不完备决策系统中发现偏好概率规则的粗糙集方法

胡明礼, 刘思峰

(南京航空航天大学 经济与管理学院, 江苏 南京 210016)

摘 要: 为了从有偏好信息但信息不完全的多属性决策系统中获取概率决策规则, 提出一种新的不完全信息的多属性粗糙决策分析方法。首先, 提出扩展优势关系下相容度的概念; 其次, 基于相容度给出知识的粗糙近似, 并证明了粗糙近似的基本性质; 再次, 给出粗糙近似的分类质量与 β- 约简的概念, 并从不完全信息的偏好决策系统中导出概率决策规则; 最后, 通过一个实例说明新方法的可行性和有效性。

关键词: 不完全信息; 多属性决策; 粗糙集; 扩展优势关系; 相容度; 概率决策规则

中图分类号: C934 **文献标识码**: A **文章编号**: 1001- 7348(2007) 10- 0114- 05

0 前 言

多属性决策问题大量存在于社会、经济、工程和管理等各个领域中, 由于环境的复杂性和不确定性以及决策者偏好的影响, 信息不完全或不确定以及含有偏好信息是实际决策系统的的显著特征。在决策过程中, 人们所面临的数据往往具有不确定性、不完整性和定性的特点, 从这类数据中发现知识, 将不确定性、定性知识进行整合和推理是一个难题[1]。

由波兰科学家 Pawlak 于 1982 年提出的粗糙集理论是一种不确定数据分析与推理的有效工具。经过 20 余年的发展, 该理论已成功应用于人工智能、数据挖掘、机器学习、决策分析、模式识别和决策支持系统等众多领域[2]。粗糙集理论的显著特点是它不需要提问题所需处理数据集合之外的任何先验信息, 所以对问题的不确定性描述或处理比较客观。经典的粗糙集方法同时也存在一局限, 例如经典粗糙集不能直接处理不完备的决策表, 不能从包含噪声数据的决策表中导出概率决策规则以及不能处理含有偏好信息的决策表。

在应用中, 不少的研究是对经典粗糙集进行拓展, 以处理实际决策系统中的不完备性、不相容性或偏好信息。Ziarko 提出的变精度粗糙集模型[3]可以从含噪声的决策表中导出概率规则。该模型不仅继承了经典粗糙集模型的数学特性, 而且具有更强的泛化能力和抗噪声能力, 然而 Ziarko 模型不能处理多属性决策表中偏好信息引起的不相容性。Greco 提出优势关系粗糙集理论[4], 用优势关系代替不可分辨关系, 可以处理有偏好信息的多属性决策问题, 但该方法不能导出偏好概率决策规则。潘郁、菅利荣等提出的偏好多属性决策表中发现概念规则的变精度粗糙集方法[5,6,7], 可以处理决策表中偏好信息引起的不相容性, 能够从包含噪声数据的事例决策表中导出偏好模型。以上的研究都是基于信息完全的情形, 即决策表中不包含空值。针对信息不完全情形下的偏好多属性决策问题, 何亚群等提出的基于扩展优势关系的粗糙决策分析方法 [8], 用扩展优势关系代替优势关系, 从不完备决策信息系统中获取偏好决策规则, 有可能漏掉了某些很有价值的较强规则。如何处理不完全信息的偏好多属性决策系统中的不相容性, 导出偏好概率决策规则, 尚未见到有关的研究文献。本文针对这一问题, 提出了基于扩展优势关系的相容度的概念, 并基于相容度和相容阈值 β 确定粗糙集的上近似、下近似和边界域, 从而获取偏好概率决策规则, 发现潜在的有价值决策知识。

1 基于扩展优势关系的多属性粗糙决策分析方法

1.1 不完全信息的偏好多属性决策系统

首先给出有偏好信息、但信息不完全的多属性决策系统的概念, 设有一个决策系统 $S=<U,A,V,f>$, 其中: U 是非空的对象全集; A 是非空的属性集合, $A=C\cup D$, C, D 分别是条件属性集和决策属性集; V 是属性值, $V_C=\{V_q|q\in C\}$, $V_D=\{V_d|d\in D\}$ 分别是条件属性值集和决策属性值集, 条件

收稿日期: 2006- 09- 26

基金项目: 国家自然科学基金项目(70473037) , 江苏省自然科学基金重点项目(BK2003211)

作者简介: 胡明礼(1979-) , 男, 汉族, 山东济宁人, 南京航空航天大学博士研究生, 研究方向为系统工程、智能决策; 刘思峰(1955-) , 男, 河南平舆人, 南京航空航天大学经济与管理学院院长、教授、博士生导师, 研究方向为系统工程、灰色系统理论。

