# 粗集决策表与决策表简化的可信度比较

王德松，舒 兰

(电子科技大学应用数学学院 成都 610054)

【摘要】根据粗集决策表提供信息的完备性，借助可信度的定义，对粗集决策表和简化的决策表的决策规则的可信度进行比较，得出了简化后的决策表的决策规则的可信度高于简化前的决策表的决策规则的可信度，为粗集理论的应用提供了有用的分析工具。

关 键 词 粗集; 决策表; 决策表简化; 可信度; 决策

中图分类号 TP18 文献标识码 A

# Comparison of the Reliability between Rough Set Decision Table and Reduction of Decision Table

Wang Desong，Shu Lan

(School of Applied Mathematics, UEST of China Chengdu 610054)

**Abstract** According to the perfection to supply information by rough set decision table, the reliabilities of decision table and reduction of decision table are compared on the definition of reliability. Then we draw the contusion that reduction of decision table is superior in reliability to decision table. It is an analytical tool for the implication of rough set theory.

**Key words** rough set; decision table; reduction of decision table; reliability; decision making

近年来，文献[1, 2]等一批科学家提出了一种分析数据的数学理论，为研究对不完整数据进行分析、推理，发现数据间的关系，提取有用属性，简化信息处理，研究不精确、不确定知识的表达、学习、归纳方法等提供了一个有力的工具，而且粗集理论能够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。虽然文献[1]理论在许多方面取得了成果，但也存在一些问题，如应用文献[1]理论来推理决策时，要求提供的决策数据是完备的；对于数据不完备的情形，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。本文的目的是要对一个不完备信息的决策表和简化后的决策表的可信度进行比较，以使在使用该结果时做到心中有数。

## 1 基本概念

决策表是一类特殊而重要的知识表达系统[4]，决策表可以用知识表达系统来描述。知识表达系统(KRS)的基本成分是研究对象的集合，关于这些对象的知识可通过指定对象的基本特征(属性)和它的特征值(属性值)来描述。一个知识表达系统可表述为 $S=(U,C,D,V,f)$，其中 $U$ 是对象的集合，$C\cup D=R$ 是属性的集合，

收稿日期：2002 – 07 – 24
基金项目：国家自然科学基金资助项目(69803007)
作者简介：王德松(1973 – )，男，博士生，主要从事粗糙集理论及应用，模糊信息处理方面的研究；舒 兰(1962 – )，女，教授，主要从事粗造理论及应用、模糊信息处理方面的研究.

子集 $C$ 和 $D$ 分别称为条件属性集和结果属性集，$V=\bigcup_{r\in R}V_r$ 是属性值的集合，$V_r$ 表示了属性 $r\in R$ 的属性范围，$f:U\times R\to V$ 是一个信息函数，它指定 $U$ 中每一对象 $x$ 的属性值。这样定义的知识表达系统可以方便地用表格表达来实现。知识的表格表达法可以看成一种特殊的形式语言，它用符号表达等价关系，这样的数据表称为知识表达系统。

决策表也是知识表达系统，根据知识表达系统作定义如下：$S=(U,A,V,f)$ 为知识表达系统，其中 $A=C\cup D$，$C\cap D\neq f$，$C$ 称为条件属性集，$D$ 称为决策属性集，具有条件属性和决策属性的知识表达系统称为决策表[1]。

决策表的简化就是简化决策表的属性[3]，化简后的决策表具有与化简前的决策表相同的功能，但是，化简后的决策表具有更少的条件属性。因此，决策表简化在工程应用中相当重要。

## 2 问题描述

为了便于叙述 现以文献[5]所用的一个知识表达系统为例，如表1所示表中，$U=\{e_1,e_2,e_3,e_4,e_5,e_6,e_7,e_8\}$，$C=\{R_1,R_2,R_3\}$，其中 $R_1$=头痛，$R_2$= 肌肉痛，$R_3$= 体温，则 $R_1$ 中包含2个元素：是和否，分别记为 $r_{11}$ 和 $r_{12}$；$R_2$ 中包含两个元素：是和否，分别记为 $r_{21}$ 和 $r_{22}$；$R_3$ 中包含3个元素：正常、高和很高，分别记为 $r_{31}$、$r_{32}$ 和 $r_{33}$。决策分类 $D$ 包含2个元素：是和否，分别记为 $d_1$ 和 $d_2$。样本集 $U$ 中有8个样本，$n$种属性，每种属性有 $m_i$ ($i$ =1，2，…，$n$)个元素，每次各抽取一个元素进行组合，则总的组合数为 $\prod_{i=1}^{n}m_i$。因此，本例中的样本数应为2×2×3=12个。

