

区间决策粗糙集

刘 盾¹ 李天瑞² 李华雄³

(西南交通大学经济管理学院 成都 610031)¹ (西南交通大学信息科学与技术学院 成都 610031)²
(南京大学工程管理学院 南京 210093)³

摘 要 考虑到实际决策问题中损失函数的“多值性”特征,从贝叶斯理论出发,将区间值损失函数引入决策粗糙集,提出区间值决策粗糙集模型。首先,讨论了在贝叶斯期望风险最小的决策语义下,区间值决策粗糙集理论基本模型的构建过程。其次,分析了区间值决策粗糙集理论的相关数学性质和准则。最后,通过一个石油投资问题来阐明区间值决策粗糙集模型的应用过程。

关键词 决策粗糙集理论,概率粗糙集理论,贝叶斯过程,区间值

中图分类号 TP18,N945.25 **文献标识码** A

Interval-valued Decision-theoretic Rough Sets

LIU Dun¹ LI Tian-rui² LI Hua-xiong³

(Department of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China)¹

(Department of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China)²

(Department of Management and Engineering, Nanjing University, Nanjing 210093, China)³

Abstract By considering the “multiple-valued” character in practical decision procedure, the interval-valued loss functions were induced to decision-theoretic rough set theory (DTRS) based on the Bayesian decision theory. With respect to the minimum Bayesian expected risk, a model of interval-valued decision-theoretic rough set theory was built. The corresponding propositions and criteria of interval-valued decision-theoretic rough set theory were also analyzed. An example of oil investment was given to illuminate the proposed model in applications.

Keywords Decision-theoretic rough set theory, Probabilistic rough set theory, Bayesian decision procedure, Interval-valued

1 引言

波兰数学家 Pawlak 于 20 世纪 80 年代初提出的粗糙集理论(Rough Set Theory)是一种能够处理不精确、不确定信息的数学方法与工具^[10]。由于其基于数据驱动且不需要主观先验知识,以及在解决属性约简和规则推导问题上的优势,使得它一经提出,便在模式识别、数据挖掘、智能信息计算等方面得到了广泛的应用。粗糙集理论是利用两个经典集——上、下近似集来刻画一个粗糙集的。上、下近似集将整个论域分为 3 个部分:正域、边界域和负域。Yao 基于 3 个区域提出了三枝决策方法,并认为从正域里获取的正规则用来接受某事物;从负域里获取的负规则用来表示拒绝某事物;落在边界域上的规则表示需要进一步观察和研究,即延迟决策^[19-21]。三枝决策方法的提出为将粗糙集理论应用到实际决策问题提供了坚实的理论基础^[28]。

对于一系列概率粗糙集模型,如 0.5-概率粗糙集模型^[16]、决策粗糙集模型^[22,23]、变精度粗糙集模型^[4,25]、粗糙隶

属函数^[12]、贝叶斯粗糙集模型^[14,15]、博弈粗糙集模型^[2]、朴素贝叶斯粗糙集模型^[18]等,它们都使用两个参数 α 和 β 来划分 3 个区域。然而,在大多概率粗糙集模型中, α 和 β 的选取都未考虑到决策过程中的语义环境^[28]。基于此,Yao 等从贝叶斯理论出发,利用损失函数 λ 来描述决策风险,讨论了在贝叶斯期望风险最小的决策语义下 α 和 β 的自动获取过程,并提出了决策粗糙集模型^[22]。此外,Yao 还进一步指出:各种概率粗糙集模型均可以由决策粗糙集导出,它们被视为决策粗糙集的特例^[23]。决策粗糙集模型很好地描述了人类在解决实际决策问题时的思维模式,其派生出的三枝决策思想已经被考虑或有可能进一步地应用到许多模式识别问题(如医学诊断、产品过程控制、环境管理、数据集选择、智能学习、邮件分类、石油投资等领域)中^[6-8,26,28]。

然而,在决策粗糙集模型中,对损失函数 λ 的刻画是基于单值的。在实际决策过程中,风险应具有一定的随机性,损失函数应具有一定的伸缩性,因而利用具有“多值”特征的区间数来描述损失函数 λ ,更接近决策事实。基于此,本文拟将区

到稿日期:2011-08-27 返修日期:2011-12-02 本文受国家自然科学基金重大项目(71090402/G1),国家自然科学基金(61175047),教育部人文社会科学青年基金(11YJC630127)资助。

刘 盾(1983—),男,博士生,讲师,主要研究方向为粗糙决策、数据挖掘等,E-mail:newton83@163.com;李天瑞(1969—),男,教授,博士生导师,主要研究方向为智能信息处理、数据挖掘等;李华雄(1977—),男,讲师,主要研究方向为粒计算、粗糙集等。

间值损失函数引入决策粗糙集,提出一种区间值决策粗糙集模型。本文首先简要介绍相关的粗糙集基本理论和研究背景;其次,讨论区间值决策粗糙集理论基本模型的构建过程,并介绍其相关语义和性质;最后,通过一个石油投资问题来给出一个模型的具体应用。

