基于概率的有序信息系统

闫新宝 王国胤 张清华
(重庆邮电大学计算机科学与技术研究所 重庆 400065)

摘 要 若信息系统中所有的条件属性都是偏好有序的,则称此信息系统为有序信息系统。首先,分析了区间值有序信息系统没有蕴含属性值区间上的概率分布信息的缺点,建立了一种基于概率的有序信息系统。然后,在这种信息系统上,研究了关于单调偏好有序属性和非单调偏好有序属性的二元偏好关系,建立了一种基于概率的优势关系,定义了基于这种优势关系的粗糙集模型。最后研究了基于概率的有序决策表及其决策规则。

关键词 区间值有序信息系统,概率,粗糙集,决策规则

Probability-based Ordered Information Systems

YAN Xin-bao WANG Guo-yin ZHANG Qing-hua
(Institute of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract An information system is called an ordered information system if the domains of all condition attributes are ordered according preference. Firstly, the defects of interval-valued ordered information systems that do not include the probability distribution were analyzed and probability-based ordered information systems were proposed. Secondly, the approaches of establishing the outranking relation between objects with respect to attributes with monotonic preference and non-monotonic preference in probability-based ordered information systems were given respectively, and a probability-based dominance relation was defined, and the rough set model based on the probability-based dominance relation was presented. Finally, the probability-based ordered decision tables were defined, and the decision rules from probability-based ordered decision tables were researched.

Keywords Interval-valued ordered information systems, Probability, Rough sets, Decision rules

1 引言

粗糙集理论[1]是波兰数学家 Pawlak 教授于 1982 年提出的一种表达和分析数据的数学工具,可以作为属性值的表达模型用来描述属性间的依赖关系,评价属性的重要性以及获取有用的决策规则。粗糙集理论由于能够用来分析和处理不精确、模糊和不确定知识和信息,因此受到广大研究者的关注,已经被成功应用在人工智能、机器学习与知识发现、决策支持与分析、模式识别与分类等领域。

多属性决策是决策理论研究的一个重要内容,信息偏好有序是实际多属性决策问题的重要特征。由于受到决策者偏好的影响,对象属性值之间往往存在着优劣关系,但经典粗糙集理论是基于不可分辨关系,并没有考虑属性域偏好有序的情况。为此,Greco 等[2-4]提出了一种基于优势关系的粗糙集方法,该方法考虑偏好属性的顺序特征,为解决具有偏好信息的多属性决策问题提供了新思路。

在一些实际问题中,一类对象的一些偏好属性的属性值为一个区间,而不是一个单值,区间值有序信息系统[5]是描述这类对象的工具之一。然而,区间值有序信息系统存在丢失属性值区间上的概率分布信息的问题,属性值区间上的概率分布信息对建立对象间的优势关系是非常重要的。本文第 2 节具体分析了属性值区间上的概率分布信息对建立对象间的优势关系的重要性;第 3 节针对区间值有序信息系统的上述缺点,建立了一种基于概率的有序信息系统,它能体现属性值区间上的概率分布,有效地克服区间值有序信息系统存在的上述问题。

优势关系是基于二元偏好关系的(用"$\geq$"来表示)。$y \geq_a x$ 表示在属性 a 下,对象 y 至少和对象 x 一样好。现实世界中,一些对象的偏好属性是单调偏好有序的,如银行的负债率越低,破产风险越小;还存在一些对象,它们的一些偏好属性是非单调有序的,例如,动物的体温、血压等不是越高或者越低越好。已有的文献中,二元偏好关系都是基于偏好属性是单调有序的,本文在提出的基于概率的信息系统上,研究了关于非单调偏好有序属性的二元偏好关系,进而建立了一种新的基于概率的优势关系。然后本文定义了基于此优势关系的粗糙集模型,最后研究了从基于概率的有序决策表中获取

到稿日期:2011-02-20 返修日期:2011-04-15 本文受国家自然科学基金项目(61073146),重庆市杰出青年科学基金项目(2008BA2041),重庆市教育委员会科学技术研究项目(KJ110512)资助。

闫新宝(1986—),男,硕士生,主要研究方向为粒计算、粗糙集理论,E-mail:yxinbao@163.com;王国胤(1970—),男,博士,教授,博士生导师,主要研究方向为粗糙集理论、粒计算、数据挖掘、知识技术等;张清华(1974—),男,博士,副教授,主要研究方向为智能信息处理、粒计算、粗糙集等。

决策规则的方法。

2 区间值有序信息系统存在的问题分析

这一节首先介绍区间值有序信息系统的概念，然后提出社区人口健康状况评估问题，并分析用区间值有序信息系统对问题进行建模存在的问题。

定义 1[5,6] 称 $S=(U,A,V,f)$ 是区间值信息系统，其中，U 为非空有限的对象集合，也称为论域，A 为非空有限的属性集合，$V=\bigcup_{a\in A}V_a$ 是属性值的集合，V_a 表示属性 $a\in A$ 的值域，$f:U\times A\to V$ 是一个信息函数，它指定 U 中每一个对象 x 的属性值，$\forall_{x\in U,a\in A}(f(x,a)\subseteq V_a\wedge f(x,a)=[a^L(x),a^U(x)])$，其中，$a^L(x),a^U(x)\in R$，$a^L(x)$ 和 $a^U(x)$ 分别表示对象 x 在属性 a 下区间取值的下边界和上边界，$a^L(x)\leqslant a^U(x)$。

