

文章编号: 1001-0920(2006)12-1421-04

## 广义信息系统及其决策规则

周军<sup>1, 2a</sup>, 张庆灵<sup>1</sup>, 佟绍成<sup>2b</sup>

(1. 东北大学理学院, 沈阳 110004; 2. 辽宁工学院 a. 计算机科学与工程学院; b. 数理科学系, 辽宁 锦州 121001)

**摘要:** 提出一类更广泛的信息系统, 称其为广义信息系统。它包含了完备信息系统、不完备信息系统和多值信息系统。给出了广义信息系统集合近似的概念、方法和相关性质。讨论了广义信息决策系统的决策描述形式。这种决策描述形式易于转化为 Skolem 标准型, 可以直接应用于人工智能的归结推理。

**关键词:** 广义信息系统; 决策表; 粗糙集; 决策规则

**中图分类号:** TP14

**文献标识码:** A

## Generalized Information Systems and Decision Rules

ZHOU Jun<sup>1, 2a</sup>, ZHANG Qing-ling<sup>1</sup>, TONG Shao-cheng<sup>2b</sup>

(1. College of Science, Northeastern University, Shenyang 110004, China; 2a. Department of Computer Science, 2b. Department of Mathematics and Physics, Liaoning Institute of Technology, Jinzhou 121001, China. Correspondent: ZHOU Jun, E-mail: lnjunzhou@sina.com)

**Abstract:** A class of more general information systems, called generalized information systems, is presented. Generalized information systems contain complete information systems, incomplete information systems and multivalued information systems. A concept of set approximation and an extension of generalized information systems are given. The methods of computing rules and some properties about generalized information systems are also discussed. Furthermore, the describing forms of decision of generalized information decision systems are given. Since this kind of describing forms of decision can be easily transformed into the Skolem's standard form, they can be applied directly to resolution inference in artificial intelligence field.

**Key words:** Generalized information systems; Decision tables; Rough sets; Decision rules

### 1 引言

粗糙集理论<sup>[1]</sup>是智能数据分析和数据挖掘的一种新的数学方法, 已成为机器学习、知识获取、决策分析、模式识别等领域的基本理论。粗糙集理论通过不可区分关系给出了对象分类的系统方法, 并在许多实际问题中得到了成功的应用。然而, 经典的粗糙集理论是建立在等价关系的基础上, 而现实世界中的许多关系并不是等价关系, 这使得这一理论的应用受到了限制。为此, 国内外学者提出了对经典粗糙集理论的推广<sup>[2~6]</sup>。其中最具代表性的是基于对象属性值存在残缺的不完备信息系统的粗糙集理

论<sup>[2]</sup>, 并形成了较为系统的理论<sup>[3,4]</sup>。文献[5]根据实际应用需要提出了多值粗糙集理论, 该理论是一类完备信息系统的推广。

本文在经典粗糙集理论、基于不完备信息系统的粗糙集理论和多值粗糙集理论的基础上, 提出了广义信息系统及其粗糙集理论, 并证明了它包含完备信息系统、不完备信息系统和多值信息系统。这些信息系统仅是广义信息系统的一些特例, 广义信息系统具有更广泛的应用范围。同时给出了广义信息系统集合近似的概念、方法和相关性质。最后讨论了广义信息决策系统的决策描述形式, 它易于将其转

收稿日期: 2005-09-19; 修回日期: 2006-01-19

基金项目: 国家自然科学基金项目(60674056); 辽宁省青年人才基金项目(2005219001); 辽宁省优秀青年骨干教师基金项目

作者简介: 周军(1966-), 女, 内蒙古通辽人, 教授, 博士生, 从事数据挖掘、知识发现等研究; 张庆灵(1956-), 男, 辽宁营口人, 教授, 博士生导师, 从事智能控制、数据处理等研究

化为子句集的形式 在人工智能的归结推理<sup>[7]</sup>中, 无论是问题证明还是问题求解, 广义决策表的决策描述都是简单和方便的

## 2 相关知识

信息系统是指一个三元组  $\mathcal{Q} = (\Theta, A_T, f)$ . 其中:  $\Theta$  是非空有限对象集,  $A_T$  是有限的非空属性集,  $f$  是映射, 满足  $\forall a \in A_T, f_a: \Theta \rightarrow V_a \cup \{*\}$ .  $V_a$  是属性  $a$  的值域, 任意属性的值域  $V_a$  可以包含特殊符号“ $*$ ”; 并用  $*$  表示对应的属性值是不确定的  $\text{Inf}(x) = \{(a, f_a(x)) \mid a \in A_T\}$  称为  $x$  的信息向量 详细内容参见文献[3, 4]

