最小化决策规则集的计算方法

裴小兵,吴　涛,陆永忠
(华中科技大学 软件学院,湖北 武汉 430074)

摘　要:在决策算法中,并不是所有的决策规则都是必要的,一些过剩的决策规则应该消去,而不影响作决策,因此,研究最小化决策规则集的计算方法是很有意义的. 传统的决策算法并没有给出最小化决策规则集的形式化计算方法,为了解决最小化决策规则集的形式化计算问题,引入了最小化决策规则可辨识矩阵概念,提供了基于可辨识矩阵的基本决策规则的最小化决策规则集的计算方法.

关键词:粗糙集;决策表;最小化决策规则集
中图分类号:TP311　**文献标识码**:A　**文章编号**:1673-4785(2007)06-0065-03

Calculating method for a minimal set of decision rules

PEI Xiao-bing, WU Tao, LU Yong-zhong
(School of Software Engineering, Huazhong University of Science & Technology, Wuhan 430074, China)

Abstract: Not all decision rules in existing decision-making algorithms are necessary. Irrelevant and superfluous decision rules should be eliminated as they do not affect decision-making, yet increase computational overhead as well as complexity, which can lead to errors. Therefore, research on a method for calculating a minimal set of decision rules is very important. But the algorithms currently available for decision rules don't present a formalized method for calculating a minimal set of decision rules. To solve this problem, in this paper, a new concept of a discernable matrix for minimal decision rules is introduced, and a theorem for judging a minimal set of decision rules is given. On this basis, we proposed a formalized calculation method for a minimal decision rules set based on a discernable matrix. To illustrate this method, an example is presented.

Keywords: rough set; decision-making table; minimal decision rules set

粗糙集理论[1]是 20 世纪 80 年代由波兰的 Pawlak 教授提出,它是一种新型的处理模糊性和不确定性知识的数学工具,在数据挖掘、决策支持系统等领域得到了应用[2-3].

决策规则获取是粗糙集理论和应用的关键问题之一,在决策算法中,并不是所有的决策规则都是必要的,或者说,具有相同决策类相结合的一些过剩的决策规则应该消去,因此,研究最小化决策规则集的计算方法是很有意义的. 计算最小化决策规则集(决策算法的极小化)是决策规则获取算法的重要组成部分[2-3]. 目前,许多学者对决策规则的获取作了深入研究,并取得了很好成果[4-5],但是这些研究都没有给出最小化决策规则集,这样就产生了过多的决策规则,影响了决策时的效率,文献[2-3]中强调了最小化决策规则集的意义及重要性,但没有给出最小化决策规则的形式化计算方法,为了解决最小化决策规则集的计算问题,引入了最小化决策规则集可辨识矩阵概念,从而提供了基于可辨识矩阵的基本决策规则的最小化决策规则集的计算方法,该方法直观、易理解.

1　最小化决策规则集的基本概念

定义 1　一个决策表系统 S 可以表示为:$S = \langle U, A, V, f \rangle$,其中,$U$ 为对象的集合,即论域;A 是属性集合,$A = C \cup D$,C 和 D 分别是条件属性集和决策属性集,$D \neq \Phi$,$V = V \bigcup_{a \in A} V_a$,$V_a$ 表示属性 a 的值域;$f: U \times A \to V$ 是一信息函数,它指的 U 中每一

收稿日期:2006-10-12.

© 1994-2008 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

个对象 x 的属性值,即 $x \in U, a \in A$,有 $f(x,a) \in V_a$.

对于任意 $B \subseteq A$,记:$\mathrm{ind}(B) = \{(x,y): f(x,a_k) = f(y,a_k), \forall a_A \in A\}$,则 $\mathrm{ind}(B)$ 是 U 上的等价关系,称为由 B 决定的不可区分关系,它们产生的 U 上的划分为:$U/\mathrm{ind}(B) = \{[x]_B : x \in U\}$,其中:$[x]_B = \{y:(x,y) \in \mathrm{ind}(B)\}$是 x 关于 B 的等价类[2-3].