定义 2[2] 给定信息系统 $S=(U,A,V,f)$，若信息系统 S 中所有的条件属性都是偏好有序的，则称信息系统 S 为有序信息系统。

定义 3[5] 给定区间值信息系统 $S=(U,A,V,f)$，若区间值信息系统 S 中所有的条件属性都是偏好有序的，则称区间值信息系统 S 为区间值有序信息系统。

区间值有序信息系统处理某些问题时存在一些问题，下面通过两个实例进行分析。

首先，介绍将要用到的区间值有序信息系统上的优势关系的定义。

定义 4 给定区间值信息系统 $S=(U,A,V,f)$，属性子集 $B(B\subseteq A)$，对象之间的优势关系定义为：

(1)上优势关系[5]

$R_B^{U\geqslant}=\{(y,x)\in U\times U|\ \forall_{a\in B}(a^U(y)\geqslant a^U(x))\}$

如果 $(y,x)\in R_B^{U\geqslant}$，则称在属性子集 B 中，对象 y 至少上优于对象 x。

(2)下优势关系[5]

$R_B^{L\geqslant}=\{(y,x)\in U\times U|\ \forall_{a\in B}(a^L(y)\geqslant a^L(x))\}$

如果 $(y,x)\in R_B^{L\geqslant}$，则称在属性子集 B 中，对象 y 至少下优于对象 x。

(3)下上优势关系[5]

$R_B^{UL\geqslant}=\{(y,x)\in U\times U|\ \forall_{a\in B}(a^L(y)\geqslant a^U(x))\}$

如果 $(y,x)\in R_B^{UL\geqslant}$，则称在属性子集 B 中，对象 y 一定优于对象 x。

(4)上下优势关系[6-8]

$R_B^{LU\geqslant}=\{(y,x)\in U\times U|\ \forall_{a\in B}(a^U(y)\geqslant a^L(x))\}$

如果 $(y,x)\in R_B^{LU\geqslant}$，则称在属性子集 B 中，对象 y 可能优于对象 x。

区间值上优势关系的特点：如果对象 x 和对象 y 满足优势关系(3)，对象 x 一定优于对象 y，但其条件太苛刻，划分粒度过大；如果对象 x 和对象 y 满足优势关系(1)、(2)、(4)，对象 x 只以一定概率优于对象 y。文献[9]提出了可变精度优势关系，在特定假设下提出了计算对象 x 优于对象 y 的概率的方法。所以利用优势关系(1)、(2)、(4)对属性域进行排序的结果以一定概率成立，得到的决策规则以一定概率成立，需要决策者承担风险。

某社区为了对自己社区的人口的健康状况进行评估，要采集人的各项重要生理指标，包括体温、血糖、血压等，然后综合各项指标对每个人的健康状况做出综合评价。因为人的体温、血糖、血压等属性是偏好有序的，并且它们的属性值不是一个单值，而是一个连续变化的区间，所以应建立区间值信息系统来描述对象集合。然而经过分析发现，用区间值有序信息系统来描述人口健康信息有很大缺点。

对人的体温进行多次测量，可以得到人的体温信息。考虑如图 1 所示的情况。对象 x 的体温属性的变化范围是区间 $[35,36.5]$，其中处于区间 $[35,36]$ 与 $[36,36.5]$ 的概率分别为 10%、90%；对象 y 的体温属性的变化范围是区间 $[35,37]$，其中处于区间 $[35,36]$ 与 $[36,37]$ 的概率分别为 90%、10%。直观上容易得到对象 x 的确定优于对象 y，进而得到确定的排序结果和确定的决策规则。如果建立区间值有序信息系统来描述这两个对象，区间值有序信息系统丢失了概率信息，对象 x 和对象 y 显然不满足下上优势关系，不能得到确定的决策规则；它们满足上优势关系、下优势关系和上下优势关系，能得到不确定的决策规则。

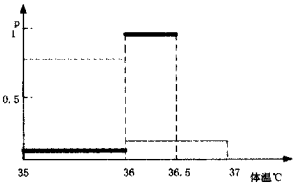

图 1 属性值区间上的概率分布(粗线段为 x，细线段为 y)

考虑如图 2 所示的另一种情况。设有两个对象 x 和 y，经过多次测量，对象 x 的体温属性的变化范围是区间 $[37,38]$，对象 y 的体温属性的变化范围也是区间 $[37,38]$，对象 x 的体温有 80%的时间处于区间 $[37,37.5]$ 内，对象 y 的体温有 80%的时间处于区间 $[37.5,38]$ 内。可以得出对象 x 的体温优于对象 y 的体温。若建立区间值有序信息系统来描述这两个对象，区间值有序信息系统丢失了概率信息，对象 x 和 y 的属性值都是区间 $[37,38]$，不满足区间值信息系统上的任何一种优势关系，对象 x 和 y 都是不可分辨的，不可避免地扩大了划分的粒度。

