

# 基于决策背景的决策蕴涵规范基

贺建英

(四川文理学院计算机学院 达州 635000)

**摘要** 主要给出决策背景下的决策蕴涵基,通过不确定的分组阈值对决策背景进行分组,找到每个分块的决策背景下的决策蕴涵规范基,并证明了每个分块上的决策蕴涵集是完备、无冗余和最优的;也证明了每个决策背景分块上的决策蕴涵基的集合在整个决策背景上也是完备、无冗余和最优的。在生成决策背景的决策蕴涵规范基的过程中,优化了最小生成子算法;同时也给出了基于决策背景的决策蕴涵规范基的生成算法。实验表明,通过分组策略和算法的优化改进,可以很好地抑制冗余决策蕴涵规范基的生成,使决策蕴涵规范基更为有效和紧凑。

**关键词** 阈值,决策背景,决策蕴涵,规范基

**中图分类号** TP182

**文献标识码** A

**DOI** 10.11896/j.issn.1002-137X.2017.08.050

## Canonical Basis Based on Decision Implication of Decision Context

HE Jian-ying

(College of Computer, Sichuan University of Arts and Science, Dazhou 635000, China)

**Abstract** The decision criterion of the decision backdrop was introduced primarily. The decision backdrop is categorized into groups by the uncertainty threshold value which can lead to findings the decision criterion subordinating to the decision backdrop, and it's further proved that every branch-patching decision implication group is complete, non-redundant and the most superior, and the set of decision criterion on the decision backdrop patch is complete, non-redundant and the most superior on the whole decision backdrop. The minimum spanning piece algorithm was optimized while generating the decision backdrop criterion, at the same time, the derived algorithm based on the decision backdrop was generated. Experiments show that improving and optimizing the grouping tactics and algorithm can greatly suppress the formation of the redundant decision criterion more efficient and sturdy.

**Keywords** Threshold, Decision making background, Decision implication, Canonical basis

## 1 引言

概念格由德国的 Wille R 教授于 1982 年提出,概念格的每一个节点是一个形式概念。形式概念分析(Formal Concept Analysis, FCA)是一种通过形式背景建立概念格来对数据进行分析 and 提取规则的一个强有力的工具。它可以根据形式背景(数据集)中对象和属性之间的相互关系(二元关系)来得到数据与数据之间隐含的概念并建立它们之间的代数结构,即概念格。目前, FCA 已被广泛地应用到机器学习、数字图书馆、软件工程、信息检索等相关领域,主要从概念格的构建、决策规则的提取、模糊概念格、FCA 的其他扩展、属性约束和应用等方面对其进行研究。传统的形式概念分析缺乏处理缺值的能力;当数据规模较大时,概念格会生成大量节点而难以分析;在概念格上提取的蕴涵规则数目庞大,无法满足用户需求。针对以上问题,设计出有效的近似模糊算法,在不完备决策背景上进行近似概念的构建<sup>[1]</sup>,通过 K-Means 算法简化形式背景<sup>[2]</sup>,增加决策蕴涵的前提或者减少决策蕴涵的结

果,从而推导出新的决策蕴涵等相关研究显得尤为重要。

蕴涵在形式概念分析中是指 IF-THEN 的规则,形如  $A \rightarrow B$  的公式表明从  $A$  可以推导出  $B$ ,若有任意的属性集  $A$ ,则必有属性集  $B$  的存在。故,蕴涵是指在给定的数据集中规则  $A \rightarrow B$  是成立的,该集合中也没有与该规则相矛盾的数据存在,在数据集中蕴涵可以组成一个蕴涵集。该集合可以是无冗余且完备的;无冗余是指该集合中没有冗余蕴涵,判断是否存在某一个或几个蕴涵可以从该蕴涵集的其他蕴涵集中推导出来,即表明该蕴涵是否紧致;完备性是指在不损失形式背景信息的前提下生成一个完整的数据集的表示。决策蕴涵可以从逻辑角度和数据角度进行研究;逻辑角度是为给定的条件集合和决策集合定义决策蕴涵;数据角度首先定义决策背景,然后再根据决策背景来定义决策蕴涵。目前对决策蕴涵的研究主要包括文献[3-9]。文献[3]主要给出了  $\alpha$  规则,并分析讨论了  $\alpha$  规则的合理性。文献[4]先给出决策前提的概念,然后以决策前提为前提,生成决策蕴涵集。这两种研究都证明了生成决策蕴涵集的完备性和无冗余性,并且给出了一种基

到稿日期:2016-06-11 返修日期:2016-09-30 本文受四川省教育厅项目:基于 Hadoop 云平台的电子政务应用研究—以达州市政府机构编制实名制管理系统为例(15ZB0324),四川文理学院智能计算与物联网工程技术中心资助项目资助。

贺建英(1979—),女,硕士,讲师,主要研究方向为软件工程、数据挖掘、概念格等,E-mail:jiabei520@163.com。

于最小生成子的规则集成算法。

本文通过研究已有策略发现,对于 IF-THEN 这一蕴涵规则,即便是较小规模的决策背景也会产生大量的决策蕴涵。文献[3-4]已经给出了较为紧凑、无冗余且完备的决策蕴涵集。然而通过分析发现,在文献[3]中,对于  $\alpha$  推理规则,如果存在另一个合理的推理规则  $\gamma$ ,并且能与  $\alpha$  规则一起构造新的推理规则集,那么  $\alpha$  极大决策规则就存在冗余,可进一步精简。本文将在已有策略的基础上研究另一个更为紧凑的决策蕴涵规范基,在文献[10]的基础上做进一步改进,从决策背景的分块选取、最小生成子及决策蕴涵的生成等方面进行优化,并证明该决策蕴涵基是完备且无冗余的。

本文还将借助数据挖掘中的聚类思想对决策背景进行抽样划分,在划分中按条件属性的不同对取值个数  $\lambda$  进行分块,同一个决定背景根据  $\lambda$  的不同取值得到不同的分块数,从而找到每个划分区域的最小决策蕴涵基,进而使得到的决策蕴涵是整个决策背景中的最小决策蕴涵基,并证明了决策蕴涵基在所有的决策蕴涵基中所含决策蕴涵数最小。分组实验结果表明,取得的决策蕴涵规范基是在决策背景下建立起来的规范基,是较优的决策规范基;算法的改进具有一定的参考价值和意义。