多值信息系统是有序对  $M = (\Theta, A_T, V)$ . 其中:  $\Theta$  是非空有限对象集,  $A_T$  是有限的非空属性集,  $V$  是  $A_T$  中属性的值域,  $V_a$  是属性  $a$  的值域 如果  $\forall x \in \Theta, \forall a \in A_T$ , 则  $a(x) \subseteq V_a$ . 详细内容参见文献[5, 6]

## 3 主要结果

### 3.1 广义信息系统

**定义 1** 广义信息系统是指一个三元组  $\mathcal{Q} = (\Theta, A_T, f)$ . 其中:  $\Theta$  是非空有限对象集,  $A_T$  是有限的非空属性集,  $f$  是映射, 满足  $f_a: \Theta \rightarrow \mathbf{R}(V_a), \forall a \in A_T, V_a$  是属性  $a$  的值域, 任意属性的值域  $V_a$  可以包含特殊符号  $*$ . 这里用  $*$  表示该属性值是残缺的和不确定, 其真实值属于  $V_a - \{*\}$ .  $\mathbf{R}(V_a)$  是属性  $V_a$  的幂集  $\text{Inf}(x) = \{(a, f_a(x)) \mid a \in A_T\}$  称为  $x$  的信息向量集 如果  $\forall x \in \Theta, \forall a \in A_T, f_a(x) \neq \emptyset$ , 且  $*$   $\notin f_a(x)$ , 则称广义信息系统  $\mathcal{Q}$  是常规的 如果  $\forall x \in \Theta, \forall a \in A_T$ , 有  $*$   $\in f_a(x)$ , 则称广义信息系统  $\mathcal{Q}$  是确定的; 否则, 称其为奇异的

**命题 1** 设广义信息系统  $\mathcal{Q} = (\Theta, A_T, f)$ , 若满足:

- 1)  $\forall a \in A_T, f_a(\Theta)$  中的元素均为单元素集;
- 2)  $\forall x \in \Theta, \forall a \in A_T$ , 有  $*$   $\notin f_a(x)$ ;
- 3)  $\forall x \in \Theta, \forall a \in A_T$ , 有  $f_a(x) \neq \emptyset$ .

则  $\mathcal{Q} = (\Theta, A_T, f)$  为完备信息系统

**证明** 由 1) 知  $\forall a \in A_T, f_a(\Theta)$  中的元素均为单元素集, 即  $\forall x \in \Theta$  有  $\text{card}(f_a(x)) = 1$ . 由 2) 和 3) 知  $*$   $\notin f_a(x), f_a(x) \neq \emptyset$ . 所以  $f_a(x)$  既不是不确定的  $*$ , 也不是空集  $\emptyset$ , 而是  $V_a$  中的一个确定的值 因此  $\mathcal{Q} = (\Theta, A_T, f)$  是完备信息系统

**命题 2** 设广义信息系统  $\mathcal{Q} = (\Theta, A_T, f)$ , 若满足:

- 1)  $\forall a \in A_T, f_a(\Theta)$  中的元素均为单元素集;
- 2)  $\exists x \in \Theta, \exists a \in A_T$ , 有  $f_a(x) = \{*\}$ ;
- 3)  $\forall x \in \Theta, \forall a \in A_T$ , 有  $f_a(x) \neq \emptyset$ .

则  $\mathcal{Q} = (\Theta, A_T, f)$  为不完备信息系统

**命题 3** 设广义信息系统  $\mathcal{Q} = (\Theta, A_T, f)$ , 若满足:

- 1)  $\forall x \in \Theta, \forall a \in A_T$ , 有  $*$   $\notin f_a(x)$ ;
- 2)  $\forall x \in \Theta, \forall a \in A_T$ , 有  $f_a(x) \neq \emptyset$ .

则  $\mathcal{Q} = (\Theta, A_T, f)$  为多值信息系统

命题 2 和命题 3 的证明类似于命题 1 的证明

广义信息系统包含了完备信息系统, 不完备信息系统和多值信息系统 下面的例子可以说明广义信息系统中确实存在既不是完备信息系统, 也不是不完备信息系统, 更不是多值信息系统的情况

**例 1** 表 1 是一个广义信息系统  $\mathcal{Q}$  的示例 其中  $*$  表示属性值是不确定的,  $\{1, 2\}$  和  $\{2, 3\}$  表示该元素的属性具有两个属性值, 空格表示属性值是不存在的 即对象 2 的属性  $a$  的值是不确定的, 属性  $c$  的值是  $\emptyset$ , 对象 3 的属性  $c$  和对象 6 的属性  $b$  的值是不确定的