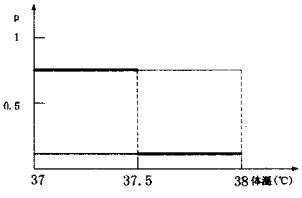

图 2 属性值区间上的概率分布(粗线段为 x，细线段为 y)

我们容易得出导致上述结果的原因是区间值有序信息系统缺少属性值区间上的概率分布信息，概率信息的丢失导致建立的对象之间的优势关系不符合实际情况，所以我们希望建立一种体现属性值区间上的概率分布信息的信息系统。本文第 3 节建立一种基于概率的有序信息系统，建立这种有序信息系统的代价和建立区间值有序信息系统的代价相当。

3 基于概率的有序信息系统

3.1 基于概率的有序信息系统

现有一个对象集，对象的属性值在一个范围内变化，我们要建立一种基于概率的信息系统，它能体现属性值区间上的概率分布。首先，假设某对象 x 的某个属性 a 在取值空间的取值为总体 X^a，对总体 X^a 做 n 次观察，得到一个样本 X_1^a，

$X_2^a,\cdots,X_n^a$；然后，以样本 $X_1^a,X_2^a,\cdots,X_n^a$ 的样本值 $x_1^a,x_2^a,\cdots,x_n^a$ 作为这个属性的值。用上述方法得到对象全集 U 中任意属性的任意属性值，这样就可以得到一个基于概率的信息系统。

假设属性值 X^a 在区间上的概率密度函数 $f(x;\theta),\theta\in\Theta$ 的形式为已知(θ 为待估参数，Θ 是 θ 的可能取值范围)，运用数理统计中估计总体参数的点估计法，可以计算出概率密度函数中待估参数的估计量和估计值，进而得到属性值 X^a 在区间上的概率密度函数。所以这样的信息系统可以体现对象的属性值在区间上的概率分布。

定义 5 称 $S=(U,A,V,f)$ 是基于概率的信息系统，其中，U 为非空有限的对象集合，也称为论域，A 为非空有限的属性集合，$V=\bigcup_{a\in A}V_a$ 是属性值的集合，V_a 表示属性 $a\in A$ 的值域，$f:U\times A\to V$ 是一个信息函数，它指定 U 中每一个对象 x 的属性值，$\forall_{x\in U,a\in A}(f(x,a)\subseteq V_a\wedge f(x,a)=(a(x)_1,a(x)_2,\cdots,a(x)_n))$，其中，$(a(x)_1,a(x)_2,\cdots,a(x)_n)$ 表示从对象 x 在属性 a 下的取值空间随机抽取的一个样本 $X_1^a,X_2^a,\cdots,X_n^a$ 的样本值。

定义 6 如果一个基于概率的信息系统 $S=(U,A,V,f)$ 的所有条件属性都是偏好属性，那么我们称之为基于概率的有序信息系统。

用基于概率的信息系统描述社区人口健康信息，根据基于概率的有序信息系统的定义可以建立如下基于概率的有序信息系统(见表 1)。为了简化工作，我们采取了 3 项生理指标。各项生理指标为 a_1：体温，a_2：血糖，a_3：血压(收缩压)。

偏好属性的值域是按二元偏好关系“$\geq$”偏好有序的，“$x\geq_a y$”意思是在属性 a 上 x 至少和 y 一样好。

表 1 基于概率的有序信息系统：社区人口健康信息表

U	a_1	a_2	a_3
x_1	(36.5,36.7,36.7,37.0,37.0,37.0,37.3)	(4.8,4.9,4.9,5.0,5.0,5.2,5.3)	(109,110,110,115,116,116,118)
x_2	(35.5,35.9,35.9,36.0,36.0,36.2,36.6)	(3.3,3.4,3.4,3.5,3.5,3.6,3.8)	(148,150,153,153,154,155,155)
x_3	(38.0,38.3,38.3,38.5,38.5,38.8,39.0)	(5.2,5.2,5.3,5.5,5.7,5.8,5.8)	(139,141,145,147,147,148,148)
x_4	(37.8,37.8,38.2,38.7,38.9,38.9,39.1)	(6.1,6.3,6.4,6.4,6.5,6.6,6.6)	(85,86,87,87,87,87,89)
x_5	(37.7,37.9,37.9,38.0,38.1,38.1,38.3)	(5.3,5.4,5.4,5.5,5.5,5.6,5.8)	(93,94,94,96,96,97,99)

在以往的研究文献中，都是假设属性是递增有序或递减有序的，我们发现此信息系统的偏好属性是非单调偏好有序的，每个属性都有一个最优值，这不同于我们以往遇到的情况。

设对象 x 在某非单调偏好有序属性 a 上的属性值为随机变量 X^a，对象 y 在非单调偏好有序属性 a 上的属性值为随机变量 Y^a。通过以下步骤，在具有非单调偏好有序属性的基于概率的有序信息系统上，建立两个对象 x 与 y 在非单调偏好有序属性 a 上的偏好关系“$\geq_a$”。