## 2 相关概念及推论

本节将直接给出格论、完备格、决策背景及决策蕴涵等相关概念。

**定义 1**<sup>[11]</sup> 假定  $M$  为一个集合,  $a, b, c \in M$ ,  $M$  的一个偏序是一个二元关系  $\leq$ , 该偏序满足自反性( $a \leq a$ )、反对称性(若  $a \leq b$  且  $b \leq a$ , 则  $a = b$ )、传递性(若  $a \leq b$  且  $b \leq c$ , 则  $a \leq c$ ), 那么称  $M$  为偏序集, 记为:  $(M, \leq)$ 。

**定义 2**<sup>[12]</sup> 称  $(M, \leq)$  为完备格, 则设  $(M, \leq)$  为一个偏序集, 且  $B \subseteq M, c \in M$ , 对任意的  $d \in B$  都有  $c \leq d$ , 则称  $c$  为  $B$  的一个下界, 若  $B$  的下界集合中存在一个最大元, 那么称该最大元为  $B$  的上确界, 记作:  $\bigwedge B$ ; 同样能找到  $B$  的一个上界, 若  $B$  的上界集合中存在一个最小元, 那么称该最小元为  $B$  的下确界, 记作:  $\bigvee B$ 。若对任意的  $m, n \in M$ , 均存在  $m \wedge n$  和  $m \vee n$ , 则称  $(M, \leq)$  为格。若对  $M$  的任意子集  $B$  都存在  $\bigwedge B$  和  $\bigvee B$ , 则称  $(M, \leq)$  为完备格。

**定义 3**<sup>[13]</sup> 三元组  $(O, A, R)$  是一个决策背景, 其中  $O = \{o_1, o_2, o_3, \dots, o_n\}$  代表对象集,  $A = \{a_1, a_2, a_3, \dots, a_n\}$  代表属性集,  $A \subseteq C \cup D$ ,  $C$  为条件属性集,  $D$  为决策属性集, 且  $C \cap D = \emptyset$ ,  $R$  为属性  $A$  上的二元关系, 其中  $R_c$  为  $O$  和  $C$  上的二元关系:  $R_c \subseteq O \times C$ ;  $R_d$  为决策二元关系  $R_d \subseteq O \times D$ 。

例 1 决策背景的二维表表示方式如表 1 所列。

表 1 形式背景表

| O     | C     |       |       |       |       | D     |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
|       | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $d_1$ | $d_2$ |
| $o_1$ | 0     | 1     | 0     | 1     | 0     | 1     | 0     |
| $o_2$ | 1     | 0     | 1     | 0     | 1     | 0     | 1     |
| $o_3$ | 0     | 1     | 1     | 1     | 1     | 1     | 0     |
| $o_4$ | 1     | 1     | 0     | 1     | 0     | 1     | 0     |
| $o_5$ | 1     | 1     | 0     | 0     | 1     | 0     | 1     |

该表中的对象为  $\{o_1, o_2, o_3, o_4, o_5\}$ , 条件属性为  $\{c_1, c_2,$

$c_3, c_4, c_5\}$ , 决策属性为  $\{d_1, d_2\}$ 。在形式概念分析中, 从该形式背景可以得到两方面的数据: 1) 概念格表示, 以代数格的方式保存数据; 2) 蕴涵, 以属性逻辑的方式保存数据。

从表 1 可以看出, 在条件集相同或相似的情况下可能有相同的蕴涵。相对地, 如果决策背景越小, 则得到的决策蕴涵就越小, 即得到的最小决策蕴涵基也就越小。根据数据挖掘的抽样划分聚类方式对决策背景作进一步划分, 得到如下定义。

**定义 4** 假设决策背景  $\{O, C, D\}$ , 可将由  $N$  个对象集  $\{o_1, o_2, o_3, \dots, o_n\}$ 、条件属性  $C$  及决策属性  $D$  构成的二维表  $\{c_1, c_2, c_3, \dots, c_i, d_1, d_2, d_3, \dots, d_j\}$  划分为  $K$  个大小不同的子二维表, 其中  $K < N$ 。

分块的划分取决于相似度的划分, 这里根据决策背景的数据量, 设相似度为  $\lambda$ , 其中  $0 < \lambda \leq N/2$ 。特别地, 当决策背景相似度极高时, 得到  $K=1$ ; 反之, 当决策背景的相似背景极低时, 得到  $K=N$ , 此时已无研究的意义; 根据定义 4, 将得到划分的子二维的数量  $K$  的范围定义为  $1 \leq K \leq N/2$ 。

**定义 5**<sup>[13]</sup> 对于每个划分分区中对象集合  $O'$  上的集合  $C'$  和  $D'$ , 决策蕴涵  $B_1' \rightarrow B_2'$  都可以从每个子划分块的决策蕴涵集  $S'$  中语义导出, 假设对于任意的  $C' \cup D'$  子集合  $L'$ ,  $L' \vdash S'$  蕴涵  $L' \vdash B_1' \rightarrow B_2'$ , 记作:  $S' \vdash B_1' \rightarrow B_2'$ 。决策蕴涵集  $S'$  是无冗余的, 若对于任意的蕴涵  $B_1' \rightarrow B_2' \in S'$ ,  $S' \setminus \{B_1' \rightarrow B_2'\}$  不能推导出  $B_1' \rightarrow B_2'$ , 则决策蕴涵集  $S'$  是封闭的; 若决策蕴涵集  $S'$  语义中导出的决策蕴涵都包含在  $S'$  中, 则封闭的决策蕴涵集  $S'$  是完备的。故对于  $S'$  的决策蕴涵集的集合也是封闭且完备的。

**定义 6** 设  $(O, A, R)$  是一个决策背景, 其中  $A \subseteq C \cup D$ ,  $R \subseteq R_c \cup R_d$ , 且  $B_{c1} \subseteq C, B_{d1} \subseteq D$ , 那么当  $B_{c1} \subseteq B_{d1}$  成立时,  $B_{c1} \rightarrow B_{d1}$  是一个决策蕴涵。