2 粗糙理论的基本概念

本节主要介绍 Pawlak 粗糙、相关概率粗糙的基本概念和定义,以及相关区间值模型的研究背景^[10,12,18-23,25,29]。

定义 1 假设 U 是一个有限的非空子集, R 是定义在 U 上的一种等价关系。记 $apr = (U, R)$ 为一粗糙近似空间, U 在等价关系 R 下的划分记为 $U/R = \{[x]_R | x \in U\}$, 其中, $[x]$ 是包含 x 的等价类所构成的集合。对 $X \subseteq U$, 其上、下近似可定义为

$$\underline{apr}(X) = \{x \in U | [x] \subseteq X\}$$

$$\overline{apr}(X) = \{x \in U | [x] \cap X \neq \emptyset\}$$

上、下近似将论域分为正域 $POS(X)$ 、边界域 $BND(X)$ 和负域 $NEG(X)$, 其定义分别为

$$POS(X) = \underline{apr}(X)$$

$$BND(X) = \overline{apr}(X) - \underline{apr}(X) \quad (1)$$

$$NEG(X) = U - \overline{apr}(X)$$

定义 2 假设 $S = (U, A, V, f)$ 是一个信息表, $\forall x \subseteq U$, $X \subseteq U$, Pawlak 和 Skowron 建议使用粗糙隶属函数来重新定义上、下近似集^[12]。粗糙隶属函数 μ_A 定义为

$$\mu_A = \Pr(X|[x]) = |[x] \cap X| / |[x]| \quad (2)$$

式中, $|\cdot|$ 表示集合中元素的基数, $\Pr(X|[x])$ 表示分类的条件概率。

根据粗糙隶属函数的定义, 令 $0 \leq \beta < \alpha \leq 1$, 则 (α, β) -下近似、 (α, β) -上近似可定义为

$$\underline{apr}_{(\alpha, \beta)}(X) = \{x \in U | \Pr(X|[x]) \geq \alpha\} \quad (3)$$

$$\overline{apr}_{(\alpha, \beta)}(X) = \{x \in U | \Pr(X|[x]) > \beta\}$$

同样地, (α, β) -上、下近似集将论域分为 3 个部分: $POS_{(\alpha, \beta)}(X)$, $BND_{(\alpha, \beta)}(X)$ 和 $NEG_{(\alpha, \beta)}(X)$, 其定义分别为

$$POS_{(\alpha, \beta)}(X) = \{x \in U | \Pr(X|[x]) \geq \alpha\}$$

$$BND_{(\alpha, \beta)}(X) = \{x \in U | \beta < \Pr(X|[x]) < \alpha\} \quad (4)$$

$$NEG_{(\alpha, \beta)}(X) = \{x \in U | \Pr(X|[x]) \leq \beta\}$$

式中, 当 $\alpha=1, \beta=0$ 时, 式(4)转化为 Pawlak 粗糙集模型; 当 $\alpha=\beta=0.5$ 时, 式(4)转化为 0.5-概率粗糙集模型; 当 $\beta=1-\alpha$ 时, 式(4)转化为对称变精度粗糙集模型; 当 $\beta \neq 1-\alpha$ 时, 式(4)转化为非对称变精度粗糙集模型。然而, 对于式(4)中的 α 和 β 取值, 大多数概率粗糙集模型并未给出其语义解释, α 和 β 的语义问题需要进一步深入讨论。

定义 3 区间值信息系统为一个四元组 $I = \{U, A, V, f\}$ 。其中, $U = \{u_1, u_2, \dots, u_n\}$ 是非空有限论域集; $A = \{a_1, a_2, \dots, a_m\}$ 是非空有限属性集; $V = \bigcup_{a_k \in A} V_{a_k}$ 表示属性 $a_k \in A$ 的值域; $f: U \times A \rightarrow V$ 是一个映射函数, 它赋予论域中 U 的每一个对象 u_i 在属性 a_k 上的区间属性值。对于 $\forall u_i \in U, \forall a_k \in A$ 有, $f(u_i, a_k) = [v_{ik}^L, v_{ik}^U]$ 。其中, v_{ik}^L 表示值域的下界, v_{ik}^U 表示值域的上界。

3 区间值决策粗糙集模型

为了更好地解释概率粗糙集模型的语义问题, Yao 等提

出了决策粗糙集模型。决策粗糙集理论的基本模型可参考文献[22, 23]。刘盾和李华雄等人对决策粗糙集近 20 年来的发展做了详细综述^[5, 7, 26, 27]。决策粗糙集模型的基本思想来源于 Duda 等人的著作“Pattern Classification and Scene Analysis”^[1]。其基本模型为: 假设 $S = (U, A, V, f)$ 是一个信息表, $\Omega = \{w_1, w_2, \dots, w_m\}$ 为 m 个有限的状态集, $A = \{a_1, a_2, \dots, a_n\}$ 为 n 个有限的行动集。 $\Pr(w_i | x)$ 表示 x 在状态 w_i 下的条件概率。令 $\lambda(a_j | w_i)$ 为在状态 w_i 下采取行动 a_j 的损失或代价。对于某元素 x , 如果采取行动 a_j , 则其期望损失为

$$R(a_j | x) = \sum_{i=1}^m \lambda(a_j | w_i) \Pr(w_i | x) \quad (5)$$

一般地, 对于给定的描述 x , 令 $\tau(x)$ 为一个决策规则, 它是描述空间到 A 的一个函数, $\tau(x) \in A$ 。令 $\mathcal{R}$ 是在给定一个决策规则 τ 下的总体期望风险, 它可表示为

$$\mathcal{R} = \sum_{x \in U} R(\tau(x) | x) \Pr(x) \quad (6)$$

式中, $\Pr(x)$ 为 x 的先验概率, $R(\tau(x) | x)$ 为采取不同行动 x 的条件风险。按照贝叶斯决策原理, 需要选取使得总体期望风险 $\mathcal{R}$ 达到最小的决策行为作为最佳行动方案。如果有多个决策使得 $\mathcal{R}$ 达到最小, 则根据实际情况选择其中之一。