属性值 V_q 和决策属性值 V_d 具有偏好次序：$f: U\times A\to V$ 是一个信息函数，表示对每一个 $q\in A, x\in U, f(x,q)\in V_q$，如果某些属性值 $V_q\in V_C=*$，* 表示空值，称 S 是一个不完全信息的偏好多属性决策系统。基于扩展优势关系的粗糙决策分析方法的研究对象就是这类决策系统。

1.2　扩展优势关系

定义 1　给定论域 U，集合 $Cl_r, Cl_s\subseteq U, r,s\in\{1,2,\cdots,n\}$，$\forall x,y\in U$，xSy 表示 x 不劣于 y，ySx 表示 x 不优于 y，有 $\{x\in Cl_r, y\in Cl_s, r>s\}\Rightarrow\{xSy\wedge\neg ySx\}$，那么，向上累积集 $Cl_t^{\geq}$ 定义为：$Cl_t^{\geq}=\bigcup_{s\geq t} Cl_s$；向下累积集 $Cl_t^{\leq}$ 定义为 $Cl_t^{\leq}=\bigcup_{s\leq t} Cl_s$，其中：$t=1,2,\cdots,n$。

由定义 1，显然可以得到 $Cl_t^{\geq}$ 和 $Cl_t^{\leq}$ 的两条性质：

（Ⅰ）$Cl_1^{\geq}=Cl_n^{\leq}=U, Cl_n^{\geq}=Cl_n, Cl_1^{\leq}=Cl_1$

（Ⅱ）$Cl_{t-1}^{\leq}=U-Cl_t^{\geq}, t=2,3,\cdots,n$

定义 2 [8]　给定决策系统 $S=\langle U,A,V,f\rangle$ 集合 $P\subseteq A$，假设 $x,y\in U$，P 上的扩展优势关系 EDOM(P) 定义为：

$EDOM(P)=\{(x,y)\in U\times U|\forall q\in P, f(y,q)\geq f(y,q)$ 或 $f(y,q)=*$ 或 $f(x,q)=*\}$。这时称“y 扩展优势于 x”，简写为 $yD_P^x x$。显然，扩展优势关系 D_P^x 满足自反性和传递性。　(1)

定义 3　对于 $P\subseteq A, x\in U, D_P^{+*}(x)=\{y\in U: yP_P^{D*}x\}$ 称为 P 扩展优势集，$D_P^{-*}(x)=\{y\in U: xP_P^{D*}y\}$ 称为 P 扩展被优势集。

1.3　基于扩展优势关系的粗糙近似

定义 4　对于 $P\subseteq C, x\in U, Cl_t^{\geq}, Cl_t^{\leq}\subseteq U, t=1,2,\cdots,n$，$U_P^*=\{x\in U: \exists q\in P, f(x,q)\neq *\}$，即 U_P^* 要求每个对象至少有一个非空属性，在扩展优势关系下，向上累积集 $Cl_t^{\geq}$ 和向下累积集 $Cl_t^{\leq}$ 的下近似、上近似和边界域分别定义为：

$$\underline{P}(Cl_t^{\geq})^x=\{x\in U_P^*: D_P^{+*}(x)\subseteq Cl_t^{\geq}\}, \bar{P}(Cl_t^{\geq})^*=\bigcup_{x\in Cl_t^{\geq}} D_P^{+*}(x),$$

$$Bn_P(Cl_t^{\geq})^*=\bar{P}(Cl_t^{\geq})-\underline{P}(Cl_t^{\geq})^*;$$

$$\underline{P}(Cl_t^{\leq})=\{x\in U_P^*: D_P^{-*}(x)\subseteq Cl_t^{\leq}\}, \bar{P}(Cl_t^{\leq})^*=\bigcup_{x\in Cl_t^{\leq}} D_P^{-*}(x),$$

$$Bn_P(Cl_t^{\leq})^*=\bar{P}(Cl_t^{\leq})^*-\underline{P}(Cl_t^{\leq})^*.$$

1.4　分类质量与约简

定义 5　对于 $P\subseteq A, x\in U, Cl_t^{\geq}, Cl_t^{\leq}\subseteq U, t=1,2,\cdots,n$，扩展优势关系下粗糙近似的分类质量定义为：

$$\gamma_P(Cl)^*=\frac{\left|U-\bigcup_{t=1}^{n}(Bnp(Cl_t^{\leq})^*)\right|}{|U|}=\frac{\left|U-\bigcup_{t=1}^{n}(Bnp(Cl_t^{\geq})^*)\right|}{|U|}\qquad(2)$$

Cl 的分类质量 $\gamma_P(Cl)^*$ 表示偏好多属性决策表中正确分类的对象与总对数的比率。

定义 6　满足 $\gamma_P(Cl)^*=\gamma_C(Cl)^*$ 的最小子集 $P\subseteq C$ 称为 C 关于 Cl 的一个约简，记作：$red_P(C,Cl)$。