根据定义 5、定义 6 及决策背景, 可以得到如下推论。

**推论 1** 若  $L' \vdash B_1' \rightarrow B_2'$  且  $B_1' \rightarrow B_2'$  是决策背景的一个决策蕴涵, 则  $L' \vdash B_1' \rightarrow B_2'$  成立, 进而  $L \vdash B_1' \rightarrow B_2'$  也成立。

**推论 2** 若决策蕴涵集  $S'$  在  $S'$  上是封闭且完备的, 则决策蕴涵集的集合  $S$  在  $S$  上也是封闭且完备的。决策蕴涵集的集合  $S$  是  $S'$  的并的集合, 如下式所示:

$$\bigvee_{L \in C \cup D} L' \vdash B_i' \rightarrow B_j'$$

**证明 1:** 假设  $B_1' \rightarrow B_2'$  是决策背景的一个决策蕴涵, 由定义 6 知  $B_2' \subseteq B_1'$ , 如果  $B_1' \subseteq L' \cap C'$ , 那么由  $L' \vdash B_1' \rightarrow B_1'^{C'D'}$  可知  $B_1'^{C'D'} \subseteq L' \cap D'$ , 故  $B_2' \subseteq B_1'^{C'D'} \subseteq L' \cap D'$ 。从而得到  $L' \vdash B_1' \rightarrow B_2'$ 。又因  $B_1' \subseteq B_1, B_2' \subseteq B_2$  且  $L' \subseteq L$ , 故  $L \vdash B_1' \rightarrow B_2'$  成立。

**证明 2:** 假设  $B_1' \rightarrow B_2'$  是决策背景的一个决策蕴涵, 并且  $L' \vdash B_1' \rightarrow B_2'$  成立, 那么在每个划分的子决策背景集中都有  $L' \vdash B_1' \rightarrow B_2'$  成立。而  $L' \subseteq L$ , 故  $L \vdash B_1' \rightarrow B_2'$ , 且对任意的  $B_i' \rightarrow B_j'$  都能满足  $L \vdash B_i' \rightarrow B_j'$ , 其中  $B_i' \subseteq B_1, B_j' \subseteq B_2$ 。故满足  $L \vdash B_1' \rightarrow B_2'$  集合的  $\cup$  运算的关系也成立。

**定义 7** 设一决策背景为  $S$ , 对于决策背景中的集合  $X \subseteq O, Y_1 \subseteq C, Y_2 \subseteq D$ , 记它们之间的关系为:

$$X^C = \{n \in C \mid \langle o, n \rangle \in R_c, \forall o \in X\}$$

$$X^D = \{n \in D \mid \langle o, n \rangle \in R_d, \forall o \in X\}$$

$$Y_1^c = \{o \in O \mid \{(o, n) \in R_c, \forall n \in Y_1\}\}$$

$$Y_2^c = \{o \in O \mid \{(o, n) \in R_d, \forall n \in Y_2\}\}$$

定义 8<sup>[15]</sup> 设  $R$  为一个封闭集合,  $M, N \subseteq R$ , 假如  $\varphi M = R$  成立, 则  $M$  为  $R$  的生成子; 进一步, 若  $\forall N \subset M$ , 则  $\varphi N = \varphi M = R$  都成立, 此时称  $M$  为  $R$  的最小生成子。

对于决策背景  $(O, A, R)$ , 若  $A_i \subseteq A$  (其中  $i=1, 2, 3, \dots, n$ ) 是属性子集,  $C \subseteq A_i$  成立, 则  $A_i$  的生成子  $C$  是满足条件  $C' = A_i'$  的  $A_i$  的一个子集,  $A_i$  的最小生成子  $C$  是使  $A_i' = C'$  的  $C$  子集的最小子集。

### 3 基于决策背景的决策蕴涵规范基

借助分块决策背景生成对应的决策规范基, 本节将证明基于决策背景的决策规范基是无冗余、完备且最优的, 同时证明分块后生成的决策蕴涵基构成的集合也是完备、无冗余且最优的。

根据定义 5 可知, 对于每个划分分区上对象集合  $O'$  上的集合  $C'$  和  $D'$ ,  $C \subseteq C'$  和  $D \subseteq D'$  满足以下两点: 1)  $C'$  相对于  $C'^{CD'}$  是最小的, 如果  $C'' \subset C'$ , 那么  $C''^{CD'} \subset C'^{CD'}$ ; 2)  $C'$  是恰当的, 即  $C'^{CD'} \neq \bigcup \{B_1'^{CD'} \mid B_1' \subset C'\}$ , 其中  $B_1'$  是决策前提。

对于划分块的任意子集  $C'' \subset C$ , 尤其对于决策前提可以得到  $C'' \rightarrow C'^{CD'}$ , 而  $C'^{CD'}$  为  $C'$  相对于决策前提子背景的闭包, 则此决策蕴涵必为决策背景集  $S$  的一个决策蕴涵<sup>[13]</sup>。进一步验证每个划分子集上的决策前提组成的决策蕴涵集合是完备的, 进而得到每个子集的决策蕴涵组成的  $\bigcup$  集合在决策背景  $S$  上是完备无冗余的。值得注意的是, 决策蕴涵中条件属性  $C'$  和决策属性  $D'$  中所组成的蕴含不一定是决策蕴涵, 因为决策蕴涵更为紧凑。

在表 1 中对其进行分块划分, 取  $\lambda=2$ , 即有 2 个相同的属性, 则把两个对象划分到一个分块区域里, 得到的子二维表如表 2 和表 3 所列, 得到的决策前提分别如表 4 和表 5 所列。

表 2 形式背景分块 1

| O     | A     |       |       |       |       | D     |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
|       | C     |       |       |       |       | $d_1$ | $d_2$ |
|       | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ |       |       |
| $o_1$ | 0     | 1     | 0     | 1     | 0     | 1     | 0     |
| $o_2$ | 0     | 1     | 1     | 1     | 1     | 1     | 0     |
| $o_3$ | 1     | 1     | 0     | 1     | 0     | 1     | 0     |

表 3 形式背景分块 2

| O     | A     |       |       |       |       | D     |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
|       | C     |       |       |       |       | $d_1$ | $d_2$ |
|       | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ |       |       |
| $o_1$ | 1     | 0     | 1     | 0     | 1     | 0     | 1     |
| $o_2$ | 1     | 1     | 0     | 0     | 1     | 0     | 1     |

