文章编号: 1001-0920(2009)03-0335-07

基于直觉模糊粗糙集的属性约简

路艳丽, 雷英杰, 华继学
(空军工程大学 导弹学院, 陕西 三原 713800)

摘 要: 针对 Jensen 下近似定义的局限性,提出一种新的等价类形式的近似算子表示,并将其推广到直觉模糊环境. 在此基础上,将相对正域、相对约简、相对核等粗糙集的知识约简概念推广到直觉模糊环境,提出一种直觉模糊信息系统的启发式属性约简算法. 实例计算表明,该方法比 Jensen 的属性约简方法更为合理有效.

关键词: 粗糙集; 模糊集; 直觉模糊集; 模糊粗糙集
中图分类号: TP18 **文献标识码**: A

Attribute reduction based on intuitionistic fuzzy rough set

LU Yan-li, LEI Ying-jie, HUA Ji-xue
(Missile Institute, Air Force Engineering University, Sanyuan 713800, China. Correspondent: LU Yan-li, E-mail: luyanlihgh@163.com)

Abstract: To the issue of lower approximation proposed by Jensen, an improved definition of lower approximation described with fuzzy equivalent classes is proposed and extended to intuitionistic fuzzy approximation space firstly. The notion of positive region and relative reduct in classical rough set are generalized to the intuitionistic fuzzy information system and a heuristic algorithm for attribute reduction based on intuitionistic fuzzy rough set is proposed. Finally, two particular experiments verify the rationality and effectiveness of the developed algorithm.

Key words: Rough set; Fuzzy set; Intuitionistic fuzzy set; Fuzzy rough set

1 引 言

属性约简是粗糙集理论的核心内容之一,其目的是在保证信息系统分类能力不变的前提下,删除其中不相关或不重要的属性. 现有的粗糙集属性约简方法[1-3]大多适用于离散属性,所能处理的知识和概念都是清晰的,对于现实中常见的模糊概念和模糊知识则无法处理. 对于连续属性的数据处理,通常先对其进行离散化,而离散化往往会丢失有用信息,易使约简和决策产生错误.

模糊集(FS)和直觉模糊集(IFS)[4,5]通过隶属函数和非隶属函数,为论域中的每一对象指定一个隶属度和非隶属度,从而可对论域中任一不精确概念进行更精细的描述(FS 是隶属度与非隶属度之和为 1 的 IFS). 将 FS 与 IFS 的软边界优势引入粗糙集,用模糊化代替离散化,对象对每个等价类都有一定的隶属度,各等价类之间没有陡峭截断,减少了信息丢失. 模糊粗糙集(FRS)[6,7]和直觉模糊粗糙集(IFRS)[8]已成为粗糙集理论的重要研究方向之一.

近年来备受关注的 FRS 模型是 Dubois 的 FRS 模型[6]和 Radzikowska 的广义 FRS 模型[7]. Jensen 基于 Dubois 的 FRS 模型,给出一种启发式属性约简算法(称为 Jensen 算法),并将其用于网络分类和复杂系统监控[9,10]. 其中等价类形式的下近似(称为 Jensen 下近似)存在一定的局限性,可能出现知识增多时,本应增大的下近似反而减小的情况. Chris 将 Radzikowska 的 FRS 模型推广到 IF 环境,给出了基于 IF 逻辑算子的 IFRS 定义[8],该定义是对粗糙集原始概念的一种推广,具有重要的理论价值,但却没有对 IFRS 的属性约简进行研究.

本文针对以上问题,首先对 Jensen 下近似进行改进,给出一种更为合理的等价类形式的下近似算子表示;然后介绍了 IFRS 模型,将改进的近似算子扩展到 IF 环境;再后给出 IF 信息系统相对约简的相关定义,在此基础上给出基于 IFRS 的启发式属

收稿日期: 2008-01-21; **修回日期**: 2008-04-08.
基金项目: 国家自然科学基金项目(60773209); 陕西省自然科学基金项目(2006F18).
作者简介: 路艳丽(1980—),女,陕西大荔人,博士,从事智能信息处理方法的研究; 雷英杰(1956—),男,陕西华阴人,教授,博士生导师,从事智能信息处理、智能决策的研究.

性约简算法;最后用具体算例验证了本文方法的有效性.

2 等价类形式的近似算子表示

本文用 FS(U) 表示论域 U 上 FS 的全体,IFS(U) 表示论域 U 上 IFS 的全体.

FRS 模型[6,7] 与经典粗糙集的不同之处在于:被近似集由经典集变为模糊集,等价关系由普通等价关系推广为模糊等价关系. Dubois 用等价类描述的方式定义了 FRS 的上下近似算子,即如下定义:

定义 1(模糊粗糙集[6]) 设 U 为有限非空论域,R 是 U 上的模糊等价关系,称 FAS = (U, R) 为模糊近似空间. R 将论域 U 进行模糊划分,所得的模糊等价类集合 $U/R = \{E_1, E_2, \cdots, E_k\}$,用 U/R 中的元素描述给定的模糊集 $X \in \mathrm{FS}(U)$,所得下近似 $\underline{R}X$ 和上近似 $\overline{R}X$ 为 U/R 上的一对模糊集,即

$$\begin{cases} \mu_{\underline{R}X}(E_i) = \inf\limits_{y\in U} \max\{1 - \mu_{E_i}(y), \mu_X(y)\}, \\ \mu_{\overline{R}X}(E_i) = \sup\limits_{y\in U} \min\{\mu_{E_i}(y), \mu_X(y)\}. \end{cases} \tag{1}$$

与经典上下近似的定义不同,式(1) 并没有明确给出单个对象 $x \in U$ 对上下近似的隶属程度. 为此,Jensen 根据定义 1 给出了下式:

$$\begin{cases} \mu_{\underline{R}X}(x) = \sup\limits_{E_i\in U/R} \min\{\mu_{E_i}(x), \\ \qquad \inf\limits_{y\in U} \max\{1 - \mu_{E_i}(y), \mu_X(y)\}\}, \\ \mu_{\overline{R}X}(x) = \sup\limits_{E_i\in U/R} \min\{\mu_{E_i}(x), \\ \qquad \sup\limits_{y\in U} \min\{\mu_{E_i}(y), \mu_X(y)\}\}. \end{cases} \tag{2}$$

用于表示一个对象 $x \in U$ 对等价类描述的上下近似的隶属度[9]. 其下近似的含义是:对象 x 属于 $\underline{R}X$,等价于 $\exists E_i \in U/R$,满足 $x \in E_i$ 且 $E_i \in \underline{R}X$.