一个决策表中可能有不止一个约简，所有约简的交集称为核，核是偏好多属性决策表中最重要的属性集，它也可能是空集。

1.5　偏好决策规则

1）根据 $Cl_t^{\geq}$ 的粗糙下近似得到确定性的偏好决策规划：

If $f(x,q_1)\geq r_{q1}, f(x,q_2)\geq r_{q2},\cdots, f(x,q_p)\geq r_{qp}$, then $x\in Cl_t^{\geq}$，其中：$P=\{q_1,q_2,\cdots,q_p\}\subseteq C, (r_{q1},r_{q2},\cdots,r_{qp})\in V_{q1}\times V_{q2}\times\cdots, V_{qp}, t=1,2,\cdots,n$

2）根据 $Cl_t^{\leq}$ 的粗糙下近似得到确定性的偏好决策规则：

If $f(x,q_1)\leq r_{q1}, f(x,q_2)\leq r_{q2},\cdots, f(x,q_p)\leq r_{qp}$, then $x\in Cl_t^{\leq}$，其中：$P=\{q_1,q_2,\cdots,q_p\}\subseteq C, (r_{q1},r_{q2},\cdots,r_{qp})\in V_{q1}\times V_{q2}\times\cdots, V_{qp}, t=1,2,\cdots,n$

3）根据 $Cl_t^{\geq}$ 和 $Cl_t^{\leq}$ 的边界域得到可能性偏好决策规则：

If $f(x,q_1)\geq r_{q1}, f(x,q_2)\geq r_{q2},\cdots, f(x,q_k)\geq r_{qk}, f(x,q_{k+1})\leq r_{qk+1}, f(x,q_{k+2})\leq r_{qk+2},\cdots, f(x,q_p)\leq r_{qp}$, then $x\in Cl_t\cup Cl_{t+1}\cup\cdots\cup Cl_s$，其中：$Q'=\{q_1,q_2,\cdots,q_k\}\subseteq C, Q''=\{q_{k+1},q_{k+2},\cdots,q_p\}\subseteq C$，$P=Q'\cup Q''$，$Q'$ 和 Q'' 可以相交，$(r_{q1},r_{q2},\cdots,r_{qp})\in V_{q1}\times V_{q2}\times\cdots, V_{qp}, s,t=\{1,2,\cdots,n\}, t<s$。

2　不完全信息的偏好多属性决策系统中发现概率规则的扩展粗糙集方法

为了处理由偏好信息引起的不相容性并导出用于概率估计的决策规则，本文借鉴 Ziarko 的变精度粗糙集模型的思想，通过设计一个相容阈值 β，将基于扩展优势关系的粗糙集方法扩展为可以从不完全信息的偏好多属性决策系统中发现概率规则的扩展粗糙集方法。

2.1　不完全信息下相容度的概念

假定多属性决策系统 $S=\langle U,A,V,f\rangle$ 中含有偏好信息和空值，给定一个偏好信息的集合 $P\subseteq C, x\in U, y\in U$，对 $Cl_t^{\geq}\subseteq U, t=2,\cdots,n$，若下列条件成立，则称为基于扩展优势关系的不相容性。

1）$x\in Cl_t^{\geq}$，但 $D_P^{+*}(x)\cap Cl_{t-1}^{\leq}\neq\phi$。即 x 属于类别 Cl_t 或比 Cl_t 好的类别，在所有考虑的偏好属性中，y 至少和 x 一样好，但在扩展优势关系下 y 却属于一个比 x 差的类别中。

2）$x\in Cl_{t-1}^{\leq}$，但 $D_P^{-*}(x)\cap Cl_{t-1}^{\geq}\neq\phi$。即 x 属于类别 Cl_t 差的类别，在所有考虑的偏好属性中，y 至少和 x 一样好，但在扩展优势关系下 y 却属于一个比 x 好的类别中。

为了刻画不完全信息下由偏好信息引起的不相容性的影响，我们给出扩展优势关系下的相容度的概念。

定义 6　假设决策系统 $S=\langle U,A,V,f\rangle, Cl_t^{\geq}, Cl_t^{\leq}\subseteq U$，

对于属性集 $P\subseteq C$，如果 $x\in Cl_t^{\geq}$，扩展优势集下的相容度 $c(x,Cl_t^{\geq})$ 定义为：

$$c(x,Cl_t^{\geq})=\frac{|D_P^{+*}(x)\cap Cl_t^{\geq}|}{|D_P^{+*}(x)|},\ (t=2,3,\cdots,n) \tag{3}$$

