

规则“真”的一般概念和“真”规则形式算法

叶小平, 吉永杰, 何 花, 沈启超
(中山大学计算机科学系, 广东 广州 510275)

摘 要: 在粗糙集理论框架内研究基于不完备信息系统的决策规则“真”的一般概念, 讨论在一般“真”意义下最优决策规则的统一获取算法, 其中首先探讨了在研究不完备决策系统时的基本考虑, 然后讨论了决策规则的“真”的一般框架, 引入了在一般“真”意义下广义决策向量概念, 以此为基础, 研究了最优决策规则的形式获取算法。

关键词: 不完备决策系统; 粗糙集; 规则的“真”; “真”规则形式算法

中图分类号: TP311.131 **文献标识码:** A **文章编号:** 0529-6579 (2005) 04-0016-04

粗糙 (Rough) 集理论和技术自上个世纪 80 年代初由波兰科学家 Z. Pawlak^[1] 提出以来, 已经发展成为一种处理不精确、不完全信息的重要工具与计算方法。经典粗糙集理论以完备信息系统为研究对象, 以等价关系 (自反性、对称性、传递性) 为理论起点^[2,3]。在现实应用中, 由于数据对象属性值缺损和数据对象有可能取多值等引发的信息系统不完备现象广泛存在, 此时难以或不宜引入等价关系。这种情况促使人们将等价关系放宽为相容关系 (自反性、对称性)、相似关系 (自反性、传递性) 甚至一般二元关系^[4-8]。

如果信息系统中一部分属性和另一部分属性有着某种意义上的联系, 则可以将它们分别考虑为条件属性和决策属性, 记为 C 和 D 。此时的信息系统就称为决策系统或决策表。如果 C 和 D 分别确定论域 U 上相应关系中至少有一个是非等价关系, 就超出了完备决策系统范畴。此时需要讨论: ①基于粗糙集理论框架, 在一般关系之下描述决策规则和决策规则的“真”; ②为了使决策规则获取能够自动完成, 需要设计相应的算法, 这就要求将涉及的概念和方法形式化。

1 不完备决策系统

1.1 相关概念与基本考虑

定义 1 (不完备决策系统) 如果决策系统 $S = \{U, C \cup D\}$ 中, 由条件属性 C 或者决策属性 D 至少确定了一个非等价关系, 则称 S 为不完备决策系统。由 C 和 D 确定的关系简称为条件关系和决策关系, 分别记为 R_c 和 R_d 。

非等价关系决策系统带来的主要问题是可能导致不协调决策规则出现, 即同一个条件关系类内的对象, 其决策值不在同一个由决策属性确定的关系类中。这在应用中与直观认识相悖; 而在理论上, 又涉及到规则“真”的概念界定, 因为规则“真”要求满足“条件”的对象也应满足“结论”。这就有下面的基本考虑 1。

基本考虑 1: 决策系统中一条相对于对象 x 的决策规则为真的充分必要条件是对象 y 如果和 x 有条件关系 R_c , 则 y 和 x 也有决策关系 R_d 。

由于非等价关系的选择余地较大, 条件关系和决策关系的“配对”就有多种不同选定, 其中最自然的情况就是条件关系和决策关系属于同一种类型。另外, 关系的自反性是一个基本要求, 因此, 还可有如下的基本考虑 2。

基本考虑 2: 决策系统中条件关系和决策关系的选取是同类型的并且具有自反性质。

定义 2 (决策规则) 在决策系统 $(U, C \cup D)$, 每一个对象 x 所在的“行”看作一条“决策”规则, 从而一个决策系统 $(U, C \cup D)$ 可以看作相对于每一个对象 x 的所有具有如下形式的决策规则集: $r(x): \bigwedge (c_k, c_k(x)) \rightarrow \bigvee (d_l, d_l(x)) \quad k \in \{1, \dots, n\}, l \in \{1, \dots, m\}$ $c_k(x)$ 和 $d_l(x)$ 分别表示对象 x 关于条件属性 c_k 和决策属性 d_l 的属性值 $\bigwedge (c_k, c_k(x))$ 称为规则的条件部分, $\bigvee (d_l, d_l(x))$ 称为规则的决策部分。

定义 3 (条件类和决策类) 称对象 x 在条件

* 收稿日期: 2004-08-30

基金项目: 国家自然科学基金资助项目 (60373081); 广东省自然科学基金资助项目 (04105503)