表 4 表 2 中的决策前提

|             |                |                |                |
|-------------|----------------|----------------|----------------|
| $\emptyset$ | $\{c_2, c_4\}$ | $\{c_3, c_4\}$ | $\{c_1, c_5\}$ |
| $\{c_1\}$   | $\{c_2, c_3\}$ | $\{c_4, c_5\}$ | $\{c_2, c_5\}$ |

表 5 表 3 中的决策前提

|             |                |                |
|-------------|----------------|----------------|
| $\emptyset$ | $\{c_1, c_5\}$ | $\{c_3, c_5\}$ |
| $\{c_1\}$   | $\{c_2, c_5\}$ |                |

根据上述分析, 得出如下定理。

定理 1 集合  $S' = \{C' \rightarrow C'^{CD'} \mid C' \text{ 是决策前提}\}$  是完备且

无冗余的, 则称其为决策背景的决策蕴涵规范基。当  $S' \subseteq S$  时, 各个子集  $S'$  的决策蕴涵基构成的集合是集合  $S$  上的决策背景的决策蕴涵规范基。

证明: 假设  $L' \vdash B_1' \rightarrow B_2'$  是决策背景  $K'$  的一个决策蕴涵, 根据完备的条件应满足  $L' \vdash B_1' \rightarrow B_2'$ , 则只需证明它成立即可。由推论 1 知其只需满足  $L' \vdash B_1' \rightarrow B_1'^{CD'}$ , 即  $L \vdash B_1 \rightarrow B_1^{CD}$  成立即可。若要证明其成立, 则可分为 3 种情况进行讨论。

(1)  $B_1' \rightarrow B_1'^{CD'}$  中  $B_1'$  相对于  $B_1'^{CD'}$  是最小的, 且是恰当的, 此时  $B_1'$  是决策前提, 并且  $B_1' \rightarrow B_1'^{CD'} \in S'$ , 因  $L' \vdash S'$ , 故  $L' \vdash B_1' \rightarrow B_1'^{CD'}$  成立, 则  $L \vdash B_1 \rightarrow B_1^{CD}$ 。

(2)  $B_1' \rightarrow B_1'^{CD'}$  中  $B_1'$  相对于  $B_1'^{CD'}$  是最小的, 且并不是恰当的, 此时存在  $B_1'^{CD'} = \bigcup \{B_3'^{CD'} \mid B_3' \text{ 是决策前提}, B_3' \subset B_1'\}$ , 设  $B_1' \subseteq L'$ , 因  $L' \vdash S'$ , 对于  $B_3' \subset B_1'$  的所有决策前提是  $B_3'$ , 有  $L' \vdash B_3' \rightarrow B_3'^{CD'}$ , 且因  $B_3' \subset B_1' \subseteq L'$ , 得  $B_3'^{CD'} \subset L'$ , 故  $B_1'^{CD'} = \bigcup \{B_3'^{CD'}\} \subseteq L'$ , 进而使得  $L' \vdash B_1' \rightarrow B_1'^{CD'}$  成立, 则  $L \vdash B_1 \rightarrow B_1^{CD}$ 。

(3)  $B_1' \rightarrow B_1'^{CD'}$  中  $B_1'$  相对于  $B_1'^{CD'}$  不是最小的, 此时只要  $B_3'^{CD'} = B_1'^{CD'}$  的最小集合  $B_3' \subset B_1'$ , 根据情况 (2) 可得  $L' \vdash B_3' \rightarrow B_3'^{CD'}$ , 从而使得  $L' \vdash B_1' \rightarrow B_1'^{CD'}$  成立, 故  $L \vdash B_1 \rightarrow B_1^{CD}$ 。

以上 3 种情况讨论了在决策背景的每个子集  $S'$  中的决策蕴涵基是完备的。那么对于各个子集  $S'$  的决策蕴涵基构成的集合在  $S$  上对于决策背景也是完备的。假设对于每个决策背景的子集  $S'$  的决策蕴涵是无冗余的, 那么对于  $S$  的决策蕴涵的并集构成的决策蕴涵集合是否是无冗余的呢? 首先证明子集  $S'$  的决策蕴涵是无冗余的。

若要证明  $S'$  是无冗余的, 就需要证明  $B_1' \rightarrow B_1'^{CD'} \in S'$ , 有  $S' \setminus \{B_1' \rightarrow B_1'^{CD'}\} \not\vdash B_1' \rightarrow B_1'^{CD'}$ 。根据文献[10]可知, 先找到一个集合  $L''$ , 满足  $L'' \vdash S' \setminus \{B_1' \rightarrow B_1'^{CD'}\}$  且  $L'' \not\vdash B_1' \rightarrow B_1'^{CD'}$  即可。

设  $L'' = B_1' \cup \bigcup \{B_3'^{CD'} \mid B_3' \text{ 是决策前提}, B_3' \subset B_1'\}$ , 这里断言:

(1)  $L'' \vdash S' \setminus \{B_1' \rightarrow B_1'^{CD'}\}$ 。设  $B_3' \rightarrow B_3'^{CD'} \in S' \setminus \{B_1' \rightarrow B_1'^{CD'}\}$ , 如果  $\{B_3' \subseteq L'' \cap C' = B_1'\}$ , 则  $B_3' \subset B_1'$ , 显然  $\{B_3' \neq B_1'\}$ , 故  $B_3'^{CD'} \subseteq \bigcup_{B' \subset B_1'} B'^{CD'} = L'' \cap D'$ 。

(2)  $L'' \not\vdash B_1' \rightarrow B_1'^{CD'}$ , 则  $S' \setminus \{B_1' \rightarrow B_1'^{CD'}\} \not\vdash B_1' \rightarrow B_1'^{CD'}$ 。因为  $B_1'$  为决策前提, 那么有  $B_1'^{CD'} \neq \bigcup \{B_3'^{CD'} \mid B_3' \text{ 是决策前提}, B_3' \subset B_1'\} = keys$ , 则存在  $K \in B_1'^{CD'} \setminus keys$  成立。又因  $B_1' \subseteq L'' \cap C' = B_1'$ , 但  $keys \in B_1'^{CD'} \not\subseteq L'' \cap D' = keys$ , 那么  $L'' \not\vdash B_1' \rightarrow B_1'^{CD'}$  成立。