如果 $x\in Cl_t^{\leq}$，扩展被优势集下的相容度 $c(x,Cl_t^{\leq})$ 定义为：

$$c(x,Cl_t^{\leq})=\frac{|D_P^{-*}(x)\cap Cl_t^{\leq}|}{|D_P^{-*}(x)|},\ (t=1,2,\cdots,n-1) \tag{4}$$

2.2 基于相容度的粗糙近似

定义 7 假设决策系统 $S=<U,A,V,f>$，$x\in U$，$Cl_t^{\geq},Cl_t^{\leq}\subseteq U$，给定属性集 $P\subseteq C$ 和一个相容阈值 $\beta\in(0.5,1]$，向上累积集 $Cl_t^{\geq}$（$t=2,3,\cdots,n$）和向下累积集 $Cl_t^{\leq}$（$t=1,2,\cdots,n-1$）的 β-下近似分别定义为：

$$\underline{P}^{\beta}(Cl_t^{\geq})^*=\{x\in Cl_t^{\geq}\mid c(x,Cl_t^{\geq})\geq\beta\} \tag{5}$$

$$\underline{P}^{\beta}(Cl_t^{\leq})^*=\{x\in Cl_t^{\leq}\mid c(x,Cl_t^{\leq})\geq\beta\} \tag{6}$$

由定义 7 可知，$Cl_t^{\geq}$ 的 β-下近似表示 U 中肯定属于 $Cl_t^{\geq}$ 的相容度不小于 β 的对象构成的集合；$Cl_t^{\leq}$ 的 β-下近似表示 U 中肯定属于 $Cl_t^{\leq}$ 的相容度不小于 β 的对象构成的集合。

定义 8 假设决策系统 $S=<U,A,V,f>$，$x\in U$，$Cl_t^{\geq},Cl_t^{\leq}\subseteq U$，给定属性集 $P\subseteq C$ 和一个相容阈值 $\beta\in(0.5,1]$，向上累积集 $Cl_t^{\geq}$（$t=2,3,\cdots,n$）和向下累积集 $Cl_t^{\leq}$（$t=1,2,\cdots,n-1$）的 β-下近似分别定义为：

$$\bar{P}^{\beta}(Cl_t^{\geq})^*=U-\bar{P}^{\beta}(Cl_{t-1}^{\leq})^* \tag{7}$$

$$Bn_P^{\beta}(Cl_t^{\geq})^*=\bar{P}^{\beta}(Cl_t^{\geq})^*-\underline{P}^{\beta}(Cl_t^{\geq})^* \tag{8}$$

$$\bar{P}^{\beta}(Cl_t^{\leq})^*=U-\bar{P}^{\beta}(Cl_{t-1}^{\geq})^* \tag{9}$$

$$Bn_P^{\beta}(Cl_t^{\leq})^*=\bar{P}^{\beta}(Cl_t^{\leq})^*-\underline{P}^{\beta}(Cl_t^{\leq})^* \tag{10}$$

定理 1 假设决策系统 $S=<U,A,V,f>$，给定相容阈值 $\beta\in(0.5,1]$ 和属性 $P\subseteq C$，以及 $Cl_t^{\geq},Cl_t^{\leq}\subseteq U$，则有

$$\bar{P}^{\beta}(Cl_t^{\geq})^*=Cl_t^{\geq}\cup\{x\in Cl_{t-1}^{\leq}\mid c(x,Cl_{t-1}^{\leq})<\beta\},\ (t=2,3,\cdots,n) \tag{11}$$

$$\bar{P}^{\beta}(Cl_t^{\leq})^*=Cl_t^{\leq}\cup\{x\in Cl_{t+1}^{\geq}\mid c(x,Cl_{t+1}^{\geq})<\beta\},\ (t=1,2,\cdots,n-1) \tag{12}$$

证明：先证明式（11），

$$\begin{aligned}\bar{P}^{\beta}(Cl_t^{\geq})^* &=U-\underline{P}^{\beta}(Cl_{t-1}^{\leq})^*\\ &=Cl_t^{\geq}\cup Cl_{t-1}^{\leq}-\{x\in Cl_{t-1}^{\leq}\mid \frac{|D_P^{-*}(x)\cap Cl_{t-1}^{\leq}|}{|D_P^{-*}(x)|}\geq\beta\}\\ &=Cl_t^{\geq}\cup\{x\in Cl_{t-1}^{\leq}\mid \frac{|D_P^{-*}(x)\cap Cl_{t-1}^{\leq}|}{|D_P^{-*}(x)|}<\beta\}\\ &=Cl_t^{\geq}\cup\{x\in Cl_{t-1}^{\leq}\mid c(x,Cl_{t-1}^{\leq})<\beta\}\end{aligned}$$