决策粗糙集模型是基于贝叶斯决策过程的, 它利用 2 个状态集和 3 个行动集来描述决策过程。状态集 $\Omega = \{X, \neg X\}$ 中的元素分别表示某事件属于 X 和不属于 X , 行动集 $A = \{a_P, a_B, a_N\}$ 中的元素分别表示接受某事件、延迟决策和拒绝某事件 3 种行动。考虑到采取不同行动会产生不同的损失, 记 $\lambda_{PP}, \lambda_{BP}, \lambda_{NP}$ 分别表示当 x 属于 X 时, 采取行动 a_P, a_B 和 a_N 下的损失。同样地, 记 $\lambda_{PN}, \lambda_{BN}, \lambda_{NN}$ 分别表示当 x 不属于 X 时, 采取行动 a_P, a_B 和 a_N 下的损失。在这里, 考虑到损失函数具有一定的波动性, 因此引入区间值损失函数来代替单值损失函数。令 $\lambda_{PP} = [\lambda_{PP}^L, \lambda_{PP}^U]$, $\lambda_{BP} = [\lambda_{BP}^L, \lambda_{BP}^U]$, $\lambda_{NP} = [\lambda_{NP}^L, \lambda_{NP}^U]$; $\lambda_{PN} = [\lambda_{PN}^L, \lambda_{PN}^U]$, $\lambda_{BN} = [\lambda_{BN}^L, \lambda_{BN}^U]$, $\lambda_{NN} = [\lambda_{NN}^L, \lambda_{NN}^U]$ 。其中, λ^L 代表各损失函数的下界, 它表示乐观主义者对风险的衡量, 因为乐观主义者追求一种高风险、高收益的决策模式, 所以往往对损失给予较小的数值; λ^U 代表各损失函数的上界, 它可以表示悲观主义者对风险的衡量, 因为悲观主义者趋于保守, 害怕犯错, 所以往往对损失给予较大的数值。基于上述分析, 根据式(1), 乐观主义决策准则和悲观主义决策准则采取 a_P, a_B 和 a_N 3 种行动下的期望损失可分别由式(7)和式(8)给出。

$$\begin{aligned} R(a_P | [x]) &= \lambda_{PP}^L \Pr(X|[x]) + \lambda_{PN}^L \Pr(\neg X|[x]) \\ R(a_B | [x]) &= \lambda_{BP}^L \Pr(X|[x]) + \lambda_{BN}^L \Pr(\neg X|[x]) \\ R(a_N | [x]) &= \lambda_{NP}^L \Pr(X|[x]) + \lambda_{NN}^L \Pr(\neg X|[x]) \end{aligned} \quad (7)$$

$$\begin{aligned} R(a_P | [x]) &= \lambda_{PP}^U \Pr(X|[x]) + \lambda_{PN}^U \Pr(\neg X|[x]) \\ R(a_B | [x]) &= \lambda_{BP}^U \Pr(X|[x]) + \lambda_{BN}^U \Pr(\neg X|[x]) \\ R(a_N | [x]) &= \lambda_{NP}^U \Pr(X|[x]) + \lambda_{NN}^U \Pr(\neg X|[x]) \end{aligned} \quad (8)$$

根据贝叶斯决策准则, 需要选择期望损失最小的行动集作为最佳行动方案, 于是可得到如下 3 条决策规则:

(P) 若 $R(a_P | [x]) \leq R(a_B | [x])$ 和 $R(a_P | [x]) \leq R(a_N | [x])$ 同时成立, 则 $x \in POS(X)$;

(B) 若 $R(a_B | [x]) \leq R(a_P | [x])$ 和 $R(a_B | [x]) \leq R(a_N | [x])$ 同时成立, 则 $x \in BND(X)$;

(N) 若 $R(a_N | [x]) \leq R(a_P | [x])$ 和 $R(a_N | [x]) \leq R(a_B | [x])$ 同时成立, 则 $x \in NEG(X)$ 。

由于 $\Pr(X|[x]) + \Pr(\neg X|[x]) = 1$, 因此上述规则只与概率 $\Pr(X|[x])$ 和相关的损失函数 λ 有关。此外, 接受正确事物的损失不大于延迟接受正确事物的损失, 且这两者都小于拒绝正确事物的损失; 同样地, 拒绝错误事物的损失不大于延迟拒绝错误事物的损失, 且这两者都小于接受错误事物的损失, 即 $0 \leq \lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}$, $0 \leq \lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}$ 。对于乐观主义者, 一个合理的假设为 $0 \leq \lambda_{BP}^L \leq \lambda_{BP}^U < \lambda_{NP}^L$, $0 \leq \lambda_{NN}^L \leq \lambda_{BN}^L < \lambda_{PN}^L$; 对于悲观主义者, 一个合理的假设为 $0 \leq \lambda_{BP}^U \leq \lambda_{BP}^L < \lambda_{NP}^U$, $0 \leq \lambda_{NN}^U \leq \lambda_{BN}^U < \lambda_{PN}^U$ 。据此, 可分别计算乐观主义者和悲观主义者 3 个阈值参数的值, 如表 1 所列。