步骤 1 求属性值随机变量 X^a 和 Y^a 的概率密度函数。假设 X^a 在区间上的概率密度函数 $f(x;\theta),\theta\in\Theta$ 的形式为已知(θ 为待估参数，Θ 是 θ 的可能取值范围)，运用数理统计中估计总体参数的点估计法，得到 X^a 在区间上的概率密度函数。

步骤 2 令 $A(X^a)=E(|X^a-a^{best}|)$，a^{best} 为非单调偏好有序属性 a 的最优值(经验值)，$A(X^a)$ 表示对象 x 在属性 a 上相对于属性最优值 a^{best} 的平均偏移量。计算 $A(X^a)$ 和 $A(Y^a)$，若 $A(X^a)<A(Y^a)$，则 $x\geq_a y$；若 $A(X^a)>A(Y^a)$，则 $y\geq_a x$；若 $A(X^a)=A(Y^a)$，则转到步骤 3。$A(X^a)$ 的计算方法：$A(X^a)=E(|X^a-a^{best}|)=|E(X^a)-a^{best}|$，$E(X^a)$ 为已知。

步骤 3 比较 $D(X^a)$ 和 $D(Y^a)$，若 $D(X^a)<D(Y^a)$，则 $x\geq_a y$；若 $D(X^a)>D(Y^a)$，则 $y\geq_a x$；若 $D(X^a)=D(Y^a)$，则认为 x 和 y 相对于 a 上的偏好关系是不可分辨的。

若在基于概率的有序信息系统上偏好属性是单调偏好有序的，设对象 x 在某单调偏好有序属性上的属性值为随机变量 X^a，对象 y 在某单调偏好有序属性上的属性值为随机变量 Y^a。通过以下步骤，建立两个对象 x 与 y 在单调偏好有序属性 a 上的偏好关系“$\geq_a$”。

步骤 1 求属性值随机变量 X^a 和 Y^a 的分布律或概率密度函数。参照在非单调偏好有序属性 a 上建立偏好关系的方法的步骤 1。

步骤 2 比较 $E(X^a)$ 与 $E(Y^a)$。若 $E(X^a)<E(Y^a)$，如果属性 a 递增有序，则 $y\geq_a x$，如果属性 a 递减有序，则 $x\geq_a y$；若 $E(X^a)>E(Y^a)$，如果属性 a 递增有序，则 $x\geq_a y$，如果属性 a 递减有序，则 $y\geq_a x$；若 $E(X^a)=E(Y^a)$，则转到步骤 3。

步骤 3 比较 $D(X^a)$ 和 $D(Y^a)$，若 $D(X^a)<D(Y^a)$，则 $x\geq_a y$；若 $D(X^a)>D(Y^a)$，则 $y\geq_a x$；若 $D(X^a)=D(Y^a)$，则认为 x 和 y 相对于 a 上的偏好关系是不可分辨的。

考虑属性子集 $B\subseteq A$，定义 $x\geq_B y\Leftrightarrow\forall a\in B,x\geq_a y$，$x\geq_a y$ 的定义如上所述。基于如此定义的 $x\geq_B y$，可以定义基于概率的有序信息系统上的优势关系。令 $R_B^{\geq}$，$(B\subseteq A)$ 表示基于概率的有序信息系统上的优势关系，我们定义 $R_B^{\geq}$ 如下。

定义 7 给定基于概率的有序信息系统 $S=(U,A,V,f)$，属性子集 $B(B\subseteq A)$，基于概率的优势关系 $R_B^{\geq}$ 与 $R_B^{\leq}$ 的定义为：

$$R_B^{\geq}=\{(y,x)\in U\times U|(\forall a\in B)(y\geq_B x)\}$$

$$R_B^{\leq}=\{(y,x)\in U\times U|(\forall a\in B)(x\geq_B y)\}$$

优势类 $[x]_B^{\leq}$ 与 $[x]_B^{\geq}$ 定义如下。

定义 8 给定基于概率的有序信息系统 $S=(U,A,V,f)$，属性子集 $B(B\subseteq A)$，基于概率的优势类 $[x]_B^{\geq}$ 与 $[x]_B^{\leq}$ 定义为：

$$[x]_B^{\geq}=\{y\in U|(x,y)\in R_B^{\geq}\}$$

$$[x]_B^{\leq}=\{y\in U|(y,x)\in R_B^{\geq}\}$$

定理 1 给定基于概率的有序信息系统 $S=(U,A,V,f)$，属性子集 $B,D\subseteq A$，则：

(1) $R_B^{\geq}=\bigcap_{a\in B}R_a^{\geq}$，$R_B^{\leq}=\bigcap_{a\in B}R_a^{\leq}$；

(2) $R_B^{\geq}$，$R_B^{\leq}$ 是自反的和传递的，不是对称的；

(3) 若 $B\subseteq D\subseteq A$，则 $R_B^{\geq}\supseteq R_D^{\geq}\supseteq R_A^{\geq}$；

(4) 若 $B\subseteq D\subseteq A$，则 $[x]_B^{\geq}\supseteq[x]_D^{\geq}\supseteq[x]_A^{\geq}$；

