文章编号:0253-2778(2015)04-0329-08

直觉模糊目标信息系统的正域约简

鲍忠奎[1,2*],杨善林[1]

(1.合肥工业大学管理学院,安徽合肥 230009;2.安徽大学数学科学学院,安徽合肥 230601)

摘要:经典的粗糙集理论对直觉模糊目标信息系统不能直接进行知识约简.为此,首先在直觉模糊目标信息系统中引入优势关系,给出了基于优势关系的直觉模糊粗糙集定义;然后将经典粗糙集理论中的相对正域、属性依赖度以及属性重要性等概念推广至直觉模糊环境中,同时证明了直觉模糊目标信息系统的相对正域具有单调性的特征;结合属性的不同特征以及正域约简的定义给出了正域约简的判定定理,从而设计出以属性重要性为启发式信息的正域约简算法,并给出算法的复杂度分析;最后通过数据实验验证了算法的有效性.

关键词:直觉模糊目标信息系统;直觉模糊粗糙集;优势关系;正域约简

中图分类号:TP18 **文献标识码**:A doi:10.3969/j.issn.0253-2778.2015.04.011

引用格式:BAO Zhongkui, YANG Shanlin. Positive domain reduction in intuitionistic fuzzy objective information systems[J]. Journal of University of Science and Technology of China, 2015,45(4):329-336.
鲍忠奎,杨善林. 直觉模糊目标信息系统的正域约简[J]. 中国科学技术大学学报,2015,45(4):329-336.

Positive domain reduction in intuitionistic fuzzy objective information systems

BAO Zhongkui[1,2], YANG Shanlin[1]

(1. *School of Management, Hefei University of Technology, Hefei* 230009, *China*;
2. *School of Mathematical Sciences, Anhui University, Hefei* 230601, *China*)

Abstract: The classical rough set theory can not be directly used to reduce knowledge for intuitionistic fuzzy objective information systems. To solve this problem, dominance relation was firstly introduced to intuitionistic fuzzy objective information systems, and intuitionistic fuzzy rough set based on dominance relation was defined. Then, the notion of the relative positive domain and the significance of attributes in classical rough set theory were generalized to intuitionistic fuzzy objective information systems, while the monotone property of the relative positive domain was investigated. According to the different characteristics of attributes and the definition of positive domain reduction, the judgment theorem for positive domain reduction was given, the positive domain reduction algorithm using attribute significance as heuristic information was presented, and the complexity analysis of the algorithm was given. Finally, the effectiveness of the proposed algorithm was illustrated with comparative experiments.

Key words: intuitionistic fuzzy objective information systems; intuitionistic fuzzy rough set; dominance relation; positive domain reduction

收稿日期:2014-09-12;**修回日期**:2014-12-29

基金项目:国家自然科学基金(61473001,71071045,71131002);安徽大学青年科学研究基金(33050054).

作者简介:鲍忠奎,男,1981年生,博士生/讲师.研究方向:粗糙集理论及其应用、智能信息处理、决策分析. E-mail:zkbao@ahu.edu.cn

通讯作者:杨善林,中国工程院院士. E-mail: hgdysl@gmail.com

0 引言

粗糙集理论是波兰数学家 Pawlak 于 1982 年提出的一种数据分析理论[1]，它是一种新的处理不确定性知识的数学工具. 目前粗糙集理论已被成功应用于决策分析、数据挖掘、模式识别和智能信息处理等领域[2]. 知识约简是粗糙集理论研究的核心问题之一. 所谓知识约简，就是在保持知识库分类能力不变的条件下，删除其中不相关或不重要的属性. 通过知识约简去掉不必要的属性，从而使得知识表示简化，又不丢失基本信息. 已有许多学者对知识约简做了深入的研究[3-6].

经典的粗糙集理论是建立在等价关系基础上，处理的是符号值，而在实际生活中遇到的更多是模糊概念和模糊知识. 为此，学者们对粗糙集理论进行了模糊环境下的推广，提出了粗糙模糊集理论和模糊粗糙集理论[7]，并对模糊环境下信息系统的知识约简进行了研究[8-10]. 近年来，Atanassov 提出的直觉模糊集[11]和区间值直觉模糊集[12]，是对模糊集理论的进一步拓展，较传统的模糊集有更强的信息表达能力，因此，将粗糙集理论推广至直觉模糊环境中，将更具有理论意义和实际应用价值. 已有不少学者对直觉模糊集和粗糙集的融合进行了研究[13-15]. 这些研究大多集中在对直觉模糊粗糙集定义和性质的研究上，对直觉模糊环境下的知识约简的研究较少. 文献[16-17]首次将优势关系引入直觉模糊目标信息系统和区间值直觉模糊目标信息系统中，给出了基于可辨识矩阵的知识约简方法. 该约简方法可以求出所有的属性约简，但计算耗时过大，而且可辨识矩阵中往往会出现大量的重复元素占用大量内存，所以，一般不适合处理海量数据. 对于论域较大、属性较多的大数据集，有必要引入一种启发式知识约简算法. 鉴于此，本文以直觉模糊目标信息系统为研究对象，首先给出优势关系下的直觉模糊粗糙集定义，然后将经典粗糙集理论中的相对正域、属性依赖度以及属性重要性等概念引入直觉模糊目标信息系统，并结合正域约简的判定定理，提出了一种基于属性重要性的启发式正域约简算法.