再证明式（12）。

$$\begin{aligned}\bar{P}^{\beta}(Cl_t^{\leq})^* &=U-\underline{P}^{\beta}(Cl_{t-1}^{\geq})^*\\ &=Cl_t^{\leq}\cup Cl_{t-1}^{\geq}-\{x\in Cl_{t-1}^{\geq}\mid \frac{|D_P^{+*}(x)\cap Cl_{t+1}^{\geq}|}{|D_P^{+*}(x)|}\geq\beta\}\\ &=Cl_t^{\leq}\cup\{x\in Cl_{t+1}^{\geq}\mid \frac{|D_P^{+*}(x)\cap Cl_{t-1}^{\geq}|}{|D_P^{+*}(x)|}<\beta\}\\ &=Cl_t^{\leq}\cup\{x\in Cl_{t+1}^{\geq}\mid c(x,Cl_{t+1}^{\geq})<\beta\}\end{aligned}$$

证毕。

由定理 1 可知，$Cl_t^{\geq}$ 的 β-上近似表示 U 中肯定属于 $Cl_{t-1}^{\leq}$ 的相容度小于 β 的对象集合与集合 $Cl_t^{\geq}$ 的并；$Cl_t^{\leq}$ 的 β-上近似表示 U 中肯定属于 $Cl_{t+1}^{\geq}$ 的相容度小于 β 的对象集合与集合 $Cl_t^{\leq}$ 的并。可以看出，这样的粗糙近似考虑了基于扩展优势关系的不相容性。

定理 2 假设决策系统 $S=<U,A,V,f>$，给定相容阈值 $\beta\in(0.5,1]$ 和属性集 $P\subseteq C$，以及 $Cl_t^{\geq},Cl_t^{\leq}\subseteq U$，则有：

$$\underline{P}^{\beta}(Cl_t^{\geq})^*\subseteq Cl_t^{\geq}\subseteq\bar{P}^{\beta}(Cl_t^{\geq})^*,\ (t=2,3,\cdots,n) \tag{13}$$

$$\underline{P}^{\beta}(Cl_t^{\leq})^*\subseteq Cl_t^{\leq}\subseteq\bar{P}^{\beta}(Cl_t^{\leq})^*,\ (t=1,2,\cdots,n-1) \tag{14}$$

证明：由式（7），（9），（11）和（12）可以很容易得到式（13）和（14）。

定理 3 假设决策系统 $S=<U,A,V,f>$，$x\in U$，$Cl_t^{\geq},Cl_t^{\leq}\subseteq U$，给定属性集 $P\subseteq C$ 和相容阈值 β,β'，若 $0.5<\beta'<\beta\leq1$，$t=2,3,\cdots,n$，有：

$$\underline{P}^{\beta}(Cl_t^{\geq})^*\subseteq\underline{P}^{\beta'}(Cl_t^{\geq})^* \tag{15}$$

$$\bar{P}^{\beta'}(Cl_t^{\geq})^*\subseteq\bar{P}^{\beta}(Cl_t^{\geq})^* \tag{16}$$

证明：

先证式（15）。因为：$\forall x\in U$，$x\in\underline{P}^{\beta}(Cl_t^{\geq})^*\Rightarrow c(D_P^{+*}(x),Cl_t^{\geq})\geq\beta\Rightarrow c(D_P^{+*}(x),Cl_t^{\geq})\geq\beta'\Rightarrow x\in\underline{P}^{\beta'}(Cl_t^{\geq})^*$，但反之不成立，所以：$\underline{P}^{\beta}(Cl_t^{\geq})^*\subseteq\underline{P}^{\beta'}(Cl_t^{\geq})^*$。

再证式（16）。因为：$\forall x\in U$，$x\in\bar{P}^{\beta'}(Cl_t^{\geq})^*\Rightarrow(x\in Cl_t^{\geq}\ \text{or}\ c(x,Cl_{t-1}^{\leq})<\beta')\Rightarrow(x\in Cl_t^{\geq}\ \text{or}\ c(x,Cl_{t-1}^{\leq})<\beta)\Rightarrow x\in\underline{P}^{\beta}(Cl_t^{\geq})^*$，但反之不成立，所以：$\bar{P}^{\beta'}(Cl_t^{\geq})^*\subseteq\underline{P}^{\beta}(Cl_t^{\geq})^*$。

定理 4 假设决策系统 $S=<U,A,V,f>$，$x\in U$，$Cl_t^{\geq},Cl_t^{\geq}\subseteq U$，给定属性集 $P\subseteq C$ 和相容阈值 β,β'，若 $0.5<\beta'<\beta\leq1$，$t=1,2,\cdots,n-1$，有：

$$\underline{P}^{\beta}(Cl_t^{\leq})^*\subseteq\underline{P}^{\beta'}(Cl_t^{\leq})^* \tag{17}$$

$$\bar{P}^{\beta'}(Cl_t^{\leq})^*\subseteq\bar{P}^{\beta}(Cl_t^{\leq})^* \tag{18}$$

