学术论文

一种模糊Rough决策方法

罗 党[1,2]
(1. 南京航空航天大学经济管理学院，南京 210016；2. 华北水利水电学院，郑州 450011)

[摘要] 利用模糊集理论和粗糙集理论在处理不确定性和不精确性问题方面侧重点的差异性，构造一种组合决策模型。该模型从问题领域内的部分不精确信息出发利用模糊聚类方法构造一个决策信息系统，利用粗糙集理论关于决策规则的约简方法从决策信息系统中提取（挖掘）决策规则，使之适用于问题的整个领域。

[关键词] 模糊聚类；粗糙集；决策表；决策规则
[中图分类号] O159 [文献标识码] A [文章编号] 1009-1742（2004）12-0032-05

1 引言

1965年计算机科学专家L. A. Zadeh教授提出了模糊集[1]，经过近40年的研究和实践，模糊集理论已发展成为具有坚实数学基础和良好结构的概念和技术体系[2]。在诸多非经典数学中，发展最快、应用最多的就是模糊数学[3]，但是模糊集主要着眼于知识的模糊性，强调的是集合边界的不分明性，模糊集是不可计算的，没有给出数学公式描述这一含糊概念，无法计算出它的边界线上的具体的含糊元素数目。模糊集着重研究属于同一类的不同对象的隶属程度[4]。在处理不确定问题时，需要提供所需处理的数据集合之外的某些先验信息，在自动控制以外的与人的认识相关的领域中成功应用很少。因此，从认识方面探索人的信息处理过程，完善模糊集理论以拓展其应用，成为十分有意义的研究课题[2]。

1982年波兰数学家Z. Pawlak提出了一种处理模糊和不确定性问题的新型数学工具，1991年Z. Pawlak的专著《Rough Sets——Theoretical Aspects of Reasoning About Data》[5]的问世，标志着粗糙集理论及其应用的研究进入了活跃时期。目前粗糙集理论在信息科学、管理科学、金融、医学等众多学科都取得了成功的应用。无疑自粗糙集问世以来，无论在理论或应用上都是一种新的、最重要的并且是迅速发展的一门既有理论又有应用的研究领域[6]。粗糙集理论在处理模糊和不确定性问题方面着眼于知识的粗糙（rough）性，强调的是集合对象间的不可分辨性，着重研究的是不同类中的对象组成的集合之间的关系。在处理不确定性问题时，不需要提供问题所需处理的数据集合之外的任何先验信息。但是粗糙集理论不具备处理模糊或不精确原始数据的机制。由于模糊集和粗糙集都可以用来描述知识的不确定性，各自的特点不同，因此2种理论有着很强的互补性，将这2种理论进行适当整合后来处理知识的模糊和不完全性，比它们各自去处理知识的不确定性和不完全性会显出更强的功能。基于上述原因，国内外已有不少学者为此而进行研究，取得了一系列的成果[2, 7~10]。粗糙集理论与模糊集等理论的融合是一个有着广阔发展前景的领域，基于此，笔者提出了一种模糊粗糙集决策模型，该模型从问题领域内的部分不精确信息

[收稿日期] 2003-11-11；修回日期 2004-01-17
[基金项目] 国家自然科学基金资助项目（10071074）；国家博士学科点科研基金资助项目（20020287001）；江苏省自然科学基金重点资助项目（BK2003211）；河南省自然科学基金资助项目（2003120001）
[作者简介] 罗 党（1959-），男，河南汝南县人，华北水利水电学院副教授，南京航空航天大学博士生

出发利用模糊聚类方法构造一个决策信息系统，利用粗糙集理论关于决策算法的约简方法，从决策信息系统中提取决策规则，使之适用于问题的整个领域。

2　基本概念

重点介绍与文中模型有关的粗糙集理论的基本概念

定义 1　一个决策表是一个信息知识表达系统 $S=(U, A, V, f)$，其中 $U=\{x_1, x_2, \cdots, x_n\}$ 为对象的非空有限集合，称为论域；$A=P\cup D$ 是属性集合，子集 $P=\{a_i|i=1, 2, \cdots, m\}$ 和 $D=\{d\}$ 分别称为条件属性集和决策属性集，$D\neq\varnothing$；$V=\bigcup_{a\in A}V_a$，V_a 是属性 a 的值域；$F: U\times A\to V$ 是一个信息函数；$a_i(x_j)$ 是样本 x_j 在属性 a_i 上的取值。$(\boldsymbol{C}_\mathrm{D})_{ij}$ 表示可辨识矩阵中第 i 行第 j 列的元素，则可辨识矩阵 $\boldsymbol{C}_\mathrm{D}$ 定义为

$$(\boldsymbol{C}_\mathrm{D})_{ij}=\begin{cases}\{a_k|a_k\in P\wedge a_k(x_i)\neq a_k(x_j)\} & d(x_i)\neq d(x_j),\\ \varnothing & d(x_i)=d(x_j),\end{cases}$$

