1000-9825/2004/15(10)1470 ©2004 Journal of Software 软 件 学 报 Vol.15, No.10

模糊目标信息系统上的知识约简方法*

管 涛+, 冯博琴

(西安交通大学 电子与信息工程学院,陕西 西安 710049)

Knowledge Reduction Methods in Fuzzy Objective Information Systems

GUAN Tao+, FENG Bo-Qin

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

+ Corresponding author: Phn: +86-29-82664160, E-mail: tguan@ctec.xjtu.edu.cn, http://www.xjtu.edu.cn

Received 2004-03-16; Accepted 2004-05-08

Guan T, Feng BQ. Knowledge reduction methods in fuzzy objective information systems. *Journal of Software*, 2004,15(10):1470~1478.

http://www.jos.org.cn/1000-9825/15/1470.htm

Abstract: Fuzzy objective information systems (FOISs) exist in many applications, and knowledge reduction in them can't be implemented by reduction methods in Pawlak information systems. This paper firstly presents new reduction methods including α distribution, α maximum distribution, α assignment, and rough distribution reductions. It then probes into their properties and the relation between them and the reduction methods on Pawlak information systems. Furthermore, this paper proposes the judgement theorems and discernibility matrixes with respect to these reductions. These reductions extend the corresponding methods in Pawlak information systems and provide a new and low computation complexity way for knowledge discovery and rough-fuzzy rule based fuzzy concept classifiers in fuzzy objective information systems.

Key words: rough set; fuzzy objective information system; knowledge reduction; distribution reduction; assignment reduction

摘 要: 模糊目标信息系统(fuzzy objective information systems,简称FOISs)在许多实际应用中存在,这种系统上的知识简化不能采用Pawlak信息系统上的约简方法.因此,提出了模糊目标信息系统上的α分布约简、α最大分布约简、α分配约简、粗糙分布约简,并给出了它们的性质以及与Pawlak信息系统上约简的关系,同时也给出了这些约简的判定定理、对应的可辨识矩阵、约简公式.这些约简推广了Pawlak信息系统上的知识约简方法,为模糊目标信息系统上的知识发现和基于粗糙模糊规则的模糊概念分类器提供了新的低复杂性手段.

关键词: 粗糙集;模糊目标信息系统;知识约简;分布约简;分配约简

中图法分类号: TP18 **文献标识码**: A

* Supported by the National High-Tech Research and Development Plan of China under Grant No.2003AA1Z2610 (国家高技术研究发展计划(863))

作者简介: **管涛**(1974－),男,河南商城人,博士生,主要研究领域为模糊集与粗糙集,机器学习,数据挖掘;**冯博琴**(1942－),男,教授,博士生导师,主要研究领域为智能决策系统,编译系统.

自从粗糙集理论[1]被提出以来,它已被广泛应用于规则生成、医疗(故障)诊断、分类等领域.在不确定信息获取领域,如知识获取和信息粒度刻画,粗糙集呈现出独特优势.知识约简是信息系统知识发现的一个重要过程,目前的知识约简主要面向 Pawlak 信息系统,如协调信息系统、不协调信息系统、可变精度粗集模型(VPRSs)等[2~10].通过研究 Pawlak 对象空间的条件划分和决策划分之间的关系,不少研究者给出了一些知识约简方法,如 Kryszkiewicz 提出了知识约简的 5 个概念[3],其中可能约简和约简(也称分布约简[4])是两个主要的方式;Beynon 提出了不协调信息系统上的 β 下近似约简[5];张文修等人发展了分布约简方法,提出了不协调信息系统上的最大分布约简的概念[6].在 VPRSs 方面,米据生等人提出了 β 下分布约简和 β 上分布约简[7].Slezak 等人在文献[8]中给出了贝叶斯 VPRSs 上的知识约简方法,进而,在文献[9]中提出了一种保持所有对象类成员分布不变的约简方法,这种方法等价于保持广义推断度量函数不变的约简方法[10].此外,通过允许在用户指定的参数范围内改变广义推断函数的值,他们引入了广义约简方法.这些约简建立在 Pawlak 信息系统或其扩展模型之上,然而,当决策属性是模糊的时候,如学生考试成绩(0~100 分)的"优"、"良"、"中"、"差"等,Pawlak 信息系统上的约简方法则不再有效.

目前,模糊目标信息系统(fuzzy objective information systems,简称 FOISs)上的知识约简方法主要是基于模糊目标的粗糙近似精度[2],它们分别是 (α,β) 约简和 (α,β) 弱约简,但这种约简只是估计了模糊目标函数上的 (α,β) 水平集,并没有考虑对象空间的模糊划分与对象特征划分之间的关系,也没有利用对象空间的目标分割特征函数水平集.另外,弱约简可能导致决策规则的改变.例如,在眼科诊断中得到的"近视度"为一系列连续的数值,而"低度"和"高度"是由"近视度"导出的对象空间的模糊函数.因此,基于目标函数"近视度"水平集的约简方法对于有意义的诊断决策意义不大,可能得不到预想的诊断推理规则和最小知识集.

这里,不直接在模糊目标上进行分割,而是对模糊目标特征产生的对象空间模糊划分水平分割,并结合 Pawlak 信息系统上的相关概念和方法,给出了几种新的模糊目标属性信息系统上的属性约简概念和形式,包括 α 分布约简、α 最大分布约简、α 分配约简、(α,β) 粗糙分布约简,进而给出了相应的判定定理、不可分辨矩阵和约简公式.