由以上两点假设成立的推理可得, 断言  $S'$  是无冗余的。对于每个  $S'$  的决策蕴涵基的集合  $S$ , 其本身是无相同的决策蕴涵基, 那么只需要证明任意的  $B_1' \rightarrow B_2'$  都不能由其他的决策蕴涵推导出来, 即在  $S$  中  $B_1 \rightarrow B_2$  都不能由  $S'$  的决策蕴涵推导出来。则该结果是无冗余的决策蕴涵基。

文献[10]已证明决策蕴涵规范基比已有的  $\alpha$  极大决策蕴涵集更为紧凑, 现将进一步证明决策蕴涵基是最优的。

定理 2 设  $S'$  的一个决策蕴涵集为  $L'$ ,  $C'$  为决策前提, 那么  $C' \rightarrow C'^{CD'} \in L'$  或者  $L' \not\vdash C' \rightarrow C'^{CD'}$  成立。则对于  $S'$  的集合

$S$  的蕴涵集  $C \rightarrow C^D \in L$  或者  $L \vdash C \rightarrow C^D$  也成立。

证明:对于  $L'$  任意的蕴涵  $B_1' \rightarrow B_2'$ , 存在  $B_1' \rightarrow B_1'^{CD}$  成立, 对于蕴涵集  $L'$ , 则能得到  $\overline{L'} \vdash C' \rightarrow C'^{CD}$ , 则  $L' \vdash C' \rightarrow C'^{CD}$ 。现只需要证明当决策蕴涵集  $T' \vdash \overline{L'}$  成立时,  $T' \vdash C' \rightarrow C'^{CD}$  成立即可。

设  $C' \subseteq T'$ ,  $C' \rightarrow C'^{CD} \in L'$  且  $T' \vdash \overline{L'}$ , 因  $C'$  为决策前提, 故  $C'^{CD} \supseteq \bigcup \{B_1'^{CD} \mid B_1' \text{ 为决策前提}, B_1' \subset C'\}$ , 根据定理 1 可以断言:  $\beta = \bigcup \{B_1'^{CD} \mid B_1' \text{ 为决策前提}, B_1' \subset C'\} \supseteq \bigcup \{B_1'^{CD} \mid B_1' \rightarrow B_1'^{CD} \in \overline{L'}, B_1' \subset C'\} = \text{Vector}$  是成立的。那么将存在  $C'^{CD} \supseteq \text{Vector}$ 。由定理 2 的第二种情况的证明可知  $keys \in C'^{CD} \setminus \text{Vector}$  成立。同理根据定理 2 的证明可得  $T' \vdash \overline{L'}$ , 但  $\overline{L'} \not\vdash C' \rightarrow C'^{CD} \in L'$ , 由前文可知  $\overline{L'} \vdash C' \rightarrow C'^{CD}$ , 故  $L' \vdash C' \rightarrow C'^{CD}$ , 因此对于  $S$  存在  $L \vdash C \rightarrow C^D$ 。

同理, 在此分 3 种情况来证明  $B_1' \rightarrow B_1'^{CD} \in L'$  成立。

(1) 当  $B_1'$  相对于  $B_1'^{CD}$  是最小的, 且是恰当的。  $B_1'$  是决策前提, 那么断言  $B_1'^{CD} \subseteq \beta$  是成立的。

(2) 当  $B_1'$  相对于  $B_1'^{CD}$  是最小的, 但不是恰当的。因有  $B_1'^{CD} = \bigcup \{B_3'^{CD} \mid B_3' \text{ 为决策前提}, B_3' \subset B_1'\}$ , 则  $B_1'^{CD}$  可以由  $\beta$  的子集合并而成。故可以断言  $B_1'^{CD} \subseteq \beta$  成立。

(3) 当  $B_1'$  相对于  $B_1'^{CD}$  不是最小的, 且不是恰当的。假设能找一个最小集  $B_3' \subset B_1'$ , 能满足  $B_3'^{CD} = B_1'^{CD}$ , 此时由上一步的证明可以断言  $B_1'^{CD} \subseteq \beta$  成立。

通过以上定理的提出及证明, 可以推导出如下的推论。

**推论 3** 在每个  $S'$  的完备集中, 其决策蕴涵数目是最小的, 那么由每个  $S'$  的集合组成的决策蕴涵集对于集合  $S$  的决策蕴涵数目也是最小的。

证明: 由定理 2 可知, 对于任意的完备集  $L'$  和决策蕴涵前提  $C'$ , 假设  $C' \rightarrow C'^{CD} \notin L'$ , 那么  $L' \not\vdash C' \rightarrow C'^{CD}$ , 此时有  $C' \rightarrow C'^{CD} \in L'$ , 因  $C'$  是任意的, 那么  $S'$  中任意决策蕴涵  $\subseteq L'$ , 故  $S'$  中的决策蕴涵数目是最小的; 又因  $S' \subseteq S$ , 故  $S'$  的决策蕴涵组成的  $S$  中的决策蕴涵集合数目也是最小的。

由上述 2 个定理和 3 个推论可知, 基于决策背景的决策蕴涵规范基是完备、无冗余且最优的。

## 4 基于决策背景的决策蕴涵规范基的生成

### 4.1 最小生成子算法的优化

在文献[16]中对最小生成子的生成采用从上到下的深度优先遍历索引树的方式, 对所包含决策属性的最小生成子进行处理从而得到决策规则集。对于决策背景  $(O, A, R)$  中的每一概念  $(X, Y)$ , 决策属性  $d \in D, M \subseteq Y$  为  $Y$  的最小生成子, 计算对应规则的支持度和可信度。进一步处理所有概念  $((X_i, Y_j))$ , 其中  $i, j = 1, 2, 3, \dots, n$  的最小生成子, 从而得到整个封闭集合的决策规则集。本文在此算法的基础上对最小生成子算法进行优化, 对于深度优先遍历的节点, 从父节点按照逐层添加子节点的方式进行最小生成子的判断, 并对其结果分别进行存储。对不能生成最小子集的节点进行标记, 防止该父节点再次访问该子节点时做重复的工作, 从而可以提高生成子的效率。算法 1 为优化后的最小生成子算法。