值得提出的是:在经典粗糙集中,每个对象 $x \in U$ 仅属于 U/R 中的某一等价类;而在 FRS 中,对象对每个模糊等价类 $E_i \in U/R$ 都有一定的隶属度. 应用 FS 的思想分析经典粗糙集的上下近似,不难发现,x 对 $\underline{R}X$ 的隶属度由 x 的 R 等价类 $[x]_R$ 包含于 X 的程度决定,x 对 $\overline{R}X$ 的隶属度由 $[x]_R$ 与 X 相交的程度决定. 式(2) 的上近似与经典上近似的内涵相一致;而式(2) 的下近似只要求存在一个等价类 E_i,满足 $x \in E_i$ 且 $E_i \in \underline{R}X$,并没有考虑 x 对其他模糊等价类的隶属度,以及这些等价类对下近似的隶属度,这样在构造下近似可能出现知识增多(减少)时,本应增大(减小) 的下近似反而减小(增大).

设 U 为有限非空论域,A = C ∪ D 为包含条件属性集 C 和决策属性集 D 的模糊属性集,对论域的划分形成模糊等价类,V 为 A 的值域,F 为一信息函数,则 FIS = (U, C ∪ D, V, F) 称为模糊信息系统. 当 D 为离散属性时,FIS 称为模糊条件信息系统.

设 $X \in \mathrm{FS}(U)$,$P_2 \subseteq P_1 \subseteq C$,$U/P_1 = \left\{\bigcap\limits_{a\in P_1} E_{ai} \mid i = 1, 2, \cdots, k\right\}$,$U/P_2 = \left\{\bigcap\limits_{a\in P_2} E_{ai} \mid i = 1, 2, \cdots, h\right\}$;记 $E_i = \bigcap\limits_{a\in P_1} E_{ai}$,$E_i' = \bigcap\limits_{a\in P_2} E_{ai}$. 因为 $P_2 \subseteq P_1$,所以 $E_i \subseteq E_i'$. 根据粗糙集的思想,P_1 包含的知识多于 P_2 包含的知识,应有 $\mu_{\underline{P_1}X}(x) \geq \mu_{\underline{P_2}X}(x)$ 成立. 然而,由式(2) 得

$$\begin{cases} \mu_{\underline{P_1}X}(x) = \sup\limits_{E_i\in U/P_1} \min\{\mu_{E_i}(x), \inf\limits_{y\in U} \max\{1 - \\ \qquad \mu_{E_i}(y), \mu_X(y)\}\}, \\ \mu_{\underline{P_2}X}(x) = \sup\limits_{E_i'\in U/P_2} \min\{\mu_{E_i'}(x), \inf\limits_{y\in U} \max\{1 - \\ \qquad \mu_{E_i'}(y), \mu_X(y)\}\}. \end{cases} \tag{3}$$

显然,式(3) 无法保证 $\mu_{\underline{P_1}X}(x) \geq \mu_{\underline{P_2}X}(x)$ 恒成立,这将直接影响后续相对正域的计算,从而很难将经典粗糙集相对约简的思想推广到 FRS 中.

本文针对这一问题,给出另一种计算每个对象对下近似隶属程度的方法,如下式所示:

$$\mu_{\underline{R}X}(x) = \inf_{E_i\in U/R} \max\{1 - \mu_{E_i}(x), \inf_{y\in U} \max\{1 - \mu_{E_i}(y), \mu_X(y)\}\}. \tag{4}$$

其含义是:当所有包含对象 $x \in U$ 的模糊等价类都包含于 X 时,x 属于 X 的 R 下近似.

同理,设 $P_2 \subseteq P_1 \subseteq C$,由式(4) 得

$$\begin{cases} \mu_{\underline{P_1}X}(x) = \inf\limits_{E_i\in U/P_1} \max\{1 - \mu_{E_i}(x), \\ \qquad \inf\limits_{y\in U} \max\{1 - \mu_{E_i}(y), \mu_X(y)\}\}, \\ \mu_{\underline{P_2}X}(x) = \inf\limits_{E_i'\in U/P_2} \max\{1 - \mu_{E_i'}(x), \\ \qquad \inf\limits_{y\in U} \max\{1 - \mu_{E_i'}(y), \mu_X(y)\}\}. \end{cases}$$

根据 max 的单调递增性质,容易证明 $\mu_{\underline{P_1}X}(x) \geq \mu_{\underline{P_2}X}(x)$ 恒成立,即可保证知识增多(减少) 时,下近似减小(增大). 另外,从式(4) 可以看出,上下近似本质上体现的是一种合取关系和蕴涵关系,因此将式(4) 进一步推广,得到如下定义:

定义 2(等价类形式的上下近似表示) 设 R 是论域 U 上的模糊等价关系,$U/R = \{E_1, E_2, \cdots, E_k\}$,$X \in \mathrm{FS}(U)$,$\Psi$ 为模糊蕴涵算子,T 为模糊三角模. 则 $\forall x \in U$,x 对 $\underline{R}X$ 和 $\overline{R}X$ 的隶属度分别为

$$\begin{cases} \mu_{\underline{R}X}(x) = \inf\limits_{E_i\in U/R} \Psi[\mu_{E_i}(x), \\ \qquad \inf\limits_{y\in U} \Psi(\mu_{E_i}(y), \mu_X(y))], \\ \mu_{\overline{R}X}(x) = \sup\limits_{E_i\in U/R} T[\mu_{E_i}(x), \\ \qquad \sup\limits_{y\in U} T(\mu_{E_i}(y), \mu_X(y))]. \end{cases} \tag{5}$$

当 $\Psi(a, b) = \max\{1 - a, b\}$ 时,式(5) 中的下近

似可转化为式(4).