1 预备知识

定义 1.1[11] 称 $A=\{<x,\mu_A(x),\nu_A(x)>|\,x\in U\}$ 为非空论域 U 上的直觉模糊集，其中，$\mu_A:U\to[0,1]$，$\nu_A:U\to[0,1]$ 分别称为元素 x 对于 A 的隶属度和非隶属度，且满足 $0\leqslant\mu_A+\nu_A\leqslant1$. 称 $\pi_A(x)=1-\mu_A(x)-\nu_A(x)$ 为 x 对于 A 的犹豫度.

定义 1.2[11] 设 A 和 B 是两个直觉模糊集，则有

① $A\subseteq B\Leftrightarrow\forall x\in U,\mu_A(x)\leqslant\mu_B(x)\wedge\nu_A(x)\geqslant\nu_B(x)$；

② $A=B\Leftrightarrow\forall x\in U,\mu_A(x)=\mu_B(x)\wedge\nu_A(x)=\nu_B(x)$；

③ $A\cap B=\{<x,\mu_A(x)\wedge\mu_B(x),\nu_A(x)\vee\nu_B(x)>|\,x\in U\}$；

④ $A\cup B=\{<x,\mu_A(x)\vee\mu_B(x),\nu_A(x)\wedge\nu_B(x)>|\,x\in U\}$.

为叙述方便，称 $\alpha=<\mu,\nu>$，$0\leqslant\mu,\nu\leqslant1$，$0\leqslant\mu+\nu\leqslant1$ 为直觉模糊数[18].

定义 1.3[11] 设 $a_i=<\mu_i,\nu_i>(1\leqslant i\leqslant n)$ 为 n 个直觉模糊数，则有

① $\bigcap_{1\leqslant i\leqslant n}a_i=<\min_{1\leqslant i\leqslant n}\mu_i,\max_{1\leqslant i\leqslant n}\nu_i>$；

② $\bigcup_{1\leqslant i\leqslant n}a_i=<\max_{1\leqslant i\leqslant n}\mu_i,\min_{1\leqslant i\leqslant n}\nu_i>$.

定义 1.4[18] 设 $\alpha=<\mu,\nu>$ 为直觉模糊数，定义 $s(\alpha)=\mu-\nu$ 为 α 的得分函数，$h(\alpha)=\mu+\nu$ 为 α 的精度，对于两个直觉模糊数 $a_i=<\mu_i,\nu_i>(i=1,2)$ 的大小排序如下：

①若 $s(a_1)>s(a_2)$，则 a_1 大于 a_2，记为 $a_1>a_2$；

②若 $s(a_1)=s(a_2)$，同时 $h(a_1)>h(a_2)$，则 a_1 大于 a_2，记为 $a_1>a_2$；

③若 $s(a_1)=s(a_2)$，同时 $h(a_1)=h(a_2)$，则 a_1 等于 a_2，记为 $a_1=a_2$.

定义 1.5 设 $S=<U,C\cup D,V,f>$ 为一信息系统，其中 U 为非空有限对象集，C 是非空有限条件属性集，$D=\{d\}$为决策属性，且 $C\cap D=\varnothing$；V 是属性值集合，f 是从 $U\times(C\cup D)$ 到 V 的一个映射，它表示每个对象在每个属性上对应一个值，称为信息函数. 若对 $\forall x\in U$，$c\in C$，$f(x,c)$ 和 $f(x,d)$ 均为直觉模糊数，则称 $S=<U,C\cup D,V,f>$ 为直觉模糊目标信息系统（intuitionistic fuzzy objective information systems，IFOIS). 表 1 即为一直觉模糊目标信息系统，其中论域 $U=\{x_1,x_2,x_3,x_4,x_5\}$，条件属性 $C=\{c_1,c_2,c_3,c_4\}$，决策属性 d 关于论域有两个直觉模糊值 D_1 和 D_2.

表 1　直觉模糊目标信息系统

Tab. 1　Intuitionistic fuzzy objective information systems

U	c_1	c_2	c_3	c_4	D_1	D_2
x_1	$<0.3,0.5>$	$<0.3,0.6>$	$<0.7,0.2>$	$<0.6,0.1>$	$<0.4,0.5>$	$<0.5,0.4>$
x_2	$<0.4,0.5>$	$<0.4,0.6>$	$<0.6,0.1>$	$<0.6,0.1>$	$<0.4,0.6>$	$<0.6,0.4>$
x_3	$<0.3,0.5>$	$<0.1,0.7>$	$<0.6,0.1>$	$<0.7,0.2>$	$<0.3,0.6>$	$<0.5,0.4>$
x_4	$<0.4,0.5>$	$<0.4,0.6>$	$<0.7,0.2>$	$<0.7,0.2>$	$<0.5,0.4>$	$<0.4,0.6>$
x_5	$<0.4,0.6>$	$<0.4,0.6>$	$<0.7,0.2>$	$<0.7,0.1>$	$<0.6,0.4>$	$<0.4,0.6>$

2　直觉模糊目标信息系统的正域约简

2.1　IFOIS 的优势关系

根据直觉模糊数的大小关系，在 IFOIS 中建立下面的优势关系.

定义 2.1　给定 IFOIS $=<U,C\cup\{d\},V,f>$，对 $\forall x,y\in U$，$B\subseteq C$，称

$$R_B^{\geqslant}=\{(x,y)\mid f(x,a)>f(y,a)\vee f(x,a)=f(y,a),\forall a\in B\}$$

为 IFOIS 上的优势关系.

