

文章编号: 1001-0920(2002)02-0255-02

粗糙集理论处理不完备信息的可信度分析

唐建国

(重庆三峡学院 电子工程系, 重庆 万州 404000)

摘要: 针对当提供的信息不完备时, 用粗糙集理论得出的决策规则其依据不充分的问题, 给出了可信度的概念和定义, 为粗糙集理论的应用提供了有用的分析工具。

关键词: 粗糙集; 不完备信息; 可信度

中图分类号: TP 202.1

文献标识码: A

Reliability analysis to deal with imperfect information by rough set theory

TAN G Jian-guo

(Electronic Engineering Department, Chongqing Three Gorges University, Wanzhou 404000, China)

Abstract: The gist of the decision rules educed by rough set theory is not sufficient when the information given is not complete. Therefore, some concepts and a definition of the reliability are given, it can offer a useful analysis tool for the application of rough set theory.

Key words: rough set; imperfect information; reliability

1 引言

由于粗糙集(RS)理论能有效地处理各种不完备的信息, 并从中发现隐含的知识, 揭示潜在的规律^[1], 因此已成为国际上数字处理和人工智能领域的一个研究热点。虽然RS理论在许多方面取得了令人鼓舞的成果, 但也存在一些问题, 如应用RS理论来推理决策规则时, 要求提供的决策数据是遍历的; 对于数据不遍历即信息不完备的情况, 可将其视为完备的^[2], 或做某种形式的处理后再进行推理决策^[3]。这样虽然简单, 但由于依据不充分, 得到的结果并不可靠, 因此不能贸然使用。本文的目的是要对这种结果给出一个可依赖程度的衡量指标, 以便在使用该结果时做到心中有数。

2 问题描述

为便于叙述, 现以文献[2, 3]所用的一个知识表达系统为例, 如表1所示。

表1 一个不完备的决策数据

样本 U	高度 R_1	头发 R_2	眼睛 R_3	分类 E
1	矮 r_{12}	黑 r_{21}	蓝 r_{31}	-
2	高 r_{11}	黑 r_{21}	蓝 r_{31}	-
3	高 r_{11}	黑 r_{21}	棕 r_{32}	-
4	高 r_{11}	红 r_{22}	蓝 r_{31}	+
5	矮 r_{12}	黄 r_{23}	蓝 r_{31}	+
6	高 r_{11}	黄 r_{23}	棕 r_{32}	-
7	高 r_{11}	黄 r_{23}	蓝 r_{31}	+
8	矮 r_{12}	黄 r_{23}	棕 r_{32}	-

收稿日期: 2001-01-11

基金项目: 重庆市教委科研基金项目(990902)

作者简介: 唐建国(1954—), 男, 重庆开县人, 副院长, 教授, 从事鲁棒控制、粗糙集理论等研究。

这里共有 3 种属性: 高度、头发、眼睛, 分别用 R_1, R_2, R_3 表示. R_1 中包含 2 个元素: 高和矮, 分别记为 r_{11} 和 r_{12} ; R_2 中包含 3 个元素: 黑、红、黄, 分别记为 r_{21}, r_{22}, r_{23} ; R_3 中包含 2 个元素: 蓝和棕, 分别记为 r_{31} 和 r_{32} . 决策分类 E 包含 2 个元素: “+”和“-”, 记为 e_1 和 e_2 . 样本集 U 中有 8 个样本, n 种属性, 每种属性有 $m_i (i = 1, 2, \dots, n)$ 个元素, 每次各抽取一个元素进行组合, 则总的组合数为 Πm_i . 因此本例中的样本数应为 $2 \times 3 \times 2 = 12$ 个. 显然, 本例的样本集还缺少 4 个样本, 这 4 个样本应为

$$U_{1'} = \{r_{12}, r_{21}, r_{32}\}, \quad U_{2'} = \{r_{11}, r_{22}, r_{32}\}$$

$$U_{3'} = \{r_{12}, r_{22}, r_{31}\}, \quad U_{4'} = \{r_{12}, r_{22}, r_{32}\}$$

这便是信息不完备的情况之一. 应用 RS 理论对表 1 进行处理, 可推理得出如下决策规则

$$[r_{21}] \quad e_2 \quad (1)$$

$$[r_{22}] \quad e_1 \quad (2)$$

$$[r_{23}, r_{31}] \quad e_1 \quad (3)$$

$$[r_{23}, r_{32}] \quad e_2 \quad (4)$$

这样得到的决策规则是否可信? 可信程度如何? 为了解答这些问题, 必须找到一个合理的衡量决策规则可信程度的指标.