3　基于 IFRS 的属性约简算法

3.1　直觉模糊粗糙集模型

将模糊等价关系拓展为 IF 等价关系 R,模糊逻辑算子拓展为 IF 蕴涵算子 Ψ和 IF 三角模 T,则模糊近似空间便扩展为 IF 近似空间 IFAS = (U, R, Ψ, T). 本文的 IF 蕴涵算子选择 S-蕴涵和 R-蕴涵. 关于 IF 蕴涵算子和 IF 三角模这里不作详述,具体参见文献[8].

定义 3(直觉模糊粗糙集[8])　设 IFAS = (U, R, Ψ, T), $\forall x \in U$, $X \in \mathrm{IFS}(U)$,则 X 的 R 下近似 $\underline{R}X$ 和 R 上近似 $\overline{R}X$ 为 U 上的一对 IFS,即

$$\begin{cases} \underline{R}X(x) = \inf\limits_{y \in U} \Psi(R(x,y), X(y)), \\ \overline{R}X(x) = \sup\limits_{y \in U} T(R(x,y), X(y)). \end{cases} \tag{6}$$

称$(\underline{R}X, \overline{R}X)$ 为 X 在近似空间 IFAS 上的一个粗糙近似.

定义 3 的 IFRS 是对 FRS[7] 的扩展. 当近似空间 IFAS 退化为模糊近似空间时,被近似集由 IFS 退化为 FS,IFRS 便退化为 FRS[7].

IF 信息系统 IFIS = $(U, \mathrm{C} \cup \mathrm{D}, V, F)$ 与模糊信息系统类似;不同之处在于:IFIS 所包含的属性 A = C ∪ D 是 IF 属性,对论域的划分形成的是 IF 等价类. 当 D 为离散属性时,称 IFIS 为 IF 条件信息系统.

下面从等价类的角度来定义上下近似. 设 IFIS = $(U, \mathrm{C} \cup \mathrm{D}, V, F)$, $R \in \mathrm{C}$, U/R 表示 R 对论域 U 的 IF 划分, $U/R = \{E_1, E_2, \cdots, E_k\}$, $E_i \in \mathrm{IFS}(U)$,并设 $E_i \in U/R$, $X \in \mathrm{IFS}(U)$. 根据定义 3, E_i 对 $\underline{R}X$ 和 $\overline{R}X$ 的隶属度和非隶属度为

$$\underline{R}X(E_i) = \inf_{x \in U} \Psi(E_i(x), X(x)),$$
$$\overline{R}X(E_i) = \sup_{x \in U} T(E_i(x), X(x)).$$

定义 4(IFRS 中等价类形式的上下近似表示)　设 IFIS = $(U, \mathrm{C} \cup \mathrm{D}, V, F)$, $U/R = \{E_1, E_2, \cdots, E_k\}$, $X \in \mathrm{IFS}(U)$, Ψ为 IF 蕴涵算子, T 为 IF 三角模. 则 $\forall x \in U$, x 对 $\underline{R}X$ 和 $\overline{R}X$ 的隶属度和非隶属度为

$$\begin{cases} \underline{R}X(x) = \inf\limits_{E_i \in U/R} \Psi[E_i(x), \inf\limits_{y \in U} \Psi(E_i(y), X(y))], \\ \overline{R}X(x) = \sup\limits_{E_i \in U/R} T[E_i(x), \sup\limits_{y \in U} T(E_i(y), X(y))]. \end{cases} \tag{7}$$

若 IFIS = $(U, \mathrm{C} \cup \mathrm{D}, V, F)$ 为 IF 条件信息系统, IF 蕴涵算子 $\Psi(a,b) = \Psi_{S_M, N}(a,b) = \max\{N(a), b\}$,其中 $a = (a_1, a_2)$ 和 $b = (b_1, b_2)$ 为 IF 值, a_1 为隶属度, a_2 为非隶属度, $N(a) = (a_2, a_1)$ 为 IF 标准否定算子. 则 x 对 $\underline{R}X$ 和 $\overline{R}X$ 的隶属度和非隶属度为

$$\begin{cases} \underline{R}X(x) = \inf\limits_{E_i \in U/R} \max\{N(E_i(x)), \\ \qquad \inf\limits_{y \in U, y \notin X} N(E_i(y))\}, \\ \overline{R}X(x) = \sup\limits_{E_i \in U/R} \min\{E_i(x), \sup\limits_{y \in U} E_i(y)\}. \end{cases} \tag{8}$$

3.2　直觉模糊信息系统的相对约简

下面将经典粗糙集中的不可区分关系、相对正域、相对约简与核等概念推广到 IF 环境.

设 IFIS = $(U, \mathrm{C} \cup \mathrm{D}, V, F)$, $\mathrm{P} \subseteq \mathrm{C}$且 $\mathrm{P} \neq \varnothing$, P 中所有 IF 等价关系的交称为 P 上的 IF 不可区分关系,记为 ind(P), ind(P) 的 IF 等价类集合为 $U/\mathrm{ind}(\mathrm{P})$,简写为 U/P, $U/\mathrm{P} = \otimes\{U/a \mid \forall a \in \mathrm{P}\}$. 若 $\forall a \in \mathrm{P}$, $U/a = \{E_{a1}, E_{a2}, \cdots, E_{ak}\}$, $E_{aj} \in \mathrm{IFS}(U)$,则 $U/\mathrm{P} = \left\{ \bigcap\limits_{a \in \mathrm{P}} E_{aj} \mid j = 1, 2, \cdots, k \right\}$.