记 $[x]_B^{\geqslant}=\{y\in U\mid(y,x)\in R_B^{\geqslant}\}$ 为对象 x 的优势类集合，$[x]_B^{\leqslant}=\{y\in U\mid(x,y)\in R_B^{\geqslant}\}$ 为对象 x 的劣势类集合，且有下列性质成立.

性质 2.1　优势关系 $R_B^{\geqslant}$ 满足自反性和传递性，且 $R_B^{\geqslant}=\bigcap_{a\in B}R_{\{a\}}^{\geqslant}$，即 $[x]_B^{\geqslant}=\bigcap_{a\in B}[x]_{\{a\}}^{\geqslant}$.

例 2.1　由定义 2.1 可得表 1 中各对象的优势类和劣势类，结果见表 2.

表 2　各对象的优势类和劣势类

Tab. 2　Dominating sets and dominated sets of objects with respect to *C*

U	$[x_i]_C^{\geqslant}$	$[x_i]_C^{\leqslant}$
x_1	$\{x_1, x_4, x_5\}$	$\{x_1\}$
x_2	$\{x_2, x_4,\}$	$\{x_2\}$
x_3	$\{x_3, x_4, x_5\}$	$\{x_3\}$
x_4	$\{x_4\}$	$\{x_1, x_2, x_3, x_4\}$
x_5	$\{x_5\}$	$\{x_1, x_3, x_5\}$

2.2　IFOIS 的直觉模糊粗糙集

利用上面的优势关系，将经典的粗糙集定义推广至直觉模糊环境中，给出如下定义.

定义 2.2　给定 IFOIS $=<U,C\cup\{d\},V,f>$，用 U/d 表示 d 对论域 U 的模糊划分，$\forall x\in U$，$B\subseteq C$，任意的直觉模糊集 $D\in U/d$，定义：

$$\underline{B}(D)(x)=\bigcap\{f(y,D):y\in[x]_B^{\geqslant}\},$$
$$\overline{B}(D)(x)=\bigcup\{f(y,D):y\in[x]_B^{\geqslant}\};$$

则 $\underline{B}(D)$ 和 $\overline{B}(D)$ 分别称为直觉模糊集 D 关于条件属性集 B 的下近似和上近似，$(\underline{B}(D),\overline{B}(D))$ 称为直觉模糊粗糙集. $\underline{B}(D)$ 可理解为对象 x 肯定属于直觉模糊集 D 的隶属程度，$\overline{B}(D)$ 可理解为对象 x 可能属于直觉模糊集 D 的隶属程度. 当 B 和 d 均为模糊集时，则 $(\underline{B}(D),\overline{B}(D))$ 退化为模糊环境下的模糊粗糙集；当 B 和 d 均为清晰集时，则 $(\underline{B}(D),\overline{B}(D))$ 退化为优势关系下的 Pawlak 粗糙集，因此直觉模糊粗糙集是已有粗糙集模型的进一步推广.

由定义 2.2 易知，直觉模糊粗糙集有下列性质.

性质 2.2　对 $\forall x\in U$，$A\subseteq B\subseteq C$，$D\in U/d$，有

① $\underline{B}(D)(x)\leqslant\overline{B}(D)(x)$ ；

② $\underline{A}(D)(x)\leqslant\underline{B}(D)(x)$ ；

③ $\overline{B}(D)(x)\leqslant\overline{A}(D)(x)$.

例 2.2　由定义 2.2 可求得表 1 中的直觉模糊集 D_1 和 D_2 关于条件属性集 C 的下、上近似，结果见表 3.

表 3　直觉模糊集 D_1 和 D_2 关于 C 的下、上近似

Tab. 3　The lower and upper approximations of intuitionistic fuzzy set D_1 and D_2

U	$\underline{C}(D_1)(x_i)$	$\overline{C}(D_1)(x_i)$	$\underline{C}(D_2)(x_i)$	$\overline{C}(D_2)(x_i)$
x_1	$<0.4,0.5>$	$<0.6,0.4>$	$<0.4,0.6>$	$<0.5,0.4>$
x_2	$<0.4,0.6>$	$<0.5,0.4>$	$<0.4,0.6>$	$<0.6,0.4>$
x_3	$<0.3,0.6>$	$<0.6,0.4>$	$<0.4,0.6>$	$<0.5,0.4>$
x_4	$<0.5,0.4>$	$<0.5,0.4>$	$<0.4,0.6>$	$<0.6,0.4>$
x_5	$<0.6,0.4>$	$<0.6,0.4>$	$<0.4,0.6>$	$<0.5,0.4>$

2.3　IFOIS 的正域约简

为便于在 IFOIS 中进行知识约简，首先将经典

粗糙集理论中的相对正域和属性依赖度等概念推广至直觉模糊环境中.

2.3.1 相对正域

定义2.3 给定IFOIS $=\langle U,C\cup\{d\},V,f\rangle$,设决策属性 d 有 r 个直觉模糊值 $D_1,D_2,\cdots,D_r$,其中 $D_i(i=1,2,...,r)$ 均为论域 U 上的直觉模糊集,记为 $U/d=\{D_1,D_2,...,D_r\}$. 定义决策属性 d 关于条件属性集 C 的相对正域为:

$$\mathrm{POS}_C(d)(x)=\bigcup_{D_i\in U/d}\underline{C}(D_i)(x),\ x\in U.$$

对于表1,由定义2.3可得决策属性关于条件属性集的相对正域为: $\mathrm{POS}_C(d)=\{\langle x_1,0.4,0.5\rangle,\langle x_2,0.4,0.6\rangle,\langle x_3,0.4,0.6\rangle,\langle x_4,0.5,0.4\rangle,\langle x_5,0.6,0.4\rangle\}$. 另外,IFOIS的相对正域有下列特征.

