第 29 卷 第 7 期
Vol. 29 No. 7

控 制 与 决 策
Control and Decision

2014 年 7 月
Jul. 2014

文章编号: 1001-0920 (2014) 07-1325-05　　DOI: 10.13195/j.kzyjc.2013.0554

基于置信优势关系的粗糙集近似模型

苟光磊[1,2,3], 王国胤[1,2], 利 节[2], 吴 迪[2]

(1. 西南交通大学 信息科学与技术学院，成都 610031；2. 中国科学院重庆绿色智能技术研究院，重庆 401122；3. 重庆理工大学 计算机科学与工程学院，重庆 400054)

摘 要: 不完备有序信息处理是现实生活中的常见问题. 多种拓展优势关系及其粗糙集模型被提出并用于解决不完全的偏好决策问题, 但均未考虑序关系特性, 与现实语义存在矛盾. 对此, 提出一种置信优势关系及其粗糙集近似模型, 讨论了基于置信优势关系的粗糙集模型与已有模型的关系. 与现有的拓展关系相比, 该置信优势关系满足序关系特性, 避免了语义上的矛盾. 定理证明和实例分析表明, 置信优势关系粗糙集近似模型的近似精度和分类精度更优.

关键词: 不完备序信息系统；粗糙集；拓展优势关系；置信优势关系

中图分类号: TP520.6080　　**文献标志码:** A

Confidential dominance relation based rough approximation model

GOU Guang-lei[1,2,3], *WANG Guo-yin*[1,2], *LI Jie*[2], *WU Di*[2]

(1. School of Information Science and Technology，Southwest Jiaotong University，Chengdu 610031，China; 2. Chongqing Institute of Green and Intelligent Technology，Chinese Academy of Sciences，Chongqing 401122，China; 3. School of Computer Science and Engineering，Chongqing University of Technology，Chongqing 400054，China. Correspondent：GOU Guang-lei，E-mail：ggl@cqut.edu.cn)

Abstract: Incomplete ordered information processing is a common problem in the real life. Various extended dominance relation rough set models are already proposed to solve the incomplete ordinal decision problem. But there exits some ambivalence over the real semantics because the characteristics of the order relation are not considered. Therefore, the confidence dominance relation and its rough set approximation model are presented. Relationships between the confidential dominance relation based rough set model and existing models are discussed. Compared with the current expansion relations, confidence dominance relation fits the characteristics of the order relation to avoid semantic contradiction. Theorem proving and example analysis show that the confidential dominance relation based rough approximation model has better approximate accuracy and classification accuracy.

Key words: incomplete ordered information systems；rough set；extended dominance relation；confidential dominance relation

0 引 言

随着计算机网络技术的快速发展, 各个领域的数据量急剧增加, 但由于数据采集技术的限制、传输故障以及一些人为因素等原因, 造成数据缺损和丢失现象时常发生. 在现实世界中, 受环境的复杂性和不确定性的影响, 人们将面对信息不确定、不完全以及具有决策偏好信息的决策问题, 因此, 不完备有序信息系统的研究受到了人们的极大关注.

粗糙集理论是波兰学者 Pawlak 等[1]提出的一种处理不确定和含糊信息的新型数学工具, 它建立在等价关系 (满足自反性、对称性及传递性) 基础上, 只能用于处理离散型的完备信息系统. Greco 等[2]利用优势关系 (满足自反性和传递性) 替代等价关系, 提出了优势关系粗糙集 (DRSA), 将经典粗糙集扩展到完备的有序信息系统. 为解决序信息系统下缺失值的问题, Greco 等[3]对优势关系进行了扩展; 在类似观点下, 何亚群等[4]提出了扩展优势关系的概念, 利用扩展优势关系得到知识的粗糙近似; Shao 等[5]讨论了扩展优势关系在不完备序信息系统下的推理, 可直接获取决策规则, 并提出了知识约简方法. 扩展优势关系虽然满足自反性和传递性, 但定义过于宽松. 为避免两个对象没有共同非空属性可以比较的情况, 胡明礼等[6]提

收稿日期: 2013-05-03；**修回日期:** 2013-07-08.

基金项目: 国家自然科学基金项目(61073146)；重庆市自然科学基金项目(cstc2012jjA40032).

作者简介: 苟光磊(1980−), 男, 讲师, 博士生, 从事粗糙集、粒计算的研究；王国胤(1970−), 男, 教授, 博士生导师, 从事粗糙集、智能信息处理等研究.