### 算法 1 getSmallSet()

输入: 决策背景闭集合  $ZC$ , 对象  $OB$ , 属性集  $AT$ , 关系  $RL$

输出: 最小生成子  $\text{small\_Set}$ , 更新后的  $RL$

getSmallSet( $ZC, OB, AT, RL$ )

```

1. small_Set = ∅
2. FOREACH RL DO{
3.   IF P(A, B) ⊆ AT THEN
4.     P = P ∪ OB // 满足最小生成子条件则更新集合 P
5.     IF ZC[i] ∩ AT = ∅ THEN
6.       NEXT ZC[i]; MASK = 0; // 若 ZC[i] 的子节点与 AT 相交为 ∅, 改变其访问标记为 0, 并访问下一个子节点
7.     ELSE ADD(new_SET_C); // 添加新增节点
8.     d_Set = set_C - new_SET_C; // 计算父节点与新增节点的属性值得差集
9.     FOREACH set_C DO // 对 set_C 的所有产生子进行遍历
10.      k_SET = set_C ∩ d_SET // 与 d_Set 的交集产生的子集合
11.      IF (set_c ∩ d_set ≠ ∅) THEN
12.        d_NEW_SET[] // 存储集合
13.        FOREACH d_NEW_SET DO // 遍历 d_NEW_SET 集合
14.          FOREACH d_SET DO // 遍历 d_SET 集合
15.            IF (d_NEW_SET[i] ∪ d_SET[i] && ∀ K_SET ∈ k_SET && k_SET ⊆ d_NEW_SET[i] ∪ {d_SET[i]}) MASK = 1) THEN
16.              IF (d_SET[i+1] ≠ ∅) THEN
17.                CONTINUE;
18.              ELSE BREAK;
19.            } // END FOREACH
20.            IF (d_NEW_SET[i+1] ≠ ∅) THEN
21.              CONTINUE;
22.            ELSE BREAK; } // END FOREACH
23.            IF (ZC[i+1] ≠ ∅) THEN
24.              CONTINUE;
25.            ELSE BREAK; } // END FOREACH
26.          } // END FOREACH
27. RETURN small_Set

```

该算法的优点是对一个封闭集逐层增加子节点, 多次使用并和交的方式找到新增节点的生成子, 使得在整个封闭集中得到的生成子最优且无冗余。

### 4.2 基于决策背景的决策蕴涵规范基的生成

根据文献[10]中的生成决策蕴涵规范基的方法, 借助文献[3]的思想, 利用一种算法来获取  $\alpha$  极大蕴涵集。本文的思想是对已有的决策背景  $S$  进行分块分区讨论, 对其在每一个子背景块  $S'$  中的决策蕴涵进行讨论, 进而推导出在决策蕴涵的集合中也能满足相同的条件。根据文献[10]可以得出如下引理。

**引理 1** 设  $S = (O, A, R)$  为决策背景, 根据  $\lambda$  的划分规则, 得到分块  $S' = (O', C' \cup D', R_C' \cup R_D')$  的决策背景, 当  $B'_{1C}$  的条件子背景  $S'_C = (O', C', R'_C)$  为最小生成子, 且  $B'_{2D} = B'_{1C}{}^{CD}$ , 同时  $B'_{1C} \subset B'_{1C}$ ,  $B'_{1C}{}^{CD} \neq B'_{1C}{}^{CD}$  也成立时,  $B'_{1C} \rightarrow B'_{2D}$  是  $S'$  的  $\alpha$  极大蕴涵集成立。则每个  $S'$  中的  $\alpha$  极大决策蕴涵的集合构成了  $S$  的  $\alpha$  极大蕴涵规范基。

在划分的子背景集  $S'$  中, 每个决策蕴涵都符合以上的引理, 因此在每个  $S'$  中的决策蕴涵都是  $\alpha$  极大决策蕴涵。在每

个  $S'$  中的所有  $\alpha$  极大决策蕴涵集合构成了  $S'$  的决策蕴涵规范基, 因此对于  $S$  的背景前提中每个  $S'$  的集合构成了  $S$  的决策蕴涵规范基。

分两个步骤来证明每个  $S'$  的集合构成的  $S$  的决策蕴涵是否满足条件。

(1) 找出并去掉每个  $S'$  中不满足  $C'^{CD'} \neq \bigcup \{B_1'^{CD'} \mid B_1' \text{ 是决策前提}, B_1' \subset C'\}$  的决策蕴涵。

(2) 找出每个  $S'$  的集合里相对于集合  $S$  中不满足  $C^D \neq \bigcup \{B_1^D \mid B_1 \text{ 是决策前提}, B_1 \subset C\}$  的决策蕴涵。

其算法描述如算法 2 所示。

算法 2 决策背景下决策蕴涵基的生成算法

输入: 决策背景  $S$

输出: 决策蕴涵规范基  $W$

```
1.  $S' = \emptyset$  // 决策背景  $S$  的划分分块
2.  $W = \emptyset$  // 决策蕴涵规范基
3.  $W' = \emptyset$  // 决策蕴涵规范基
4.  $\Pi = \emptyset$  // 最小生成子集合
5.  $\Pi' = \emptyset$  // 每个  $S'$  最小生成子集合
6.  $S = \text{sort}(S)$ ; // 对决策背景  $S$  进行排序, 对对象按照属性值从少到多的顺序排序
7. 取划分决策背景分块的阈值  $\lambda$  来进行决策背景的分块以得到  $S'$  的集合  $\{S_1', S_2', S_3', \dots, S_K'\}$ 
8. FOREACH  $S_i' \in S'$  DO // 对每个子分块进行遍历
9.  $\Pi' = \text{getSmallSet}(S_i', O', C', R'^{CD'})$  //  $\Pi'$  为  $S_i'$  分块的所有最小生子
10.  $S_i' = \text{sortByDictionary}(\Pi')$  // 以字典序对进行排序
11. FOREACH  $B'^{CD'} \in \Pi'$  DO
12. 检查  $W' = \emptyset$  // 用于检验  $B'^{CD'} = B_{C_1'}^{CD'}$  条件
13. FOREACH 满足  $B'^{CD'} \subset B_{C_1'}^{CD'}$  的  $B'^{CD'} \rightarrow B_{C_2'}^{CD'}$  DO {
14. IF  $B_{C_1'}^{CD'} \neq B_{C_2'}^{CD'}$  THEN
15.  $W' = W' \cup \{B_{C_2'}^{CD'}\}$ 
16. ELSE
17.  $W' = B_{C_2'}^{CD'}$ 
18. BREAK
19. END IF} // END FOREACH
20. deleteSet( $\Pi'$ ) // 从  $\Pi'$  中移除  $B_{C_2'}^{CD'}$ 
21. IF  $B_{C_1'}^{CD'} \neq W'$  THEN
22.  $W'. \text{ADD}(B_{C_2'}^{CD'} \rightarrow B_{C_1'}^{CD'})$  // 生成决策蕴涵  $B_{C_2'}^{CD'} \rightarrow B_{C_1'}^{CD'}$  并添加到  $W'$ 
23. END IF} // END FOREACH
24.  $W. \text{ADD}(W')$  // 把  $W'$  添加到  $W'$  中
25. } // END FOREACH
26. 对  $W$  中的决策蕴涵基再次调用步骤(20)一步骤(24)
27. Return 0
```