根据定义 4,可以很自然地将经典相对正域的思想推广到 IF 环境.

定义 5(直觉模糊相对正域)　设 IFIS = $(U, \mathrm{C} \cup \mathrm{D}, V, F)$, Ψ为 IF 蕴涵算子, $\mathrm{P} \subseteq \mathrm{C}$, D 的 P 正域 $\mathrm{POS}_{\mathrm{P}}(\mathrm{D})$ 为 U 上的 IFS, $\forall x \in U$,

$$\mathrm{POS}_{\mathrm{P}}(\mathrm{D})(x) = (\mu_{\mathrm{POS}_{\mathrm{P}}(\mathrm{D})}(x), \gamma_{\mathrm{POS}_{\mathrm{P}}(\mathrm{D})}(x)) = \sup_{X_j \in U/\mathrm{D}} \inf_{E_i \in U/\mathrm{P}} \Psi[E_i(x), \inf_{y \in U} \Psi(E_i(y), X(y))]. \tag{9}$$

当 IFIS = $(U, \mathrm{C} \cup \mathrm{D}, V, F)$ 为 IF 条件信息系统,IF 蕴涵算子 $\Psi = \Psi_{S_M, N}$ 时,由式(8)得

$$\mathrm{POS}_{\mathrm{P}}(\mathrm{D})(x) = \sup_{X_j \in U/\mathrm{D}} \inf_{E_i \in U/\mathrm{P}} \max\{N(E_i(x)), \inf_{y \in U, y \notin X_j} N(E_i(y))\}. \tag{10}$$

根据 IF 蕴涵算子的单调性,可得如下推论:

推论 1　设 IFIS = $(U, \mathrm{C} \cup \mathrm{D}, V, F)$, $\mathrm{P}_2 \subseteq \mathrm{P}_1 \subseteq \mathrm{C}$, 则 $\forall x \in U$, $\mu_{\mathrm{POS}_{\mathrm{P}_1}(\mathrm{D})}(x) \geq \mu_{\mathrm{POS}_{\mathrm{P}_2}(\mathrm{D})}(x)$, $\gamma_{\mathrm{POS}_{\mathrm{P}_1}(\mathrm{D})}(x) \leq \gamma_{\mathrm{POS}_{\mathrm{P}_2}(\mathrm{D})}(x)$.

设 $\mathrm{P} \subseteq \mathrm{C}$, $R \in \mathrm{P}$,若 $\mathrm{POS}_{\mathrm{P}}(\mathrm{D}) = \mathrm{POS}_{\mathrm{P}-\{R\}}(\mathrm{D})$,则称 R 为 P 中 D 不必要的;否则,称 R 为 P 中 D 必要的. $\forall R \in \mathrm{P}$,若 R 均为 D 必要的,则称 P 为 D 独立的,否则称为依赖的.

定义 6(直觉模糊信息系统的相对约简)　设 IFIS = $(U, \mathrm{C} \cup \mathrm{D}, V, F)$, $\mathrm{S} \subseteq \mathrm{P} \subseteq \mathrm{C}$,称 S 为 P 的 D 约简,当且仅当 S 是 P 的 D 独立子族,且 $\mathrm{POS}_{\mathrm{S}}(\mathrm{D}) = \mathrm{POS}_{\mathrm{P}}(\mathrm{D})$.

一般情况下, P 的 D 约简集不唯一,所有 P 的 D 约简的交称为 P 的 D 核. 所有 P 的 D 约简中维数最小的约简称为最小约简.

在实际应用中,人们往往希望找到信息系统的最小约简,但由于属性的组合爆炸,使得求解最小约简成为 NP-hard 问题. 解决这类问题的一般方法是

采用启发式信息找出最优或次优约简,而依赖度和属性重要性通常作为启发式信息.

下面给出 IF 信息系统的依赖度及属性重要性的度量.信息系统 IFIS = $(U, C \cup D, V, F)$ 的分类能力可由知识 D 对知识 C 的依赖度 $\nu_C(D)$ 来度量,即

$$\nu_C(D) = |\mathrm{POS}_C(D)| / |U|, \tag{11}$$

其中 $|\mathrm{POS}_C(D)|$ 为 $\mathrm{POS}_C(D) \in \mathrm{IFS}(U)$ 的基数.与 FS 不同,$A \in \mathrm{IFS}(U)$ 的基数 $|A| = \left(\sum_{x\in U}\mu_A(x), \sum_{x\in U}(1-\gamma_A(x))\right)$[11],因此可将 $\nu_C(D)$ 表示为一 IF 值,即

$$\nu_C(D) = (\mu_{\nu_C(D)}, \gamma_{\nu_C(D)}) = \left(\sum_{x\in U}\mu_{\mathrm{POS}_C(D)}(x)/|U|, \sum_{x\in U}\gamma_{\mathrm{POS}_C(D)}(x)/|U|\right). \tag{12}$$

$\forall a \in C$, a 对于信息系统的重要性 $\sigma(a)$,可通过计算去掉 a 后信息系统分类能力的变化来度量.由于 $v_C(D)$ 为一 IF 值,本文引入 IF 值的相异度 ds 来度量信息系统分类能力的变化.属性 a 的重要性 $\sigma(a)$ 为

$$\sigma(a) = \mathrm{ds}(v_C(D), v_{C-\{a\}}(D)) = \frac{1}{3}(|\mu_{\nu_C(D)} - \mu_{C-\{a\}}| + |\gamma_{\nu_C(D)} - \gamma_{C-\{a\}}| + |\pi_{\nu_C(D)} - \pi_{C-\{a\}}| + |\tau_{\nu_C(D)} - \tau_{C-\{a\}}|). \tag{13}$$

其中

$$\pi = 1 - \mu - \gamma,\ \tau = (1 - \gamma + \mu)/2.$$

根据式(13),$\sigma(a) \in [0,1]$,$\sigma(a)$ 越大(越小),表明条件属性 a 对信息系统的分类能力影响越大(越小),从而重要性越大(越小).