出了有限扩展优势关系, 限定两个对象至少有一个共同非空属性可进行比较. 有限扩展模型满足自反性, 在属性较多的情况下, 定义仍然过分宽松. 对此, 文献 [7] 提出了广义扩展优势关系, 使用阈值λ来限定两个对象可比较的共同非空属性的个数, 可以看出, 广义扩展优势关系是有限扩展优势关系的特例, 但广义扩展模型并不满足自反性. Chen 等[8]在广义扩展优势关系基础上, 将缺失值细化为"丢失"和"不关心"两种情况, 提出了k度扩展特性优势关系, 讨论了该关系粗集模型下的动态更新近似原理及算法. Yang 等[9]提出了相似优势关系定义, 将缺失值视为不确定、不存在的值, 无法对其进行描述, 仅允许比较已知值部分, 缺失值部分不允许比较. 相似优势关系满足自反性和传递性, 但优势关系和劣势关系采用两种定义形式, 使得$y \succeq x$不一定能得到$x \preceq y$. 骆公志等[10]提出了限制优势关系, 将缺失值对象的优势关系限定在属性最大值和最小值的情况下, 并结合限制优势关系和相似关系提出了限制相似优势关系[11], 这两个定义都需知道属性值的最大值和最小值, 而且定义过于严格, 即缺失值对象只能在属性最大值或最小值情况下才能确定限制优势关系. Yang 等[12]提出了量化优势关系以表示对象间满足优势关系的概率, 关系定义都是基于属性值之间存在统一的概率分布. 上述的各种拓展优势关系被提出并用于处理不完备序信息, 但均欠考虑"序"系统的特性, 存在某些序特征下的语义矛盾.

本文认为对于不完备序信息系统, 拓展的优势关系应首先遵从序关系特性, 即满足自反性、传递性和序对称性. 此外, 考虑到既然是对优势关系的拓展, 假定对象x_1优于对象x_2, 则在x_2已知的信息上, x_1不劣于x_2, 而且x_1的已知信息不少于x_2, 这符合现实情况. 例如在分析监测数据时, 人们更愿意使用信息更优、更详细的数据. 基于此, 本文提出一种新的拓展优势关系, 即置信优势关系 (CDR), 它满足序关系特性, 避免了语义矛盾; 并提出了置信优势关系的粗集近似模型, 讨论其与已有的拓展优势关系及其粗集模型的关系. 定理证明和实例验证表明, 本文提出的方法在几种方法中, 可以得到最大的正域和最小的边界域, 因此在近似精度和分类精度上均优于已有方法.

1 基础理论

定义 1 (不完备有序决策系统) 设有一个决策系统$\mathrm{DS} = (U, A, V, f)$. 其中: U是论域, 即非空的对象集合; A是属性集合, $A = C \bigcup D$, C和D分别表示条件属性集和决策属性集; V是属性值域, 具有偏好; $f: U \times A \to V$是信息函数, $f = \{f(x_i, a) | f(x_i, a) : x_i \to v_a, a \in C, x_i \in U, 1 \leqslant i \leqslant |U|\}$.

定义 2 (优势关系) 设$P \subseteq C, x, y \in U, \forall a \in P$, $f(y, a) \succeq f(x, a)$, 满足这样的关系称为优势关系, 用$yD_P x$表示.

以下定义均满足假设$x, y \in U$, $P \subseteq C$, $B_P = \{b | b \in P \bigwedge f(x, b) \neq *\}$.

定义 3 (扩展优势关系) $\mathrm{EDom}(P) = \{(x, y) \in U \times U | \forall q \in P, f(y, q) \geqslant f(x, q) \bigvee f(x, q) = * \bigvee f(y, q) = *\}$, 用$yD_P^{\mathrm{EDom}} x$表示. 限制优势关系[10]是定义 3 的特殊形式, 仅允许在属性最大值和最小值进行扩展.

定义 4 (有限扩展优势关系) $\mathrm{LEDom}(P) = \{(x, y) \in U \times U | \forall q \in P, (B_P(x) \bigcap B_P(y) \neq \varnothing) \bigwedge ((f(x, q) = * \bigwedge f(y, q) = *) \bigvee (f(x, q) = * \bigwedge f(y, q) \neq *) \bigvee (f(x, q) \neq * \bigwedge f(y, q) = *) \bigvee ((f(x, q) \neq * \bigwedge f(y, q) \neq *) \to f(y, q) \geqslant f(x, q)))\}$, 用$yD_P^{\mathrm{LEDom}} x$表示.