1 粗糙集概念和模糊目标信息系统

1.1 粗糙集概念

在 Pawlak 粗糙集理论中[1],近似空间 (U,A) 中每一个非空子集 $B\subset A$ 决定一个不可分辨关系 $R_B=\{(x,y)\in U\times U:a(x)=a(y),\forall a\in B\}$,$R_B$ 分割 U 为一些等价类:$U/R_B=\{[x]_B\}:x\in U\}$,其中,$[x]_B$ 是相对于 B 的包含 x 的等价类,即$[x]_B=\{y\in U:(x,y)\in R_B\}$,$[x]_B$ 也可等价地表示为特征函数

$$f_{[x]_B}(y)=\begin{cases}1, & y\in[x]_B\\ 0, & y\notin[x]_B\end{cases},\ x\in U .$$

因此,$U=\bigcup_{y\in U}\{f_{[x]_B}(y):x\in U\}$.对于 U 中的概念 $X(X\subset U)$,可用其在(U,R_B)中的下、上近似解释:

$$\underline{R_B}=\bigcup\{[x]_B:[x]_B\subset X,x\in U\},\ \overline{R_B}=\bigcup\{[x]_B:[x]_B\cap X\neq\varnothing,x\in U\} .$$

$\underline{R_B}X$ 表示确定性属于 X 的对象集合;$\overline{R_B}X$ 表示可能属于 X 的对象集合. $(\underline{R_B}X,\overline{R_B}X)$ 称为 X 相对于 B 的粗糙集.

1.2 模糊目标信息系统

首先引述目标信息系统的相关概念[2].

定义 1. 目标信息系统具有如下的一般形式: (U,A,F,D,G) ,其中:

(1) U 表示有限对象的论域,即 $U=\{x_1,x_2,\ldots,x_n\}$;

(2) A 表示有限符号属性集合,即 $A=\{a_1,a_2,\ldots,a_p\},a_i:U\to V_i,i\le p$;

(3) F 为特征映射集合,$f\in F,f:U\times A\to V,V$ 表示符号特征值域,$V=\cup_{1\le i\le p}V_i$;

(4) D 表示有限目标属性集合,即 $D=\{d_1,d_2,\ldots,d_p\},D_i:U\to V'_i,i\le q$;

(5) G 表示模糊目标映射集合, $g\in G$, $g:U\times D\to V'$, $V'=\bigcup_{1\le i\le q}V'_i$.

例 1:考虑如下的信息系统,见表 1.

Table 1 An evaluation system

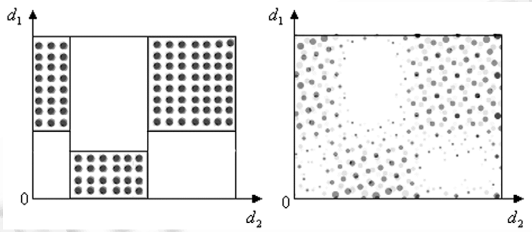
表 1 评价系统

U	a_1	a_2	a_3	d_1	d_2
x_1	2	1	3	1	1
x_2	3	2	1	0	1
x_3	2	1	3	0	0

对象空间目标特征函数表示为

$$D_1=\begin{cases}1, & x\in\{x_1\}\\ 0, & x\in\{x_2,x_3\}\end{cases},D_2=\begin{cases}1, & x\in\{x_2\}\\ 0, & x\in\{x_1,x_3\}\end{cases},D_3=\begin{cases}1, & x\in\{x_3\}\\ 0, & x\in\{x_1,x_2\}\end{cases}.$$

图 1 给出了两个目标特征(d_1,d_2)分割对象空间的示例.利用 2d 图给出了两种对象空间的划分,其中点的大小代表接近于 1 的程度,点越大,则目标特征值越接近于 1.



(a) Equivalence partition (b) Fuzzy partition

(a) 等价划分 (b) 模糊划分

Fig.1

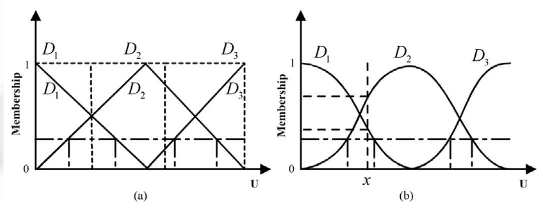
图 1

定义 2. 若 V'_i=[0,1]为模糊特征函数集合,则称目标信息系统(U,A,F,D,G)为模糊目标信息系统.

2 模糊目标信息系统的知识约简

FOISs 上的 α 约简利用了对象空间特征函数的 α 水平集.假设 $\{D_i:i\le r\}$ 是 U 上由 D 导出的模糊分割,则对 $0<\alpha\le 1$,得到水平集 $\{D_i^\alpha:i\le r\}$,显然 $\bigcup_{i=1}^r D_i^\alpha\subseteq U$.若 $U/R=\bigcup\{[x]_B:x\in U\}$,则粗糙成员关系函数为 $D(D_i^\alpha/[x]_B)=|D_i^\alpha\cap[x]_B|/|[x]_B|$.

图 2 给出了论域 U 的两类划分的特征函数表示及其水平集.图 2(a)的虚线表示符号值决策特征构成的等价划分,图 2(a)的实线表示标准三角模糊函数划分,图 2(b)的实线表示标准钟形模糊函数划分.点划线表示它们的 α 水平分割.图 2(b)同时给出了数据点 x 对 D_1,D_2,D_3 的隶属度表示.