表 1 乐观主义者和悲观主义者 3 个阈值参数取值表

参数	乐观主义者	悲观主义者
α	$\alpha_1 = \frac{(\lambda_{PN}^L - \lambda_{BN}^L)}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)}$	$\alpha_2 = \frac{(\lambda_{PN}^U - \lambda_{BN}^U)}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)}$
β	$\beta_1 = \frac{(\lambda_{BN}^L - \lambda_{NN}^L)}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}$	$\beta_2 = \frac{(\lambda_{BN}^U - \lambda_{NN}^U)}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)}$
γ	$\gamma_1 = \frac{(\lambda_{PN}^L - \lambda_{NN}^L)}{(\lambda_{PN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}$	$\gamma_2 = \frac{(\lambda_{PN}^U - \lambda_{NN}^U)}{(\lambda_{PN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)}$

由规则 (B) 可知, $\alpha_1 > \beta_1$ 且 $\alpha_2 > \beta_2$, 则 $\frac{\lambda_{BP}^L - \lambda_{PP}^L}{\lambda_{PN}^L - \lambda_{BN}^L} < \frac{\lambda_{NP}^L - \lambda_{BP}^L}{\lambda_{BN}^L - \lambda_{NN}^L}$, $\frac{\lambda_{BP}^U - \lambda_{PP}^U}{\lambda_{PN}^U - \lambda_{BN}^U} < \frac{\lambda_{NP}^U - \lambda_{BP}^U}{\lambda_{BN}^U - \lambda_{NN}^U}$ 。此外, 由不等式 $\frac{b}{a} > \frac{d}{c} \Rightarrow \frac{b}{a} > \frac{b+d}{a+c} > \frac{d}{c}$ ($a, b, c, d > 0$), 有 $\frac{\lambda_{BP}^L - \lambda_{PP}^L}{\lambda_{PN}^L - \lambda_{BN}^L} < \frac{\lambda_{NP}^L - \lambda_{BP}^L}{\lambda_{BN}^L - \lambda_{NN}^L} < \frac{\lambda_{NP}^U - \lambda_{BP}^U}{\lambda_{BN}^U - \lambda_{NN}^U}$ 。因而, $0 \leq \beta_1 < \gamma_1 < \alpha_1 \leq 1$ 且 $0 \leq \beta_2 < \gamma_2 < \alpha_2 \leq 1$ 。这样, 决策规则 (P), (B), (N) 可重写如下。

对于乐观主义者:

(P1) 如果 $\Pr(X|[x]) \geq \alpha_1$, 则 $x \in \text{POS}(X)$;

(B1) 如果 $\beta_1 < \Pr(X|[x]) < \alpha_1$, 则 $x \in \text{BND}(X)$;

(N1) 如果 $\Pr(X|[x]) \leq \beta_1$, 则 $x \in \text{NEG}(X)$ 。

对于悲观主义者:

(P2) 如果 $\Pr(X|[x]) \geq \alpha_2$, 则 $x \in \text{POS}(X)$;

(B2) 如果 $\beta_2 < \Pr(X|[x]) < \alpha_2$, 则 $x \in \text{BND}(X)$;

(N2) 如果 $\Pr(X|[x]) \leq \beta_2$, 则 $x \in \text{NEG}(X)$ 。

当 $\lambda_{BP}^L = \lambda_{BP}^U$, $\lambda_{BN}^L = \lambda_{BN}^U$, $\lambda_{NP}^L = \lambda_{NP}^U$, $\lambda_{PN}^L = \lambda_{PN}^U$, $\lambda_{BN}^L = \lambda_{BN}^U$, $\lambda_{NN}^L = \lambda_{NN}^U$ 时, 决策规则 (P1) - (N1) 与 (P2) - (N2) 等价, 区间值决策粗糙集退化为 Yao 提出的决策粗糙集。

在实际决策问题中, 人们往往采用群决策的方式来求解问题。由于不同专家具有不同的偏好, 因此最终决策结果往往是集结各个专家的意见。在区间值决策粗糙集模型中, 必须考虑方案集结后带来的影响, 这在阈值函数 α 和 β 的设置上表现为 α 和 β 也应为区间值。另外, 考虑到模型的实际语义性, 其需要满足:

$$\begin{aligned} 0 \leq \lambda_{BP}^L \leq \lambda_{BP}^U < \lambda_{NP}^L < \lambda_{NP}^U \\ 0 \leq \lambda_{NN}^L \leq \lambda_{NN}^U < \lambda_{BN}^L < \lambda_{BN}^U < \lambda_{PN}^L < \lambda_{PN}^U \end{aligned} \quad (9)$$