证明：证明过程与定理 3 类似，故省略。

定理 5 当 $\beta=1$ 时，$Cl_t^{\geq}$ 的 β-下近似和 β-上近似及 $Cl_t^{\leq}$ 的 β-下近似和 β-上近似等价于基于扩展优势关系的粗糙

近似，即有

$$\underline{P}^{\beta}(Cl_t^{\geq})^*\Big|_{\beta=1}=\underline{P}(Cl_t^{\geq})^*,\ \underline{P}^{\beta}(Cl_t^{\leq})^*\Big|_{\beta=1}=\underline{P}(Cl_t^{\leq})^*$$

$$\bar{P}^{\beta}(Cl_t^{\geq})^*\Big|_{\beta=1}=\bar{P}(Cl_t^{\geq})^*,\ \bar{P}^{\beta}(Cl_t^{\leq})^*\Big|_{\beta=1}=\bar{P}(Cl_t^{\leq})^*$$

证明：当$\beta=1$时，$x\in\underline{P}^{\beta}(Cl_t^{\geq})^*|_{\beta=1}\Leftrightarrow\frac{|D_P^{+*}(x)\cap Cl_t^{\geq}|}{|D_P^{+*}(x)|}=1\Leftrightarrow D_P^{+*}(x)\subseteq Cl_t^{\geq}\Leftrightarrow x\in\underline{P}(Cl_t^{\geq})^*$，所以$\underline{P}^{\beta}(Cl_t^{\geq})^*|_{\beta=1}=\underline{P}(Cl_t^{\geq})^*$。同理可证：$\underline{P}^{\beta}(Cl_t^{\leq})^*|_{\beta=1}=(Cl_t^{\leq})^*$。

当$\beta=1$时，$x\in\bar{P}^{\beta}(Cl_t^{\geq})^*|_{\beta=1}\Leftrightarrow x\in(U-\underline{P}(Cl_{t-1}^{\leq})^*)\Leftrightarrow x\in\{U-\{x\in U,D_P^{-*}(x\subseteq Cl_{t-1}^{\leq})\}\Leftrightarrow x\in\{x\in U,D_P^{-*}(x\cap Cl_t^{\geq}\neq\phi\}\Leftrightarrow x\in\bar{P}(Cl_t^{\geq})^*$，所以$\bar{P}^{\beta}(Cl_t^{\geq})^*|_{\beta=1}$。同理可证：$\bar{P}^{\beta}(Cl_t^{\leq})^*|_{\beta=1}=\bar{P}(Cl_t^{\leq})^*$.
证毕。

2.3　分类质量与近似约简

2.3.1　分类质量

定义 9　假设决策系统 $S=<U,A,V,f>$, $Cl\subseteq U$，给定相容阈值$\beta\in(0.5,1]$ 和属性集 $P\subseteq C$，粗糙近似的分类质量 $\gamma_P^{\beta}(Cl)^*$ 的定义为：

$$\gamma_P^{\beta}(Cl)^*=\frac{\left|U-\bigcup_{t=1}^{n}(Bn_P^{\beta}(Cl_t^{\leq})^*)\right|}{|U|}=\frac{\left|U-\bigcup_{t=1}^{n}(Bn_P^{\beta}(Cl_t^{\geq})^*)\right|}{|U|}\quad(19)$$

分类质量$\gamma_P^{\beta}(Cl)^*$度量了扩展优势关系下论域中给定某一相容阈值β，不完备的偏好多属性决策系统中可以正确分类的知识在现有知识中所占的比例。

2.3.2　近似约简

定义 10　假设决策系统 $S=<U,A,V,f>A=C\cup D,Cl\subseteq U$，给定相容阈值$\beta\in(0.5,1]$，满足$\gamma_P^{\beta}(Cl)^*=\gamma_C^{\beta}(Cl)^*$的最小子集 $P\subseteq C$ 称为 C 关于 Cl 的一个β-约简，记作 $red_P^{\beta}(C,Cl)^*$。

在扩展优势关系下，一个偏好多属性决策系统可能有不止一个β-约简，所有β-约简的交集称为β-核，记作 $Core_P^{\beta}(C,Cl)^*$。β-核是不完全信息的偏好多属性决策系统中最重要的属性集，它也可能是空集。

2.4　获取偏好概率决策规则

得到基于扩展优势关系的知识约简后，便可由决策系统中的偏好属性导出偏好概率决策规则集，规则集中包含两类规则：$D_{\geq}$概率决策规则和 $D_{\leq}$概率决策规则。