(a) (b)

Fig.2 Three types of the characteristic functions for decision features in object space

图 2 对象空间 U 上的 3 种划分的特征函数

下面利用 $\{D_i^\alpha:i\le r\}$ 给出 FOISs 上的知识约简方法.

定义 3. 设 (U,A,F,D,G) 是 FOIS, $0<\alpha\le 1$, $B\subset A$,记 $\mu_B^\alpha(x)=(D(D_1^\alpha/[x]_B),D(D_2^\alpha/[x]_B),\ldots,D(D_r^\alpha/[x]_B))$,若 $\forall x\in U$, $\mu_B^\alpha(x)=\mu_A^\alpha(x)$,则 B 称为 α 分布协调集.进而,若 $\forall B'\subset B$ 真子集, $\mu_{B'}^\alpha(x)\ne\mu_B^\alpha(x)$,则 B 称为 α 分布约简.

α 分布协调集是保持对象在每个 α 水平决策类的隶属度不变的属性集, $\mu_B^\alpha(x)$ 给出了 x 在 α 水平决策集中的可能性分布. α 分割产生的水平决策集可能重叠,对象 x 可能隶属不同的决策集,但隶属度不同,因此考虑具有最大隶属度的决策集更有意义.

定义 4. 设 (U,A,F,D,G) 是 FOIS, $0<\alpha\le 1$, $B\subset A$, 记 $\gamma_B^\alpha(x)=\{D_k^\alpha : D(D_k^\alpha/[x]_B)=\max_i D(D_i^\alpha/[x]_B)\}$, 若 $\forall x\in U$, $\gamma_B^\alpha(x)=\gamma_A^\alpha(x)$, B 称为 α 最大分布协调集.进而,若 $\forall B'\subset B$ 真子集, $\gamma_{B'}^\alpha(x)\ne\gamma_B^\alpha(x)$, 则 B 称为 α 最大分布约简.

α 最大分布协调集保持每个对象的 α 最大分布决策类不变.

定理 1. 设 (U,A,F,D,G) 是 FOIS, $0<\alpha\le 1$, 则 α 分布协调集必是 α 最大分布协调集.

证明:根据定义可知 α 决策类保持不变,故可证. □

定义 5. 设 (U,A,F,D,G) 是 FOIS, $0<\alpha\le 1$, $B\subset A$, 记 $\delta_B^\alpha(x)=\{D_j^\alpha : [x]_B\cap D_j^\alpha\ne\varnothing\}, x\in U$, $j\le r$, 如果 $\forall x\in U$, $\delta_B^\alpha(x)=\delta_A^\alpha(x)$, B 称为 α 分配协调集.进而,若 $\forall B'\subset B$ 真子集, $\delta_{B'}^\alpha(x)\ne\delta_B^\alpha(x)$, 则 B 称为 α 分配约简.

α 分配协调集保持所有对象的可能 α 决策类不变.

定理 2. 设 (U,A,F,D,G) 是 FOIS, $0<\alpha\le 1, B\subset A$, 则 α 分布协调集必是 α 分配协调集.

证明:设 B 为 α 分布协调集,则 $\mu_B^\alpha(x)=\mu_A^\alpha(x)$, 即 $D(D_i^\alpha/[x]_B)=D(D_i^\alpha/[x]_A), i\le r$, 若 $D_i^\alpha\in\delta_B^\alpha$, 则 $[x]_B\cap D_i^\alpha\ne\varnothing\Rightarrow D(D_i^\alpha/[x]_B)\ne 0\Rightarrow D(D_i^\alpha/[x]_A)\ne 0\Rightarrow [x]_A\cap D_i^\alpha\ne\varnothing\Rightarrow D_i^\alpha\in\delta_A^\alpha$, 因此, $\delta_B^\alpha(x)\subseteq\delta_A^\alpha(x)$. 另一方面,若 $D_B^\alpha(x)\in\delta_A^\alpha$, 则由 $[x]_A\subset[x]_B$ 和 $[x]_A\cap D_i^\alpha\ne\varnothing$ 得 $[x]_B\cap D_i^\alpha\ne\varnothing$, 从而 $D_B^\alpha\in\delta_B^\alpha$, 因此, $\delta_B^\alpha(x)\supseteq\delta_A^\alpha(x)$, 从而 $\delta_B^\alpha(x)=\delta_A^\alpha(x)$. □

从模糊集合的粗糙上、下近似考虑,可以得到以下约简定义.

定义 6. 设 (U,A,F,D,G) 是 FOIS, $B\subset A$, $0<\beta\le\alpha\le 1$, 记

$$L\mu_B^\alpha(x,A)=(D(\underline{R_A}(D_1)_\alpha/[x]_B),D(\underline{R_A}(D_2)_\alpha/[x]_B),\ldots,D(\underline{R_A}(D_r)_\alpha/[x]_B)),$$

$$H\mu_B^\beta(x,A)=(D(\overline{R_A}(D_1)_\beta/[x]_B),D(\overline{R_A}(D_2)_\beta/[x]_B),\ldots,D(\overline{R_A}(D_r)_\beta/[x]_B)),$$

则有:

(1) 若 $\forall x\in U$, $L\mu_B^\alpha(x,A)=L\mu_A^\alpha(x,A)$, B 称为(α,A)下近似分布协调集(或称α下近似分布协调集).进而,若 $\forall B'\subset B$ 真子集,

$$L\mu_{B'}^\alpha(x,A)\ne L\mu_B^\alpha(x,A),$$

则 B 称为 A 的 (α,A) 下近似分布约简.

