

基于模糊数风险最小化的拓展决策粗糙集模型

袁锦仪 叶东毅

(福州大学数学与计算机科学学院 福州 350108)

摘要 决策粗糙集模型中损失函数一般是基于单值的。考虑到实际决策问题中损失函数的不确定特征,为了处理一般的情形,引入模糊数来表示损失函数。从模糊数学的角度出发,通过一系列模糊运算得出决策阈值 α, β 的模糊分布,并据此给出决策规则。同时,对比区间决策粗糙集模型,给出获得更紧凑的阈值 α, β 上、下确界的方法。最后,通过一个石油投资的例子来阐明该模型的应用过程。

关键词 决策粗糙集理论, 概率粗糙集理论, 贝叶斯过程, 模糊数

中图分类号 TP18, N945.25

文献标识码 A

Extended Decision-theoretic Rough Set Models Based on Fuzzy Minimum Cost

ZHONG Jin-yi YE Dong-yi

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China)

Abstract The loss function in decision-theoretic rough set theory is generally a single-valued function. Considering the “uncertainty” character in practical decision-making, we introduced a fuzzy-number based loss function to deal with a more general decision-making problem under uncertainty. The fuzzy distributions of the decision thresholds α, β were calculated through series of fuzzy operations, and the corresponding decision rules were given. A method for getting more compact supremum and infimum of the thresholds α, β was also presented. 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, Fuzzy number

1 引言

波兰学者 Pawlak 于 20 世纪 80 年代提出的粗糙集理论^[1,2]是一种可以处理不精确、不确定信息系统的数学方法与工具,该理论方法由于具有知识约简功能和不需要主观先验知识等优点,在数据挖掘、模式分类、智能信息处理等领域得到广泛的应用。

粗糙集理论引入上近似、下近似等概念刻画知识的不确定性和模糊性。上、下近似将整个论域分为正区域、边界域和负区域 3 个部分。基于这 3 个区域, Yao 等人在文献^[3-5]中提出了三枝决策规则,探讨了粗糙集的一个新的语义。从正区域里获取的正规则用来接受某事物(acceptance);从负区域里获取的负规则用来表示拒绝某事物(rejection);落在边界域上的规则表示需要进一步观察,即延迟决策(deferment)。这种将论域分为 3 部分的决策方式,很好地描述了人类在解决实际决策问题时的思维模式。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。

一系列概率粗糙集模型,如 0.5-概率粗糙集模型^[6]、决策粗糙集模型^[7,8]、变精度粗糙集模型^[9]、贝叶斯粗糙集模型^[10,11]等,都使用两个参数 α, β 来划分 3 个区域。然而在大多数概率粗糙集模型中, α, β 是作为原始信息由用户给定的,

未考虑到实际决策过程中的语义环境。基于此, Yao 等提出决策粗糙集模型^[8],该模型从贝叶斯理论出发,通过最小化贝叶斯风险值来计算 α, β ,为 α, β 的计算提供了方法和充分的理论依据。

然而,在决策粗糙集模型中,损失函数仍然需要由用户给定,而在此之前的研究中,对损失函数的刻画一般是基于单值的。在实际决策过程中,考虑到人的判断模糊性,单值损失函数存在很大的误差性,损失函数应当具有一定的伸缩性。刘盾、李天瑞、李华雄提出区间决策粗糙集^[12],讨论了用区间值来刻画损失函数,这仍具有一定的误差性。特别地,在群决策中,不同的专家具有不同的偏好,往往给出多个区间值。此外,区间值只是模糊数的一种特例。为了处理更一般的情形,本文对此做进一步的拓展,将损失函数当作一个模糊数,从模糊数学的角度出发,通过一系列模糊运算得出阈值 α, β 的模糊分布,据此给出决策规则。