定义 5 (广义扩展优势关系) $\mathrm{GEDom}(P) = \{(x, y) \in U \times U | \forall q \in P, (|B_P(x) \bigcap B_P(y)| / |C| \geqslant \lambda) \bigwedge ((f(x, q) = * \bigwedge f(y, q) = *) \bigvee (f(x, q) = * \bigwedge f(y, q) \neq *) \bigvee (f(x, q) \neq * \bigwedge f(y, q) = *) \bigvee ((f(x, q) \neq * \bigwedge f(y, q) \neq *) \to f(y, q) \geqslant f(x, q)))\}\}$, 用$yD_P^{\mathrm{GEDom}} x$表示. k度扩展特性优势关系[8]是定义 5 的特殊形式, 将缺失值细分为两种情况进行扩展.

定义 6 (相似优势关系) 即

$$\mathrm{SDR}^+(P) =$$
$$\{(x, y) \in U^2 : \forall q \in B_P, f(x, q) \geqslant f(y, q)\},$$
$$\mathrm{SDR}^-(P) =$$
$$\{(x, y) \in U^2 : \forall q \in B_P, f(x, q) \leqslant f(y, q)\},$$

用$yD_P^{\mathrm{SDR}+} x$和$yD_P^{\mathrm{SDR}-} x$表示. 限制相似关系[11]则是将限制关系与定义 6 结合的特殊形式.

2 置信优势关系的粗糙近似

2.1 置信优势关系

定义 7 (置信优势关系) $\mathrm{CDR}(P) = \{(x, y) \in U^2 | |B_P(x) \bigcap B_P(y)| / |B_P(x)| = 1) \bigwedge \forall_{q \in P} ((f(x, q) = * \bigwedge f(y, q) = *) \bigvee (f(x, q) = * \bigwedge f(y, q) \neq *) \bigvee (f(y, q) \succeq f(x, q)\}$, 用$yD_P^{\mathrm{CDR}} x$表示"$y$置信优势于$x$".

从置信优势关系的定义可以看出, 在不完备序信息系统下考虑y置信优势于x, 需满足: 1) y中包含的信息不少于x的信息; 2) y的非空属性均优于x.

定义 8 (序对称性) 任意两个对象x, y之间存在序关系, 若$y \succeq x$, 则一定存在$x \preceq y$, 满足这样的性质, 称为序对称性.

定义 9 (置信优势集、置信劣势集) 假定$\mathrm{DS} = (U, A, V, f)$是一个 IODS, $x \in U$, 则x的置信优势集和置信劣势集定义如下:

$$D_P^{\mathrm{CDR}+}(x) = \{y \in U | y D_P^{\mathrm{CDR}} x\},$$

$$D_P^{\mathrm{CDR}-}(x) = \{y \in U | x D_P^{\mathrm{CDR}} y\}.$$

性质 1

1) 置信优势关系满足自反性、传递性及序对称性.

2) $G = \{D_P^{\mathrm{CDR}+}(x) | x \in U\}$, $G = \{D_P^{\mathrm{CDR}-}(x) | x \in U\}$ 均是 U 的覆盖.

3) 如果 $x_j \in D_P^{\mathrm{CDR}+}(x_i)$, 则 $D_P^{\mathrm{CDR}+}(x_j) \subseteq D_P^{\mathrm{CDR}+}(x_i)$; 如果 $x_j \in D_P^{\mathrm{CDR}-}(x_i)$, 则 $D_P^{\mathrm{CDR}-}(x_j) \subseteq D_P^{\mathrm{CDR}-}(x_i)$.

2.2　基于置信优势关系的粗糙近似

假定 $\mathrm{DS} = (U, A, V, f)$ 是一个 IODS, 决策属性 D 将论域 U 划分为 n 个类 $\mathrm{Cl} = \{\mathrm{Cl}_t | t \in \{1, 2, \cdots, n\}\}$. 其中: $\mathrm{Cl}_n \succ \mathrm{Cl}_{n-1} \succ \cdots \succ \mathrm{Cl}_1$; t 类的上联合定义为 $\mathrm{Cl}_t^{\geqslant} = \bigcup\limits_{s \geqslant t} \mathrm{Cl}_s$, 下联合定义为 $\mathrm{Cl}_t^{\leqslant} = \bigcup\limits_{s \leqslant t} \mathrm{Cl}_s$. $x \in \mathrm{Cl}_t^{\geqslant}$ 表明 x 至少属于类 Cl_t, $x \in \mathrm{Cl}_t^{\leqslant}$ 表明 x 至多属于类 Cl_t. 基于置信优势关系的上下近似、边界域、近似精度及分类精度定义如下.