3 主要结果

将衡量决策规则可信程度的指标定义为可信度, 用 η 表示. 由上面分析知, 造成我们对推理的决策规则不信任的原因是由于提供的推理决策信息不完备, 因此, 合理的可信度 η 定义应与信息的完备程度有关. 如果信息是完备的, 则可信度 η 应为 1, 即 100% 可信. 另外, 缺损的信息对每条规则的影响是不一样的, 各条规则的可信度可能也不一样, 因此, 可信度只能对每一条规则单独定义.

每一条决策规则确定了由一个样本子集到一个决策分类的映射, 例如规则 (1), $[r_{21}] \quad e_2$, 表示只要样本中含有 r_{21} , 对应的决策分类就一定是 e_2 . $[r_{21}]$ 表示表 1 中含有 r_{21} 的所有样本组成的一个子集, 即 $\{U_1, U_2, U_3\}$. 显然, 这是一个不完备的子集, 因为包含 r_{21} 的样本应有 4 个, 完备的子集应为 $\{U_1, U_2, U_3, U_{1'}\}$. 因此可将规则 (1) 的可信度定义为

$$\eta([r_{21}] \quad e_2) = \frac{\{U_1, U_2, U_3\} \text{ 中样本个数}}{\{U_1, U_2, U_3, U_{1'}\} \text{ 中样本个数}} \quad (5)$$

将其推广到一般情况, 任一规则 $[\bullet] \quad e_i$ 的可信度定义为

$$\eta([\bullet] \quad e_i) = \frac{\text{不完备子集中样本个数}}{\text{完备子集中样本个数}} \quad (6)$$

根据该定义, 可以计算出前面推理所得 4 条规则的可信度分别为

$$\eta([r_{21}] \quad e_2) = 3/4, \quad \eta([r_{22}] \quad e_1) = 1/4$$

$$\eta([r_{23}, r_{31}] \quad e_1) = 1, \quad \eta([r_{23}, r_{32}] \quad e_2) = 1$$

其中 $[r_{23}, r_{31}]$ 是指同时含有 r_{23} 和 r_{31} 的样本子集. 显然, 表 1 中这样的样本子集没有缺损, 故规则 (3) 的可信度为 1, 是完全可信的. 同理, 规则 (4) 也是完全可信的.

下面给出以上定义的可信度的一些性质.

定理 1 用 RS 理论从完备的数据表推理得出的所有决策规则都是完全可信的, 可信度为 1.

定理 1 是显然的, 不必证明.

定理 2 用 RS 理论从数据表推理得出的所有决策规则, 其可信度均为 1, 并不能说明给出的数据表一定是完备的.

证明 很容易找出说明定理 2 成立的例子, 例如: 表 2 与表 1 对照可知, 表 2 即为表 1 的 U_5, U_6, U_7, U_8 4 个样本组成.

表 2 一个不完备的决策数据

样本 U	高度 R_1	头发 R_2	眼睛 R_3	分类 E
1	矮 r_{12}	黄 r_{23}	蓝 r_{31}	+
2	高 r_{11}	黑 r_{23}	棕 r_{32}	-
3	高 r_{11}	黄 r_{23}	蓝 r_{31}	+
4	矮 r_{12}	黄 r_{23}	棕 r_{32}	-

用 RS 理论对表 2 进行推理可得到 2 条决策规则

$$[r_{23}, r_{31}] \quad e_1, \quad [r_{23}, r_{32}] \quad e_2$$

且两条规则的可信度均为 1, 但显然表 2 不是完备的, 故定理 2 成立.

由此可见, 以上可信度定义满足前面所说的要求, 是一个比较合理的定义, 而且由于它是一个构造性的定义, 十分便于计算和应用, 物理意义也很清楚.

参考文献 (References):

[1] Paw lak Z Rough sets[J] Commun of ACM, 1995, 38 (11): 89-95.
 [2] 曾黄麟 粗集理论及其应用[M] 重庆: 重庆大学出版社, 1998
 [3] 唐建国 粗糙集理论决策推理时样本缺损的处理方法[J] 重庆三峡学院学报(J of Chongqing Three Gorges Univ), 2000, 16(6): 78-80