表1 粗集决策表

| 样本 $U$ | 条件属性 | | | 决策属性 |
|---|---|---|---|---|
| | 头痛 $R_1$ | 肌肉痛 $R_2$ | 体温 $R_3$ | 流感 $D$ |
| $e_1$ | 是 $r_{11}$ | 是 $r_{21}$ | 正常 $r_{31}$ | 否 $d_2$ |
| $e_2$ | 是 $r_{11}$ | 是 $r_{21}$ | 高 $r_{32}$ | 是 $d_1$ |
| $e_3$ | 是 $r_{11}$ | 是 $r_{21}$ | 很高 $r_{33}$ | 是 $d_1$ |
| $e_4$ | 否 $r_{12}$ | 是 $r_{21}$ | 正常 $r_{31}$ | 否 $d_2$ |
| $e_5$ | 否 $r_{12}$ | 否 $r_{22}$ | 高 $r_{32}$ | 否 $d_2$ |
| $e_6$ | 否 $r_{12}$ | 是 $r_{21}$ | 很高 $r_{33}$ | 是 $d_1$ |
| $e_7$ | 否 $r_{12}$ | 否 $r_{22}$ | 高 $r_{32}$ | 是 $d_1$ |
| $e_8$ | 否 $r_{12}$ | 是 $r_{21}$ | 很高 $r_{33}$ | 否 $d_2$ |

显然，本例中的样本集为

$$\{U_1,U_2,U_3,U_4,U_5,U_6\}$$

其中

$$U_1=\{r_{11},r_{21},r_{31}\},U_2=\{r_{11},r_{21},r_{32}\},U_3=\{r_{11},r_{21},r_{33}\}$$
$$U_4=\{r_{12},r_{21},r_{31}\},U_5=\{r_{12},r_{22},r_{32}\},U_6=\{r_{12},r_{21},r_{33}\}$$

还缺少6个样本(因为条件属性完全相同归为一类，只能算一个，即 $e_5$ 与 $e_7$ 相同，$e_6$ 与 $e_8$ 相同，这里只取 $e_5$、$e_6$)，这缺少的6个样本应为

$$U_{1^*}=\{r_{11},r_{22},r_{31}\}，U_{2^*}=\{r_{11},r_{22},r_{32}\}，U_{3^*}=\{r_{11},r_{22},r_{33}\}$$
$$U_{4^*}=\{r_{12},r_{22},r_{31}\}，U_{5^*}=\{r_{12},r_{21},r_{32}\}，U_{6^*}=\{r_{12},r_{22},r_{33}\}$$

根据可信度的定义[6]：任一规则 $[\cdot]\to d_i$ 的可信度定义为

$$h([\cdot]\to d_i)=\frac{\text{不完备子集中样本个数}}{\text{完备子集中样本个数}}$$

式中 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)$[r_{11},r_{32}]\to d_1$，表示只要样本中同时含有 $r_{11}$ 和 $r_{32}$，对应的决策分类就一定是 $d_1$。$[r_{11},r_{32}]$ 表示表1中同时含有 $r_{11}$ 和 $r_{32}$ 的所有样本组成的一个子集，即 $\{U_2\}$。显然，这是一个不完备的子集，因为同时包含 $r_{11}$ 和 $r_{32}$ 的样本应有2个，完备的子集应为 $\{U_2,U_{2^*}\}$，因此其可信度 $h([r_{11},r_{32}]\to d_1)=\frac{\{U_2\}\text{中样本个数}}{\{U_2,U_{2^*}\}\text{中样本个数}}=\frac{1}{2}$。再如规则(2)$[r_{31}]\to d_2$，表示只要样本中含有 $r_{31}$，对应的决策分类就一定是 $d_2$。$[r_{31}]$表示表1中含有 $r_{31}$ 的所有样本组成的一个子集，即 $\{U_1,U_4\}$。显然，这是一个不完备的子集，因为同时包含 $r_{31}$ 的样本应有4个，完备的子集应为 $\{U_1,U_{1^*},U_4,U_{4^*}\}$，其可信度

$$h([r_{31}]\to d_2)=\frac{\{U_1,U_4\}\text{中样本个数}}{\{U_1,U_{1^*},U_4,U_4\}\text{中样本个数}}=\frac{2}{4}=\frac{1}{2}$$