其中 $i, j=1, 2, \cdots, n$。

定义 2　称布尔函数 $\Delta=\prod_{(x,y)\in U\times U}\sum(\boldsymbol{C}_\mathrm{D})_{xy}$ 为可分辨函数，其中 $(\boldsymbol{C}_\mathrm{D})_{xy}$ 是决策表信息系统 $S=(U, P\cup D, V, f)$ 的可分辨矩阵 $\boldsymbol{C}_\mathrm{D}$ 中的任一元素。当 $(\boldsymbol{C}_\mathrm{D})_{xy}=\{a_1, a_2, \cdots, a_k\}\neq\varnothing$ 时，$\sum(\boldsymbol{C}_\mathrm{D})_{xy}=a_1\vee a_2\vee\cdots\vee a_k$；当 $(\boldsymbol{C}_\mathrm{D})_{xy}=\varnothing$ 时，$\sum(\boldsymbol{C}_\mathrm{D})_{xy}=1$。

命题 1　决策表信息系统 $S=(U, P\cup D, V, f)$ 的可分辨函数 Δ 的极小析取范式中的所有合取式是属性集 $P\cup D$ 的所有约简。

定义 3　令决策表 $S=(U, C\cup D, V, f)$，其中 C 为条件属性集，D 为决策属性集。称公式 $(a_1, v_1)\wedge(a_2, v_2)\wedge\cdots\wedge(a_n, v_n)$ 为 P 基本公式，其中 $v_i\in V_{a_i}$，$a_i\in P(i=1, 2, \cdots, n)$，$P\subset C$。

定义 4　如果 $A\to B$ 为决策规则，且 A 是 P 基本公式，$B=(d, d_i)$，则称 $A\to B$ 为基本决策规则。又如果 $\varphi_1\to\psi$，$\varphi_2\to\psi$，$\cdots$，$\varphi_n\to\psi$ 是基本决策规则，则决策规则 $\varphi_1\vee\varphi_2\vee\cdots\vee\varphi_n\to\psi$ 被称为基本决策规则 $\varphi_1\to\psi$，$\varphi_2\to\psi$，$\cdots$，$\varphi_n\to\psi$ 的组合。

定义 5　在决策表 $S=(U, C\cup D, V, f)$ 中如果决策规则 $A\to B$ 为真，即决策表中的所有实例都满足决策规则 $A\to B$，则称决策规则 $A\to B$ 在决策表 S 中是一致的；否则称决策规则 $A\to B$ 在决策表 S 中是不一致的。

命题 2　设 $\varphi\to\psi$ 是决策表上的一条决策规则，属性值 v 是一可被约去的，当且仅当 $\vdash(\varphi\to\psi)\to(\varphi\setminus\{(a, v)\}\to\psi)$，其中 φ 和 ψ 均为决策表上的基本公式。

命题 3　设 d_x 是一条被消去所有过剩条件属性值的决策规则，条件属性集 C 的等价类 $[x]_\mathrm{C}$ 中任何最小属性 A 的等价类 $[x]_\mathrm{A}$ 的交集包含于相应决策类 $[x]_\mathrm{D}$ 中，则由此而得到的最小条件属性 A 组成的相应于 d_x 的新决策规则 $d_{x'}$ 是 d_x 的一个决策规则约简。

命题 4　任何决策都可以分解成为一个或多个等价的基本决策规则。

3　模糊 Rough 集决策方法

3.1　模糊 Rough 集决策方法的基本思想

模糊聚类分析有模糊等价关系法和模糊 ISODATA 法等。模糊聚类所处理的问题是已知待分类对象（样本）集 $X=(X_1, X_2, \cdots, X_n)$，每个对象（样本）都有 m 个特征（属性）且每个对象的 m 个特征值是已知的，按照问题的目的要求将对象划分为若干类。但是模糊聚类结果只是对已知样本作出的抉择，不具有柔性，不能通过已知的信息及聚类结果对问题所涉及领域内的新样本的类别作出决策。克服这一不足之处，显然是十分有意义的。