(5) 若 $x_j\subseteq[x_i]_B^{\geq}$，则 $[x_j]_B^{\geq}\subseteq[x_i]_B^{\geq}$，$[x_i]_B^{\geq}=\bigcup\{[x_j]_B^{\geq}:x_j\in x_i]_B^{\geq}\}$。

限于篇幅，证明略。

令 $U/R_B^{\succcurlyeq}$ 代表对论域 U 的分类，$U/R_B^{\succcurlyeq}$ 即是集合簇 $F=\{[x]_B^{\succcurlyeq}\mid x\in U\}$。$U/R_B^{\succcurlyeq}$ 中的任何元素都是优势类，一般地，$U/R_B^{\succcurlyeq}$ 中的全部优势类并不是集合 U 的一个划分。实际上，$F=\{[x]_B^{\succcurlyeq}\mid x\in U\}$ 是 U 的一个覆盖，$\bigcup_{x\in U}[x]_B^{\succcurlyeq}=U$。

3.2 基于概率的有序信息系统的粗糙集模型

本小节引入基于概率的有序信息系统的粗糙集方法，并分析其重要性质。

定义 9 令 $S=(U,A,V,f)$ 为一个基于概率的信息系统，$B\subseteq A$，$X\subseteq U$，X 关于优势关系 $R_B^{\succcurlyeq}$ 的上近似集与下近似集的定义如下：

$$\underline{R_B^{\succcurlyeq}}(X)=\{x\in U\mid [x]_B^{\succcurlyeq}\subseteq X\}$$

$$\overline{R_B^{\succcurlyeq}}(X)=\{x\in U\mid [x]_B^{\succcurlyeq}\cap X\neq\Phi\}$$

式中，$\underline{R_B^{\succcurlyeq}}(X)$ 是根据知识 B，U 中一定能归入已知集合 X 的对象的集合，$\overline{R_B^{\succcurlyeq}}(X)$ 是根据知识 B，U 中可能归入已知集合 X 的对象的集合。$Bn(X)=\overline{R_B^{\succcurlyeq}}(X)-\underline{R_B^{\succcurlyeq}}(X)$ 是集合 X 的边界域。

定理 2 令 $S=(U,A,V,f)$ 为一个基于概率的信息系统，$B\subseteq A$，$X,Y\subseteq U$，则：

(1) $\underline{R_B^{\succcurlyeq}}(\Phi)=\overline{R_B^{\succcurlyeq}}(\Phi)=\Phi$，$\underline{R_B^{\succcurlyeq}}(U)=\overline{R_B^{\succcurlyeq}}(U)=U$；

(2) $\underline{R_B^{\succcurlyeq}}(X)\subseteq X\subseteq\overline{R_B^{\succcurlyeq}}(X)$；

(3) $\underline{R_B^{\succcurlyeq}}(\underline{R_B^{\succcurlyeq}}(X))=\underline{R_B^{\succcurlyeq}}(X)$，$\overline{R_B^{\succcurlyeq}}(\overline{R_B^{\succcurlyeq}}(X))=\overline{R_B^{\succcurlyeq}}(X)$；

(4) $\underline{R_B^{\succcurlyeq}}(X)=\sim\overline{R_B^{\succcurlyeq}}(\sim X)$，$\overline{R_B^{\succcurlyeq}}(X)=\sim\underline{R_B^{\succcurlyeq}}(\sim X)$；

(5) $\underline{R_B^{\succcurlyeq}}(X)\subseteq\underline{R_A^{\succcurlyeq}}(X)$，$\overline{R_B^{\succcurlyeq}}(X)\supseteq\overline{R_A^{\succcurlyeq}}(X)$。

限于篇幅，证明略。

4 基于概率的有序决策表及其决策规则

本节介绍基于概率的有序决策表和基于概率的有序决策表决策规则。

4.1 基于概率的有序决策表

一个基于概率的有序决策表是一个基于概率的有序信息系统 $S=(U,C\cup d,V,f)$，其中 $d(d\notin C$ 且 $f(x,d)(x\in U)$ 是单值）称为决策属性，是体现对象总体性质的偏好属性，C 是条件属性。

决策属性 d 对论域 U 划分为有限个类，设 $\boldsymbol{D}=\{D_1,D_2,\cdots,D_r\}$ 是这些类的集合，$\boldsymbol{D}$ 中的元素是经过排序的，$\forall i,j\leqslant r$，如果 $i\geqslant j$，则 D_i 中的任意元素优于 D_j 中的任意元素。

向上累积集 $D_i^{\succcurlyeq}$ 定义为 $D_i^{\succcurlyeq}=\bigcup_{j\geqslant i}D_j$；向下累积集 $D_i^{\preccurlyeq}$ 定义为 $D_i^{\preccurlyeq}=\bigcup_{j\leqslant i}D_j$，其中 $1\leqslant i\leqslant r$。