定义 10 (上下近似)　假定 $P \subseteq C$, $x \in U$, $\mathrm{Cl}_t^{\geqslant}$, $\mathrm{Cl}_t^{\leqslant} \subseteq U$, $t = 1, 2, \cdots, n$. $\mathrm{Cl}_t^{\geqslant}$ 和 $\mathrm{Cl}_t^{\leqslant}$ 的上下近似分别定义如下:

$$\underline{P}(\mathrm{Cl}_t^{\geqslant}) = \{x \in U | D_P^{\mathrm{CDR}+}(x) \subseteq \mathrm{Cl}_t^{\geqslant}\},$$

$$\overline{P}(\mathrm{Cl}_t^{\geqslant}) = \bigcup_{x \in \mathrm{Cl}_t^{\geqslant}} D_P^{\mathrm{CDR}+}(x) = \left\{x | D_P^{\mathrm{CDR}-}(x) \bigcap \mathrm{Cl}_t^{\geqslant} \neq \varnothing\right\};$$

$$\underline{P}(\mathrm{Cl}_t^{\leqslant}) = \{x \in U | D_P^{\mathrm{CDR}-}(x) \subseteq \mathrm{Cl}_t^{\leqslant}\},$$

$$\overline{P}(\mathrm{Cl}_t^{\leqslant}) = \bigcup_{x \in \mathrm{Cl}_t^{\leqslant}} D_P^{\mathrm{CDR}-}(x) = \left\{x | D_P^{\mathrm{CDR}+}(x) \bigcap \mathrm{Cl}_t^{\leqslant} \neq \varnothing\right\}.$$

定义 11 (边界域)　即

$$\mathrm{Bn}(\mathrm{Cl}_t^{\geqslant}) = \overline{P}(\mathrm{Cl}_t^{\geqslant}) - \underline{P}(\mathrm{Cl}_t^{\geqslant}),$$

$$\mathrm{Bn}(\mathrm{Cl}_t^{\leqslant}) = \overline{P}(\mathrm{Cl}_t^{\leqslant}) - \underline{P}(\mathrm{Cl}_t^{\leqslant}).$$

定义 12 (近似精度)　$\mathrm{Cl}_t^{\geqslant}$ 和 $\mathrm{Cl}_t^{\leqslant}$ 的粗糙近似的近似精度分别定义如下:

$$\alpha(\mathrm{Cl}_t^{\geqslant}) = |\underline{P}(\mathrm{Cl}_t^{\geqslant})| / |\overline{P}(\mathrm{Cl}_t^{\geqslant})|,$$

$$\alpha(\mathrm{Cl}_t^{\leqslant}) = |\underline{P}(\mathrm{Cl}_t^{\leqslant})| / |\overline{P}(\mathrm{Cl}_t^{\leqslant})|.$$

定义 13 (分类精度)　定义如下:

$$\gamma(\mathrm{Cl}^{\geqslant}) = \left|U - \bigcup_{t=1}^{n} \mathrm{Bn}(\mathrm{Cl}_t^{\geqslant})\right| / |U|,$$

$$\gamma(\mathrm{Cl}^{\leqslant}) = \left|U - \bigcup_{t=1}^{n} \mathrm{Bn}(\mathrm{Cl}_t^{\leqslant})\right| / |U|.$$

3　几种拓展优势关系粗糙近似的对比

3.1　几种拓展优势关系的对比

在序信息系统下, 序关系特性应具有自反性、传递性和序对称性. 几种拓展优势关系满足自反性、传递性和序对称性的情况如表 1 所示.