式 (9) 表示接受正确事物的最大损失小于延迟接受正确事物的最小损失, 延迟接受正确事物的最大损失小于接受错误事物的最小损失, 拒绝错误事物的最大损失小于延迟拒绝错误事物的最小损失, 延迟拒绝错误事物的最大损失小于接受错误事物的最小损失。

下面, 给出两个定理来确定 α 和 β 的区间上、下确界。

定理 1 在满足式 (9) 给定的条件下, $\alpha \in$

$$\left[\frac{(\lambda_{PN}^L - \lambda_{BN}^L)}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)}, \min \left(\frac{(\lambda_{PN}^U - \lambda_{BN}^U)}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)}, 1 \right) \right].$$

证明: 由于 $\lambda_{PN} - \lambda_{BN} \geq 0$, $\lambda_{BP} - \lambda_{PP} \geq 0$, 因此 $\alpha =$

$$\frac{(\lambda_{PN}^L - \lambda_{BN}^L)}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)}。由于 \lambda_{PN}^L - \lambda_{BN}^L \leq \lambda_{PN} - \lambda_{BN} \leq \lambda_{PN}^U - \lambda_{BN}^U, \lambda_{BP}^L - \lambda_{PP}^L \leq \lambda_{BP} - \lambda_{PP} \leq \lambda_{BP}^U - \lambda_{PP}^U, 则 (\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L) \leq (\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP}) \leq (\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U), 因此 \frac{(\lambda_{PN}^L - \lambda_{BN}^L)}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)} \leq \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} \leq \frac{(\lambda_{PN}^U - \lambda_{BN}^U)}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)}。此外, 由 \alpha \in [0, 1] 有, \frac{(\lambda_{PN}^L - \lambda_{BN}^L)}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)}, \frac{(\lambda_{PN}^U - \lambda_{BN}^U)}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)} \in [0, 1], 即 \alpha \in \left[\frac{(\lambda_{PN}^L - \lambda_{BN}^L)}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)}, \min \left(\frac{(\lambda_{PN}^U - \lambda_{BN}^U)}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)}, 1 \right) \right]。$$

定理 2 在满足式 (9) 给定的条件下, $\beta \in$

$$\left[\frac{(\lambda_{BN}^L - \lambda_{NN}^L)}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}, \min \left(\frac{(\lambda_{BN}^U - \lambda_{NN}^U)}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)}, 1 \right) \right].$$

证明: 由于 $\lambda_{NP} - \lambda_{BP} \geq 0$, $\lambda_{BN} - \lambda_{NN} \geq 0$, 因此 $\beta =$

$$\frac{(\lambda_{BN}^L - \lambda_{NN}^L)}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}。由于 \lambda_{NP}^L - \lambda_{BP}^L \leq \lambda_{NP} - \lambda_{BP} \leq \lambda_{NP}^U - \lambda_{BP}^U, \lambda_{BN}^L - \lambda_{NN}^L \leq \lambda_{BN} - \lambda_{NN} \leq \lambda_{BN}^U - \lambda_{NN}^U, 则 (\lambda_{NP}^L - \lambda_{BP}^L) + (\lambda_{BN}^L - \lambda_{NN}^L) \leq (\lambda_{NP} - \lambda_{BP}) + (\lambda_{BN} - \lambda_{NN}) \leq (\lambda_{NP}^U - \lambda_{BP}^U) + (\lambda_{BN}^U - \lambda_{NN}^U), 因此 \frac{(\lambda_{BN}^L - \lambda_{NN}^L)}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)} \leq \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} \leq \frac{(\lambda_{BN}^U - \lambda_{NN}^U)}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)}。此外, 由 \beta \in [0, 1] 有, \frac{(\lambda_{BN}^L - \lambda_{NN}^L)}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}, \frac{(\lambda_{BN}^U - \lambda_{NN}^U)}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)} \in [0, 1]。即$$

$$\beta \in \left[\frac{(\lambda_{BN}^L - \lambda_{NN}^L)}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}, \right.$$

$$\left. \min \left(\frac{(\lambda_{BN}^U - \lambda_{NN}^U)}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)}, 1 \right) \right]$$

明显地, 有

$$\alpha_1, \alpha_2 \in [\alpha_L, \alpha_U] = \left[\frac{(\lambda_{PN}^L - \lambda_{BN}^L)}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)}, \right.$$

$$\left. \min \left(\frac{(\lambda_{PN}^U - \lambda_{BN}^U)}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)}, 1 \right) \right]$$

$$\beta_1, \beta_2 \in [\beta_L, \beta_U] = \left[\frac{(\lambda_{BN}^L - \lambda_{NN}^L)}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}, \right.$$

$$\left. \min \left(\frac{(\lambda_{BN}^U - \lambda_{NN}^U)}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)}, 1 \right) \right] \quad (10)$$

式 (10) 说明, 单纯的乐观决策准则和悲观决策准则只是区间值决策粗糙集的一个特例。

4 案例分析

下面利用一个石油投资的例子来阐明区间值决策粗糙集的应用过程。

在一个实际的投资问题中, 是否执行一个项目, 主要看它的期望开采结果。一个待开发的油田的结果可能有两种情况: 发现石油和未发现石油。对于一块油田而言, 决策粗糙集里的损失函数主要来源于其开采风险。Xie 等人总结了石油开采过程中的风险主要分为 5 大类: 开采能力风险、地质和技