算法 2 调用了算法 1 即最小生成子算法  $\text{getSmallSet}()$ , 另外使用了排序方法  $\text{sort}()$  和字典排序算法  $\text{sortByDictionary}()$  对每个划分分子块的最小生成子  $\Pi'$  进行字典排序。

通过此算法得到的决策蕴涵比文献[10]更为紧凑, 因为其两次判断并删除不满足决策蕴涵条件的冗余的决策蕴涵规范基。对决策背景  $S$  进行分块的时间复杂度为  $O(N/2)$ 。对每个分块  $S'$  提取并排序最小生成子的时间复杂度为  $O(|C'| \cdot |TimeC' + (\lfloor \frac{|C'|}{2} \rfloor) \cdot |S'| \cdot |C'|)$ 。其中,  $TimeC'$  为子划

分块  $S'$  的形式背景访问的时间,  $|C'|$  为对决策前提进行判断所用的时间,  $TimeN$  为集合  $S$  根据阈值  $\lambda$  的取值得到的分块  $S'$  的个数。最坏情况下本文算法处理并生成每个子集合  $S'$  的最小生成子的数目所需的时间复杂度为  $O(|\Pi|^2)$ , 其中  $\Pi$  是每个  $S'$  生成最小子的数目。最终该算法的时间复杂度为  $O(N/2 + (|C'| \cdot |TimeC' + (\lfloor \frac{|C'|}{2} \rfloor) \cdot |S'| \cdot |C'| + |\Pi|^2)^\wedge TimeN)$ 。

5 相关实验对比

根据算法 2 在 MatLab 中进行实现。本文实验通过对阈值  $\lambda$  的不同取值来进行实验对比。最好情况下的分块数为  $N = \lambda/2 + 1$ , 最坏情况下的分块数为 1。本实验所采用的数据集来源于政府编制实名制管理系统, 因涉及到信息的保密性, 这里只选取 3 组较小规模的数据集进行实验对比。在不同的数据集中首先对数据集做预处理操作, 包括移除空缺的值和排序操作等, 并在不同的背景下取相同的阈值  $\lambda$ ,  $\lambda$  取多个值。在这 3 组实验数据集中, 针对相同的  $\lambda$  值, 取得不同的分块数, 从而得到不同的决策蕴涵基。测试数据集的相关信息如表 6 所列。

表 6 测试数据简表

| 数据名称     | 对象数 | 属性数 |
|----------|-----|-----|
| NumList1 | 218 | 12  |
| NumList2 | 185 | 15  |
| NumList3 | 380 | 15  |

对于数据集中的数据, 依照前文所述分别对每个数据集生成 5 个决策背景, 生成的方法参见文献[10]。

对于每一个数据集, 采用算法 2,  $\lambda$  分别取值为 1, 3, 5, 7, 把数据集分成多个子集合, 再为每个子集合生成条件子背景的形式概念和最小生成子。表 7 列出了当  $\lambda = 1$  时对 NumList1 数据集中的一个分块进行处理的结果, 其中 1, 2, 4, 6, 7, 9, 11 是条件属性。对其他分块的处理与此相同。

表 7  $\lambda = 1$  时分块子集合  $S_1'$  中的规范基

| 数据     | 条件属性 | 概念数 | 最小生成子 | 规范基所含蕴涵数 |
|--------|------|-----|-------|----------|
| $S_1'$ | 1    | 1   | 1     | 1        |
| $S_1'$ | 2    | 2   | 2     | 1        |
| $S_1'$ | 4    | 2   | 2     | 1        |
| $S_1'$ | 6    | 4   | 4     | 1        |
| $S_1'$ | 7    | 8   | 10    | 2        |
| $S_1'$ | 9    | 12  | 14    | 4        |
| $S_1'$ | 11   | 16  | 20    | 4        |

每个数据集最终的最小生成子和决策蕴涵函数如表 8—表 10 所列。

表 8 NumList1 中规范基的生成比较

| 阈值 $\lambda$ | 分块数 | 最小生成子集合 | 决策蕴涵数集合 |
|--------------|-----|---------|---------|
| 1            | 20  | 527     | 62      |
| 3            | 50  | 501     | 58      |
| 5            | 80  | 480     | 50      |
| 7            | 92  | 465     | 40      |

表 9 NumList2 中规范基的生成比较

| 阈值 $\lambda$ | 分块数 | 最小生成子集合 | 决策蕴涵数集合 |
|--------------|-----|---------|---------|
| 1            | 25  | 585     | 75      |
| 3            | 55  | 520     | 62      |
| 5            | 82  | 475     | 51      |
| 7            | 101 | 470     | 50      |

表 10 NumList3 中规范基的生成比较

| 阈值 $\lambda$ | 分块数 | 最小生成子集合 | 决策蕴涵数集合 |
|--------------|-----|---------|---------|
| 1            | 30  | 675     | 88      |
| 3            | 62  | 580     | 75      |
| 5            | 102 | 490     | 60      |
| 7            | 112 | 482     | 58      |