Rough 集理论的主要研究对象就是决策表，表中包含了领域内一定数量的样本（实例）信息。决策表中的一个样本就代表一条基本决策规则，如果把所有这样的决策规则罗列出来，就可以得到一个决策规则集合。但是，这样的决策规则集合是没有多大价值的，因为其中的基本决策规则没有适应性，只是机械地记录了一个样本的情况，不能适应新的、其他的情况。为了从领域内局部（或部分）样本信息形成的决策表中抽取得到适应度大的规则，需要对决策表进行约简，使得经过约简处理的决策表中的一个记录就代表一类具有相同规律特征的样本，这样得到的决策规则就具有较高的适应性。

当前常用的属性约简方法有数据分析方法及可分辨矩阵方法。经过决策表属性约简，省略了对决

策分类不必要的属性，从而实现决策表的简化，这有利于从决策表中分析发现对决策分类起作用的属性。但属性约简只是在一定程度上去掉了决策表中的冗余属性，还没有充分去掉决策表中的冗余信息，需要进一步对决策表进行处理得到更加简化的决策表，这个过程就是对决策表进行属性值的约简。以命题2和命题3为依据的数据分析方法[6]是当前常用的属性值约简方法。

对于一个经过属性约简和属性值约简的决策表，再从中消去所有过剩的决策规则。这样得到的决策表中每一个记录就代表一条基本决策规则。依据命题4将这些基本决策进行适当的组合，即得到具有适应性的一组或多组决策规则。

3.2 模糊 Rough 集决策方法的步骤

归纳上述基本思想可得模糊 Rough 集决策方法的步骤如下：

Step 1 将已知样本（实例）及样本的特征（属性）值组成一个信息矩阵；

Step 2 特征预处理即选择最有代表性的特征，区分出派出特征和相关特征，然后进行无量纲化的规格化处理，使各特征在[0，1]中取值；

Step 3 对已做了特征预处理的信息矩阵建立模糊相似关系 R；

Step 4 用平方法求 R 的等价闭包 $t(R)$；

Step 5 做 λ 截集，以此分类；

Step 6 选取适合聚类目标的分类，将各个样本（实例）所在的类作为决策属性值与已知信息矩阵组合成决策信息系统；

Step 7 对决策信息系统进行属性约简，对所有约简结果写出相应的决策信息系统；

Step 8 分别对 Step 7 中得到的每一个决策表进行决策规则约简，即属性值的约简；

Step 9 从经过属性约简和属性值约简的决策表中消去所有过剩的决策规则；

Step 10 将 Step 1 得到的决策表中基本决策规则进行组合。

4 应用实例

取文献[3]中的一个实例来说明上述决策方法。

已知环境单元分类信息：每个环境单元可以包括空气、水分、土壤、作物4个要素（特征）。环境单元的污染状况由污染物在4要素中含量的超限度来描写。假设有5个单元（样本）x_1，x_2，x_3，x_4，x_5，其污染数据如表1所示。

表1 污染数据

Table 1 Contaminate data

	空气	水分	土壤	作物
x_1	5	5	3	2
x_2	2	3	4	5
x_3	5	5	2	3
x_4	1	5	3	1
x_5	2	4	5	1

取论域 $U = \{x_1, x_2, x_3, x_4, x_5\}$，按下述公式对各单元的要素值 x_{ij}^0 进行规格化预处理(取 $C_1 = 0.1$)。$x_{ij} = C_1 x_{ij}^0 (i = 1, 2, 3, 4, 5, j = 1, 2, 3, 4)$，再按公式 $r_{ij} = 1 - C_2 \sum_{k=1}^{4} \left| x_{ik} - x_{jk} \right| (i, j = 1, 2, 3, 4, 5,)$ 求相似系数(取 $C_2 = 1$)，得模糊相似矩阵

$$\underset{\sim}{\boldsymbol{R}} = (r_{ij})_{5\times5} = \begin{bmatrix} 1 & 0.1 & 0.8 & 0.5 & 0.3 \\ 0.1 & 1 & 0.1 & 0.2 & 0.4 \\ 0.8 & 0.1 & 1 & 0.3 & 0.1 \\ 0.5 & 0.2 & 0.3 & 1 & 0.6 \\ 0.3 & 0.4 & 0.1 & 0.6 & 1 \end{bmatrix}。$$

利用平方法求 $\underset{\sim}{\boldsymbol{R}}$ 的传递闭包

$$t(\underset{\sim}{\boldsymbol{R}}) = \underset{\sim}{\boldsymbol{R}}^8 = \underset{\sim}{\boldsymbol{R}}^4 = \begin{bmatrix} 1 & 0.4 & 0.8 & 0.5 & 0.5 \\ 0.4 & 1 & 0.4 & 0.4 & 0.4 \\ 0.8 & 0.4 & 1 & 0.5 & 0.5 \\ 0.5 & 0.4 & 0.5 & 1 & 0.6 \\ 0.5 & 0.4 & 0.5 & 0.6 & 1 \end{bmatrix},$$