术风险、财政风险、环境风险和政策风险^[17]。根据这 5 类风险, Yusgiantoro 和 Hsiao 指出反映石油储量多少, 较为重要的指标之一是矿床碳氢化合物的品质和含量。而所有的油田可以分为 3 个类型: 具有高含量且较少被开采的矿床需要被开采; 具有中等含量且具有少量油储存的矿床需要谨慎开采; 具有贫乏的碳氢化合物储藏矿床不值得开采^[24]。Macmillan 在其博士论文中也认为石油的投资决策可分为 3 个部分: 钻井、不钻井和延迟决策^[9]。

根据决策粗糙集理论的思想, 石油投资问题可用 2 个状态集和 3 个行动集来描述决策过程。对于某块油田 o_i , 状态集 $\Omega=\{O, \neg O\}$ 包含某油田有油和无油两种状态; 行动集 $A=\{a_P, a_B, a_N\}$ 包含投资、延迟决策和不投资 3 种行动。各个专家通过对油田 5 类风险的仔细评估, 可以分别给出油田的损失函数。由于不同专家有不同的偏好, 对 5 类风险的权重偏

好也不尽相同, 因此最终各损失函数是由区间值构成的。 $[\lambda_{BP}^L, \lambda_{BP}^U], [\lambda_{BP}^L, \lambda_{BP}^U], [\lambda_{NP}^L, \lambda_{NP}^U]$ 分别表示油田有油时采取行动 a_P, a_B 和 a_N 下的损失区间; $[\lambda_{BN}^L, \lambda_{BN}^U], [\lambda_{BN}^L, \lambda_{BN}^U], [\lambda_{NN}^L, \lambda_{NN}^U]$ 分别表示油田无油时采取行动 a_P, a_B 和 a_N 下的损失区间。表 2 给出了 6 块油田的损失情况, 其数值是综合相关专家意见得到的结果。通过表 1 和式(10)计算可得 6 块油田相应的 α 和 β 值, 其计算结果如表 3 所列。

表 2 6 块油田的综合评估损失情况(u 为单位损失)

O	λ_{BP}^L	λ_{BP}^U	λ_{BP}^L	λ_{BP}^U	λ_{NP}^L	λ_{NP}^U	λ_{NP}^L	λ_{NP}^U	λ_{BN}^L	λ_{BN}^U	λ_{BN}^L	λ_{BN}^U	λ_{PN}^L	λ_{PN}^U
o_1	1u	1.5u	2u	2.5u	3.5u	6.5u	1u	1.5u	1.95u	2u	5u	5.25u		
o_2	1u	2u	3u	4u	5u	6u	0.5u	1u	1.8u	2u	3.5u	4u		
o_3	2u	4u	5u	7u	8u	9u	0	0.5u	1u	1.2u	3u	3.5u		
o_4	3u	4u	5u	6.5u	8u	10u	1u	1.5u	2.5u	3u	4u	4.2u		
o_5	0.5u	1u	1.5u	2u	3u	6u	1u	1u	3.5u	4u	5.5u	6u		
o_6	4u	5u	6u	6.5u	8u	9u	2u	3u	4.5u	6u	9u	10u		

表 3 6 块油田的相关阈值计算结果

油田	乐观决策准则			悲观决策准则			阈值下界		阈值上界	
	α_1	β_1	γ_1	α_2	β_2	γ_2	α_L	β_L	α_U	β_U
o_1	0.7531	0.3878	0.6154	0.7647	0.1111	0.4286	0.6250	0.0818	0.9429	0.6897
o_2	0.4595	0.3939	0.4286	0.5000	0.3333	0.4286	0.2885	0.1778	0.8800	0.8333
o_3	0.4000	0.2500	0.3333	0.4340	0.2593	0.3750	0.2400	0.0962	0.8929	0.8000
o_4	0.4286	0.3333	0.3750	0.3243	0.3000	0.3103	0.1923	0.1429	0.8500	0.8000
o_5	0.6667	0.6250	0.6429	0.6667	0.4286	0.5000	0.3750	0.3333	1.0000	0.8571
o_6	0.6923	0.5556	0.6364	0.7273	0.5455	0.6364	0.3750	0.2143	1.0000	1.0000

表 3 清晰地给出了分别在乐观决策准则、悲观决策准则下, α, β 和 γ 的取值情况, 并给出了 α 和 β 的阈值下界和阈值上界。可以看到, $\alpha_1, \alpha_2 \in [\alpha_L, \alpha_U], \beta_1, \beta_2 \in [\beta_L, \beta_U]$ 。通过比较本案例 6 块油田可发现, α_2 较 α_1 的取值普遍要大, β_2 较 β_1 的取值普遍要小, 且区间长度 $[\beta_2, \alpha_2]$ 要普遍大于 $[\beta_1, \alpha_1]$ 。这说明悲观主义者厌恶风险, 它通过较大的 α 值和较小的 β 值来避免犯错的概率; 而乐观主义者偏好风险, 它通过较小的 α 值和较大的 β 值来博取更高的收益, 但这样会增大犯错的概率。在实际问题中, 需要比较条件概率 $\Pr(O|o_i)$ 与 α 和 β 值的大小来做出最终决策。 $\Pr(O|o_i)$ 的计算需要先验信息, 其具体计算过程可参见文献[18]。在这里, 为方便描述, 表 4 给出了条件概率 $\Pr(O|o_i)$ 从 0 到 1、步长为 0.1 时, 根据决策准则(P)–(N), 分别考虑 $(\alpha_1, \beta_1), (\alpha_2, \beta_2), (\alpha_L, \beta_L), (\alpha_U, \beta_U)$ 4 种情况下, 值得投资油田(进入决策正域)的情况。