(2) 若 $\forall x\in U$, $H\mu_B^\beta(x,A)=H\mu_A^\beta(x,A)$, B 称为 (β,A) 上近似分布协调集(或称β下近似分布协调集).进而,若 $\forall B'\subset B$ 真子集,

$$H\mu_{B'}^\beta(x,A)\ne H\mu_B^\beta(x,A),$$

则 B 称为 A 的 (β,A) 上近似分布约简.

(3) 若 B 是 A 的 (α,A) 下近似分布协调集和 (β,A) 上近似分布协调集,则称 B 为 (α,β,A) 粗糙分布协调集;若 B 是 A 的 (α,A) 下近似分布约简和 (β,A) 上近似分布约简,则称 B 为 A 的 (α,β,A) 粗糙分布约简. (α,A) 下近似分布协调集((β,A) 上近似分布协调集、(α,β,A) 粗糙分布协调集)是保持对象在每个决策类的 A 粗糙 α 下近似(β 上近似、 α 下近似和 β 上近似)的隶属度不变的属性集.

性质 1. 设 (U,A,F,D,G) 是 FOIS, $0<\beta\le\alpha\le 1$, $\forall x\in U$, 则

(1) $\forall A_i\subset A$, $L\mu_B^\alpha(x,A_i)\le H\mu_B^\beta(x,A_i)$;

(2) $\forall A_1\subset A_2\subset A$, $L\mu_B^\alpha(x,A_1)\le L\mu_B^\alpha(x,A_2)$ 和 $H\mu_B^\beta(x,A_2)\le H\mu_B^\beta(x,A_1)$.

证明:(1) 对于 $\forall x\in U, A_i\subset A$ 由 $\underline{R_{A_i}}(D_i)_\alpha\subset\overline{R_{A_i}}(D_i)_\beta$, 显然可证.

(2) 由定义 6、粗糙成员关系函数以及粗糙近似的单调性质显然可证. □

性质 1 表明了 (α,A) 下近似分布约简和 (β,A) 上近似分布约简之间的关系.

性质 2. 设 (U,A,F,D,G) 是 FOIS, $0<\beta\le\alpha\le 1$, $B\subset A$. 假设 $D_i(x)=c_{ij}, x\in C_j$; $i\le r$, $C_j\in U/A$, 其中 c_{ij} 为常数,则

(1) B 是α分布约简等价于 B 是(α,A)下近似分布约简;

(2) B 是β分布约简等价于 B 是(β,A)上近似分布约简.

证明:由已知 $\underline{R_A}(D_i)_\alpha = D_i^\alpha$, $\overline{R_A}(D_i)_\beta = D_i^\beta$,则 $\forall x \in U$, $L\mu_B^\alpha(x,A) = \mu_B^\alpha(x)$, $H\mu_B^\beta(x,A) = \mu_B^\beta(x)$,从而(1)(2)得证. □

当 $D_i, i \le r$ 为 U 上关于 A 的粗糙模糊集时,由性质 2 表明,粗糙分布约简与分布约简等价.

定义 7. 设 (U,A,F,D,G) 是 FOIS, $B \subset A$,则

(1) 若 $\forall x \in U, \forall \alpha \in (0,1], \mu_B^\alpha(x) = \mu_B^\alpha(x)$,则 B 称为分布协调集;进而,若 $\forall B' \subset B, B' \ne B, B'$ 不是分布协调集,则 B 称为分布约简.

(2) 若 $\forall x \in U, \forall \alpha \in (0,1], \gamma_B^\alpha(x) = \gamma_B^\alpha(x)$,则 B 称为最大分布协调集;进而,若 $\forall B' \subset B, B' \ne B, B'$ 不是最大分布协调集,则 B 称为最大分布约简.

(3) 若 $\forall x \in U, \forall \alpha \in (0,1], \delta_B^\alpha(x) = \delta_B^\alpha(x)$,则 B 称为分配协调集;进而,若 $\forall B' \subset B, B' \ne B, B'$ 不是分配协调集,则 B 称为分配约简.

(4) 若 $\forall x \in U, \forall \alpha \in (0,1]$, $L\mu_A^\alpha(x,A) = L\mu_B^\alpha(x,A)$,则 B 称为 A 下近似分布协调集;进而,若 $\forall B' \subset B, B' \ne B, B'$ 不是 A 下近似分布协调集,则 B 称为 A 下近似分布约简.

(5) 若 $\forall x \in U, \forall \beta \in (0,1]$, $H\mu_A^\beta(x,A) = H\mu_B^\beta(x,A)$,则 B 称为 A 上近似分布协调集;进而,若 $\forall B' \subset B, B' \ne B, B'$ 不是 A 上近似分布协调集,则 B 称为 A 上近似分布约简.

(6) 若 $\forall x \in U, \forall \alpha \in (0,1]$, $L\mu_A^\alpha(x,A) = L\mu_B^\alpha(x,B)$,则 B 称为下近似分布协调集;进而,若 $\forall B' \subset B, B' \ne B, B'$ 不是下近似分布协调集,则 B 称为下近似分布约简.

(7) 若 $\forall x \in U, \forall \beta \in (0,1]$, $H\mu_A^\beta(x,A) = H\mu_B^\beta(x,B)$,则 B 称为上近似分布协调集;进而,若 $\forall B' \subset B, B' \ne B, B'$ 不是下近似分布协调集,则 B 称为下近似分布约简.