定义2　设决策表系统 $S = <U,A,V,f>$,AS 是 S 上的一个基本算法,AS 中有相同后件 ϕ 的所有基本规则的集合被表示为 AS_ϕ,并且用 AP_ϕ 表示属于 AS_ϕ 的决策规则的所有前件的集合. 如果 $\models_S \vee AP_\phi \leftrightarrow \vee(AP_\phi - \{\phi\})$,则在 AS 中的基本规则 $\phi \to \phi$ 是可约去的,其中 $\vee AP_\phi$ 表示 AP_ϕ 中所有公式的析取,否则这条规则在 AS_ϕ 中是不能被约去的;如果 AS_ϕ 中所有决策规则都是不能被约去的,则规则集合 AS_ϕ 被称为独立的;如果 AS_ϕ' 中所有决策规则集都是独立的,并且 $\models_S \vee AP_\phi \leftrightarrow \vee AP_\phi'$,则称决策规则集 AS_ϕ 的子集 AS_ϕ' 是 AS_ϕ 的一个约简,AS_ϕ' 称做 AS_ϕ 的一个最小化决策规则集. 显然,AS_ϕ 可以有多个最小化决策规则集[2-3].

2　最小化决策规则集的计算方法

定义3　设决策表系统 $S = <U,A,V,f>$,AS 是 S 上的一个基本算法,AS 中有相同后件 ϕ 的所有基本规则的集合为 $AS_\phi = \{\phi_1 \to \phi, \phi_2 \to \phi, \cdots, \phi_n \to \phi\}$,$AP_\phi$ 表示属于 AS_ϕ 的决策规则的所有前件的集合 $AP_\phi = \{\phi_1, \phi_2, \cdots, \phi_n\}$. 记 $[\phi]$ 表示满足所有后件 ϕ 的对象集合,$[\phi_k]$ 表示满足所有前件的 ϕ_k 对象集合,$[\phi] = \bigcup_{k=1}^{n}[\phi_k]$,对任意的 $x_i, x_j \in [\phi]$,当 $i \neq j$ 时,$C_{ij} = \{\delta : \delta \in AP_\phi$ 且 $x_i, x_j \in [\delta]$,或 $\delta = \phi_k \wedge \phi_l$ 且 $x_i \in [\phi_k]$、$x_j \in [\phi_l]$,ϕ_k、$\phi_l \in AP_\phi\}$;当 $i = j$ 时,$C_{ij} = \Phi$. 称 C_{ij} 为 x_i, x_j 的最小化决策规则集;$[C_{ij}]$ 为决策规则集 $AS_\phi = \{\phi_1 \to \phi, \phi_2 \to \phi, \cdots, \phi_n \to \phi\}$ 的最小化决策规则集可辨识矩阵.

定义4　设决策表系统 $S = <U,A,V,f>$,$AS_\phi = \{\phi_1 \to \phi, \phi_2 \to \phi, \cdots, \phi_n \to \phi\}$ 为具有相同后件 ϕ 的所有基本规则的集合,$[C_{ij}]$ 为 AS_ϕ 的最小化决策规则可辨识矩阵,则称 $M = \bigwedge_{i,j}(\vee\{a_k : a_k \in C_{ij}\})$ 为可辨识公式.

命题1　$[C_{ij}]$ 为决策规则集 $AS_\phi = \{\phi_1 \to \phi, \phi_2 \to \phi, \cdots, \phi_n \to \phi\}$ 的最小化决策规则可辨识矩阵,则 $[C_{ij}]$ 具有如下性质:

1) $[C_{ij}]$ 是对称矩阵;2) $[C_{ij}]$ 主对角线上的元素为空.

性质1)、2)可以直接根据最小化决策规则可辨识矩阵的定义得到. 根据性质1),计算最小化决策规则可辨识矩阵时,只需计算矩阵的下三角矩阵即可.

实例1　考虑如表1的决策表,条件属性集 $C = \{a,b,c\}$,决策属性集 $D = \{d\}$,$U = \{\#1, \#2, \#3, \#4, \#5\}$.