定义 10 令 $S=(U,C\cup d,V,f)$ 为一个基于概率的有序决策表，$A\subseteq C$ 且 $\boldsymbol{D}=\{D_1,D_2,\cdots,D_r\}$ 是由决策属性导出的决策，$D_i^{\succcurlyeq}(i\leqslant r)$ 关于优势关系 $R_A^{\succcurlyeq}$ 的上、下近似集定义为：

$$\underline{R_A^{\succcurlyeq}}(D_i^{\succcurlyeq})=\{x\in U\mid [x]_A^{\succcurlyeq}\subseteq D_i^{\succcurlyeq}\}$$

$$\overline{R_A^{\succcurlyeq}}(D_i^{\succcurlyeq})=\bigcup_{x\in D_i^{\succcurlyeq}}[x]_A^{\succcurlyeq}$$

定义 11 令 $S=(U,C\cup d,V,f)$ 为一个基于概率的有序决策表，$A\subseteq C$ 且 $\boldsymbol{D}=\{D_1,D_2,\cdots,D_r\}$ 是由决策属性导出的决策，$D_i^{\preccurlyeq}(i\leqslant r)$ 关于优势关系 $R_A^{\succcurlyeq}$ 的上、下近似集定义为：

$$\underline{R_A^{\succcurlyeq}}(D_i^{\preccurlyeq})=\{x\in U\mid [x]_A^{\succcurlyeq}\subseteq D_i^{\preccurlyeq}\}$$

$$\overline{R_A^{\succcurlyeq}}(D_i^{\preccurlyeq})=\bigcup_{x\in D_i^{\preccurlyeq}}[x]_A^{\succcurlyeq}$$

自然地，得出 $D_i^{\succcurlyeq}(i\leqslant r)$ 和 $D_i^{\preccurlyeq}(i\leqslant r)$ 的 A 边界域定义为：$Bn_A(D_i^{\succcurlyeq})=\overline{R_A^{\succcurlyeq}}(D_i^{\succcurlyeq})-\underline{R_A^{\succcurlyeq}}(D_i^{\succcurlyeq})$，$Bn_A(D_i^{\preccurlyeq})=\overline{R_A^{\succcurlyeq}}(D_i^{\preccurlyeq})-\underline{R_A^{\succcurlyeq}}(D_i^{\preccurlyeq})$。

下近似集 $\underline{R_A^{\succcurlyeq}}(D_i^{\succcurlyeq})$ 和 $\underline{R_A^{\succcurlyeq}}(D_i^{\preccurlyeq})$ 可以用于从基于概率的有序信息系统中提取确定性决策规则，边界域 $Bn_A(D_i^{\succcurlyeq})$ 和 $Bn_A(D_i^{\preccurlyeq})$ 可以用于从基于概率的有序信息系统中提取可能性决策规则。

4.2 基于概率的有序决策表的决策规则

从基于概率的信息系统中，可以提取以下两类的决策规则。

第一类规则：

(1) if $\delta_1\wedge\cdots\wedge\delta_m\wedge\gamma_{m+1}\wedge\cdots\wedge\gamma_n\wedge\lambda_{n+1}\wedge\cdots\wedge\lambda_p$

then $x\in D_i^{\succcurlyeq}$；

(2) if $\delta_1\wedge\cdots\wedge\delta_m\wedge\gamma_{m+1}\wedge\cdots\wedge\gamma_n\wedge\lambda_{n+1}\wedge\cdots\wedge\lambda_p$

then x could belong to $D_i^{\succcurlyeq}$。

其中：

$\delta_l=(E(X^{a_l})\geqslant\mu_{a_l})\vee((E(X^{a_l})=\mu_{a_l})\wedge(D(X^{a_l})\leqslant\sigma^2_{a_l}))$，$(1\leqslant l\leqslant m)$；

$\gamma_k=(E(X^{a_k})\leqslant\mu_{a_k})\vee((E(X^{a_k})=\mu_{a_k})\wedge(D(X^{a_k})\leqslant\sigma^2_{a_k}))$，$(m+1\leqslant k\leqslant n)$；

$\lambda_s=(|E(X^{a_s})-a_s^{best}|\leqslant|\mu_{a_s}-a_s^{best}|)\vee((|E(X^{a_s})-a_s^{best}|=|\mu_{a_s}-a_s^{best}|)\wedge(D(X^{a_s})\leqslant\sigma^2_{a_s}))$，$(n+1\leqslant s\leqslant p)$。

第二类规则：

(1) if $\delta_1\wedge\cdots\wedge\delta_m\wedge\gamma_{m+1}\wedge\cdots\wedge\gamma_n\wedge\lambda_{n+1}\wedge\cdots\wedge\lambda_p$

then $x\in D_i^{\preccurlyeq}$；

(2) if $\delta_1\wedge\cdots\wedge\delta_m\wedge\gamma_{m+1}\wedge\cdots\wedge\gamma_n\wedge\lambda_{n+1}\wedge\cdots\wedge\lambda_p$

then x could belong to $D_i^{\preccurlyeq}$。

其中：

$\delta_l=(E(X^{a_l})\leqslant\mu_{a_l})\vee((E(X^{a_l})=\mu_{a_l})\wedge(D(X^{a_l})\geqslant\sigma^2_{a_l}))$，$(1\leqslant l\leqslant m)$；

$\gamma_k=(E(X^{a_k})\geqslant\mu_{a_k})\vee((E(X^{a_k})=\mu_{a_k})\wedge(D(X^{a_k})\geqslant\sigma^2_{a_k}))$，$(m+1\leqslant k\leqslant n)$；