(8) 若 B 既是 A 下近似分布约简,又是 A 上近似分布约简,则 B 是 A 粗糙分布约简.

(9) 若 B 既是下近似分布约简,又是上近似分布约简,则 B 是粗糙分布约简.

定理 4. 设 $\Phi = (U,A,F,D,G)$ 是 FOIS, $0 < \alpha \le 1$,令 $\{I_j : j \le r\}$ 构成了 U 上的一个等价分割,若 $D_i(x) \ge \alpha_2, x \in I_i$; $D_i(x) \le \alpha_1, x \in I_j; j \ne i, i \le r$,则对 $\forall \alpha \in (\alpha_1, \alpha_2]$, Φ 退化为 Pawlak 信息系统 Φ^α,且有性质成立:

(1) (α, A) 下近似分布约简等价于 (α, A) 上近似分布约简;

(2) (α, A) 下近似分布约简等价于 α 分布约简.

证明:由已知, $\{D_1^\alpha, D_2^\alpha, ..., D_r^\alpha\}$ 构成了 U 上的一个等价划分,结论显然. □

定理 4 表明,FOIS 上的约简是 Pawlak 信息系统约简的一种概括形式.根据 Pawlak 信息系统约简的相关概念[2~4,11],同样可得 FOIS 上的判定定理、约简逻辑公式.

定理 5. 设 (U,A,F,D,G) 是 FOIS, $0 < \beta \le \alpha \le 1$, $B \subset A$,则

(1) B 是 α 分布协调集当且仅当 $\forall x, y \in U$,当 $\mu_A^\alpha(x) \ne \mu_A^\alpha(y)$ 时, $[x]_B \cap [y]_B = \varnothing$.

(2) B 是 α 最大分布协调集当且仅当 $\forall x, y \in U$,当 $\gamma_A^\alpha(x) \ne \gamma_A^\alpha(y)$ 时, $[x]_B \cap [y]_B = \varnothing$.

(3) B 是 α 分配协调集当且仅当 $\forall x, y \in U$,当 $\delta_A^\alpha(x) \ne \delta_A^\alpha(y)$ 时, $[x]_B \cap [y]_B = \varnothing$.

(4) B 是 β 上近似分布协调集当且仅当 $\forall x, y \in U$,当 $H\mu_A^\beta(x) \ne H\mu_A^\beta(y)$ 时, $[x]_B \cap [y]_B = \varnothing$.

(5) B 是 α 下近似分布协调集当且仅当 $\forall x, y \in U$,当 $L\mu_A^\alpha(x) \ne L\mu_A^\alpha(y)$ 时, $[x]_B \cap [y]_B = \varnothing$.

(6) B 是 (α, β) 粗糙分布协调集当且仅当 $\forall x, y \in U$,当 $L\mu_A^\alpha(x) \ne L\mu_A^\alpha(y)$, $H\mu_A^\beta(x) \ne H\mu_A^\beta(y)$ 时, $[x]_B \cap [y]_B = \varnothing$.

证明:根据 Pawlak 信息系统上的约简方法[2,4,6,11]以及解模糊过程,(1)~(3)证明显然.下面仅给出(4)的证明,(5)(6)类似.记 $\vartheta([x]_B) = \{[y]_A : [y]_A \subseteq [x]_B\}$,由于 $B \subset A$,所以 $\vartheta([x]_B)$ 构成了 $[x]_B$ 的一个划分,同时可知 $[x]_B = \bigcup\{[y]_A : [y]_A \in \vartheta([x]_B)\}$.

(4)的证明:设 B 是 β 上近似分布协调集,即

$$D(\overline{R_A}(D_i)_\beta / [x]_B) = \sum\left\{ |[y]_A \cap \overline{R_A}(D_i)_\beta| : [y]_A \in \vartheta([x]_B) \right\} \Big/ |[x]_B|$$

$$= \sum\left\{ \frac{|[y]_A \cap \overline{R_A}(D_i)_\beta|}{|[y]_A|} \cdot \frac{|[y]_A|}{|[x]_B|} : [y]_A \in \vartheta([x]_B) \right\} = \sum\left\{ D(R_A(D_i)_\beta / [y]_A) \cdot \frac{|[y]_A|}{|[x]_B|} : [y]_A \in \vartheta([x]_B) \right\}$$

$$=\sum\left\{D(R_A(D_i)_\beta/[x]_A)\cdot\frac{|[y]_A|}{|[x]_B|}:[y]_A\in\vartheta([x]_B)\right\}=D(\overline{R_A}(D_i)_\beta/[x]_A),(i\le r).$$

由已知, $\forall x,y\in U$,若 $[x]_B\cap[y]_B\ne\varnothing$,则 $[x]_B=[y]_B$,从而 $H\mu_A^\beta(x)=H\mu_A^\beta(y)$,矛盾.反之,由 $[x]_B=\cup\{[y]_A:[y]_A\in\vartheta([x]_B)\}$ 知 $D(\overline{R_A}(D_i)_\beta/[x]_B)=D(\overline{R_A}(D_i)_\beta/[x]_A)$,因此 $H\mu_A^\beta(x)=H\mu_A^\beta(y)$,从而 B 是 β 上近似分布协调集. □

定义 8. 设 (U,A,F,D,G) 是 FOIS, $U/R_A=\{C_1,C_2,...,C_m\}$, $0<\beta\le\alpha\le1$,记

$D_1^{*\alpha}=\{([x]_A,[y]_A):\mu_A^\alpha(x)\ne\mu_A^\alpha(y)\}$;