由表 8—表 10 可知,对于同一数据集,当对分块的阈值  $\lambda$  取不同值时,得到的分块数不同,且每个分块数得到的最小生成子个数的集合也不同。分块数越多,所获得的最小生成子的集合相对减少,决策蕴涵集合数也随之减少,虽然差距不大,但也可以说明分块数越多,其生成的决策蕴涵越紧凑。算法 2 证明了决策规范基的完备性和无冗余性,通过以上 3 组数据的对比也进一步证明了此算法生成的决策蕴涵规范基是最优的。在最坏情况下,一个对象进行一个分块,从而得到最冗余的决策蕴涵规范基,此时已无意义。

基于表 6 中的 3 组数据,本文还根据实名制编制管理系统的算法在 Matlab 中进行模拟实验,通过对 3 组数据集不设置  $\lambda$  的值以及取不同  $\lambda$  的值,分别得到不同的数据分组,再依照不同的分组对这些数据集进行访问和操作,其时间对比如表 11 所列。

表 11 不同分组访问时间的对比

| 数据集      | 对数据集的访问耗时/s |             |             |             |             |
|----------|-------------|-------------|-------------|-------------|-------------|
|          | 不分组         | $\lambda=1$ | $\lambda=3$ | $\lambda=5$ | $\lambda=7$ |
| NumList1 | 1.324       | 0.835       | 0.623       | 0.426       | 0.321       |
| NumList2 | 1.210       | 0.875       | 0.621       | 0.475       | 0.301       |
| NumList3 | 1.850       | 0.901       | 0.678       | 0.498       | 0.318       |

在模拟实验中,本文还继续对  $\lambda$  的取值进行实验,结果表明  $\lambda$  的取值越接近  $N/2(N$  为数据集的对象个数),其分块数差距越小,因此可以认为  $\lambda$  的最优值接近  $N/2$ 。

**结束语** 本文主要给出了一个基于决策背景的决策规范基。该规范基首先对决策背景进行分块研究,将每个分块上的决策蕴涵相对于分块的决策子背景作为决策蕴涵集,再对每个分块的决策蕴涵集进行整合,剔除不满足条件的决策蕴涵,从而得到决策蕴涵集。证明了分块得到的决策蕴涵基集合相对于决策背景是完备、无冗余、最优的。并且在分组阈值  $\lambda$  取不同的值时得到的决策规范基的蕴涵数是不相同的, $\lambda$  的取何值时最优还有待进一步研究。本文还给出了对应的基于最小生成子的算法,通过该算法使用 MatLab 进行实验,并进行模拟对比。在该算法中根据  $\lambda$  的不同取值,得到不同的决策规范基。该算法中是否还可以进一步改进生成最小生成子的算法、是否可以进一步减少最小生成子的个数等还有待研究。

参 考 文 献

[1] LI L F,ZHANG D X. The Application of Concept Lattice Theory in the Reduction of the Proposition Set in Two-Valued Propositional Logic[J]. Acta Electronica Sinica,2007,35(8):1538-1542. (in Chinese)  
李立峰,张东晓. 概念格在二值命题逻辑命题集约简中的应用[J]. 电子学报,2007,35(8):1538-1542.  
[2] KUMAR C A,SRINIVAS S. Concept lattice reductiong using

fuzzy K-Means clustering[J]. Expert Systems With Application,2010,37(3):2696-2704.  
[3] KAI S Q,YAN H Z,JI Y L, et al. Study of decision implications based on formal concept analysis[J]. International Journal of General Systems,2007,36(2):147-156.  
[4] LI J,MEI C, LV Y. A heuristic knowledge-reduction method for decision formal contexts[J]. Computers & Mathematics With Applications,2011,61(4):1096-1106.  
[5] JIN H L,CHANG L M, LV Y J. Knowledge reduction in decision formal contexts[J]. Knowledge Based Systems,2011,24(5):709-715.  
[6] JIN H L,CHANG L M, LV Y J. Knowledge reduction in formal decision contexts based on an order-preserving mapping[J]. International Journal of General Systems,2012,41(2):143-161.  
[7] JIN H L,CHANG L M, LV Y J. Incomplete decision contexts: Approximate concept constrution, rule acquisition and knowledge reduction[J]. International Journal of Approximate Reasoning,2013,54(1):149-165.  
[8] JIN H L,CHANG L M, LV Y J. Knowledge reduction in real decision formal contexts[J]. Information Sciences,2012,189(7):191-207.  
[9] JIN H L,CHANG L M,XIAO Z, et al. On rule acquisition in decision formal Contexts[J]. International Journal of Machine Learning and Cybernetics,2013,4(6):1-11.  
[10] ZHAI Y H,LI D Y,QU K S. Canonical Basis for Decision Implications[J]. Acta Electronica Sinica,2015,43(1):18-23. (in Chinese).  
翟岩慧,李德玉,曲开社. 决策蕴涵规范基[J]. 电子学报,2015,43(1):18-23.  
[11] ZHAI Y H. Decision Logic in Formal Concept Analysis[D]. Taiyuan:Shanxi University,2015:12-39. (in Chinese)  
翟岩慧. 基于形式概念分析的决策逻辑研究[D]. 太原:山西大学,2015:12-39.  
[12] GANTER B,WLLLE R. Formal Context Analysis; Mathematical Foundations[M]. Springer,1999.  
[13] QU K S,ZHAI Y H,LIANG J Y, et al. Study of decision implications based on formal concept analysis[J]. International Journal of General Systems,2007,36(2):147-156.  
[14] GUO X B,ZHAO S L,WANG C B, et al. Multi-valued Attribute Association Rules Mining Based on Concept Lattice[J]. Computer Science,2014,41(3):267-271. (in Chinese)  
郭晓波,赵书良,王长宾,等. 基于概念格的多值属性关联规则挖掘[J]. 计算机科学,2014,41(3):267-271.  
[15] ZHANG Y C. Reseach on the Method of Decision Rule Acquisition in Concept Lattice [D]. Taiyuan:Shanxi University,2008:27-33. (in Chinese).  
张倚弛. 概念格中决策规则的获取方法研[D]. 太原:山西大学,2008:27-33.  
[16] STUMME G,TAOUIL R,BASTIDE Y, et al. Fast computation of concept lattices using data mining techniques[C]// Proceedings of the 7th International Workshop on Knowledge Representation meets Databases. Berlin:Springer,2000:129-139.