表 4 条件概率变化下各油田是否投资情况

$\Pr(O o_i)$	(α_1, β_1)	(α_2, β_2)	(α_L, β_L)	(α_U, β_U)
0.1	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
0.2	$\emptyset$	$\emptyset$	$\{o_4\}$	$\emptyset$
0.3	$\emptyset$	$\emptyset$	$\{o_2, o_3, o_4\}$	$\emptyset$
0.4	$\{o_3\}$	$\{o_4\}$	$\{o_2, o_3, o_4, o_5, o_6\}$	$\emptyset$
0.5	$\{o_2, o_3, o_4\}$	$\{o_2, o_3, o_4\}$	$\{o_2, o_3, o_4, o_5, o_6\}$	$\emptyset$
0.6	$\{o_2, o_3, o_4\}$	$\{o_2, o_3, o_4\}$	$\{o_2, o_3, o_4, o_5, o_6\}$	$\emptyset$
0.7	$\{o_2, o_3, o_4, o_5, o_6\}$	$\{o_2, o_3, o_4, o_5\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\emptyset$
0.8	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\emptyset$
0.9	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\{o_2, o_3, o_4\}$
1.0	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6\}$

从表 4 可以得到, 在 4 种特殊情况下值得投资油田在不同条件概率情况下的变动情况。在实际应用中, 相关阈值参数的取值应集结各专家的意见得出。一旦相关损失函数确定, 则可视作区间值决策粗糙集模型的一个特殊情形, 而相应的决策区域也随之确定。此外, 对于不同风险偏好者对阈值参数影响的仿真实验可参见文献[4], 这能够帮助决策者对实际问题进行灵敏度、鲁棒性等技术分析。综上所述, 所提方法能在实际决策中快速、直观地帮助决策者做出合理的选择, 具有一定的应用价值。

结束语 本文从决策粗糙集出发, 利用区间值来设定损失函数, 首先提出一种区间值决策粗糙集模型来处理损失函数是“多值”特征的问题。损失函数的上、下界分别对应了相关悲观主义和乐观主义的决策准则, 区间决策粗糙集导出的相关阈值参数也应为区间值。其次, 从数学上, 决策粗糙集模型可看作区间决策粗糙集模型的一个特例, 并给出阈值参数 α 和 β 的上、下确界。最后通过一个石油投资案例说明了模型的应用过程。然而, 本文对区间值决策粗糙集模型的研究尚在初级阶段, 其相关数学性质的进一步挖掘, 以及在群决策环境下如何利用区间值决策粗糙集来解决实际工程管理问题的关键技术, 将在后续研究中逐步展开。

参 考 文 献

[1] Duda R O, Hart P E. Pattern Classification and Scene Analysis [M]. New York: Wiley Press, 1973

[2] Herbert J P, Yao J T. Criteria for choosing a rough set model [J]. Computer and Mathematics with Application, 2009, 57: 908-918

[3] Katzberg J D, Ziarko W. Variable precision rough sets with asymmetric bounds [M]// Ziarko W. ed. Rough Sets, Fuzzy Sets and Knowledge Discovery. London: Springer, 1994: 167-177

更新算法。算法将对原决策表的核属性更新转换为对简化决策表的核属性更新,并且只需扫描一次变化后的决策表便可对核属性进行快速更新,有效地提高了核属性的更新效率。理论分析和实验结果表明,本文算法在效率和可操作性上优于已有的同类算法。求解核属性是属性约简的重要步骤,如何利用更新的核属性对决策表进行动态属性约简,将是下一步的研究重点。