$D_2^{*\alpha}=\{([x]_A,[y]_A):\gamma_A^\alpha(x)\ne\gamma_A^\alpha(y)\}$;

$D_3^{*\alpha}=\{([x]_A,[y]_A):\delta_A^\alpha(x)\ne\delta_A^\alpha(y)\}$;

$D_4^{*\alpha}=\{([x]_A,[y]_A):L\mu_A^\alpha(x)\ne L\mu_A^\alpha(y)\}$;

$D_5^{*\alpha}=\{([x]_A,[y]_A):H\mu_A^\beta(x)\ne H\mu_A^\beta(y)\}$;

定义

$$D_l^\theta(C_i,C_j)=\begin{cases}\{a_k\in A:f_k(C_i)\ne f_k(C_j)\}, & (C_i,C_j)\in D_l^{*\theta}\\ A, & (C_i,C_j)\notin D_l^{*\theta}\end{cases},$$

其中, $f_k(C_i)$ 表示属性 a_k 关于 C_i 中对象的取值, $\theta=\alpha,l=1,2,3,4$, $\theta=\beta,l=5$.分别称 $D_l^\alpha(C_i,C_j)$ ($\theta=\alpha,l=1,2,3,4$; $\theta=\beta,l=5$)为 α 分布、 α 最大分布、 α 分配、 α 下近似分布、 β 上近似分布可辨识属性集.称 $D_k^\theta=\{D_k^\alpha(C_i,C_j):i,j\le m\}$, $\theta=\alpha,l=1,2,3,4$; $\theta=\beta,l=5$ 分别为 FOIS 的 α 分布、 α 最大分布、 α 分配、 α 下近似分布、 β 上近似分布可辨识属性矩阵.

性质 3. α 分布、 α 最大分布、 α 分配、 α 下近似分布、 β 上近似分布可辨识属性矩阵满足以下性质:

(1) D_l^θ 对称;

(2) D_l^θ 主对角元为 A;

(3) $D_l^\theta(C_i,C_j)\subseteq D_l^\theta(C_i,C_s)\cup D_l^\theta(C_s,C_j)$ $(\forall i,j,s\le m)$,其中 $\theta=\alpha,l=1,2,3,4$; $\theta=\beta,l=5$.

证明: (1)(2)显然.(3)根据文献[2,7]中变精度粗糙集模型上的相应结果,同理可得. □

定义 9. 设 (U,A,F,D,G) 是 FOIS, $0<\beta\le\alpha\le1$,记

$$M_l^\theta=\wedge\{\vee\{a_k:a_k\in D_l^\theta(C_i,C_j)\}:i,j\le m\}=\wedge\{\vee\{a_k:a_k\in D_l^\theta(C_i,C_j)\}:(C_i,C_j)\in D_l^{*\theta}\},$$

其中 $\theta=\alpha,l=1,2,3,4$; $\theta=\beta,l=5$,则 M_l^α ($l=1,2,3,4$)分别称为 α 分布、 α 最大分布、 α 分配、 α 下近似分布辨识公式; M_5^β 称为 β 上近似分布辨识公式.

定理 6. 设 (U,A,F,D,G) 是 FOIS, $0<\beta\le\alpha\le1$, $B\subseteq A$,则

(1) B 是 α 分布协调集等价于 $\forall(C_i,C_j)\in D_1^{*\alpha}\Rightarrow B\cap D_1^\alpha(C_i,C_j)\ne\varnothing$;

(2) B 是 α 最大分布协调集等价于 $\forall(C_i,C_j)\in D_2^{*\alpha}\Rightarrow B\cap D_2^\alpha(C_i,C_j)\ne\varnothing$;

(3) B 是 α 分配协调集等价于 $\forall(C_i,C_j)\in D_3^{*\alpha}\Rightarrow B\cap D_3^\alpha(C_i,C_j)\ne\varnothing$;

(4) B 是 α 下近似分布协调集等价于 $\forall(C_i,C_j)\in D_4^{*\alpha}\Rightarrow B\cap D_4^\alpha(C_i,C_j)\ne\varnothing$;

(5) B 是 β 上近似分布协调集等价于 $\forall(C_i,C_j)\in D_5^{*\alpha}\Rightarrow B\cap D_5^\alpha(C_i,C_j)\ne\varnothing$.

证明:根据 Pawlak 信息系统上的约简方法[2,4,6]以及解模糊过程,可证. □

定理 7. 设 (U,A,F,D,G) 是 FOIS, $0<\beta\le\alpha\le1$,辨识公式 M_l^θ 的极小析取范式为

$$M_l^\theta=\bigvee_{k=1}^{t}(\bigwedge_{s=1}^{q_k}a_{l_s})\ (\theta=\alpha,l=1,2,3,4;\theta=\beta,l=5),$$

记 $\{B_{lk}:k=1,2,...,t\},(l=1,2,3,4,5)$ 分别是所有 α 分布、 α 最大分布、 α 分配、 α 下近似分布、 β 上近似分布约简形成的集合.

证明:基于 FOIS 上水平集的辨识公式极小范式证明过程与 Pawlak 信息系统过程同构,同理可得. □

3 在模糊概念分类器模型中的应用

通常,从符号表产生规则是 NP 难解的,一般采用启发式算法减少时间、空间复杂性,但是这些算法不考虑系统中是否存在冗余属性.除了增加计算复杂性与规则长度以外,冗余属性对于分类和推理没有任何益处,因此,在分类与提取规则之前对特征进行约简会大大提高分类的效率.