作出截集，由此得其动态分类：

当 $\lambda = 1$ 时，分为五类 $\{x_1\}$，$\{x_2\}$，$\{x_3\}$，$\{x_4\}$，$\{x_5\}$；

当 $\lambda = 0.8$ 时，分为四类 $\{x_1, x_3\}$，$\{x_2\}$，$\{x_4\}$，$\{x_5\}$；

当 $\lambda = 0.6$ 时，分为三类 $\{x_1, x_3\}$，$\{x_2\}$，$\{x_4, x_5\}$；

当 $\lambda = 0.5$ 时，分为二类 $\{x_1, x_3, x_4, x_5\}$，$\{x_2\}$；

当 $\lambda = 0.4$ 时，分为一类 $\{x_1, x_2, x_3, x_4, x_5\}$。

假设按问题的目标要求需将环境单元（样本）划分为三类，则聚类结果应取为 $\{x_1, x_3\}$，$\{x_2\}$，$\{x_4, x_5\}$，由已知数据信息和聚类结果构造决策表见表2。

表 2　环境单元分类决策表

Table 2　Decision table for environment cell sort

U	空气 k	水分 s	土壤 t	作物 z	决策 d
x_1	5	5	3	2	1
x_2	2	3	4	5	2
x_3	5	5	2	3	1
x_4	1	5	3	1	3
x_5	2	4	5	1	3

利用可辩识矩阵方法对决策 2 进行属性约简，由定义 1 计算决策表 2 的可辨识矩阵见表 3。

表 3　决策表 2 的可辨识矩阵

Table 3　Distinguishable matrix for table 2

	x_1	x_2	x_3	x_4	x_5
x_1	$\varnothing$	$kstz$	$\varnothing$	kz	$kstz$
x_2		$\varnothing$	$kstz$	$kstz$	stz
x_3			$\varnothing$	ktz	$kstz$
x_4				$\varnothing$	$\varnothing$
x_5					$\varnothing$

由定义 2 经计算可得表 3 对应的可分辨函数 $\Delta = (k \wedge t) \vee (k \wedge s) \vee z$ 于是根据命题 2 得到属性的 3 个约简结果，分别如表 4、表 5 和表 6 所示（合并重复记录后的决策表）

表 4　约简结果 $\{k, t\}$

Table 4　Reducing result $\{k, t\}$

U	x_1	x_2	x_3	x_4	x_5
k	5	2	5	1	2
t	3	4	2	3	5
d	1	2	1	3	3

表 5　约简结果 $\{k, s\}$

Table 5　Reducing result $\{k, s\}$

U	x_1	x_2	x_3	x_4	x_5
k		2	5	1	2
s		3	5	5	4
d		2	1	3	3

表 6　约简结果 $\{z\}$

Table 6　Reducing result $\{z\}$

U	x_1	x_2	x_3	x_4	x_5
z	2	5	3	1	
d	1	2	1	3	

用数据分析法分别对决策表 4、表 5、表 6 进行属性值约简。

对于一个经过属性约简而得到的决策表，可以对应其中的每一个样本形成一条决策规则。因此，可以将决策表中的样本用规则来表示，这样，约简后的决策表实际上就是规则集合。称这样的规则集合为决策算法。

对于决策表 4 其决策规则为

a. $k_5\ t_3 \to d_1$；**b**. $k_2\ t_4 \to d_2$；**c**. $k_5\ t_2 \to d_1$；**d**. $k_1\ t_3 \to d_3$；**e**. $k_2\ t_5 \to d_3$。

首先计算第一条决策规则的约简，其决策类 $[x_1]d = \{x_1, x_3\}$，$[x_1]k = \{x_1, x_3\}$，$[x_1]t = \{x_1, x_4\}$，由于 $[x_1]_k \subset [x_1]_d$，$[x_1]_t \not\subset [x_1]_d$，得约简的决策规则，即 **a**. $k_5 \to d_1$。计算第二条决策规则的约简，其决策类 $[x_2]_d = \{x_2\}$，$[x_2]_k = \{x_2, x_5\}$，$[x_2]_t = \{x_2\}$，由于 $[x_2]_k \not\subset [x_2]_d$，$[x_2]_t \subset [x_2]_d$，得约简的决策规则，即 **b**. $t_4 \to d_2$。计算第三条决策规则的约简，其决策类 $[x_3]_d = \{x_1, x_3\}$，$[x_3]_k = \{x_1, x_3\}$，$[x_3]_t = \{x_3\}$，由于 $[x_3]_k \subset [x_3]_d$，$[x_3]_t \subset [x_3]_d$，所以得 2 条约简的决策规则，**c**. $k_5 \to d_1$；$t_2 \to d_1$。同理可得 **d**，**e** 的约简，所有的约简决策规则列于表 7。