定理2.1 给定IFOIS $=\langle U,C\cup\{d\},V,f\rangle$,对 $B\subseteq C$,有 $\mathrm{POS}_B(d)\subseteq\mathrm{POS}_C(d)$.

证明 对于 $B\subseteq C$,由性质2.2知,对 $\forall D_i\in U/d$,有 $\underline{B}(D_i)(x)\leqslant\underline{C}(D_i)(x)$,所以,对 $\forall x\in U$, $\mathrm{POS}_B(d)(x)=\bigcup_{D_i\in U/d}\underline{B}(D_i)(x)\leqslant\bigcup_{D_i\in U/d}\underline{C}(D_i)(x)=\mathrm{POS}_C(d)(x)$.

定理2.1表明,IFOIS的相对正域关于条件属性具有单调性,随着条件属性集元素的增加而增大. 接下来,给出属性的必要性和独立性的定义.

定义2.4 给定IFOIS $=\langle U,C\cup\{d\},V,f\rangle$,对 $a\in C$,若 $\mathrm{POS}_{C-\{a\}}(d)=\mathrm{POS}_C(d)$,则称属性 a 在 C 中相对于 d 是不必要的;否则,称 a 在 C 中相对于 d 是必要的.

定义2.5 给定IFOIS $=\langle U,C\cup\{d\},V,f\rangle$,若 $\forall a\in C$ 在 C 中相对于 d 都是必要的,则称 C 相对于 d 是独立的,否则称 C 相对于 d 是相依的. 由 C 中所有相对于 d 的必要属性组成的集合称为 C 相对于 d 的核,记为 $\mathrm{Core}_d(C)$.

鉴于IFOIS的相对正域的单调性特征,给出以下正域约简的定义.

定义2.6 给定IFOIS $=\langle U,C\cup\{d\},V,f\rangle$,对 $B\subseteq C$,若 $\mathrm{POS}_B(d)=\mathrm{POS}_C(d)$,则称 B 是 C 相对于 d 的正域协调集. 若 B 是 C 相对于 d 的正域协调集,且 B 的任何真子集均不是 C 相对于 d 的正域协调集,则称 B 是 C 相对于 d 的一个正域约简.

所有 C 相对于 d 的正域约简构成的集合记作 $\mathrm{Red}_d(C)$,且有 $\mathrm{Core}_d(C)=\bigcap\mathrm{Red}_d(C)$,即 C 的相对核存在于 C 的每个正域约简中.

由上面的定义,IFOIS的正域约简就是保持决策属性关于条件属性集的相对正域不变的独立条件属性子集. 实际应用中,由于属性的组合爆炸,使得寻找最小正域约简成为NP-hard问题. 一般可通过启发式信息搜索来获得最优或次优的正域约简. 为便于获得启发式信息,下面先给出属性的依赖度和重要性的定义.

2.3.2 属性的依赖度和重要性

定义2.7 给定IFOIS $=\langle U,C\cup\{d\},V,f\rangle$,决策属性 d 对条件属性集 C 的依赖度定义为:

$$\gamma_C(d)=\frac{|\mathrm{POS}_C(d)|}{|U|}.$$

其中, $|\sim|$ 表示集合 $\sim$ 的基数.

采用文献[19]中给出的计算直觉模糊集基数的方法,即 $|A|=\frac{1}{2}\sum_{x\in U}(1+\mu_A(x)-\nu_A(x))$. 所以,对直觉模糊集 $\mathrm{POS}_C(d)=\{\langle x_1,0.4,0.5\rangle,\langle x_2,0.4,0.6\rangle,\langle x_3,0.4,0.6\rangle,\langle x_4,0.5,0.4\rangle,\langle x_5,0.6,0.4\rangle\}$ 来说,其基数 $|\mathrm{POS}_C(d)|=2.4$,所以表1中的决策属性对条件属性集的依赖度为 $\gamma_C(d)=0.48$.

由定义2.7知, $0\leqslant\gamma_C(d)\leqslant 1$. 当 $\gamma_C(d)=1$ 时,则称属性 d 完全依赖于条件属性集 C;当 $\gamma_C(d)=0$,则称 d 完全不依赖于 C;若 $\gamma_C(d)=k\ (0<k<1)$,则称 d 依赖于 C 的程度为 k.

一般地,IFOIS的不同条件属性相对于决策属性有不同的重要性. 这种重要性可通过从IFOIS中去掉该属性后,条件属性集相对于决策分类能力的变化来衡量. 变化越大,说明该属性相对于决策越重要;变化越小,说明该属性相对于决策越不重要. 为此,给出下面属性重要性的定义.

定义2.8 给定IFOIS $=\langle U,C\cup\{d\},V,f\rangle$,任意属性 $a\in C$ 对条件属性 C 相对于决策属性 d 的重要性定义为:

$$\mathrm{sig}^{\mathrm{inner}}(a,C,d)=\gamma_C(d)-\gamma_{C-\{a\}}(d).$$

这里,利用属性集 C 中的属性 a 在去掉前后 d 对 C 依赖度的变化来度量 a 在 C 中相对于 d 的重要性.