表1　决策表

Table 1　Decision table

U	a	b	c	d
#1	1	0	2	1
#2	2	0	3	1
#3	1	2	1	1
#4	2	1	3	0
#5	1	0	3	0

在表1中求得决策类1(相同后件 $(d,1)$)的所有基本决策规则的集合为 $AS_\phi = \{(c,2) \to (d,1), (a,2)(b,0) \to (d,1), (b,2) \to (d,1), (c,1) \to (d,1)\}$,其对应的最小化决策规则可辨识矩阵 $[C_{ij}]$ 为 $C_{11} = C_{12} = C_{13} = C_{22} = C_{23} = C_{33} = \Phi$;$C_{21} = \{(a,2)(b,0)\}$;$C_{31} = \{(c,2) \wedge (c,1), (c,2) \wedge (b,2), (a,2)(b,0) \wedge (b,2), (a,2)(b,0) \wedge (c,1)\}$;$C_{32} = \{(a,2)(b,0) \wedge (b,2), (a,2)(b,0) \wedge (c,1)\}$.

命题2　设决策表系统 $S = <U,A,V,f>$,$AS_\phi = \{\phi_1 \to \phi, \phi_2 \to \phi, \cdots, \phi_n \to \phi\}$ 为具有相同后件 ϕ 的所有基本规则的集合,$[C_{ij}]$ 为决策规则集 AS_ϕ 的最小化决策规则可辨识矩阵,AS_ϕ' 为决策规则集 AS_ϕ 的子集,AP_ϕ' 表示属于 AS_ϕ' 的决策规则的所有前件的集合,则 $\bigcup_{\delta \in AP_\phi'}[\delta] = [\phi] \Leftrightarrow$ 对任意的 $C_{ij} \neq \Phi$,都存在 $\gamma \in C_{ij}$ 使得 $\gamma \subseteq AP_\phi'$.

证明　假设存在 $C_{ij} \neq \Phi$,不存在 $\gamma \in C_{ij}$ 使得 $\gamma \subseteq AP_\phi'$,则 $x_i \in [\phi]$ 且 $x_i \notin \bigcup_{\delta \in AP_\phi'}[\delta]$ 或 $x_j \in [\phi]$ 且 $x_j \notin \bigcup_{\delta \in AP_\phi'}[\delta]$,这与 $\bigcup_{\delta \in AP_\phi'}[\delta] = [\phi]$ 矛盾. 所以,对任意的 $C_{ij} \neq \Phi$,都存在 $\gamma \in C_{ij}$ 使得 $\gamma \subseteq AP_\phi'$.

反证:假设存在 $x_i \in [\phi]$ 且 $x_i \notin \bigcup_{\delta \in AP_\phi'}[\delta]$,对任意的 $x_j \in [\phi]$,记 C_{ij} 为 x_i, x_j 的最小化决策规则集,则不存在 $\gamma \in C_{ij}$ 使得 $\gamma \subseteq AP_\phi'$,这与已知条件矛盾,所以,$\bigcup_{\delta \in AP_\phi'}[\delta] \supseteq [\phi]$. 又 $\bigcup_{\delta \in AP_\phi'}[\delta] \subseteq [\phi]$,即 $\bigcup_{\delta \in AP_\phi'}[\delta] = [\phi]$.

命题3　设决策表系统 $S = <U,A,V,f>$,$AS_\phi = \{\phi_1 \to \phi, \phi_2 \to \phi, \cdots, \phi_n \to \phi\}$ 为具有相同后件 ϕ 的所有基本规则的集合,$[C_{ij}]$ 为决策规则集 AS_ϕ 的最小化决策规则可辨识矩阵,其对应的可辨识公式 M 的极小析取范式为 M_1,则 M_1 包含 AS_ϕ 的全体

© 1994-2008 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

最小化决策规则集.

在命题 2 的基础上,借鉴文献[6 - 8]的思想,易于证明命题 3.

命题 3 直观上提供了最小化决策规则集的计算方法.

类似于基于可辨识矩阵的属性约简算法[2 - 3,6],下面根据上述结果提供基于可辨识矩阵的基本决策规则集的最小化决策规则集的计算方法.

算法 1　最小化决策规则集的计算方法

输入:决策表系统 $S = \langle U, A, V, f \rangle$,具有相同后件 ϕ 的基本决策规则集 $AS_\phi = \{\phi_1 \to \phi, \phi_2 \to \phi, \cdots, \phi_n \to \phi\}$;

输出:AS_ϕ 的全体最小化决策规则集.