表 1　几种关系的性质对比

拓展关系	自反性	传递性	序对称性
D_P^{EDom}	√	√	√
D_P^{LEDom}	√	×	√
D_P^{GEDom}	×	×	√
D_P^{SDR}	√	√	×
D_P^{CDR}	√	√	√

下面通过举例来说明几种拓展关系对于序信息系统下的语义矛盾和序关系特性的不满足. 假设有 5 个对象 $x_1 = [1, *, 2, *]$, $x_2 = [2, *, *, 1]$, $x_3 = [2, *, 1, *]$, $x_4 = [2, *, 2, 1]$, $x_5 = [*, 2, *, 2]$.

1) 从定义中可得到 $x_1 D_P^{\mathrm{EDom}} x_5$ 且 $x_5 D_P^{\mathrm{EDom}} x_1$, 即 $x_5 \succeq x_1$ 且 $x_5 \preceq x_1$, 从序的特性上不难得到 $x_5 = x_1$, 但从数据对象上很难得到这个结论.

2) $x_2 D_P^{\mathrm{LEDom}} x_1$, $x_3 D_P^{\mathrm{LEDom}} x_2$, 但 $x_3 D_P^{\mathrm{LEDom}} x_1$ 并不成立, 因此 D_P^{LEDom} 不满足传递性. 若 $x_3 \succeq x_2$ 且 $x_2 \succeq x_1$, 在序关系中可以得到 $x_3 \succeq x_1$, 而在有限扩展优势关系中无法得到 $x_3 \succeq x_1$ 的结论.

3) 尽管 D_P^{LEDom} 是 D_P^{GEDom} 的特殊形式, 即 D_P^{GEDom} 的阈值 λ 取零的情况, 但 D_P^{GEDom} 却有可能无法满足自反性. 若设 $\lambda > 0.5$, 在序关系中存在的 $x_1 \succeq x_1$, 在广义扩展优势关系中却不一定满足, 即 $x_1 D_P^{\mathrm{GEDom}} x_1$ 不成立.

4) 从相似优势关系定义可得到 $x_4 D_P^{\mathrm{SDR}+} x_1$, 但 $x_1 D_P^{\mathrm{SDR}-} x_4$ 不满足, 这在序信息系统下呈现语义矛盾, 即满足 $x_4 \succeq x_1$, 却不能得到 $x_1 \preceq x_4$.

置信优势关系满足自反性、传递性和序对称性 3 个序关系特性. “y 置信优势于 x”, 即 $y D_P^{\mathrm{CDR}} x$, 在 x 的非空属性上, y 均优于 x, 而且 y 的信息量不少于 x, 这个定义满足了序的特性, 符合现实的优势关系语义, 避免了前面几种拓展关系产生的语义矛盾.

3.2　几种基于拓展优势关系的粗糙近似的对比

这里, 用 “Δ” 来表示 “$\leqslant$” 或 “$\geqslant$”, 用 $\underline{P}(\mathrm{Cl}_t^{\Delta})^{\mathrm{EDom}}$、$\underline{P}(\mathrm{Cl}_t^{\Delta})^{\mathrm{GEDom}}$、$\underline{P}(\mathrm{Cl}_t^{\Delta})^{\mathrm{LEDom}}$、$\underline{P}(\mathrm{Cl}_t^{\Delta})^{\mathrm{CDR}}$ 分别表示 EDom、GEDom、LEDom 及 CDR 关系下的关于 Cl_t^{Δ} 的下近似. 相类似的标记也应用于上近似、近似精度及分类精度的表示.

定理 1

1) $\underline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{EDom}} \subseteq \underline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{GEDom}} \subseteq \underline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{CDR}}$,

2) $\overline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{CDR}} \subseteq \overline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{LEDom}} \subseteq \overline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{EDom}}$.

证明　1) 因为 $\forall x, y \in U$, 有 $y D_P^{\mathrm{CDR}} x \Rightarrow$

$yD_P^{\mathrm{GEDom}}x \Rightarrow yD_P^{\mathrm{EDom}}x$, 反之不成立, 所以可得到 $D_P^{\mathrm{CDR}+}(x) \subseteq D_P^{\mathrm{GEDom}+}(x) \subseteq D_P^{\mathrm{EDom}+}(x)$. 由下近似定义可知, $\forall x \in \underline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{CDR}} \Rightarrow x \in \underline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{GEDom}} \Rightarrow x \in \underline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{EDom}}$, 反之不成立, 所以 $\underline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{EDom}} \subseteq \underline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{GEDom}} \subseteq \underline{P}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{CDR}}$ 成立.