根据模糊目标概念的 α 水平集,α 分布约简产生了与全部属性具有相同数据成员分布特征的最小属性集合,因而不会改变对象相对于目标概念的分布特征;而粗糙分布约简产生了两种类型的分布约简形式.下近似分布约简保持了对象类的一定具有的目标性质分布不变,而上近似分布约简保持了对象类的可能具有的目标性质分布不变;α 最大分布约简决定了对象具有的最大可能的目标性质,如流感;α 分配约简决定了对象所具有的所有可能目标概念,如感冒、非典等.在如图 3 所示的模糊概念分类器中,利用训练数据产生系统的 α 系列约简,得到特定的粗糙模糊分类规则.利用这些规则构成的推理知识,可以对未知对象分类,以推断他们(可能)隶属的概念域,如患何病、近视度等等.

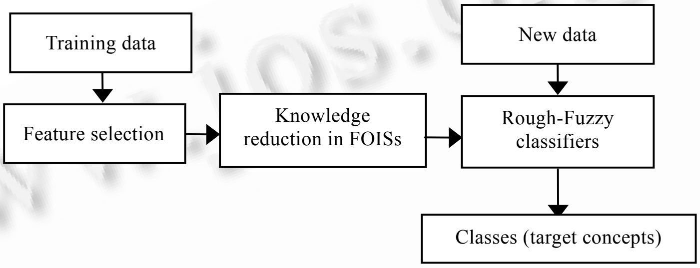


Fig.3 The rough-fuzzy rule classifier model including knowledge reduction

图 3 包含约简过程的粗糙模糊规则分类器模型

结合一个实例,下面计算几种约简结果.假定训练包含表 2 中的特征和数据,则可计算如下约简属性集.

Table 2 A fuzzy decision table

表 2 模糊决策表

U	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
a_1	1	1	1	1	0	0	0	0	0
a_2	1	1	0	0	1	1	1	1	0
a_3	1	1	1	1	1	1	0	1	1
a_4	1	1	1	0	0	0	1	0	1
d	0.20	0.30	0.40	0.50	0.55	0.65	0.70	0.80	0.90

首先由特征 $\{d\}$ 构造对象空间模糊分割 $D=\{D_1,D_2\}$,其中 $D_1(x)=d(x)$, $D_2(x)=1-d(x)$,例如 d 表示“近视”,则 D_1 表示“低度”, D_2 表示“高度”.令 $\alpha=0.6,0.7$,则

$$D_1^{0.6}(x)=\{x_6,x_7,x_8,x_9\};\ D_2^{0.6}(x)=\{x_1,x_2,x_3\};\ D_1^{0.7}(x)=\{x_7,x_8,x_9\};\ D_2^{0.7}(x)=\{x_1,x_2\}.$$

从而

$$\mu_A^{0.6}(x_i)=(0,1),i=1,2,3;\ \mu_A^{0.6}(x_4)=(0,0);\ \mu_A^{0.6}(x_i)=(2/3,0),i>4,$$

因此,

$D_1^{*0.6}=\{(C_1,C_3),(C_1,C_4),(C_1,C_5),(C_1,C_6),(C_2,C_3),(C_2,C_4),(C_2,C_5),(C_2,C_6),(C_3,C_4),(C_3,C_5),(C_3,C_6),(C_4,C_5),(C_4,C_6)\}$;

$D_1^{0.6}(C_1,C_3)=\{a_2,a_4\};\ D_1^{0.6}(C_1,C_4)=\{a_1,a_4\};\ D_1^{0.6}(C_1,C_5)=\{a_1,a_3\};\ D_1^{0.6}(C_1,C_6)=\{a_1,a_2\};\ D_1^{0.6}(C_2,C_3)=\{a_4\};$

$D_1^{0.6}(C_2,C_4)=\{a_1,a_2,a_4\};\ D_1^{0.6}(C_2,C_5)=\{a_1,a_2,a_3\};\ D_1^{0.6}(C_2,C_6)=\{a_1\};\ D_1^{0.6}(C_3,C_4)=\{a_1,a_2,a_4\};$

$D_1^{0.6}(C_3,C_5)=\{a_1,a_2,a_3,a_4\};\ D_1^{0.6}(C_3,C_6)=\{a_1,a_4\};\ D_1^{0.6}(C_4,C_5)=\{a_1,a_3,a_4\};\ D_1^{0.6}(C_4,C_6)=\{a_2,a_4\},$

则

$$M_1^{0.6}=(a_2\vee a_4)\wedge(a_1\vee a_4)\wedge(a_1\vee a_3)\wedge(a_1\vee a_2)\wedge a_4\wedge(a_2\vee a_4\vee a_1)\wedge(a_3\vee a_2\vee a_1)\wedge a_1\wedge$$
$$(a_1\vee a_2\vee a_4)\wedge(a_1\vee a_2\vee a_3\vee a_4)\wedge\ (a_1\vee a_4)\vee(a_1\vee a_3\vee a_4)\wedge(a_2\vee a_4)=a_1\wedge a_4\ .$$

同理,$M_1^{0.7}=a_1\wedge a_2$,从而可知$\{a_1,a_4\}$是 0.6 分布约简,$\{a_1,a_2\}$是 0.7 分布约简.