作者简介: 叶小平 (1955 年生), 男, 博士, 副教授; E-mail: mcxyxp@zsu.edu.cn

关系 R_c 之下的关系类 $R_c(x)$ 为决策规则 $r(x)$ 的前件, 而 x 在决策关系 R_d 之下的关系类 $R_d(x)$ 为 $r(x)$ 的后件, 同时也分别称 $R_c(x)$ 和 $R_d(x)$ 为关于对象 x 的条件关系类和决策关系类。

定义 4^[1] (近似集合) 设 $X \subseteq U$, R 是论域 U 上给定的具有自反性质的关系, 定义下述 2 个 U 中子集分别为 X 的下近似集合和上近似集合:

$$\underline{R}(X) = \{x | R(x) \subseteq X\}$$

$$\overline{R}(X) = \{x | R(x) \cap X \neq \emptyset\} = \bigcup \{R(x) | x \in X\}$$

不完备决策系统规则提取的两种思路: 直接方法和间接方法, 下面先讨论间接方法, 具体见文 [4, 8]。

1.2 间接方法

间接方法主要是通过一些技术方法, 将不完备转化为完备的情形, 其中一种主要方式是借助于广义决策函数, 将不协调转化为协调情形, 然后按等价关系情况处理。

设 $S = (U, C \cup D)$ 是不完备决策系统, 其中 $C \neq \emptyset$, $D \neq \emptyset$, $C \cap D = \emptyset$ 。

定义 5^[4] (广义决策函数) 设 $C = \{a_1, a_2, \dots, a_n\}$, $D = \{d\}$, 定义映射 $\partial_{C,D}: U \rightarrow P(V_d)$ 如下: 对任意 $x (x \in U)$, $\partial_{C,D}(x) = \{i | i = d(y), y \in R_c(x)\}$ 。 $\partial_{C,D}$ 称为决策系统 S 基于条件属性 C 和决策属性域 V_d 的广义决策函数, 其中 $R_c(x)$ 为对象 x 关于条件关系 R_c 的关系类, $P(V_d)$ 为 V_d 的幂集。

对于每一个对象 x , 其广义决策函数值 $\partial_{C,D}(x)$ 就是一个决策属性值的集合。

定义 6 (诱导决策系统) 如果将每一个对象 x 和其广义决策函数值 $\partial_{C,D}(x)$ 对应, 就得到系统 S 的一个新属性, 记为 D' , 其值域就是广义决策函数在 U 上的值域。用 D' 取代 S 中原决策属性 D , 得到一个新的决策系统, 记为 $S' = (U, C \cup D')$, 称其为由 $S = (U, C \cup D)$ 得到的诱导决策系统。

定义 7 (由等价类确定“真”) 在诱导系统 $S' = (U, C \cup D')$ 中, 分别取 U 上由 C 和 D' 确定的等价关系 R_c 和 R_d' , 而设 $R_c(x)$ 和 $R_d'(x)$ 分别为对象 x 对应的等价类, 此时相对于对象 x 的决策规则: $r(x): \bigwedge (c_k, c_k(x)) \rightarrow \bigvee (d_l, d_l(x)) k \in \{1, \dots, n\}, l \in \{1, \dots, m\}$ 为“真”当且仅当 $R_c(x) \subseteq R_d'(x)$ 。

2 决策规则的“ $\bar{ij}$ -真”

2.1 决策规则的“真”

所谓直接方式, 就是直接考虑条件属性和决策

属性上的非等价的一般关系, 而不是“迂回”的返还到“等价”关系。

设有不完备决策系统 $S(U, C \cup D)$, R_c 和 R_d 为同类型的条件关系和决策关系, 对于任意 $X \subseteq U$, $\underline{R_c}(X)$ 和 $c(X)$ 分别表示 X 在关系 $\bar{R_c}$ 之下的下、上近似集合。

定义 8 定义 $A_i(C)(X)$ 为 X 的关于属性集合 C 的 i 阶近似集合, $0 \leq i \leq 2$, 其中 $A_0(C)(X) = \bar{R_c}(X)$, $A_1(C)(X) = X$, $A_2(C)(X) = \underline{R_c}(X)$ 。如果 $C = \{c_k\}$, 则将 $A_i(C)(X)$ 简记为 $A_i(c_k)(X) (0 \leq i \leq 2)$ 。