2) 同理可证. □

定理 2

1) $\underline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{EDom}} \subseteq \underline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{LEDom}} \subseteq \underline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{CDR}}$,

2) $\overline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{CDR}} \subseteq \overline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{GEDom}} \subseteq \overline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{EDom}}$.

定理 2 的证明方法与定理 1 的相同.

推论 1

1) $\mathrm{Bn}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{CDR}} \subseteq \mathrm{Bn}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{LEDom}} \subseteq \mathrm{Bn}(\mathrm{Cl}_t^{\geqslant})^{\mathrm{EDom}}$,

2) $\mathrm{Bn}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{CDR}} \subseteq \mathrm{Bn}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{LEDom}} \subseteq \mathrm{Bn}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{EDom}}$.

证 明 文献[6]已经证得, $\underline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{LEDom}} \subseteq \underline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{GEDom}}$, $\overline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{GEDom}} \subseteq \overline{P}(\mathrm{Cl}_t^{\leqslant})^{\mathrm{LEDom}}$, 结合定理 1 和定理 2 即可直接得到. □

3.3 几种拓展优势关系的近似分类性能对比

定理 3

1) $\alpha(\mathrm{Cl}_t^{\geqslant})^{\mathrm{CDR}} \geqslant \alpha(\mathrm{Cl}_t^{\geqslant})^{\mathrm{LEDom}} \geqslant \alpha(\mathrm{Cl}_t^{\geqslant})^{\mathrm{EDom}}$,

2) $\alpha(\mathrm{Cl}_t^{\leqslant})^{\mathrm{CDR}} \geqslant \alpha(\mathrm{Cl}_t^{\leqslant})^{\mathrm{LEDom}} \geqslant \alpha(\mathrm{Cl}_t^{\leqslant})^{\mathrm{EDom}}$.

定理 3 的证明可通过定理 1 和定理 2 直接得到.

定理 3 说明, 置信优势关系在粗糙近似精度上优于扩展优势关系、有限扩展优势关系和广义扩展优势关系下的粗糙近似精度.

定理 4

1) $\gamma(\mathrm{Cl}_t^{\geqslant})^{\mathrm{CDR}} \geqslant \gamma(\mathrm{Cl}_t^{\geqslant})^{\mathrm{LEDom}} \geqslant \gamma(\mathrm{Cl}_t^{\geqslant})^{\mathrm{EDom}}$,

2) $\gamma(\mathrm{Cl}_t^{\leqslant})^{\mathrm{CDR}} \geqslant \gamma(\mathrm{Cl}_t^{\leqslant})^{\mathrm{LEDom}} \geqslant \gamma(\mathrm{Cl}_t^{\leqslant})^{\mathrm{EDom}}$.

定理 4 的证明可由推论 1 直接得到.

定理 4 说明, 置信优势关系在分类精度上优于扩展优势关系、有限扩展优势关系和广义扩展优势关系下的分类近似精度.

4 实例分析

这里采用文献[7]的实例, 表 2 为某校教学管理信息系统的一个信息表, 显然是一个 IODS. 其中, 课程分别由 C_1、C_2 及 C_3 表示, 表 2 中 1、2、3 分别代表优、中、差, 属性 d 表示对学生的总评.

表 2 某校教学总评表

学生 ID	C_1	C_2	C_3	d
1	2	1	1	1
2	3	2	*	3
3	2	*	*	1
4	*	2	2	3
5	*	3	1	1
6	3	2	1	3

从置信优势关系及其粗糙近似的定义容易得到:

置信优势集

$D_P^{\mathrm{CDR}+}(1) = \{1,6\}$, $D_P^{\mathrm{CDR}+}(2) = \{2,6\}$,

$D_P^{\mathrm{CDR}+}(3) = \{1,2,3,6\}$, $D_P^{\mathrm{CDR}+}(4) = \{4\}$,

$D_P^{\mathrm{CDR}+}(5) = \{5\}$, $D_P^{\mathrm{CDR}+}(6) = \{6\}$;

置信劣势集

$D_P^{\mathrm{CDR}-}(1) = \{1,3\}$, $D_P^{\mathrm{CDR}-}(2) = \{2,3\}$,

$D_P^{\mathrm{CDR}-}(3) = \{3\}$, $D_P^{\mathrm{CDR}-}(4) = \{4\}$,

$D_P^{\mathrm{CDR}-}(5) = \{5\}$, $D_P^{\mathrm{CDR}-}(6) = \{1,2,3,6\}$;

上联合

$\mathrm{Cl}_3^{\geqslant} = \{2,4,6\}$, $\mathrm{Cl}_1^{\geqslant} = \{1,2,3,4,5,6\}$;

下联合

$\mathrm{Cl}_1^{\leqslant} = \{1,3,5\}$, $\mathrm{Cl}_3^{\leqslant} = \{1,2,3,4,5,6\}$.