另外得到$\{a_1,a_4\}$是 FOIS 上的 0.6 最大分布约简、分配约简.下面考虑粗糙近似约简.

$$\underline{R_A}D_1=0.2/C_1+0.4/C_2+0.5/C_3+0.55/C_4+0.7/C_5+0.9/C_6,$$
$$\overline{R_A}D_1=0.3/C_1+0.4/C_2+0.5/C_3+0.8/C_4+0.7/C_5+0.9/C_6,$$
$$\underline{R_A}D_2=0.7/C_1+0.6/C_2+0.5/C_3+0.2/C_4+0.3/C_5+0.1/C_6,$$
$$\overline{R_A}D_2=0.8/C_1+0.6/C_2+0.5/C_3+0.45/C_4+0.3/C_5+0.1/C_6,$$

得到

$$\underline{R_A}^{0.6}D_1=0.7/C_5+0.9/C_6;$$
$$\overline{R_A}^{0.6}D_1=0.8/C_4+0.7/C_5+0.9/C_6;$$
$$\underline{R_A}^{0.7}D_2=0.7/C_1;$$
$$\overline{R_A}^{0.7}D_2=0.8/C_1+0.6/C_2.$$

因此,为简化同上可得:

$$M_4^{0.7}=(a_1\wedge a_2\wedge a_3)\vee(a_1\wedge a_2\wedge a_4),$$
$$M_5^{0.6}=a_1\wedge a_4,$$

从而,$\{a_1,a_2,a_3\},\{a_1,a_2,a_4\}$是 0.7 下近似分布约简,$\{a_1,a_4\}$是 0.6 上近似分布约简.进而,$\{a_1,a_2,a_4\}$是分布约简、最大分布约简、分配约简、粗糙分布约简.显然,α约简集包含于相应的约简集中.

4 结 论

等价关系属性约简方法现在研究得较为充分,如张文修等人最近先后给出了不协调信息系统、可变精度粗糙集模型上的约简方法[4,6,7],Slezak 的近似约简方法,Bayesian 约简模型[8~10]等等,但它们不能解决由于对象集合模糊性带来的约简问题,并且,粗糙和模糊特性融合是不确定信息领域研究的一个方向[14],其中包括粗糙模糊性、模糊粗糙性.FOISs 是这两种特性融合的实例系统,具有粗糙模糊特性.连续值信息系统的数据离散化方法将模糊系统转化为 Pawlak 系统,易受离散方法、断点等因素影响.FOISs 上的约简方法研究较少[2],这里提及的 FOISs 上基于相似度α,β的系列约简给出了相对于模糊目标概念的条件属性约简方法,发展了 Pawlak 信息系统约简方法.它们利用了粗糙和模糊理论的特性,在不同目标概念层次上产生最小条件属性集合,有效地降低了分类与推理的计算复杂性,并得到最短粗糙模糊决策(或分类)规则.

另外,在最小析取范式中包含大量的冗余信息,通过简化逻辑公式得到最小属性集.不可分辨矩阵中的单个属性称为核属性,必包含在约简属性核之中.矩阵的包含该属性的属性集可在逻辑公式中省略,从而通过核属性逻辑运算得到最短逻辑表达式[11].

References:

[1] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Boston: Kluwer Academic Publishers, 1991.

[2] Zhang WX, Liang Y, Wu WZ. Information Systems and Knowledge Discovery. Beijing: Science Press, 2003 (in Chinese).

[3] Kryszkiewicz M. Comparative Study of Alternative Type of knowledge reduction in inconsistent systems. Int'l Journal of Intelligent Systems, 2001,16(1):105~120.

[4] Zhang WX, Mi JS, Wu WZ, Approaches to knowledge reductions in inconsistent systems. Int'l Journal of Intelligent Systems, 2003,18(10):989~1000.

[5] Beynon M. Reducts within the vaiable precision rough sets model: A further investigation. European Journal of Operational Research, 2001,134(3):92~605.

[6] Zhang WX, Mi JS, Wu WZ. Knowledge reductions in inconsistent information systems. Chinese Journal of Computers, 2003,26(1):12~18 (in Chinese with English abstract).

[7] Mi JS, Wu WZ, Zhang WX. Approaches to knowledge reduction based on variable precision rough set model. Information Sciences, 2004,159(3-4):255~272.

[8] Slezak D, Ziarko W. Attribute reduction in the bayesian version of variable precision rough set model. Electronic Notes Theoretical Computer Science, 2003,82(4):1~11.

[9] Slezak D. Searching for dynamic reducts in inconsistent decision tables. In: Bouchon-Meunier B, Yager RR, eds. Proc. of the 7th Int'l Conf. on Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU'98),Vol 2. 1998. 1362~1369.

[10] Slezak D. Approximate reducts in decision tables. In: Bouchon-Meunier B, Delgado M, Verdegay JL, Vila MA, Yager RR, eds. Proc. of the 6th Int'l Conf., Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'96), Vol 3. 1996. 159~164. http://alfa.mimuw.edu.pl/logic/prace/1996/E63/

[11] Wang GY. Rough Set Theories and Knowledge Acquirement. Xi'an: Xi'an Jiaotong University Press, 2001 (in Chinese).

[12] Wu WZ, Zhang WX. Constructive and axiomatic approaches of fuzzy approximation operators. Information Sciences, 2004, 159(3-4):233~254.

附中文参考文献:

[2] 张文修,梁怡,吴伟志.信息系统与知识发现.北京:科学出版社,2003.

[6] 张文修,米据生,吴伟志.不协调目标信息系统的知识约简.计算机学报,2003,26(1):12~18.

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