（1）$D_{\geq}$概率决策规则：概据向上累积集$Cl_t^{\geq}$的β-下近似，可得到这类规则，形式如下：

If　$f(x,q_1)\geq r_{q1}\wedge f(x,q_2)\geq r_{q2}\wedge\cdots\wedge f(x,q_p)\geq r_{qp}$, then $x\in Cl_t^{\geq}$　with 置信度 α.其中：$P=\{q_1,q_2,\cdots,q_p\}\subseteq C$, $(rq_1,rq_2,\cdots,r_{qp})\in V_{q1}\times V_{q2}\times\cdots\times V_{qp}$, $t=2,3,\cdots,n$, $\alpha=\frac{|Cover(\rho)\cap Cl_t^{\leq}|}{|Cover(\rho)|}$, $Cover(\rho)$是指支持该规则前件的对象集合。

（2）$D_{\leq}$概率决策规则．根据向下累积集$Cl_t^{\leq}$的β-下近似，可得到这类规则，形式如下：

If　$f(x,q_1)\leq r_{q1}\wedge f(x,q_2)\leq r_{q1}\wedge\cdots\wedge f(x,q_p)\leq r_{qp}$, then $x\in Cl_t^{\leq}$　with 置信度 α'.其中：$P=\{q_1,q_2,\cdots,q_p\}\subseteq C$, $(rq_1,rq_2,\cdots,r_{qp})\in V_{q1}\times V_{q2}\times\cdots\times V_{qp}$, $t=1,2,\cdots,n-1$, $\alpha'=\frac{|Cover(\rho)\cap Cl_t^{\leq}|}{|Cover(\rho)|}$, $Cover(\rho)$ 是指支持该规则前件的对象集合。

3　实　例

为了说明本文方法便于比较分析，我们采用文献[8]中的实例。表1 是一个不完全信息的偏好多属性决策表，其中有 8 个商店作为对象，4 个偏好条件属性和 1 个偏好决策属性，条件属性 q_1表示销售人员的能力，q_2 表示商品的质量，q_3 表示交通条件，q_4 表示商场的环境，决策属性 d 表示商店的盈利或亏损，表示“*”是未知的偏好信息，现要求由表1 获取相应的偏好决策规则。

表 1　一个不完全信息的偏好多属性决策表

商店	属性				
	q_1	q_2	q_3	q_4	d
1	强	中	*	好	盈利
2	*	差	差	好	亏损
3	中	中	差	差	盈利
4	强	好	好	好	盈利
5	弱	*	好	差	亏损
6	中	好	差	*	盈利
7	中	中	*	好	亏损
8	弱	中	差	差	亏损

由表 1，有偏好次序的各属性的属性值集分别为：V_{q1}={强，中，弱}；V_{q2}={好，中，差}；V_{q3}={好，差}；V_{q4}={好，差}；V_d={盈利，亏损}={2, 1}。决策属性将对象集分成两类，亏损的对象集表示为：$Cl_1=\{2,5,7,8\}$，盈利的对象集表示为：$Cl_2=\{1,3,4,6\}$，这时有 $Cl_1^{\leq}=Cl_1=\{2,5,7,8\}$, $Cl_2^{\leq}=Cl_2=\{1,3,4,6\}$。

令相容阈值$\beta=0.8$, $P=\{q_1,q_2,q_3,q_4\}$，由式(5)-(10)得到扩展优势关系下$Cl_1^{\leq}$和$Cl_2^{\geq}$的β-下近似、β-上近似和β-边界域分别为：

$$\underline{P}^{0.8}(Cl_1^{\leq})^*=\{2,5,7,8\},\ \bar{P}^{0.8}(Cl_1^{\leq})^*=\{2,5,7,8\},\ Bn_P^{0.8}(Cl_1^{\leq})^*=\phi;$$

$$\underline{P}^{0.8}(Cl_2^{\leq})^*=\{1,3,4,6\},\ \bar{P}^{0.8}(Cl_2^{\leq})^*=\{1,3,4,6\},\ Bn_P^{0.8}(Cl_2^{\leq})^*=\phi;$$

由式(19)可知，在相容阈值$\beta=0.8$时，粗糙近似的分类质量 $\gamma_P^{\beta}(Cl)=1$，即决策表中所有知识都能够正确分类。在文

献[8]中，分类质量仅为 0.75，即决策表中有 25%的知识不能正确分类。这说明本文方法能够处理该决策系统中的不相容性，对更多的知识正确分类。

根据定义 10，求得决策表的β-约简为：$\{q_1, q_2\}$，从而获得偏好概率决策规则集，如表 2 所示，其中第一行中的规则是 $D_{\geq}$概率决策规则，下面 3 行中的规则是 $D_{\leq}$概率决策规则。