$\lambda_s=(|E(X^{a_s})-a_s^{best}|\geqslant|\mu_{a_s}-a_s^{best}|)\vee((|E(X^{a_s})-a_s^{best}|=|\mu_{a_s}-a_s^{best}|)\wedge(D(X^{a_s})\geqslant\sigma^2_{a_s}))$，$(n+1\leqslant s\leqslant p)$。

在上面两类 4 种决策规则中：$O_1=\{a_1,a_2,\cdots,a_m\}$，$O_2=\{a_{m+1},a_{m+2},\cdots,a_n\}$，$O_3=\{a_n,a_{n+1},\cdots,a_p\}$，$C=O_1\cup O_2\cup O_3$，$O_1$ 是单调递增有序的属性集，O_2 是单调递减有序的属性集，O_3 是非单调有序的属性集，$a_i^{best}(i=1,2,\cdots p)$ 是 a_i 的最优值（经验值），$(\mu_{a_1},\mu_{a_2},\cdots,\mu_{a_p})\in\cup E(X_i^{a_1})\times\cup E(X_i^{a_2})\times\cdots\times\cup E(X_i^{a_p})$，$(\sigma^2_{a_1},\sigma^2_{a_2},\cdots\sigma^2_{a_n})\in\cup D(X_i^{a_1})\times\cup D(X_i^{a_2})\times\cdots\cup D(X_i^{a_n})$。

根据社区人口健康信息表建立社区人口健康信息决策表（见表 2）。各项生理指标为 a_1：体温（摄氏度），a_2：血糖（mmol/L），a_3：血压（收缩压）（mmHg）。各属性都为非单调偏好有序属性，$a_1^{best}=37$，$a_2^{best}=5$，$a_3^{best}=115$（注：属性最优值为经验值）。

下面是从社区人口健康信息决策表提取规则的过程。

因为正态分布是自然界中最常见的分布，所以我们假设表中的各个属性的属性值在区间或集合上服从正态分布，根

据在基于概率的有序信息系统上建立偏好关系的步骤，得到以下偏好关系：$x_1 \succeq_{a_1} x_5 \succeq_{a_1} x_4 \succeq_{a_1} x_3 \succeq_{a_1} x_2$，$x_1 \succeq_{a_2} x_5 \succeq_{a_2} x_3 \succeq_{a_2} x_4 \succeq_{a_2} x_2$，$x_1 \succeq_{a_3} x_5 \succeq_{a_3} x_4 \succeq_{a_3} x_3 \succeq_{a_3} x_2$。

从而得到优势类：

$[x_1]_C^{P\geqslant}=\{x_1\}$，$[x_2]_C^{P\geqslant}=\{x_1,x_2,x_3,x_4,x_5\}$

$[x_3]_C^{P\geqslant}=\{x_1,x_3,x_5\}$

$[x_4]_C^{P\geqslant}=\{x_1,x_4,x_5\}$

$[x_5]_C^{P\geqslant}=\{x_1,x_5\}$

$[x_1]_C^{P\leqslant}=\{x_2,x_3,x_4,x_5\}$

$[x_2]_C^{P\leqslant}=\{x_2\}$

$[x_3]_C^{P\leqslant}=\{x_2,x_3\}$

$[x_4]_C^{P\leqslant}=\{x_2,x_4\}$

$[x_5]_C^{P\leqslant}=\{x_2,x_3,x_4,x_5\}$

$\boldsymbol{D}=\{D_1,D_2\}$；$D_1=\{x_1,x_4,x_5\}$，$D_2=\{x_2,x_3\}$

$D_1^{\geqslant}=D_1=\{x_1,x_4,x_5\}$，$D_2^{\leqslant}=D_2=\{x_2,x_3\}$

表 2　基于概率的有序决策表：社区人口健康信息表

U	a_1	a_2	a_3	d
x_1	(36.5,36.7,36.7,37.0,37.0,37.0,37.3)	(4.8,4.9,4.9,5.0,5.0,5.2,5.3)	(109, 110, 110, 115,116,116,118)	健康
x_2	(37.8,37.8,38.2,38.7,38.9,38.9,39.1)	(3.3,3.4,3.4,3.5,3.5,3.6,3.8)	(148, 150, 153, 153,154,155,155)	不健康
x_3	(38.0,38.3,38.3,38.5,38.5,38.8,39.0)	(5.2,5.2,5.3,5.5,5.7,5.8,5.8)	(139, 141, 145, 147,147,148,148)	不健康
x_4	(35.5,35.9,35.9,36.0,36.0,36.2,36.6)	(6.1,6.3,6.4,6.4,6.5,6.6,6.6)	(85,86,87,87,87,87,89)	健康
x_5	(37.7,37.9,37.9,38.0,38.1,38.1,38.3)	(5.2,5.4,5.4,5.5,5.5,5.6,5.8)	(93,94,94,96,96,97,99)	健康

$\underline{R_C^{P\geqslant}}(D_1^{\geqslant})=\{x_1,x_4,x_5\}$，$\overline{R_C^{P\geqslant}}(D_1^{\geqslant})=\{x_1,x_4,x_5\}$