性质2.3 给定IFOIS $=\langle U,C\cup\{d\},V,f\rangle$, $a\in C$,有 $0\leqslant\mathrm{sig}^{\mathrm{inner}}(a,C,d)\leqslant 1$.

定理2.2 给定IFOIS $=\langle U,C\cup\{d\},V,f\rangle$, $a\in C$,则 a 是 C 中相对于 d 必要的属性 $\Leftrightarrow\mathrm{sig}^{\mathrm{inner}}(a,C,d)>0$.

证明　“$\Rightarrow$” a 是 C 中相对于 d 必要的，则 $POS_{C-\{a\}}(d) \neq POS_C(d)$，由定义 1.2 知，$\exists x \in U$，使得 $\mu_{POS_{C-\{a\}}(d)}(x) \neq \mu_{POS_C(d)}(x) \vee \nu_{POS_{C-\{a\}}(d)}(x) \neq \nu_{POS_C(d)}(x)$. 另外，由正域的单调性知，$POS_{C-\{a\}}(d) \subseteq POS_C(d)$，即对 $\forall x \in U$ 有 $\mu_{POS_{C-\{a\}}(d)}(x) \leqslant \mu_{POS_C(d)}(x) \wedge \nu_{POS_{C-\{a\}}(d)}(x) \geqslant \nu_{POS_C(d)}(x)$，所以，$\sum_{x\in U}(1+\mu_{POS_{C-\{a\}}(d)}(x)-\nu_{POS_{C-\{a\}}(d)}(x)) < \sum_{x\in U}(1+\mu_{POS_C(d)}(x)-\nu_{POS_C(d)}(x))$，即

$|POS_{C-\{a\}}(d)| < |POS_C(d)|$，

故 $sig^{inner}(a,C,d) > 0$.

“$\Rightarrow$”若 $sig^{inner}(a,C,d) > 0$，

则 $|POS_C(d)| > |POS_{C-\{a\}}(d)|$，所以，$POS_{C-\{a\}}(d) \neq POS_{C\}}(d)$（这由反证法易得）. 由定义 2.4 知，$a$ 是 C 中相对于 d 必要的.

由定理 2.2 和定义 2.5 易得下面的结论.

定理 2.3　给定 $IFOIS = <U, C\cup\{d\}, V, f>$，有 $Core_d(C) = \{a \in C \mid sig^{inner}(a,C,d) > 0\}$.

定理表明，IFOIS 的相对核 $Core_d(C)$ 是由 C 中那些属性重要性大于 0 的属性（即必要属性）构成的.

由上面的定义和定理可得出正域约简的判定定理.

定理 2.4　给定 $IFOIS = <U, C\cup\{d\}, V, f>$，$B \subseteq C$，$B$ 是 C 的正域约简 $\Leftrightarrow \gamma_B(d) = \gamma_C(d)$，且对 $\forall a \in B$，有 $sig^{inner}(a,B,d) > 0$.

判定定理提供了一个求 C 的正域约简的方法，即找到 C 的一个子集 B，使得对 $\forall a \in B$，$sig^{inner}(a,B,d) > 0$. 由于 IFOIS 的相对核 $Core_d(C)$ 是唯一的且为所有正域约简的子集，故可将 $Core_d(C)$ 作为最小正域约简的基础，然后将满足 $a \in B - Core_d(C)$，且 $sig^{inner}(a,B,d) > 0$ 的属性添加到 $Core_d(C)$ 中，从而获得正域约简. 为方便搜索满足条件的属性，下面给出属性集 B 之外即 $C-B$ 中的属性 a 关于属性集 B 的重要性的定义.

定义 2.9　给定 $IFOIS = <U, C\cup\{d\}, V, f>$，$B \subseteq C$，则对于任意属性 $a \in C-B$ 对于属性集 B 相对于 d 的重要性定义为：

$$sig^{outer}(a,B,d) = \gamma_{B\cup\{a\}}(d) - \gamma_B(d).$$

$sig^{outer}(a,B,d)$ 的值越大，表明属性 a 关于属性集 B 越重要. 所以，可以 $sig^{outer}(a,B,d)$ 为启发式信息，来寻找符合条件的属性.

2.4　IFOIS 的正域约简算法

根据以上分析，可在 IFOIS 中设计基于属性重要性的正域约简算法. 即采用 bottom-up 的方式，以 IFOIS 的相对核为起点，依据属性重要性的定义，逐次选择最重要的属性添加到相对核中，直至终止条件满足. 具体步骤如下：

输入：$IFOIS = <U, C\cup\{d\}, V, f>$

输出：正域约简 B

Step1　计算 IFOIS 的决策属性 d 关于条件属性集 C 的依赖度 $\gamma_C(d)$；

Step2　令 $Core_d(C) = \varnothing$，计算每个属性 $a \in C$ 在 C 中关于 d 的重要性 $sig^{inner}(a,C,d)$. 若 $sig^{inner}(a,C,d) > 0$，则 $Core_d(C) = Core_d(C) \cup \{a\}$，得到条件属性集 C 的相对核 $Core_d(C)$；

Step3　计算决策属性集 d 关于相对核 $Core_d(C)$ 的依赖度. 若 $\gamma_{Core_d(C)}(d) = \gamma_C(d)$，则输出核 $Core_d(C)$，即为 IFOIS 的正域约简；否则，进入下一步；

Step4　令 $Core_d(C) = B$，对属性集 C-B，重复执行：

①对 $\forall a \in C-B$，计算 $sig^{outer}(a,B,d)$；

②选择属性 a，使其满足 $sig^{outer}(a,B,d) = \max_{a'\in C-B} sig^{outer}(a',B,d)$，令 $B = B \cup \{a\}$；

③若 $\gamma_B(d) = \gamma_C(d)$，则输出正域约简 B；否则，转入①.