因 $\underline{P}(\mathrm{Cl}_1^{\geqslant})^{\mathrm{CDR}} = \overline{P}(\mathrm{Cl}_1^{\geqslant})^{\mathrm{CDR}} = U$, $\underline{P}(\mathrm{Cl}_3^{\leqslant})^{\mathrm{CDR}} = \overline{P}(\mathrm{Cl}_3^{\leqslant})^{\mathrm{CDR}} = U$, 故下面主要考察 $\mathrm{Cl}_3^{\geqslant}$ 和 $\mathrm{Cl}_1^{\leqslant}$ 的上下近似、边界域及近似精度:

$\underline{P}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{CDR}} = \{2,4,6\}$, $\overline{P}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{CDR}} = \{2,4,6\}$,

$\mathrm{Bn}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{CDR}} = \varnothing$;

$\alpha(\mathrm{Cl}_3^{\geqslant})^{\mathrm{CDR}} = 1$, $\gamma(\mathrm{Cl}^{\geqslant})^{\mathrm{CDR}} = 1$;

$\underline{P}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{CDR}} = \{1,3,5\}$, $\overline{P}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{CDR}} = \{1,3,5\}$,

$\mathrm{Bn}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{CDR}} = \varnothing$;

$\alpha(\mathrm{Cl}_1^{\leqslant})^{\mathrm{CDR}} = 1$, $\gamma(\mathrm{Cl}^{\leqslant})^{\mathrm{CDR}} = 1$.

同样可以得到扩展优势关系、有限扩展优势关系和广义扩展优势关系的上下近似集、边界域、近似精度及分类精度. 特别地, 广义扩展优势关系的阈值 λ 取值若大于 1/3, 则学生 ID 为 3 的对象将无法满足自反性, 该对象的广义优势关系类为空集. 因此, 该处阈值设置为 1/3. 则有

$\underline{P}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{EDom}} = \{1\}$, $\overline{P}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{EDom}} = \{1,2,3,4,5,6\}$,

$\mathrm{Bn}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{EDom}} = \{2,3,4,5,6\}$, $\alpha(\mathrm{Cl}_1^{\leqslant})^{\mathrm{EDom}} = 0.17$;

$\underline{P}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{EDom}} = \varnothing$, $\overline{P}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{EDom}} = \{2,3,4,5,6\}$,

$\mathrm{Bn}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{EDom}} = \{2,3,4,5,6\}$, $\alpha(\mathrm{Cl}_3^{\geqslant})^{\mathrm{EDom}} = 0$;

$\underline{P}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{LEDom}} = \{1,3\}$, $\overline{P}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{LEDom}} = \{1,2,3,5,6\}$,

$\mathrm{Bn}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{LEDom}} = \{2,5,6\}$, $\alpha(\mathrm{Cl}_1^{\leqslant})^{\mathrm{LEDom}} = 0.4$;

$\underline{P}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{LEDom}} = \{4\}$, $\overline{P}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{LEDom}} = \{2,4,5,6\}$,

$\mathrm{Bn}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{LEDom}} = \{2,5,6\}$, $\alpha(\mathrm{Cl}_3^{\geqslant})^{\mathrm{LEDom}} = 0.25$;

$\underline{P}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{GEDom}} = \{1,3\}$, $\overline{P}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{GEDom}} = \{1,3,5,6\}$,

$\mathrm{Bn}(\mathrm{Cl}_1^{\leqslant})^{\mathrm{GEDom}} = \{5,6\}$, $\alpha(\mathrm{Cl}_1^{\leqslant})^{\mathrm{GEDom}} = 0.5$;

$\underline{P}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{GEDom}}=\{2,4\}$, $\overline{P}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{GEDom}}=\{2,4,5,6\}$,

$\mathrm{Bn}(\mathrm{Cl}_3^{\geqslant})^{\mathrm{GEDom}}=\{5,6\}$, $\alpha(\mathrm{Cl}_3^{\geqslant})^{\mathrm{GEDom}}=0.5$.