3.3 基于 IFRS 的启发式约简算法 IFRS-ARE

算法 IFRS-ARE 的初始约简集为一空集,每次迭代选择使依赖度增加最多的条件属性添加到约简集中,直至依赖度不再增加.该算法的关键步骤是计算依赖度,而计算依赖度的核心是相对正域.因此相对正域的计算对于约简算法的效率有直接影响.算法 1 给出了依赖度的简化计算方法.

算法 1 计算依赖度

输入:IF 条件信息系统 IFIS = $(U, C \cup D, V, F)$, U/R, U/D, $R \subset C$, $a_i \in C$;

输出:依赖度 $v_{R\cup\{a_i\}}(D) = (\mu_{\nu_R(D)}, \gamma_{\nu_R(D)})$.

Step1:计算 $R \cup \{a_i\}$ 对论域 U 的 IF 划分 $U/(R \cup \{a_i\}) = \{E_1, E_2, \cdots, E_m\}$;

Step2:$\forall E_i \in U/(R \cup \{a_i\})$, $\forall X_j \in U/D$,循环计算 $t_i = \sup_{X_j\in U/D}\inf_{y\in U, y\notin X_j} N(E_i(y))$,得到 $t = \{t_1, t_2, \cdots, t_m\}$;

Step3:$\forall x \in U$, $\forall E_i \in U/(R \cup \{a_i\})$,计算 $N(E(x)) = \{N(E_1(x)), \cdots, N(E_m(x))\}$;

Step4:$\forall x \in U$,循环计算 $\mathrm{POS}_{R\cup\{a_i\}}(D)(x) = \inf_{E_i\in U/R\cup\{a_i\}} \max\{N(E_i(x)), t_i\}$;

Step5:计算 $v_{R\cup\{a_i\}}(D)$,算法终止,输出 $v_{R\cup\{a_i\}}(D) = (\mu_{\nu_R(D)}, \gamma_{\nu_R(D)})$.

算法 1 对相对正域计算公式进行优化处理,根据式(10),可得 $\forall x \in U$,

$$\mathrm{POS}_{R\cup\{a_i\}}(D)(x) = \sup_{X_j\in U/D}\inf_{E_i\in U/R\cup\{a_i\}} \max\{N(E_i(x)), \inf_{y\in U, y\notin X_j} N(E_i(y))\} = \inf_{E_i\in U/R\cup\{a_i\}} \max\{N(E_i(x)), \sup_{X_j\in U/D}\inf_{y\in U, y\notin X_j} N(E_i(y))\}. \tag{14}$$

分析发现,式(14)中 $\sup_{X_j\in U/D}\inf_{y\in U, y\notin X_j} N(E_i(y))$ 是定值,且参与 $N(E_i(y))$ 计算的并不是论域中的所有对象,而是不包含于 X 的对象 $y \in U$,因此参与计算的对象数目远小于原先的 $|U/D||U|$.在已知 U/R 的情况下,设 $|U/R| = m$,而 $|U/D| \leqslant |U|$,a_i 有 h 个 IF 等价类,则 Step1 ~ Step5 的时间复杂度分别为 $O(mh|U|)$, $O(m|U|^2)$, $O(m|U|)$, $O(m^2)$, $O(|U|)$.若不计 $|U/(R \cup \{a_i\})|$,则算法 1 的时间复杂度为 $O(|U|^2)$.

算法 2 基于 IFRS 的属性约简算法 IFRS-ARE

输入:IF 条件信息系统 IFIS = $(U, C \cup D, V, F)$;

输出:一个近似最小约简 R.

Step1:令 $R = \varnothing$, $U/R = U$, $v_R(D) = (0,1)$, $C' = C$.

Step2:依据决策属性 D 的取值对 U 中对象排序,计算 D 对论域 U 的划分 $U/D = \{X_i \mid X_i \in U/D\}$.

Step3:$\forall a_i \in C'$,根据算法 1 计算 $v_{R\cup\{a_i\}}(D)$.

Step4:$\forall a_i \in C'$,根据式(13)计算 $\sigma(a_i) = \mathrm{dS}(v_R(D), v_{R\cup\{a_i\}}(D))$.

Step5:$\forall a_i \in C'$,求 $\sigma(a_i)$ 的最大值 m:

1) 若 $m = 0$,则算法终止,输出约简集 R;

2) 若满足 $\sigma(a_i) = m$ 的 a_i 有多个,则从中任选一 a_k 作为候选属性,令 $R = R \cup \{a_k\}$, $v_R(D) = v_{R\cup\{a_k\}}(D)$, $C' = C - \{a_k\}$,返回 Step3.

算法 2 依据决策属性对论域中的所有对象进行排序,使每次求取相对正域时,只需以固定的模式使对象参与计算,减少了计算量.使用快速排序,Step2

的时间复杂度为 $O(|U|\log|U|)$. Step3 调用算法 1 计算 $v_{R\cup\{a_i\}}(D)$,由于算法 2 始终保留上一次 U/R 的计算结果,从而可实现相对正域的渐增式计算,最差情况下,Step3 每次所要考虑的条件属性数依次为 $|C|,|C|-1,\cdots,1$,故计算依赖度的总次数为 $(|C|^2+|C|)/2$. 设 $|C|=k$ 为一常数,则算法 2 的时间复杂度为 $O(|U|^2)$.

4　实例计算

下面通过实例对本文的相对约简算法作进一步验证,并与 Jensen 算法[9,10] 进行比较.