表 1 广义信息系统的示例  $\mathcal{Q}$

对象 $\theta$	属性 $A_T$			
	$a$	$b$	$c$	$d$
1	1	2	1	1
2	*	1, 2		2
3	1	2	*	1
4	2	1	1, 2	1
5	1, 2	1	2	2
6	2, 3	*	2	1, 2
7	3	1	1	1, 2

由完备信息系统, 不完备信息系统和多值信息系统的定义可知, 表 1 的广义信息系统  $\mathcal{Q}$  既不是完备信息系统, 也不是不完备信息系统, 更不是多值信息系统 这说明广义信息系统是一类比完备信息系统, 不完备信息系统和多值信息系统更广泛的信息系统

综合命题 1~ 命题 3 和例 1, 可得如下定理:

**定理 1** 广义信息系统是一类比完备信息系统, 不完备信息系统和多值信息系统更广泛的不确定信息系统 不完备信息系统, 多值信息系统只是广义信息系统的特例

**定义 2** 设广义信息系统  $\mathcal{Q} = (\Theta, A_T, f)$ ,  $\mathcal{Q}' = (\Theta', A_T', f')$ , 并且满足: 1)  $\Theta = \Theta', A_T = A_T'$ ; 2)  $\forall a \in A_T, \forall x \in \Theta$  如果  $f_a(x) = \{*\}$ ,  $f'_a(x) = f_a(x)$ , 则称系统  $\mathcal{Q}'$  为系统  $\mathcal{Q}$  的一个扩张 系统  $\mathcal{Q}$  的所有扩张的集合记为  $\text{extn}(\mathcal{Q})$ . 如果  $\forall y \in \Theta - \{x\}, \text{Inf}(y) = \text{Inf}(y)$ , 且  $\forall a \in A_T, * \notin f_a(x)$ , 则  $\mathcal{Q}'$  称为系统  $\mathcal{Q}$  的  $x$ -扩张 所有  $x$ -扩张的集合记为  $\text{extn}(\mathcal{Q}, x)$ .

**定义 3** 设广义信息系统  $\mathcal{Q} = (\Theta, A_T, f)$ , 如果  $\mathcal{Q} = (\Theta, A_T, f)$  是  $\mathcal{Q}$  的扩张, 且  $\mathcal{Q}$  是确定的广义信息系统(即  $\mathcal{Q}$  满足  $\forall a \in A_T, * \in V_a$ ), 则系统  $\mathcal{Q}$  称为系统  $\mathcal{Q}$  的一个实例 系统  $\mathcal{Q}$  的所有实例的集合记为  $\text{exam}(\mathcal{Q})$ .

**3.2 元素的相似关系与集合的近似**

在广义信息系统  $\mathcal{Q} = (\Theta, A_T, f)$  中, 对于属性  $a \in A_T$ , 定义关系  $S_R(a)$  如下:

$$S_R(a) = \{(x, y) \in \Theta \times \Theta \mid f_a(x) = f_a(y) \text{ 或 } * \in f_a(x) \text{ 或 } * \in f_a(y) \text{ 或 } f_a(x) = \emptyset \text{ 或 } f_a(y) = \emptyset\}.$$

对于  $A \subseteq A_T$ , 则定义

$$S_R(A) = \{(x, y) \in \Theta \times \Theta \mid f_a(x) = f_a(y) \text{ 或 } * \in f_a(x) \text{ 或 } * \in f_a(y) \text{ 或 } f_a(x) = \emptyset, \forall a \in A\}.$$

**命题 4**  $S_R(a)$  满足自反性和对称性, 但不满足传递性 因此  $S_R(a)$  是一个相容关系

证明可由定义直接得出

**命题 5** 在广义信息系统  $\mathcal{Q} = (\Theta, A_T, f)$  中, 对于  $A \subseteq A_T, S_R(A)$  也是一个相容关系, 且满足  $S_R(A) = \bigcap_{a \in A} S_R(a)$ .