方法步骤为

1) 计算决策规则集 AS_ϕ 的最小化决策规则集可辨识矩阵 $[C_{ij}]$;

2) 计算最小化决策规则集可辨识矩阵 $[C_{i,j}]$ 的可辨识公式 M;

3) 计算可辨识公式 M 的极小析取范式 M,它将给出全体最小化决策规则集.

实例 1　以表 1 的决策表为实例,其中条件属性集 $C = \{a, b, c\}$,决策属性集 $D = \{d\}$,对象集 $U = \{\#1, \#2, \#3, \#4, \#5\}$.

在表 1 中求得决策类 1(相同后件 $(d,1)$)的所有基本规则的集合为 $AS_\phi = \{(c,2) \to (d,1), (a,2)(b,0) \to (d,1), (b,2) \to (d,1), (c,1) \to (d,1)\}$,其对应的最小化决策规则可辨识矩阵 $[C_{ij}]$ 为 $C_{11} = C_{12} = C_{13} = C_{22} = C_{23} = C_{33} = \Phi$; $C_{21} = \{(a,2)(b,0)\}$; $C_{31} = \{(c,2) \wedge (c,1), (c,2) \wedge (b,2), (a,2)(b,0) \wedge (b,2), (a,2)(b,0) \wedge (c,1)\}$; $C_{32} = \{(a,2)(b,0) \wedge (b,2), (a,2)(b,0) \wedge (c,1)\}$.

可辨识矩阵 $[C_{ij}]$ 的可辨识公式 M 的极小析取范式为

$$M = \{(a,2)(b,0) \wedge (b,2)\} \vee \{(a,2)(b,0) \wedge (c,1)\}.$$

因此,基本规则的集 AS_ϕ 有 2 个最小化决策规则集:

1) $(a,2)(b,0) \wedge (b,2) \to (d,1)$;

2) $(a,2)(b,0) \wedge (c,1) \to (d,1)$.

3　结束语

介绍了一种直观、易理解的最小化决策规则集的计算方法,该方法基于构造最小化决策规则集的可辨识矩阵来获得最小化决策规则集.如何提高计算最小化决策规则集的效率将是下一步的研究任务.

参考文献:

[1] PAWLAK Z. Rough sets[J]. International Journal of Computer and Information Science, 1982, 11(5): 341 - 356.

[2] 刘　清. Rough 集及 Rough 推理[M]. 北京:科学出版社, 2001.

[3] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法[M]. 北京:科学出版社, 2001.

[4] KRYSZKIEWICZ M. Rules in incomplete information systems[J]. Information Sciences, 1999, 113: 271 - 292.

[5] MOLLESTAD T, SKOWRON A. A rough set framework for data mining of propositional default rules[A]. Foundations of Intelligent Systems of the 9th International Symposium[C]. Zakopane, Poland, 1996.

[6] SKOWRON, RAUSZER C. Intelligent decision support handbook of application and advances of the rough sets theory[M]. Dordreecht: Kluwer Academic Publishers, 1992.

[7] MIJ S, WU W Z, ZHANG W X. Approaches to knowledge reduction based on variable precision rough set model[J]. Information Sciences, 2004, 159: 255 - 272.

[8] 张文修,米据生,吴伟志. 不协调决策表信息系统的知识约简[J]. 计算机学报, 2003, 26(1): 12 - 18.
ZHANG Wenxiu, MI Jusheng, WU Weizhi. Knowledge reductions in inconsistent information systems[J]. Chinese Journal of Computers, 2003, 26(1): 12 - 18.

作者简介:

裴小兵,男,1972 年生,博士,主要研究方向为数据挖掘、数据库、软件工程.

E-mail: xiaobingp@tom.com.

吴　涛,男,1972 年生,副教授,主要研究方向为软件工程.

E-mail: wutaooptal@tom.com.

陆永忠,男,1969 年生,副教授,主要研究方向为机器学习、软件工程.

E-mail: hotmailuser@126.com.

© 1994-2008 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net