算例 1　原始信息系统如表 1 所示. 其中: $U=\{x_1,x_2,\cdots,x_6\}$,条件属性集 $C=\{A,B,C\}$ 的值域为 $[-1,1]$,决策属性 $D=\{d\}$ 的值域为 $\{1,2\}$.

表 1　原始信息系统

U	A	B	C	d
x_1	-0.2	-0.3	0	1
x_2	-0.4	0.5	-0.1	2
x_3	0.3	-0.5	0	2
x_4	-0.3	-0.4	-0.3	1
x_5	0.2	-0.3	0	2
x_6	0.1	0	0	1

首先对条件属性进行 IF 化,将 A,B,C 划分为 2 个 IF 等级 $\{L,H\}$,即 $U/A=\{A_L,A_H\}$, $U/B=\{B_L,B_H\}$, $U/C=\{C_L,C_H\}$. 隶属函数如图 1 所示.

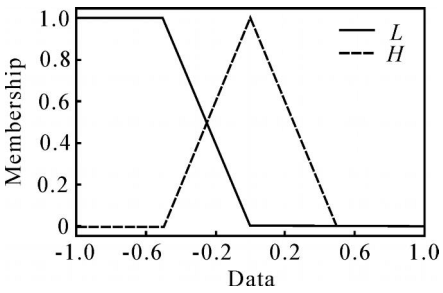


图 1　条件属性模糊化的隶属函数

为了与 Jensen 算法进行比较,非隶属函数取 $\gamma(x)=1-\mu(x)$,得到模糊条件信息系统. 按照 d 的取值对所有对象排序,得到 $U/d=\{X_1,X_2\}=\{\{x_1,x_2,x_3\},\{x_4,x_5,x_6\}\}$ (x_i 为排序后的对象编号).

首先按照 Jensen 算法[9,10] 进行约简.

第 1 层: 计算决策属性 d 对条件属性 A,B,C 的依赖度 $v_A(d)$, $v_B(d)$, $v_C(d)$. 对于 $v_A(d)$,计算 $x\in U$ 对 d 的 A 正域的隶属度为

$$\mu_{POS_{A(d)}}(x_1)=0.4,\ \mu_{POS_{A(d)}}(x_2)=0.4,$$
$$\mu_{POS_{A(d)}}(x_3)=0.4,\ \mu_{POS_{A(d)}}(x_4)=0.4,$$
$$\mu_{POS_{A(d)}}(x_5)=0.4,\ \mu_{POS_{A(d)}}(x_6)=0.4.$$

因此 $v_A(d)=2.4/6$.

同理, $v_B(d)=1.8/6$, $v_C(d)=0.8/6$. 显然, d 对 A 的依赖度最大,因此将 A 加入约简集, $R=\{A\}$,本次最大依赖度 $v_{best}=2.4/6$.

第 2 层: 计算 d 对 $\{A,B\}$, $\{A,C\}$ 的依赖度 $v_{\{A,B\}}(d)=2.4/6$, $v_{\{A,C\}}(d)=2.8/6$. 显然, d 对 $\{A,C\}$ 的依赖度最大, $R=\{A,C\}$,本次最大依赖度 $v_{best}=2.8/6$.

第 3 层: 计算 d 对 $\{A,B,C\}$ 的依赖度 $v_{\{B,C,A\}}(d)=2.4/6$.

可以看出,按照 Jensen 算法计算的 $v_{\{B,C,A\}}(d)=2.4/6<v_{\{A,C\}}(d)=2.8/6$,即出现条件属性增多而依赖度减小的异常情况.

下面按照 IFRS-ARE 算法来求解相对约简.

第 1 层: 计算 d 对 A,B,C 的依赖度. 由算法 1 得

$$\mu_{POS_{A(d)}}(x_1)=0.4,\ \mu_{POS_{A(d)}}(x_2)=0.4,$$
$$\mu_{POS_{A(d)}}(x_3)=0.4,\ \mu_{POS_{A(d)}}(x_4)=0.4,$$
$$\mu_{POS_{A(d)}}(x_5)=0.6,\ \mu_{POS_{A(d)}}(x_6)=0.4.$$

因此 $v_A(d)=2.6/6$.

同理, $v_B(d)=2.8/6$, $v_C(d)=0.8/6$. 显然, d 对 B 的依赖度最大, $R=\{B\}$,本次最大依赖度 $v_{best}=2.8/6$.

第 2 层: $v_{\{B,A\}}(d)=3.6/6$, $v_{\{B,C\}}(d)=3.4/6$. 可以看出, d 对 $\{B,A\}$ 的依赖度最大, $R=\{A,B\}$,本次最大依赖度 $v_{best}=3.6/6$.

第 3 层: 计算 $v_{\{A,B,C\}}(d)=3.6/6$. 显然, d 对 $\{A,B,C\}$ 的依赖度等于上一次得到的最大依赖度,因此算法终止,输出约简结果 $R=\{A,B\}$.

算例 2　原始决策表如表 2 所示. 其中: $U=\{x_1,x_2,\cdots,x_{10}\}$,条件属性集 $C=\{A,B,C\}$ 的值域为 $[0,3]$,决策属性 d 的值域为 $\{1,2,3\}$.

