果非常明显,尤其对于数据集 ionosphere,规范基所含的冗余决策蕴涵比 α 极大决策蕴涵集减少 90% 以上;(4) 大致说来,最小生成子越多,规范基的抑制效果就越好. 比较 4 个数据库,ionosphere 中含有的最小生成子要远大于其余 3 个数据集,因此规范基的抑制效果也远好于其余 3 个数据集;而 cleverland 和 wine 所含的最小生成子数大于 housing,因此其效果也较好;而 housing 数据集中含有的最小生成子数最少,因而其效果也最不稳定;(5) 实验也说明了算法 1 在生成规范基方面的效率不是很好,尤其对于数据集 ionosphere 来说,生成的最小生成子中有 90% 上都是非决策前提,因此如何不通过生成最小生成子来直接得到决策前提,是一个值得研究的问题.

5 结束语

本文主要给出一个决策蕴涵下的规范基. 该规范基由决策前提作为该决策蕴涵集的前提,由决策蕴涵相对于决策子背景的闭包作为该决策蕴涵集的结论. 我们还证明了该决策蕴涵规范基是完备的,无冗余的,并且在所有相对于给定形式背景完备的决策蕴涵集中,其决策蕴涵数是最小的,还给出了一个基于最小生成子的蕴涵基的生成算法.

正如我们所看到的,决策前提是采用递归方式定义的,因此,如何更有效地计算决策前提是我们下一步的工作. 对算法 1 进行改进的方向包括:(1)进一步研究决策前提的性质,考虑如何直接生成决策前提;(2)减少生成非决策前提的最小生成子个数;(3)采用有效的算法生成最小生成子(参考文献[15,16]).

参考文献

[1] R Wille. Restructuring lattice theory: an approach based on hierarchies of concepts[A]. Ordered Sets[C]. Dordrecht: Reidel, 1982. 445 – 470.

[2] Ganter B, Wille R. Formal Concept Analysis: Mathematical Foundations[M]. New York: Springer, 1999. 17 – 36, 79 – 90.

[3] Qu K S, Zhai Y H, Liang J Y, Chen M. Study of decision implications based on formal concept analysis[J]. International Journal of General Systems, 2007, 36(2): 147 – 156.

[4] 智慧来, 智东杰, 刘宗田. 概念格合并原理与算法[J]. 电子学报, 2010, 38(2): 455 – 460.

Zhi Hui-lai, Zhi Dong-jie, Liu Zong-tian. Theory and algorithm of concept lattice union[J]. Acta Electronica Sinica, 2010, 38(2):455 – 460. (in Chinese)

[5] 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, 2005, 35(06):628 – 639.

Zhang W X, Wei L, Qi J J. Attribute reduction theory and approach to concept lattice[J]. Science in China Serie E: Information Sciences, 2005, 35(06):628 – 639. (in Chinese)

[6] 曲开社,翟岩慧. 偏序集、包含度与形式概念分析[J]. 计算机学报, 2006, 29(2):219 – 226.

Qu K S, Zhai Y H. Posets, inclusion degree theory and FCA[J]. Chinese Journal of Computers, 2006, 29(2):219 – 226. (in Chinese)

[7] Xia H, Chen Y, Gao H, Li Z, Chen Y. Concept lattice-based semantic web service matchmaking[A]. Communication Software and Networks[C]. Los Alamitos: IEEE Computer Society, 2010. 439 – 443.

[8] Carpineto C, Romano G. Concept Data Analysis: Theory and Applications[M]. England: John Wiley & Sons, 2004. 85 – 104.

[9] Tilley T, Cole R, Becker P, Eklund P W. A survey of formal concept analysis support for software engineering activities[A]. Formal Concept Analysis, Foundations and Applications[C]. Berlin: Springer, 2005, 3626. 250 – 271.

[10] Roth C, Obiedkov S A, Kourie D G. Towards concise representation for taxonomies of epistemic communities[A]. Fourth International Conference on Concept Lattices and Their Applications[C]. Berlin: Springer, 2006, 4923. 240 – 255.

[11] Zhai Y H, Li D Y, Qu K S. Fuzzy decision implications[J]. Knowledge-Based Systems, 2013, 37:230 – 236.

[12] 马垣,张学东,迟呈英. 紧致依赖与内涵亏值[J]. 软件学报, 2011, 22(05):962 – 972.

Ma Y, Zhang X D, Chi C Y. Compact dependencies and intent waned values[J]. Journal of Software, 2011, 22(05):962 – 972. (in Chinese)

[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, 2008, 38(02):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 Ser: E Information Sciences, 2008, 38(02):195 – 208. (in Chinese)

[14] G Stumme, R Taouil, Y Bastide, N Pasquier, L Lakhal. Fast computation of concept lattices using data mining techniques[A]. Proceedings of the 7th International Workshop on Knowledge Representation meets Databases[C]. Berlin: Springer, 2000. 129 – 139.

[15] C Frambourg, P Valtchev, R Godin. Merge-based computation of minimal generators[A]. Conceptual Structures: Common Semantics for Sharing Knowledge[C]. Berlin: Springer-Verlag, 2005. 181 – 194.

[16] G Dong, C Jiang, J Pei, J Li, L Wong, Mining succinct systems of minimal generators of formal concepts[A]. Database Systems for Advanced Applications[C]. Berlin: Springer-Verlag, 2005. 175 – 187.

作者简介

翟岩慧　男, 1981 年 2 月出生, 山西文水人, 讲师. 2003 年、2006 年在山西大学计算机与信息技术学院分别获得工学学士和工学硕士学位. 现为山西大学计算机与信息技术学院在读博士, 主要研究方向为概念格和粗糙集.

李德玉　男, 1965 年 10 月出生, 山西临汾人. 教授, 博士生导师. 现为山西大学计算机与信息技术学院院长, 主要研究方向为粗糙集、多标签学习等.

E-mail: lidysxu@163.com

曲开社　男, 1954 年 5 月出生, 山西运城人. 教授, 主要研究方向为概念格和粗糙集等.