算法的时间复杂度分析如下：

Step1 计算 $\gamma_C(d)$ 的时间复杂度为：

$$O(|U|^2|C||U/d|);$$

Step2 要对每个属性计算 $sig^{inner}(a,C,d)$，所以时间复杂度为：

$$O(|U|^2|C|^2|U/d|);$$

Step4 每添加一个属性到正域约简集 B 中，需计算所有的 $sig^{outer}(a',B,d)$，其时间复杂度为：

$$O(|U|^2|B||C-B||U/d|);$$

所以，最坏情形下，Step4 的时间复杂度为：

$$O(|U|^2[|C|+(|C|-1)\cdot 2+\cdots+3\cdot(|C|-2)+2\cdot(|C|-1)+|C|]|U/d|),$$

而该式的时间复杂度不超过 $O(|U|^2|C|^3|U/d|)$；因此，整个约简算法时间复杂度不超过 $O(|U|^2|C|^3|U/d|)$.

当 IFOIS 的直觉模糊决策属性仅取单个直觉模糊集（即 $|U/d|=1$）时，本文的正域约简算法的复杂度不超过 $O(|U|^2|C|^3)$，而文献[16]给出的基于可辨识矩阵的属性约简算法的时间复杂度为

$O(|U|^2|C|\min\{2^{|C|},|U|^2\})$. 对比可以看出，本文约简算法的时间复杂度要低很多. 另外，基于可辨识矩阵的属性约简算法还需要大量的存储空间，其空间复杂度为 $O(|U|^2|C|)$，而本文约简算法的空间复杂度仅为 $O(|U||C|)$.

3 算例分析

首先，选用文献[16]中关于计算机审计风险评估的实例来阐述本文提出的正域约简算法. 表 4 为计算机审计风险评估决策表，是一个直觉模糊目标信息系统，$U=\{x_1,x_2,\cdots,x_{10}\}$ 为 10 个审计对象，$C=\{c_1,c_2,c_3,c_4,c_5\}$ 为条件属性集，其中 c_1 表示“优良的系统环境”，c_2 表示“较好的系统控制”，c_3 表示“安全的财务数据”，c_4 表示“可信的审计软件”，c_5 表示“标准化操作”，d 表示“可接收的最大计算机审计风险”.

表 4 计算机审计风险评估决策表[16]

Tab. 4 A computer audit risk assessment decision table

U	c_1	c_2	c_3	c_4	c_5	d
x_1	$<0.4,0.5>$	$<0.3,0.5>$	$<0.8,0.2>$	$<0.4,0.5>$	$<0.7,0.1>$	$<0.3,0.6>$
x_2	$<0.3,0.5>$	$<0.4,0.5>$	$<0.6,0.1>$	$<0.4,0.5>$	$<0.7,0.3>$	$<0.4,0.6>$
x_3	$<0.3,0.5>$	$<0.1,0.8>$	$<0.8,0.1>$	$<0.4,0.5>$	$<0.7,0.3>$	$<0.2,0.7>$
x_4	$<0.1,0.8>$	$<0.1,0.8>$	$<0.4,0.5>$	$<0.1,0.8>$	$<0.8,0.2>$	$<0.2,0.8>$
x_5	$<0.7,0.3>$	$<0.4,0.5>$	$<0.9,0.1>$	$<0.4,0.6>$	$<0.8,0.1>$	$<0.4,0.6>$
x_6	$<0.3,0.6>$	$<0.4,0.6>$	$<0.7,0.2>$	$<0.5,0.5>$	$<0.8,0.2>$	$<0.4,0.5>$
x_7	$<0.4,0.5>$	$<0.4,0.5>$	$<0.8,0.2>$	$<0.4,0.5>$	$<0.8,0.2>$	$<0.6,0.4>$
x_8	$<0.4,0.6>$	$<0.4,0.5>$	$<0.9,0.1>$	$<0.7,0.3>$	$<0.8,0.2>$	$<0.6,0.4>$
x_9	$<0.4,0.6>$	$<0.7,0.3>$	$<0.9,0.1>$	$<0.4,0.5>$	$<0.9,0.0>$	$<0.8,0.2>$
x_{10}	$<0.7,0.3>$	$<0.7,0.3>$	$<0.8,0.2>$	$<0.9,0.0>$	$<0.4,0.5>$	$<0.8,0.2>$

Step1 计算每个对象在每个属性$\{c_i\}$下的优势类(结果见表 5). 所以对象关于属性集 C 的优势类为：$[x_1]_C^{\geqslant}=\bigcap_{c_i\in C}[x_1]_{\{c_i\}}^{\geqslant}=\{x_1,x_7\}$，$[x_2]_C^{\geqslant}=\{x_2,x_7,x_8,x_9\}$，$[x_3]_C^{\geqslant}=\{x_3,x_8,x_9\}$，$[x_4]_C^{\geqslant}=\{x_4,x_5,x_7,x_8,x_9\}$，$[x_5]_C^{\geqslant}=\{x_5\}$，$[x_6]_C^{\geqslant}=\{x_6,x_8\}$，$[x_7]_C^{\geqslant}=\{x_7\}$，$[x_8]_C^{\geqslant}=\{x_8\}$，$[x_9]_C^{\geqslant}=\{x_9\}$，$[x_{10}]_C^{\geqslant}=\{x_{10}\}$. 从而得到决策属性 d 关于 C 的依赖度 $\gamma_C(d)=0.485$.