同理，$h([r_{33}]\to d_1)=\dfrac{\{U_3,U_6\}\text{中样本个数}}{\{U_3,U_{3'},U_6,U_6\}\text{中样本个数}}=\dfrac{2}{4}=\dfrac{1}{2}$。

$h([r_{12},r_{33}]\to d_1)=\dfrac{\{U_6\}\text{中样本个数}}{\{U_6,U_6\}\text{中样本个数}}=\dfrac{1}{2}$，$h([r_{11},r_{31}]\to d_2)=\dfrac{\{U_1\}\text{中样本个数}}{\{U_1,U_1\}\text{中样本个数}}=\dfrac{1}{2}$，

$h([r_{12},r_{32}]\to d_2)=\dfrac{\{U_5\}\text{中样本个数}}{\{U_5,U_5\}\text{中样本个数}}=\dfrac{1}{2}$。

根据决策表简化理论，化简后的决策表如表2所示，从表中看出简化后的决策表是一个完备信息的决策表，根据可信度的定义，其全部决策的可信度都为1，如下所示

$$h([r_{11},r_{32}]\to d_1)=\frac{\{U_2\}\text{中样本个数}}{\{U_2\}\text{中样本个数}}=\frac{1}{1}=1$$

$$h([r_{31}]\to d_2)=\frac{\{U_1,U_4\}\text{中样本个数}}{\{U_1,U_4\}\text{中样本个数}}=\frac{2}{2}=1$$

$$h([r_{12},r_{33}]\to d_1)=\frac{\{U_6\}\text{中样本个数}}{\{U_6\}\text{中样本个数}}=\frac{1}{1}=1$$

$$h([r_{33}]\to d_1)=\frac{\{U_3,U_6\}\text{中样本个数}}{\{U_3,U_6\}\text{中样本个数}}=\frac{2}{2}=1$$

表2　简化后的粗集决策表

| 样本 $U$ | 条件属性 | | 决策属性 |
|---|---|---|---|
| | 头痛 $R_1$ | 体温 $R_3$ | 流感 $D$ |
| $e_1$ | 是 $r_{11}$ | 正常 $r_{31}$ | 否 $d_2$ |
| $e_2$ | 是 $r_{11}$ | 高 $r_{32}$ | 是 $d_1$ |
| $e_3$ | 是 $r_{11}$ | 很高 $r_{33}$ | 是 $d_1$ |
| $e_4$ | 否 $r_{12}$ | 正常 $r_{31}$ | 否 $d_2$ |
| $e_5$ | 否 $r_{12}$ | 高 $r_{32}$ | 否 $d_2$ |
| $e_6$ | 否 $r_{12}$ | 很高 $r_{33}$ | 是 $d_1$ |

## 3　主要结果

简化后的决策表的决策规则的可信度高于简化前的决策表的决策规则的可信度。其原因是由于条件属性的减少，更易于决策，从而提高了决策规则的可信度。

## 4　结 束 语

粗集理论是一种新颖、有效的软科学方法，能够分析和处理不完备信息，对不确定信息处理的方式及与其他软科学和软计算方法的结合，是人工智能领域的进一步发展方向，尤其在机器学习、知识获取、决策分析及其他各个方面的应用，粗集理论为之提供了一种有效的新的数学方法。本文根据粗集决策表提供信息的完备性，借助可信度的定义，对粗集决策表和简化的决策表的决策规则的可信度进行了比较，以使在使用该结果时做到心中有数，同时为粗集理论的应用提供了有用的分析工具。

## 参 考 文 献

[1] Pawlak Z. Rough sets[J]. Communication 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]. 重庆山峡学院学报, 2000, 16(6):78-80

[4] Pawlak Z. Rough classification[J]. Int. J. Man-Machine Studies, 1984,20:469-483

[5] 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

[6] 唐建国. 粗糙集理论处理不完备信息的可信度分析[J]. 控制与决策, 2002, 17: 255-256

编　辑　刘文珍