证明  $\forall (x, y) \in S_R(A) \Leftrightarrow S_R(A)$  的定义, 可得  $\forall a \in A, (x, y)$  满足

$$f_a(x) = f_a(y) \text{ 或 } * \in f_a(x) \text{ 或 } * \in f_a(y) \Leftrightarrow \forall a \in A, (x, y) \in S_R(a) \Leftrightarrow (x, y) \in S_R(A).$$

记  $S_{R_A}(x) = \{y \in \Theta \mid (x, y) \in S_R(A)\}$ ,  $S_{R_A}(x)$  表示在  $\Theta$  中根据属性  $A$  可能与  $x$  不可区分的对象的集合

由广义信息系统  $\mathcal{Q} = (\Theta, A_T, f)$  中相似关系  $S_R(A)$  的定义, 显然有如下结论:

**命题 6** 在不完备信息系统中,  $S_R(A)$  退化为一个相容关系  $SM(A)$ ; 在完备信息系统中,  $S_R(A)$  退化为一个等价关系

**定义 4** 设广义信息系统  $\mathcal{Q} = (\Theta, A_T, f), X \subseteq \Theta, A \subseteq A_T$ , 则定义

$$\begin{aligned} \underline{\Delta}X &= \{x \in \Theta \mid S_{R_A}(x) \subseteq X\} = \{x \in X \mid S_{R_A}(x) \subseteq X\}, \\ \overline{\Delta}X &= \{x \in \Theta \mid S_{R_A}(x) \cap X \neq \emptyset\} = \{S_{R_A}(x) \mid x \in X\}. \end{aligned}$$

其中  $\underline{\Delta}X$  和  $\overline{\Delta}X$  分别称为  $X$  的下近似和上近似  $\underline{\Delta}X$  是确定属于  $X$  的对象的集合,  $\overline{\Delta}X$  是可能属于  $X$  的

对象的集合

集合的相似度、近似度等定义, 类似于完备信息系统和不完备信息系统中的定义

**3.3 决策表与决策规则**

(完备的、不完备的、广义的) 决策表  $(D_T)$  是一个 (完备的、不完备的、广义的) 信息系统, 即  $D_T = (\Theta, A_T \cup d, f)$ . 其中:  $d \in A_T$ , 且  $* \in V_d$ , 称为决策属性;  $A_T$  中的元素称为条件属性

定义函数  $\hat{a}: \Theta \rightarrow \mathbf{R}(V_d), A \subseteq A_T$ , 则有

$$\hat{a}(x) = \{f_d(x) \mid y \in S_{R_A}(x)\}.$$

$\hat{a}$  称为  $D_T$  的一般化决策,  $\hat{a}_T(x)$  是  $x$  所在相似类对应的所有可能的决策值的集合

**命题 7** 设  $x \in \Theta, A \subseteq A_T$ , 则  $\text{card}(\hat{a}(x)) = 1$ , 当且仅当  $S_{R_A}(x) \subseteq [x]_d$  其中  $[x]_d$  表示元素  $x$  关于决策属性  $d$  所在的等价类

在广义信息系统中, 决策规则的一般表示形式为  $t \rightarrow s$ , 即

$$t = \left( \bigcap_{a \in C} (a, v) \right) \quad s = (d, w).$$

其中:  $C$  是规则的条件部分存在的所有属性的集合,  $B_a$  是规则的条件部分中属性  $a$  存在的属性值的集合,  $(a, v)$  表示原子属性,  $w \in V_d$ ,  $t$  和  $s$  分别称为规则的条件部分和决策部分

一个决策部分带有单一决策值的规则称为明确的; 否则称为不明确的 一个元素  $x \in \Theta$  支撑规则  $t \rightarrow s$ , 定义为  $x$  在  $\mathcal{Q}$  中有两个属性  $t$  和  $s$

设  $X$  是具有属性  $t = \left( \bigcap_{a \in A_T} (a, v) \right)$  的集

合,  $Y$  是具有属性  $s = (d, w)$  的集合, 决策表中规则  $t \rightarrow s$  是真的, 当且仅当  $\overline{\Delta}X \subseteq Y$ . 其中  $C$  是规则的条件部分存在的所有属性的集合

**例 2** 表 2 是模拟有关人体健康状况的广义决策表 其中职业类型、工作强度和体育活动是条件属性(体育活动是指每周两次以上, 每次半小时以上的活动项目类别), 身体状况是决策属性 各个属性的值域分别为

- $V_{\text{职业}} = \{\text{教师, 科研人员, 职员, 工人}\},$
- $V_{\text{职业类型}} = \{\text{脑力劳动, 体力劳动}\},$
- $V_{\text{工作强度}} = \{\text{强, 中, 弱}\},$
- $V_{\text{体育活动}} = \{\text{球类, 田径}\},$
- $V_{\text{身体状况}} = \{\text{好, 一般, 不好}\}.$