表 5 各对象在属性$\{c_i\}$下的优势类

Tab. 5 Dominating sets of objects with respect to $\{c_i\}$

U	$[x_i]_{\{c_1\}}^{\geqslant}$	$[x_i]_{\{c_2\}}^{\geqslant}$	$[x_i]_{\{c_3\}}^{\geqslant}$	$[x_i]_{\{c_4\}}^{\geqslant}$	$[x_i]_{\{c_5\}}^{\geqslant}$
x_1	{1,5,7,10}	{1,2,5～10}	{1,3,5,7～10}	{1～3,6～10}	{1,4～9}
x_2	{1～3,5,7～10}	{2,5,7～10}	{1～3,5～10}	{1～3,6～10}	{1～9}
x_3	{1～3,5,7～10}	{1～10}	{3,5,8,9}	{1～3,6～10}	{1～9}
x_4	{1～10}	{1～10}	{1～10}	{1～10}	{4～9}
x_5	{5,10}	{2,5,7～10}	{5,8,9}	{1～3,5～10}	{5,9}
x_6	{1～3,5～10}	{2,5～10}	{1,3,5～10}	{6,8}	{4～9}
x_7	{1,5,7,10}	{2,5,7～10}	{1,3,5,7～10}	{1～3,6～10}	{4～9}
x_8	{1,5,7～10}	{2,5,7～10}	{5,8,9}	{8,10}	{4～9}
x_9	{1,5,7～10}	{9,10}	{5,8,9}	{1～3,6～10}	{9}
x_{10}	{5,10}	{9,10}	{1,3,5,7～10}	{10}	{1～10}

注：表中的数字 1,2,…,10 分别表示对象 $x_1,x_2,\cdots,x_{10}$.

Step2 计算属性 $a \in C$ 在 C 中关于 d 的重要性：

$\mathrm{sig}^{\mathrm{inner}}(c_1,C,d) = \gamma_C(d) - \gamma_{C-\{c_1\}}(d) = 0$，$\mathrm{sig}^{\mathrm{inner}}(c_3,C,d) = 0$，$\mathrm{sig}^{\mathrm{inner}}(c_2,C,d) = 0.015$，$\mathrm{sig}^{\mathrm{inner}}(c_4,C,d) = 0.045$，$\mathrm{sig}^{\mathrm{inner}}(c_5,C,d) = 0$. 由定理 2.3 知，表 4 的相对核($\mathrm{Core}_d(C)=\{c_2,c_4\}$. 而 $\gamma_{\mathrm{Core}_d(C)}(d) = 0.445 \neq \gamma_C(d)$，进入下一步；

Step3 令 $\mathrm{Core}_d(C) = \{c_2,c_4\} = B$. 对 $C-B=\{c_1, c_3, c_5\}$，计算各属性关于 B 的重要性：

$\mathrm{sig}^{\mathrm{outer}}(c_1,B,d) = \gamma_{B\cup\{c_1\}}(d) - \gamma_B(d) = 0.04$，$\mathrm{sig}^{\mathrm{outer}}(c_3,B,d) = 0.04$，$\mathrm{sig}^{\mathrm{outer}}(c_5,B,d) = 0.04$. 所以，$B_1=\{c_1,c_2,c_4\}$，$B_2=\{c_2,c_3,c_4\}$，$B_3=\{c_2,c_4,c_5\}$，经验证，$\gamma_{B_1}(d) = \gamma_C(d)$，$\gamma_{B_2}(d) = \gamma_C(d)$，$\gamma_{B_3}(d) = \gamma_C(d)$，算法终止. 所以，IFOIS 的正域约简为：$\{c_1,c_2,c_4\}$或$\{c_2,c_3,c_4\}$或$\{c_2,c_4,c_5\}$. 这与文献[16]中利用可辨识矩阵的约简方法所得的结果一致，即本文提出的基于属性重要性的启发式约简算法可以有效地获得 IFOIS 的正域约简.

为了进一步验证本文所提约简算法的有效性和优越性，从 UCI 数据集中挑选了 4 组数据集来进行属性约简实验，各数据集的描述见表 6，这些数据集的属性值全部是数值型的.

表 6 数据描述

Tab. 6 Data description

serial number	data set	abbreviation	samples	numerical attributes
1	Housing	Housing	506	13
2	Concrete compressive strength	Concrete	1 030	9
3	Wine-quality-red	Wine-r	1 599	12
4	Wine-quality-white	Wine-w	4 898	12

为便于两种算法的比较，从每个数据集中分别选取三组容量不同的数据集，用这两种算法来进行属性约简实验，实验结果汇总在表 7(表中 N_1 和 N_2 分别表示约简前后属性的个数，PR 代表本文的约简算法，DMR 代表文献[16]的约简算法). 整个约简算法是在装有 windows XP 系统的台式机(Intel(R) Core(TM)2 Duo CPU E7500 2.93GHz 2.00GB 内存)上通过 Microsoft Visual C＋＋ 2008 编程实现的.