定义 9 ($\bar{ij}$ -真) 设 $A_i(C)(R_c(x))$ 为对象 x 的条件类 $R_c(x)$ 的 i 阶近似集合, $A_j(C)(R_d(x))$ 为决策类 $R_d(x)$ 的 j 阶近似集合, 其中 $0 \leq i, j \leq 2$, 相对于对象 x 的决策规则 $r(x)$ 为“ $\bar{ij}$ -真”当且仅当 $A_i(C)(R_c(x)) \subseteq A_j(C)(R_d(x))$ 。如果一个决策规则 $r(x)$ 是 $\bar{ij}$ -真的, 就称其为 $\bar{ij}$ -真规则。

定义 10 (最优决策规则) 在决策系统 $(U, C \cup D)$ 中, 决策规则 $r(x): \bigwedge (c_k, c_k(x)) \rightarrow \bigvee (d_l, d_l(x)) (k \in \{1, \dots, n\}, l \in \{1, \dots, m\})$ 称为 $\bar{ij}$ -真意义下的最优决策规则当且仅当该规则为“ $\bar{ij}$ -真”且出现在 $r(x)$ 中前件的任意合取真子集与后件的任意析取真子集构成的规则均不为“ $\bar{ij}$ -真”。

下面是常用的几个 $\bar{ij}$ -真的概念。

11—(常义真): $R_c(x) \subseteq R_d(x)$

21—真(广义真): $\bar{R_c}(R_c(x)) \subseteq R_d(x)$

01—真(确定真): $\underline{R_c}(R_c(x)) \subseteq R_d(x)$

22—真(粗糙真): $\bar{R_c}(R_c(x)) \subseteq \bar{R_d}(R_d(x))$

12—真(弱粗糙真): $R_c(x) \subseteq \bar{R_c}(R_d(x))$

各种“真”之间有以下基本关系。

定理 1 若 20-真成立, 则任意 $\bar{ij}$ -真成立, 即 20-真意义最“强”。

证明 事实上, $\forall i = 0, 1, 2, A_i(R_c(x)) \subseteq \bar{R_c}(R_c(x)) = A_2(R_c(x)); \forall j = 0, 1, 2, A_0(R_d(x)) = \bar{R_d}(R_d(x)) \subseteq A_j(R_d(x))$, 所以当 $\bar{R_c}(R_c(x)) = A_2(R_c(x)) \subseteq A_0(R_d(x)) = \bar{R_d}(R_d(x))$ 时, 成立 $A_i(R_c(x)) \subseteq A_j(R_d(x)) (i, j = 0, 1, 2)$ 。

定理 2 对于任意 $i, j = 0, 1, 2$, 若 $\bar{ij}$ -真成立, 则必有 02-真成立, 即 02-真意义最“弱”。

证明 事实上, $\forall i = 0, 1, 2, A_0(R_c(x)) = \bar{R_c}(R_c(x)) \subseteq A_i(R_c(x)); \forall j = 0, 1, 2, A_j(R_d(x)) \subseteq \bar{R_d}(R_d(x)) = A_2(R_d(x))$, 所以对任意 $i, j = 0, 1, 2$, 当 $A_i(R_c(x)) \subseteq A_j(R_d(x))$, 必有 $A_0(R_c(x)) \subseteq A_2(R_d(x))$ 。

定理 3 对于给定的 $i (i = 0, 1, 2)$, 若 $i0$ -

真成立则对所有 j ($j=0, 1, 2$), $\bar{ij}$ -真成立。

证明 对于每一个 $j=0, 1, 2$, $\underline{R}_i(R_d(x)) = A_0(R_d(x)) \subseteq A_j(R_d(x))$, 所以对给定的 $i=0, 1, 2$, 若 $A_i(R_c(x)) \subseteq A_0(R_d(x)) = \underline{R}_i(R_d(x))$, 则 $A_i(R_c(x)) \subseteq A_j(R_d(x))$, 定理成立。

定理4 对于给定的 j ($j=0, 1, 2$), 若 $2j$ -真成立则所有的 i ($i=0, 1, 2$), $\bar{ij}$ -真成立。

证明 对于每一个 $i=0, 1, 2$, $A_i(R_c(x)) \subseteq \bar{R}_i(R_c(x)) = A_2(R_c(x))$, 所以对给定的 $j=0, 1, 2$, 若 $A_2(R_c(x)) \subseteq A_j(R_d(x))$, $A_i(R_c(x)) \subseteq A_j(R_d(x))$, 定理成立。