表 2 基于β-约简$\{q_1,q_2\}$的概率决策规则集

规则	覆盖的对象	置信度
$q_1\geq$中$\wedge q_2$中 $\xrightarrow{80\%} d\in Cl^{\geq}_2$	{1, 3, 4, 6, 7}	80%
$q_2\leq$差 $\xrightarrow{100\%} d\in Cl^{\leq}_1$	{2, 5}	100%
$q_1\leq$弱 $\xrightarrow{100\%} d\in Cl^{\leq}_1$	{2, 5, 8}	100%
$q_1\leq$中$\wedge q_2$中 $\xrightarrow{80\%} d\in Cl^{\leq}_1$	{2, 3, 5, 7, 8}	80%

根据文献[8]的方法，决策表的最小约简集是$\{q_1, q_2\}$，由最小约简导出偏好决策规则集，如表 3 所示。

表 3 基于约简$\{q_1,q_2\}$的最小偏好决策规则集

规则	覆盖的对象	置信度
$q_1\geq$强$\wedge q_2$中 $\xrightarrow{100\%} d\in Cl^{\geq}_2$	{1, 4}	100%
$q_1\geq$中$\wedge q_2\geq$好 $\xrightarrow{100\%} d\in Cl^{\geq}_2$	{4, 6}	100%
$q_2\leq$差 $\xrightarrow{100\%} d\in Cl^{\leq}_1$	{2, 5}	100%
$q_1\leq$弱 $\xrightarrow{100\%} d\in Cl^{\leq}_1$	{2, 5, 8}	100%

比较表 2 和表 3 可以看出，本文方法可以从决策表中获取偏好概率决策规则集，概率规则集中既包含了基于扩展优势关系的方法得到的决策规则，又包含了该方法无法获得的某些较强的规则，如规则 If $q_1\leq$中$\wedge q_2\leq$中 then $d\in Cl^{\leq}_1$ with 置信度 80%，从而获得更多有价值的决策知识，这将丰富决策信息，提高决策的合理性和准确性。

4 结束语

本文提出的新方法具有以下几个特点：

(1) 能够获取偏好概率决策规则。新的扩展粗糙集方法既继承了基于扩展优势关系的决策分析方法的优点，可导出强的决策规则，又能够发现非强的偏好概率决策规则，使得某些有价值的潜在知识被发掘出来。

(2) 规则集的泛化能力好。通过设计相容阈值，新方法可以处理不完全信息的偏好多属性决策系统中的不相容性，得到的概率决策规则具有较好的泛化能力。

(3) 适用范围广。新方法不仅适用于不完备的情形，对于完备的偏好多属性决策系统同样有效，因为新方法是已有方法在不完全信息情形下的推广。

总之，通过用扩展优势关系代替优势关系，并设定相容阈值，新方法能够处理不完全信息的偏好多属性决策系统中的不相容性，导出概率决策规则，从而发现某些有价值的潜在知识。这为解决有偏好信息，但信息不完全的多属性决策问题提供了一种有效的方法。

参考文献：

[1] 杨善林.智能决策方法与智能决策支持系统[M].北京：科学出版社，2005.

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

[3] Ziarko W. Variable precision rough set model [J].Journal of Computer and System Sciences,1993,46(1):39-59.

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

[5] Greco S.,Matarazzo B.,Slwinski R.,Stefanowski J.Variable consistency model of dominance-based rough sets approach[A].Rough Sets and Current Trends in Computing Lecture Notes in Artificial Intelligence (RSCTC2000)[C],Berlin:Springer-Verlag, 2001:170-181.

[6] 潘郁，菅利荣，达庆利.多标准决策表中发现概率规则的变精度粗糙集方法[J].中国管理科学.2005,13(1):95-100.

[7] 菅利荣，刘思峰.偏好多属性决策表概率决策的扩展粗糙集方法[J].南京航空航天大学学报.2005, 37(4):535-540.

[8] 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.

(责任编辑：董小玉)

Rough Sets Methodology for Preferential Probabilistic Rules Discovery from Incomplete Decision Systems

Abstract:In order to discover probabilistic decision rules in preferential multiple attribute decision system with incomplete information,an extension of the rough sets model is proposed in the paper.Firstly,the concept of consistency degree based on extended dominance relation is presented;Secondly,rough approximations of knowledge based on consistency degree are defined and basic properties of rough approximations are proved.Thirdly,the classification quality of rough approximations and β-educe of knowledge are discussed and the probabilistic sorting decision rules are given.Finally,the feasibility and effectiveness of the method are demonstrated by a real example.

Key Words:incomplete information;multiple attribute decision making;rough sets;extended dominance relation;consistency degree;probabilistic decision rules