表 7 PR 和 DMR 算法的属性约简和约简时间

Tab. 7 The time and attribute reduction of the algorithms PR and DMR

data set	samples	numerical attributes		time(s)	
		N_1	N_2	PR	DMR
Housing	50	13	8	0.000 0	0.000 0
	100	13	11	0.031 0	0.071 0
	200	13	12	0.078 0	0.547 0
Concrete	100	9	6	0.015 0	0.063 0
	200	9	7	0.078 0	0.641 0
	400	9	9	0.266 0	11.391 0
Wine-r	100	12	9	0.062 0	0.083 0
	200	12	10	0.094 0	0.219 0
	400	12	11	0.234 0	6.265 0
Wine-w	100	12	8	0.016 0	0.131 0
	200	12	11	0.079 0	0.375 0
	400	12	12	0.328 0	6.078 0

从表 7 可以看出，本文提出的启发式约简算法不仅可以获得数据集的属性约简，而且所花费的时间要远远少于基于可辨识矩阵的属性约简方法，并且随着数据集样本个数的增加，优势变得更加明显. 当表中数据集 Housing 的样本个数取 50 时，两种约简算法所花费时间都大约为 0.000 0，这表示当数据集很小时，约简算法在很短的时间内获得属性约简. 另外，由表 7 可以看出，PR 约简算法随着数据集样本个数的增多，所得约简集中属性的个数会增多，约简也要花费更多的时间，这是符合实际情况的. 因为这里研究的是数值型属性约简问题，而且数据集中属性的个数也只有有限的几个，所以随着数据集的增大，每个属性都会变得越发重要，从而导致可以约简的属性越来越少，而增大的数据集必然要花费更多的时间来进行属性约简. 通过以上的数据实验可以看出，本文的约简算法要优于文献[16]的基于可辨识矩阵的属性约简方法.

4 结论

经典的粗糙集理论对直觉模糊目标信息系统不能直接进行属性约简. 本文在直觉模糊目标信息系统中引入优势关系，给出了直觉模糊粗糙集的定义，然后将经典粗糙集理论中的相对正域、属性依赖度以及属性重要性等概念引入直觉模糊环境中，并结

合正域约简的概念给出正域约简的判定定理,从而设计出以属性重要性为启发式信息的正域约简算法.最后通过数据实验,验证了方法的有效性.下一步将对直觉模糊目标信息系统的规则提取进行研究.

参考文献(References)

[1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356.

[2] Pawlak Z, Skowron A. Rough sets: Some extensions [J]. Information Sciences, 2007, 177(1): 28-40.

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

[4] Yang X B, Yang J Y, Wu C, et al. Dominance-based rough set approach and knowledge reductions in incomplete ordered information systems [J]. Information Sciences, 2008, 178(4): 1219-1234.

[5] Qian Y H, Dang C Y, Liang J Y, et al. Set-valued ordered information systems[J]. Information Sciences, 2009, 179(16): 2809-2832.

[6] Li H X, Zhou X Z, Zhao J B, et al. Non-monotonic attribute reduction in decision-theoretic rough sets[J]. Fundamenta Informaticae, 2013, 126(4): 415-432.

[7] Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets[J]. International Journal of General System, 1990, 17(2-3): 191-209.

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

[9] CornelisC, Jensen R, Hurtado. G, et al. Attribute selection with fuzzy decision reducts[J]. Information Sciences, 2010, 180(2): 209-224.

[10] Chen D G, Zhao S Y. Local reduction of decision system with fuzzy rough sets[J]. Fuzzy Sets and Systems, 2010, 161(13): 1871-1883.

[11] Atanassov K T. Intuitionistic fuzzy sets[J]. Fuzzy Set and Systems, 1986, 20(1): 87-96.

[12] Atanassov K T, Gargor G. Interval-valued intuitionistic fuzzy sets[J]. Fuzzy Sets and Systems, 1989, 31(3): 343-349.

[13] Zhou L, Wu W Z, Zhang W X. On characterization of intuitionistic fuzzy rough sets based on intuitionistic fuzzy implicators[J]. Information Sciences, 2009, 179(7): 883-898.

[14] Zhou L, Wu W Z. Characterization of rough set approximations in Atanassov intuitionistic fuzzy set theory [J]. Computers and Mathematics with Applications, 2011, 62(1): 282-296.

[15] 徐小来,雷英杰,谭巧英. 基于直觉模糊三角模的直觉模糊粗糙集[J]. 控制与决策,2009,23(8):900-904.
Xu X L, Lei Y J, Tan Q Y. Intuitionistic fuzzy rough sets based on intuitionistic fuzzy triangle norm[J]. Control and Decision, 2009, 23(8): 900-904.

[16] Huang B, Li H X, Wei D K. Dominance-based rough set model in intuitionistic fuzzy information systems [J]. Knowledge-Based Systems, 2012, 28: 115-123.

[17] Huang B, Wei D K, Li H X, et al. Using a rough set model to extract rules in dominance-based interval-valued intuitionistic fuzzy information systems [J]. Information Sciences, 2013, 221: 215-229.

[18] Xu Z S. Intuitionistic preference relations and their application in group decision making[J]. Information Sciences, 2007, 177(11): 2363-2379.

[19] Vlachos I K, Sergiadis G D. Subsethood, entropy, and cardinality for interval-valued fuzzy sets-A algebraic derivation[J]. Fuzzy Sets and Systems, 2007, 158(12): 1384-1396.