2.2 广义决策向量和广义决策规则函数

定义11 ($\bar{ij}$ -广义决策向量) 设 $S=(U, C \cup D)$ 为不完备决策系统, x 是 U 中给定对象, $Y=U-A_j(C)(R_d(x))$, 则称如下形式的 $|Y|$ -维向量为关系类 $R_d(x)$ 对应的 $\bar{ij}$ -广义决策向量 $M_{\bar{ij}}(S)=(m(x, y_p))$, 其中 $|Y|$ 表示 Y 中元素个数, 第 p 个分量 $m(x, y_p)$ 是由对象 x 和 y_p 确定的 C 的一个属性子集, 即 $m(x, y_p) \subseteq C, (y_p \in Y, 0 \leq p \leq |Y|)$ 。确定方法为: 设 $c_k \in C, A_i(c_k)(R_c(x))$ 是对象 x 对应条件类 $R_c(x)$ 关于关系 $\{c_k\}$ 的第 i 阶近似集合, $A_j(c_k)(R_d(x))$ 是 x 对应决策类 $R_d(x)$ 关于关系 $\{c_k\}$ 的 j 阶近似集合。如果 $y_p \in A_i(c_k)(R_c(x))$, 则取 c_k 为 $m(x, y_p)$ 中元素, 否则, 取空集元素 ϕ , 即 $m(x, y_p) = \{c_k | y_p \in A_i(c_k)(R_c(x))\} (y_p \in Y, 0 \leq p \leq |Y|)$ 。

由定义11可知, $\bar{ij}$ -广义决策向量是完备决策系统^[7]的推广, 在决策矩阵情形中, 实际计算也是一个一个“行向量”地进行, 但由于基于等价关系, 每一个等价类中元素 x 等价类相同, 所以同一等价类中所有对象 x 对应的“行向量”是同维的, 形式上可以作为一个矩阵。

在前述决策规则为“真”的各个定义下, 对应的广义决策矩阵中一般元素分别为下所示。

11-真(常义真): $m(x, y) = \{c_k | y \notin R_{c_k}(x)\}, c_k \in C$, 此时 $y \in Y$, 而 $Y = U - A_1(C)(R_d(x)) = U - R_d(x)$ 。

21-真(广义真): $m(x, y) = \{c_k | y \notin \bar{R}_{c_k}(R_{c_k}(x)), c_k \in C\}$, 此时 $y \in Y$, 而 $Y = U - A_1(C)(R_d(x)) = U - (R_d(x))$ 。

01-真(确定真): $m(x, y) = \{c_k | y \notin \underline{R}_{c_k}(R_{c_k}(x)), c_k \in C\}$, 此时 $y \in Y$, 而 $Y = U - A_1(C)(R_d(x)) = U - R_d(x)$ 。

22-真(粗糙真): $m(x, y) = \{c_k | y \notin \bar{R}_{c_k}(R_{c_k}(x)), c_k \in C\}$, 此时 $y \in Y$, 而 $Y = U - A_2(C)(R_d(x)) = U - C(R_d(x))$ 。

12-真(弱粗糙真): $m(x, y) = \{c_k | y \geq \notin R_{c_k}(x), c_k \in C\}$, 此时 $y \in Y$, 而 $Y = U - A_2(C)(R_d(x)) = U - \bar{R}_i(R_d(x))$ 。

定义12 (广义决策规则函数) 取定对象 x 在 $\bar{ij}-M(S)$ 广义决策矩阵中对应的行, 现将其中的矩阵元素 $m(x, y)$ 含有的属性析取, 再将如此的析取子式合取, 得到的关于属性的合取范式称为决策系统 S 关于 $\bar{ij}$ -真的 x 类广义决策规则函数, 记为 $B_{\bar{ij}}(\bar{ij}-M(S))$, 在不引起混淆的情况下, 简记为 $B_{\bar{ij}}(x)$, 即 $B_{\bar{ij}}(x) = \bigwedge y \bigvee m(x, y) (y \in Y)$ 。其中, 如果有 $m(x, y)$ 为空集, 则在 $B_{\bar{ij}}(x)$ 中相应取 Boolean 值“1”。