参 考 文 献

- [1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Science, 1982, 11 (5): 341-356
- [2] Pawlak Z, Skowron A. Rudiments of rough sets[J]. Information Sciences, 2007, 117(1): 3-37
- [3] Hu X H, Cercone N. Learning in relational databases: A rough set approach[J]. Computational Intelligence, 1995, 11(2): 323-337
- [4] Skowron A, Rauszer C. The discernibility matrices and functions in information systems[Z]. Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht: Kluwer Academic Publisher, 1991: 331-362
- [5] 叶东毅, 陈昭炯. 一个新的差别矩阵及其求核方法[J]. 电子学报, 2002, 30(7): 1086-1088
- [6] 王国胤. 决策表核属性的计算方法[J]. 计算机学报, 2003, 26(5): 611-615
- [7] 杨明, 孙志挥. 改进的差别矩阵及其求核方法[J]. 复旦学报: 自然科学版, 2004, 43(5): 865-868
- [8] 冯林, 李天瑞. 基于 SQL 的属性核与约简高效计算方法[J]. 计算机科学, 2010, 37(1): 236-238
- [9] 徐章艳, 刘作鹏, 杨炳儒, 等. 一个复杂度为 $\max\{O(|C||U|), O(|C|^2|U/C|)\}$ 的快速属性约简算法[J]. 计算机学报, 2006, 29(3): 391-399
- [10] 钱文彬, 徐章艳, 杨炳儒, 等. 一种基于决策表的核增量式高效更新算法[J]. 小型微型计算机系统, 2010, 31(4): 739-743
- [11] 杨明. 一种基于改进差别矩阵的核增量式更新算法[J]. 计算机学报, 2006, 29(3): 407-413
- [12] 冯少荣, 张东站. 一种高效的增量式属性约简算法[J]. 控制与决策, 2011, 26(4): 495-500
- [13] 杨明, 杨萍. 基于差别矩阵的属性核快速更新算法[J]. 控制与决策, 2007, 22(4): 453-456
- [14] 冯少荣, 赖桃桃, 张东站. 一种改进的核增量式更新算法[J]. 计算机工程与应用, 2010, 46(20): 96-98
- [15] (上接第 181 页)
- [4] Li H X, Zhou X Z. Risk decision making based on decision-theoretic rough set: a three-way view decision model [J]. International Journal of Computational Intelligence Systems, 2011, 4: 1-11
- [5] Liu D, Li T R, Ruan D. Probabilistic model criteria with decision-theoretic rough sets [J]. Information Sciences, 2011, 181(17): 3709-3722
- [6] Liu D, Li T R, Hu P, et al. Multiple-category classification with decision-theoretic rough sets [J]. LNAI, 6401, 2010: 703-710
- [7] Liu D, Li H X, Zhou X Z. Two decades' research on decision-theoretic rough sets [C]//Proceedings of 9th IEEE International Conference on Cognitive Informatics, 2010: 968-973
- [8] Liu D, Yao Y Y, Li T R. Three-way investment decisions with decision-theoretic rough sets [J]. International Journal of Computational Intelligence Systems, 2011, 4: 66-74
- [9] Macmillan F. Risk, Uncertainty and Investment Decision-making in the Upstream Oil and Gas Industry [D]. UK: University of Aberdeen, 2000
- [10] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356
- [11] Pawlak Z, Wong S K M, Ziarko W. Rough sets: probabilistic versus deterministic approach [J]. International Journal of Man-Machine Studies, 1988, 29: 81-95
- [12] Pawlak Z, Skowron A. Rough membership functions [M]. Advances in the Dempster-Shafer Theory of Evidence. New York: John Wiley and Sons, 1994: 251-271
- [13] Skowron A, Stepaniuk J. Tolerance approximation spaces [J]. Fundamenta Informaticae, 1996, 27: 245-253
- [14] Slezak D. Rough sets and bayes factor [J]. LNCS Transactions on Rough Sets III, 2005: 202-229
- [15] Slezak D, Ziarko W. The investigation of the bayesian rough set model [J]. International Journal of Approximate Reasoning, 2005, 40: 81-91
- [16] Wong S K M, Ziarko W. Comparison of the probabilistic approximate classification and the fuzzy set model [J]. Fuzzy Sets and Systems, 1987, 21: 357-362
- [17] Xie G, Yue W, Wang S Y, et al. Dynamic risk management in petroleum project investment based on a variable precision rough set model [J]. Technological Forecasting & Social Change, 2010, 77: 891-901
- [18] Yao Y Y, Zhou B. Naive Bayesian Rough Sets [J]. LNAI, 2010, 6401: 719-726
- [19] Yao Y Y. Three-way decision: an interpretation of rules in rough set theory [J]. LNAI, 2009, 5589: 642-649
- [20] Yao Y Y. Three-way decisions with probabilistic rough sets [J]. Information Sciences, 2010, 180: 341-353
- [21] Yao Y Y. The superiority of three-way decision in probabilistic rough set models [J]. Information Sciences, 2011, 181: 1080-1096
- [22] Yao Y Y, Wong S K M. A decision theoretic framework for approximating concepts [J]. International Journal of Man-machine Studies, 1992, 37: 793-809
- [23] Yao Y Y. Decision-theoretic rough set models [J]. LNAI, 2007, 4481: 1-12
- [24] Yusgiantoro P, Hsiao F S T. Production-sharing contracts and decision making in oil production [J]. Energy Economics, 1993, 10: 245-256
- [25] Ziarko W. Variable precision rough set model [J]. Journal of Computer and System Sciences, 1993, 46: 39-59
- [26] 李华雄, 刘盾, 周献中. 决策粗糙集模型研究综述[J]. 重庆邮电大学学报: 自然科学版, 2010, 22(5): 624-630
- [27] 李华雄, 周献中, 李天瑞, 等. 决策粗糙集理论及其研究进展 [M]. 北京: 科学出版社, 2011
- [28] 刘盾, 姚一豫, 李天瑞. 三枝决策粗糙集[J]. 计算机科学, 2011, 38(1): 246-250
- [29] 张楠, 苗夺谦, 岳晓冬. 区间值信息系统的知识约简[J]. 计算机研究与发展, 2010, 47(8): 1362-1371