通过实例分析, 显然满足上节所讨论的定理1～定理3及推论1, 置信优势关系粗糙集的近似精度更高.

为了验证置信优势关系粗糙集能够获得更高的分类精度, 下面选用UCI中具有缺失值的Breast Cancer Wisconsin (Original)、Hepatitis、horse-colic、heart-disease-switzerland数据集以及本文实例进行仿真实验, 实验结果如表3所示.

表 3　分类精度对比表

数据集	γEDom	γLEDom	γCDR
Breast Cancer Wisconsin (Original)	0.955 7	0.955 7	**0.977 1**
Hepatitis	0.741 9	0.741 9	**0.819 4**
horse-colic	0.646 7	0.646 7	**0.700 3**
heart-disease-switzerland	0.357 7	0.357 7	**0.463 4**
本文实例	0.166 7	0.500 0	**1.000 0**

实例分析和仿真实验表明, 置信优势关系粗糙近似与已有方法相比, 可得到更大的正域空间和更小的边界域, 同时具有更高的近似精度和分类精度.

5　结　论

本文提出了新的拓展优势关系和置信优势关系, 并探讨了基于置信优势关系的粗糙近似方法. 与已有的方法相比, 本文所提出的置信优势关系更能满足序信息系统优势关系的特性, 具有自反性、传递性和序对称性, 避免了在序关系下的语义矛盾. 另外, 与现有方法比较, 本文方法具有更大的下近似空间和最小的边界区域, 因此, 置信优势关系的粗糙近似方法具有更高的近似精度和分类精度.

参考文献(References)

[1] Pawlak Z, Skowron A. Rudiments of rough sets[J]. Information Sciences, 2007, 177(1): 3-27.

[2] Greco S, Matarazzo B, Slowinski R. Rough approximation of a preference relation by dominance relations[J]. European J of Operational Research, 1999, 117(1): 63-83.

[3] Greco S, Matarazzo B, Slowinski R. Rough sets theory for multicriteria decision analysis[J]. European J of Operational Research, 2001, 129(1): 1-47.

[4] 何亚群, 胡寿松. 不完全信息的多属性粗糙决策分析方法[J]. 系统工程学报, 2004, 19(2): 117-120.
(He Y Q, Hu S S. Rough analysis method of multi-attribute decision making with incomplete information[J]. J of Systems Engineering, 2004, 19(2): 117-120.)

[5] Shao M W, Zhang W X. Dominance relation and rules in an incomplete ordered information system[J]. Int J of Intelligent Systems, 2005, 20(1): 13-27.

[6] 胡明礼, 刘思峰. 基于有限扩展优势关系的粗糙决策分析方法[J]. 系统工程, 2006, 24(4): 106-110.
(Hu M L, Liu S F. An analysis method of rough multi-attribute decision making based on limited extended dominance relation[J]. Systems Engineering, 2006, 24(4): 106-110.)

[7] 胡明礼, 刘思峰. 基于广义扩展优势关系的粗糙决策分析方法[J]. 控制与决策, 2007, 22(12): 1347-1351.
(Hu M L, Liu S F. Rough analysis method of multi-attribute decision making based on generalized extended dominance relation[J]. Control and Decision, 2007, 22(12): 1347-1351.)

[8] Chen H M, Li T R, Ruan D. Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining[J]. Knowledge-based Systems, 2012, 31: 140-161.

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

[10] 骆公志, 杨晓江, 刘思峰. 基于限制优势关系的粗糙决策分析模型[J]. 中国管理科学, 2009, 17(5): 127-132.
(Luo G Z, Yang X J, Liu S F. Rough analysis model of multi-attribute decision making based on limited dominance relation[J]. Chinese J of Management Science, 2009, 17(5): 127-132.)

[11] 骆公志, 杨晓江. 基于限制相似优势关系的粗糙决策分析模型[J]. 系统工程理论与实践, 2009, 29(9): 134-140.
(Luo G Z, Yang X J. Rough analysis model of multi-attribute decision making based on limited similarity dominance relation[J]. Systems Engineering-Theory and Practice, 2009, 29(9): 134-140.)

[12] Yang X B, Dou H L. Valued dominance-based rough set approach to incomplete information system[J]. Trans on Computer Science, 2011, LNCS 6750: 92-107.

(责任编辑: 李君玲)