按广义决策向量的定义, 取 $B_{\bar{ij}}(x)$ 的极小析取范, 则其中的每一个合取子式就是 $\bar{ij}$ -真决策规则 $r(x)$ 的最小条件属性集, 由此即得在 $\bar{ij}$ -真意义下最优决策规则。

3 $\bar{ij}$ -真最优决策规则获取算法

最优决策规则的获取是一种基于广义决策向量和广义决策规则函数的操作过程。若 $S=(U, C \cup D)$ 为一不完备决策系统, $|U|=n$ 。通过决策矩阵获取不完备决策表的所有最优决策规则的统一算法如下: ① 扫描不完备决策系统 S , 计算 S 中所有决策类 $\{R_d(x) | x \in U\}$ 。② 对于每个决策关系类 $R_d(x)$, 计算对应的广义决策向量 $M(R_d(x))$ 。③ 对每个广义决策向量 $M(R_d(x))$, 可得到函数 $B_{\bar{ij}}(x)$ 。④ 利用幂等率($a \wedge a = a$)和吸收率($a \wedge (a \vee b) = a$)将决策函数 $B_{\bar{ij}}(x)$ 化简为极小析取范式。⑤ 对于广义决策函数 $B_{\bar{ij}}(x) (x \in U)$, 当 $B_{\bar{ij}}(x) \neq 1$ 时, 该析取范式中的每一合取子式对应一条最优决策规则, 充当这条规则的条件部分, 而最优规则的决策部分等于 $(d_1, d_1(x)) \vee \cdots \vee (d_q, d_q(x))$ 。当 $B_{\bar{ij}}(x) = 1$ 时, 表示任意一个条件属性都可作为一条最优决策规则的条件部分, 且该规则的条件部分只有一个条件属性, 而规则的决策部分等于 $(d_1, d_1(x)) \vee \cdots \vee (d_q, d_q(x))$ 。然后对各 $B_{\bar{ij}}(x) (x \in U)$ 得出的规则进行合并, 则得出该不完备决策系统的最优决策规则集。

参考文献:

- [1] PAWLAK Z. Rough Sets[J]. Int J Computer and Information

- Sci, 1982(11):341 – 356.
- [2] 刘清著. Rough 集及 Rough 推理[M]. 北京: 科学出版社, 2001.
- [3] 张文修, 梁怡, 吴伟志. 信息系统与知识发现[M]. 北京: 科学出版社, 2003.
- [4] KRYSZKIEWICZ M. Rules in incomplete information systems [J]. Information Sciences, 1999, 113: 171 – 192.
- [5] 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238 – 1243.
- [6] SLOWINSKI R, VANDERPOOTEN D. A generalized definition of rough approximations based on similarity [J]. IEEE Transaction on knowledge and data engineering. 2000, 12(2): 331 – 336.
- [7] NING S, WOJCIECH Z. Data based acquisition and incremental modification of classification rules [J]. Computational Intelligence, 1995, 11(2): 357 – 370.
- [8] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112: 39 – 49.

General Discussion about the True Property of Rules and Formalized Algorithm for True Rules

YE Xiao-ping, JI Yong-jie, HE Hua, SHEN Qi-chao

(Department of Computer Science, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: The general concepts of the true property of decision rules based on incomplete information systems are studied and an unified algorithm for extracting optimal decision rules which are generally true is discussed in this paper. At first, the basic considerations about incomplete information systems are researched, and then the general frame of the true property of decision rules is discussed and the formalized concept of generalized decision vector is inducted based on the generally true meaning of rules. At last, a formalized algorithm for extracting optimal decision rules is proposed based on the above concepts.

Key words: incomplete decision-making systems; rough set; true property of rules; formalized algorithm for true rules

(上接第 15 页)

A Multiple Ant Colonies Optimization Algorithm for Finding Different Optimal Tours in the TSP

WU Hong, WANG Yuan-shi

(School of Mathematics and Computational Sciences, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: A multiple ant colonies optimization algorithm for the TSP is presented. While each ant is attracted by the pheromone of its own colony, there exists repulsion in the colony-level interactions; the first colony is not affected by other colonies; the second colony is only repelled by the first colony; the third colony is only repelled by the first and the second colonies, ... Inductively, the final colony is repelled by all the colonies before it. By introducing a time – dependent evaporation rate into the algorithm, it is presented that if there exist different optimal tours, each of the colonies is guaranteed with probability one to converge and find a different optimal tour.

Key words: ant colony optimization; travelling salesman problem; network routing