用  $*$  表示不确定的属性值, 用  $\#$  表示该属性的值为空 表 2 包含一扩展列  $\hat{a}$ , 它表示系统中每个对象的一般决策值

表 2 的决策表所包含的规则如表 3 所示

表2 人体健康状况的广义决策表

论域	属性值域					决策值 $\hat{a}$
	职业	职业类型	工作强度	体育活动	身体状况	
$u_1$	教师	脑力	强	#	不好	{不好}
$u_2$	科研人员	脑力	强	#	不好	{不好}
$u_3$	科研人员	脑力	中	球类 田径	好	{好}
$u_4$	*	*	*	球类	一般	{好, 一般}
$u_5$	工人	体力	弱	*	一般	{好, 一般}
$u_6$	工人	体力	中	球类	好	{好, 一般}
$u_7$	工人	体力	弱	田径	好	{好, 一般}

表3 决策表所包含的规则

规则	条件	结论
$r_1$	(职业, 教师) (职业类型, 脑力) (工作强度, 强) (体育活动, #)	(不好)
$r_2$	(职业, 科研人员) (职业类型, 脑力) (工作强度, 强) (体育活动, #)	(不好)
$r_3$	(职业, 科研人员) (职业类型, 脑力) (工作强度, 中) ((体育活动, 球类) (体育活动, 田径))	(一般)
$r_4$	(体育活动, 球类)	(一般)
$r_5$	(职业, 工人) (职业类型, 体力) (工作强度, 弱)	(一般)
$r_6$	(职业, 工人) (职业类型, 体力) (工作强度, 中) (体育活动, 球类)	(好)
$r_7$	(职业, 工人) (职业类型, 体力) (工作强度, 弱) (体育活动, 田径)	(好)

#### 4 结 语

广义信息系统具有更广泛的应用范围。基于广义信息系统的粗糙集理论是经典的粗糙集理论和基于不完备信息系统的粗糙集理论的推广。本文提出了应用范围更广泛的数据处理方法和技術, 并给出

了决策规则的一般描述形式。与(完备的、不完备的)决策表规则的描述形式不同, 规则的条件部分的描述形式由原子属性的合取变为原子属性值的合取。这种规范的规则描述形式有利于进行推理和计算最佳规则。

广义信息系统在各个领域的应用, 如机器学习、模式识别、数据挖掘等领域, 都是需要进一步研究的课题。

#### 参考文献 (References)

- [1] Pawlak Z. Rough Sets[J]. *Int J of Computer and Information Science*, 1982, 11: 341-356
- [2] Leung Y, Li D. Maximal Consistent Block Technique for Rule Acquisition in Incomplete Information Systems[J]. *Information Sciences*, 2003, 153: 85-106
- [3] Kryszkiewicz M. Rules in Incomplete Information Systems[J]. *Information Sciences*, 1999, 113: 271-292
- [4] Kryszkiewicz M. Rough Set Approach to Incomplete Systems[J]. *Information Sciences*, 1998, 112: 39-49
- [5] 胡可云, 石纯一. 多值粗糙集模型[J]. *计算机科学*, 2001, 28(1): 1-4  
(Hu K Y, Shi C Y. Multivalued Rough Set Model[J]. *Computer Science*, 2001, 28(1): 1-4)
- [6] 周军, 佟玉军. 多值粗糙集的一般化[J]. *辽宁工学院学报*, 2002, 22(6): 57-58, 61  
(Zhou J, Tong Y J. Generalization of Multivalued Rough Set[J]. *J of Liaoning Institute of Technology*, 2002, 22(6): 57-58, 61)
- [7] 蔡自兴, 徐光祐. *人工智能原理与应用*[M]. 第3版. 北京: 清华大学出版社, 2004: 32-66  
(Cai Z X, Xu G Y. *Artificial Intelligence: Principles and Applications*[M]. Third Edition. Beijing: Tsinghua University Publisher, 2004: 32-66)

(上接第1420页)

- [7] Yue D, Han Q L, Lam J. Network-based Robust H Control of Systems with Uncertainty[J]. *Automatica*, 2005, 41(6): 999-1007
- [8] 唐功友, 刘毅敏. 时滞离散系统最优输出跟踪控制的灵敏度法[J]. *控制与决策*, 2005, 20(11): 1279-1282  
(Tang G Y, Liu Y M. Sensitivity Approach of Optimal Output Tracking Control for Discrete Systems with

Time-delay[J]. *Control and Decision*, 2005, 20(11): 1279-1282)

- [9] Tang G Y, Wang H H. Suboptimal Control for Discrete Linear Systems with Time-delay: A Non-delay Conversion Approach[J]. *Acta Automatica Sinica*, 2005, 31(3): 419-426