表 7　包含（表 4）所有约简的决策规则

Table 7　All minimal solution for table 4

U	x_1	x_2	x_3	$x_{3'}$	x_4	x_5
k	5		5		1	
t		4		2		5
d	1	2	1	1	3	3

用同样的方法可计算并列出表 5 对应的所有约简的决策规则于表 8。

表 8　包含（表 5）所有约简的决策规则

Table 8　All minimal solution for table 5

U	x_1	x_2	x_3	x_4	x_5
k		5		1	
s			3		4
d		1	2	3	3

决策表 6 中每一条规则都是由最小条件属性组成的决策规则。分别对表 6、表 7、表 8 消去所有过剩的决策规则。表 6、表 8 中没有过剩的决策规则。对于表 7，其中第一条决策规则是冗余的，它与第二条决策规则相同。删除表 7 中过剩决策规则后得表 9。

表9 删去（表7）过剩约简决策规则的决策表

Table 9 Requiring minimal number of gates for table 7

U	x_1，x_3	x_2	$x_{3'}$	x_4	x_5
k	5			1	
t		4	2		5
d	1	2	1	3	3

分别将表6、表8、表9中的决策规则进行组合。

表6对应的决策规则组合为

$$z_2 \vee z_3 \rightarrow d_1$$
$$z_5 \rightarrow d_2 \qquad (1)$$
$$z_1 \rightarrow d_3$$

表8对应的决策规则组合为

$$k_5 \rightarrow d_1$$
$$s_3 \rightarrow d_2 \qquad (2)$$
$$k_1 \vee s_4 \rightarrow d_3$$

表9对应的决策规则组合为

$$k_5 \vee t_2 \rightarrow d_1$$
$$t_4 \rightarrow d_2 . \qquad (3)$$
$$k_1 \vee t_5 \rightarrow d_3$$

注：1）本例得到由式（1）至式（3）式所示的3个决策规则组合，在实际应用中可由决策者根据具体情况进行选择。

2）应用模糊Rough集决策方法解决实际问题时，须将得到的组合决策规则转换成自然语言。

5 结论

分析了模糊聚类分析和粗糙决策算法约简的优势及其互补性。并给出了一种模糊粗糙决策方法，丰富和发展了模糊决策理论，拓展了模糊决策的适应范围。

参考文献

[1] Zadeh L A. Fuzzy sets [J]. Information and Control, 1965, (8): 338～353

[2] 陈有安，郑丕谔，李光泉. 粗糙度量和模糊信息颗粒化 [J]. 系统工程学报，2001，16(4)：241～247

[3] 朱剑英. 智能系统非经典数学方法 [M]. 武汉：华中科技大学出版社，2001，138～141

[4] 刘叙华. 模糊逻辑与模糊推理 [M]. 长春：吉林大学出版社，1989

[5] Pawlak Z. Rough Sets-Theoretical Aspects of Reasoning About Data [M]. Kluwer Academic Publishers, Dordrecht, 1991

[6] 刘 清. Rough集及Rough推理 [M]. 北京：科学技术出版社，2001

[7] Tanaka T, Ishibuchi H, Matsuda N. Fuzzy expert system based on rough sets and its application to medical diagnosis [J]. International Journal of General Systems, 1992, 21: 83～97

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

[9] Adam M, Leszek P. Knowledge representation in fuzzy and rough controllers [J]. Fundamenta Informaticae 1994, 30: 299～311

[10] 王国胤. Rough集理论与知识获取 [M]. 西安：西安交通大学出版社，2001

A Decision-making Method Based on Fuzzy Sets and Rough Sets Theory

Luo Dang[1,2]

(1. *College of Economics and Management*, *Nanjing University of Aeronautics and Astronautics*, *Nanjing* 210016, *China*; 2. *North China Institute of Water Conservancy and Hydroelectric Power*, *Zhengzhou* 450011, *China*)

[**Abstract**] In this paper, a combined decision-making model is presented for dealing with uncertain and imprecise problems, based on the difference between fuzzy sets and rough sets theories. The model is firstly to classify the uncertain examples given using fuzzy clustering analysis, to make a decision table specified, and then to simplify the decision table by means of the rough sets theory. It gives all possible minimal decision algorithms associated with the decision table. These decision algorithms have found application in larger field.

[**Key words**] fuzzy clustering; rough sets; decision table; decision rule