$Bn_C(D_1^{\geqslant})=\Phi$

$\underline{R_C^{P\geqslant}}(D_2^{\leqslant})=\{x_2,x_3\}$，$\overline{R_C^{P\geqslant}}(D_2^{\leqslant})=\{x_2,x_3\}$

$Bn_C(D_2^{\leqslant})=\Phi$

得到的决策规则如下：

1. if $((|E(X^{a_1})-37|<0.1)\vee((|E(X^{a_1})-37|=0.1)\wedge(D(X^{a_1})<0.05))\wedge((|E(X^{a_2})-5|=0)\wedge(D(X^{a_2})<0.03))\wedge((|E(X^{a_3})-115|<2)\vee((|E(X^{a_3})-115|=2)\wedge(D(X^{a_3})<11.6)))$ then，$x\in D_1^{\geqslant}$，Supported by x_1。

2. if $((|E(X^{a_1})-37|<1)\vee((|E(X^{a_1})-37|=1)\wedge(D(X^{a_1})<0.1)))\wedge((|E(X^{a_2})-5|<1.4)\vee((|E(X^{a_2})-5|=1.4)\wedge(D(X^{a_2})<0.03)))\wedge((|E(X^{a_3})-115|<26)\vee((|E(X^{a_3})-115|=26)\wedge(D(X^{a_3})<1.3)))$ then $x\in D_1^{\geqslant}$，Supported by x_4。

3. if $((|E(X^{a_1})-37|<1)\vee((|E(X^{a_1})-37|=1)\wedge(D(X^{a_1})<0.03)))\wedge((|E(X^{a_2})-5|<0.5)\vee((|E(X^{a_2})-5|=0.5)\wedge(D(X^{a_2})<0.02)))\wedge((|E(X^{a_3})-115|<17)\vee((|E(X^{a_3})-115|=17)\wedge(D(X^{a_3})<3.1)))$ then $x\in D_1^{\geqslant}$，Supported by x_5。

4. if $((|E(X^{a_1})-37|<1.5)\vee((|E(X^{a_1})-37|=1.5)\wedge(D(X^{a_1})<0.3)))\wedge((|E(X^{a_2})-5|<1.5)\vee((|E(X^{a_2})-5|=1.5)\wedge(D(X^{a_2})<0.02)))\wedge((|E(X^{a_3})-115|<40)\vee((|E(X^{a_3})-115|=40)\wedge(D(X^{a_3})<6.1)))$ then $x\in D_2^{\leqslant}$，supported by x_2。

5. if $((|E(X^{a_1})-37|<1.5)\vee((|E(X^{a_1})-37|=1.5)\wedge(D(X^{a_1})<0.1)))\wedge((|E(X^{a_2})-5|<0.5)\vee((|E(X^{a_2})-5|=0.5)\wedge(D(X^{a_2})<0.06)))\wedge((|E(X^{a_3})-115|<30)\vee((|E(X^{a_3})-115|=30)\wedge(D(X^{a_3})<11.1)))$ then $x\in D_2^{\leqslant}$，Supported by x_3。

结束语　基于概率的有序信息系统本身包含了属性值区间上的概率信息，与区间值有序信息系统相比，基于概率的有序信息系统上的优势关系对全集 U 的划分具有确定性且粒度较小，从基于概率的有序信息系统中提取的决策规则具有确定性。

本文在建立基于概率的有序信息系统上的优势关系时，把属性在属性值区间上的期望作为评价属性值优劣的主要标准，具有一定的合理性。在一些实际问题中，需要综合衡量属性在其取值范围内的期望与方差对属性取值优劣的影响，这将是下一步的研究内容之一。

参 考 文 献

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

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

[3] Greco S, Matarazzo B, Slowinski R. Generalizing rough set theory through dominance-based rough set approach [C]// 10th Int Conf on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Heidelberg: Springer-Verlag, 2005: 1-11

[4] Greco S, Matarazzo B, Slowinski R. Dominance-based rough set approach to case-based reasoning [C]// 3th Int Conf on Modeling Decisions for Artificial Intelligence. Heidelberg: Springer-Verlag, 2006: 7-18

[5] Qian Y H, Liang J Y, Dang C Y. Interval Ordered information systems [J]. Computers and Mathematics with Applications, 2008, 56: 1994-2009

[6] Yang X B, Yu D J, Yang J Y, et al. Dominance-based Rough Set Approach to Incomplete Interval-valued Information System [J]. Data & Knowledge Engineering, 2009, 68(11): 1331-1347

[7] Dembczyński K, Greco S, Słowiński R. Second-order Rough Approximations in Multi-criteria Classification with Imprecise Evaluations and Assignments [C]// Proceedings of the 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC 2005). 2005: 54-63

[8] Dembczyński K, Greco S, Słowiński R. Rough Set Approach to Multiple Criteria Classification with Imprecise Evaluations and Assignments [J]. European Journal of Operational Research, 2009, 198(2): 626-636

[9] 杨青山，王国胤，张清华，等. 基于优势关系的区间值粗糙集扩充模型[J]. 山东大学学报：理学版，2010，45(9)：7-13