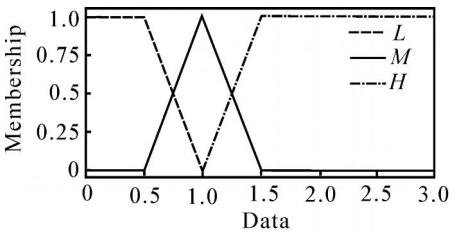
表 2　原始信息系统

U	A	B	C	d	U	A	B	C	d
x_1	0.5	1.5	1.4	1	x_6	1.4	0.85	2.5	2
x_2	0.5	0.4	1.1	3	x_7	0.55	0.8	1.8	2
x_3	0.35	1.6	1.45	1	x_8	1.23	0.4	1.5	3
x_4	0.8	0.5	2	2	x_9	0.7	0.9	1.3	3
x_5	0.6	0.6	1.4	3	x_{10}	0.62	1.4	1.8	1

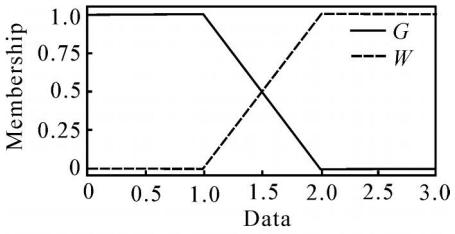
首先对条件属性进行 IF 化,将 A 和 B 划分为 3 个 IF 等级 $\{L,M,H\}$,将 C 划分为 2 个 IF 等级 $\{G,W\}$. 即 $U/A=\{A_L,A_M,A_H\}$, $U/B=\{B_L,B_M,B_H\}$, $U/C=\{C_G,C_W\}$. 隶属函数如图 2 所示.

非隶属函数为

$$\gamma_H(x)=\begin{cases}1, & 0\le x\le 1;\\ 2.9-2x, & 1<x<1.45;\\ 0, & x\ge 1.45.\end{cases}$$



(a) 划分为3个模糊等级L,M,H



(b) 划分为2个模糊等级G和W

图 2 条件属性 IF 化的隶属函数

$$\mu_M(x)=\begin{cases}1, & 0\le x\le 0.5 \text{ or } x\ge 1.5;\\ 1.9-2x, & 0.5<x<0.95;\\ 0, & 0.95\le x\le 1.05;\\ 2x-2.1, & 1.05<x<1.5.\end{cases}$$

$$\mu_L(x)=\begin{cases}0, & 0\le x\le 0.55;\\ 2x-1.1, & 0.55<x<1;\\ 1, & x\ge 1.\end{cases}$$

$$\mu_G(x)=\begin{cases}0, & 0\le x\le 1.1;\\ x-1.1, & 1.1<x<2;\\ 1, & x\ge 2.\end{cases}$$

$$\mu_W=\begin{cases}1, & 0\le x\le 1;\\ 1.9-x, & 1<x<1.9;\\ 0, & x\ge 1.9.\end{cases}$$

按照 d 的取值对所有对象排序,得到

$$U/d=\{X_1,X_2\}=\{\{x_1,x_2,x_3\},\{x_4,x_5,x_6\},\{x_7,x_8,x_9,x_{10}\}\}.$$

下面根据算法 IFRS-ARE 求解相对约简.

第 1 层: 计算 $v_A(d)$, $v_B(d)$, $v_C(d)$. 由算法 1 得

$$POS_A(d)(x_1)=(0,1),$$
$$POS_A(d)(x_2)=(0.14,0.76),$$
$$POS_A(d)(x_3)=(0,1),$$
$$POS_A(d)(x_4)=(0.36,0.54),$$
$$POS_A(d)(x_5)=(0,0.9),$$
$$POS_A(d)(x_6)=(0.44,0.46),$$
$$POS_A(d)(x_7)=(0.1,0.8),$$
$$POS_A(d)(x_8)=(0.36,0.54),$$
$$POS_A(d)(x_9)=(0.3,0.6),$$
$$POS_A(d)(x_{10})=(0,1).$$

因此 $v_A(d)=(1.7/10,7.6/10)$.

同理, $v_B(d)=(3.5/10,6/10)$, $v_C(d)=(2.65/10,6.35/10)$. 根据式(13), 可得 $\sigma(B)>\sigma(C)>\sigma(A)$,即条件属性 B 使依赖度获得最大增长,因此 $R=\{B\}$, 本次最大依赖度为 $v_{best}=(3.5/10,6.0/10)$;

第 2 层: $v_{\{B,A\}}(d)=(5.62/10,3.58/10)$, $v_{\{B,C\}}(d)=(6.4/10,2.8/10)$. 根据式(13), 可得 $\sigma(C)>\sigma(A)$,因此 $R=\{B,C\}$,本次最大依赖度为 $v_{best}=(6.4/10,2.8/10)$.

第 3 层: $v_{(B,C,A\}}(d)=(6.4/10.2.8/10)$. 可以看出, $v_{\{B,C,A\}}(d)=v_{\{B,C\}}(d)$,算法终止,约简结果 $R=\{B,C\}$.

从算例 1 和算例 2 不难得出:

1) 本文方法并未出现条件属性增多而依赖度减小的异常情况,这说明所提出的下近似计算方法比 Jensen 下近似更为合理,约简结果具有较高的可信性.

2) 依赖度越大说明信息丢失越少[12]. 从算例 1 可以看出,采用相同的模糊化方法,本文算法所得最大依赖度 $v_{\{B,A\}}(d)=3.6/6$,大于 Jensen 算法所得最大依赖度 $v_{\{A,C\}}(d)=2.8/6$,因此能更完整地反映出信息系统的分类能力,信息丢失更少.

3) 本文算法通过将对象按决策值进行排序,减小了约简的搜索范围,并采用适当的优化策略,使得相对正域的计算变得简便,从而提高了整个算法的效率.

对于相同的信息系统,设 $|C|=k$,穷尽搜索算法须计算 2^k 次依赖度,且依赖度的计算过程很难简化,而本文算法最多只需计算 $(k^2+k)/2$ 次依赖度,因而在实际应用中具有更好的性能和效率.

5 结 论

属性约简是粗糙集理论研究的核心内容之一. 本文对 Jensen 等价类形式的下近似表示进行改进,新的下近似定义可保证对象对下近似的隶属度随着知识的增多(减少)而增大(减小). 在此基础上,将相对正域、相对约简、相对核等经典粗糙集的知识约简概念推广到直觉模糊环境,提出一种基于直觉模糊粗糙集的启发式属性约简算法,并采用适当的优化策略减少了算法的计算量. 实例计算表明,本文方法比 Jensen 方法更为有效. 下一步工作将结合实际应用,对算法作进一步改进和完善.

参考文献(References)

[1] 刘启和, 李凡, 闵帆, 等. 一种基于新的条件信息熵的高效知识约简算法[J]. 控制与决策, 2005, 20(8): 878-882.

(Liu Q H, Li F, Min F, et al. An efficient knowledge reduction algorithm based on new conditional information entropy[J]. Control and Decision, 2005, 20

(8): 878-882.)

[2] 徐章艳, 刘作鹏, 杨炳儒, 等. 一个复杂度为 max($O\times(|C||U|)$, $O(|C|^2|U/C|)$)的快速属性约简算法[J]. 计算机学报, 2006, 29(3): 391-399.
(Xu Z Y, Liu Z P, Yang B R, et al. A quick attribute reduction algorithm with complexity of max($O(|C|\times|U|)$, $O(|C|^2|U/C|)$)[J]. Chinese J of Computers, 2006, 29(3): 391-399.)

[3] 杨明. 一种基于改进差别矩阵的属性约简增量式更新算法[J]. 计算机学报, 2007, 30(5): 815-822.
(Yang M. An incremental updating algorithm for attribute reduction based on improved discernibility matrix[J]. Chinese J of Computers, 2007, 30(5): 815-822.)

[4] Atanassov K. Intuitionistic fuzzy sets: Theory and applications[M]. Heidelberg: Physica-Verlag, 1999.

[5] 雷英杰, 王宝树. 基于直觉模糊逻辑的近似推理方法[J]. 控制与决策, 2006, 21(3): 305-310.
(Lei Y J, Wang B S. Techniques for approximate reasoning based on intuitionistic fuzzy logic[J]. Control and Decision, 2006, 21(3): 305-310.)

[6] Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets[J]. Int J of General Systems, 1990, 17(2): 191-209.

[7] Radzikowska A M, Kerre E E. A comparative study of fuzzy rough sets[J]. Fuzzy Sets and Systems, 2002, 126(2): 137-155.

[8] Chris C, Cock M D, Kerre E E. Intuitionistic fuzzy rough sets: At the crossroads of imperfect knowledge [J]. Expert Systems, 2003, 20(5): 260-270.

[9] Jensen R, Shen Q. Fuzzy-rough attribute reduction with application to web categorization[J]. Fuzzy Sets and Systems, 2004, 141(3): 469-485.

[10] Jensen R, Shen Q. Fuzzy-rough sets assisted attribute selection[J]. IEEE Trans on Fuzzy Systems, 2007, 15(1): 73-89.

[11] Szmidt E, Kacprzyk J. Entropy for intuitionistic fuzzy sets[J]. Fuzzy Sets and Systems, 2001, 118(3): 467-477.

[12] 刘震宇, 郭宝龙, 杨林耀. 一种新的用于连续值属性离散化的约简算法[J]. 控制与决策, 2002, 17(5): 545-549.
(Liu Z Y, Guo B L, Yan L Y. A new reduction algorithm for discretization of continuous features[J]. Control and Decision, 2002, 17(5): 545-549.)

(上接第 334 页)

[9] Su B L, Chen Z Q, Yuan Z Z. A novel algorithm of constrained multivariable fussy generalized predictive control for non-linear systems[J]. Int J of Modeling, Identification and Control, 2007, 2(2): 120-129.

[10] 苏佰丽, 陈增强, 袁著祉. 多变量非线性系统的有约束模糊预测解耦控制[J]. 系统工程学报, 2007, 22(5): 546-550.
(Su B L, Chen Z Q, Yuan Z Z. Constrained fuzzy predictive control for multivariable nonlinear systems [J]. J of Systems Engineering, 2007, 22(5): 546-550.)

[11] 陈增强, 赵天航, 袁著祉. 基于 Tank-Hopfield 神经网络的有约束多变量广义预测控制器[J]. 控制理论与应用, 1998, 15(6): 847-852.
(Chen Z Q, Zhao T H, Yuan Z Z. The constrained multivariable predictive controller based on Tank-Hopfield neural network[J]. Control Theory and Applications, 1998, 15(6): 847-852.)

[12] Cheng L, Hou Z G, Tan M. Constrained multi-variable generalized predictive control using a dual neural network [J]. Neural Computing & Applications, 2007, 16(6): 505-512.

[13] 李少远, 席裕庚, 王群仙. 小波变换在有约束广义预测控制中的应用[J]. 控制理论与应用, 2001, 18(2): 166-170.
(Li S Y, Xi Y G, Wang Q X. Applications of wavelet to constrained generalized predictive control [J]. Control Theory and Applications, 2001, 18(2): 166-170.)

[14] Maniar V M, Shan S L, Fisher D G, et al. Multivariable constrained adaptive GPC: Theory and experimental evaluation[J]. Int J of Adaptive Control and Signal Processing, 1997, 11(4): 343-365.

[15] 李奇安, 褚健. 对角 CARIMA 模型多变量广义预测控制[J]. 浙江大学学报, 2006, 40(1): 541-545.
(Li Q A, Chu J. Multivariable generalized predictive control for diagonal CARIMA model[J]. J of Zhejiang University, 2006, 40(1): 541-545.)

[16] Jorge Nocedal, Stephen J Wright. Numerical optimization[M]. Beijing: Science Press, 2006.

[17] Fletcher R. Practical methods of optimization [M]. New York: Wiley, 1987.

[18] 李奇安. 广义预测控制算法简化实现方法研究[D]. 杭州: 浙江大学, 2005.
(Li Q A. Study on simplified implementation of generalized predictive control[D]. Hangzhou: Zhejiang University, 2005.)