本文还给出获得更为紧凑的阈值 α, β 上、下确界的方法。

2 粗糙集理论的基本概念

为了便于描述本文的工作,首先简要介绍 Pawlak 粗糙集、相关概率粗糙集的基本概念和定义以及贝叶斯决策过程和决策粗糙集模型。

到稿日期:2013-05-15 返修日期:2013-08-16 本文受国家自然科学基金项目(71231003),福建省自然科学基金项目(2012J01262)资助。

袁锦仪(1989—),男,硕士生,主要研究方向为计算智能,E-mail:897912820@qq.com;叶东毅(1964—),男,教授,博士生导师,主要研究方向为计算智能、数据挖掘等。

一个信息表可表示为四元组 $S = (U, At, \{V_a | a \in At\}, \{I_a | a \in At\})$ 。其中, U 是论域, At 是属性, V_a 是属性值, I_a 是论域到属性值的一个映射。

粗糙集模型中,上、下近似是基本的运算子。在经典粗糙集模型中,对于论域的任意一个非空子集 $X \subseteq U$,对象 x 的等价类表示为 $[x]$,其上、下近似可定义为:

$$\begin{aligned} \underline{apr}(X) &= \{x \in U | [x] \subseteq X\} \\ \overline{apr}(X) &= \{x \in U | [x] \cap X \neq \emptyset\} \end{aligned} \quad (1)$$

基于 X 集合的上、下近似,可以把论域划分为 3 个不相交的集合:正区域 $POS(X)$ 、边界域 $BND(X)$ 和负区域 $NEG(X)$,其定义分别为:

$$\begin{aligned} POS(X) &= \underline{apr}(X) \\ BND(X) &= \overline{apr}(X) - \underline{apr}(X) \\ NEG(X) &= U - \overline{apr}(X) \end{aligned} \quad (2)$$

概率粗糙集提供了容错性,给定阈值 (α, β) , $0 \leq \beta < \alpha \leq 1$ 。 $P(X|[x]) = |[x] \cap X| / |[x]|$ 表示规则的正确度或置信度。其上、下近似定义为:

$$\begin{aligned} \underline{apr}_{(\alpha, \beta)}(X) &= \{x \in U | P(X|[x]) \geq \alpha\} \\ \overline{apr}_{(\alpha, \beta)}(X) &= \{x \in U | P(X|[x]) > \beta\} \end{aligned} \quad (3)$$

同样地, (α, β) -上、下近似将论域分为正区域 $POS_{(\alpha, \beta)}(X)$ 、边界域 $BND_{(\alpha, \beta)}(X)$ 和负区域 $NEG_{(\alpha, \beta)}(X)$ 3 个部分,其定义分别为:

$$\begin{aligned} POS_{(\alpha, \beta)}(X) &= \underline{apr}_{(\alpha, \beta)}(X) \\ BND_{(\alpha, \beta)}(X) &= \overline{apr}_{(\alpha, \beta)}(X) - \underline{apr}_{(\alpha, \beta)}(X) \\ NEG_{(\alpha, \beta)}(X) &= U - \overline{apr}_{(\alpha, \beta)}(X) \end{aligned} \quad (4)$$

式中,当 $\alpha=1, \beta=0$ 时,式(4)转化为 Pawlak 粗糙集模型。然而,对于式(4)中 α 和 β 取值,大多数概率粗糙集模型并未给出其语义解释。

Yao 等提出决策粗糙集模型^[8],其通过引入贝叶斯决策过程,为 α, β 的计算提供了方法和理论依据。贝叶斯决策的基本模型为:假定 $\Omega = \{\omega_1, \omega_2, \dots, \omega_s\}$ 表示为 s 个有限的状态集, $A = \{a_1, a_2, \dots, a_m\}$ 表示为 m 个有限的行动集。 $P(w_j | x)$ 为 x 属于状态 w_j 的条件概率。 $\lambda(a_i | w_j)$ 为在状态 w_j 下采取行动 a_i 的损失。对于一个对象 x ,采取行动 a_i 的期望损失为:

$$R(a_i | x) = \sum_{j=1}^s \lambda(a_i | w_j) P(w_j | x) \quad (5)$$

贝叶斯决策过程的目标是使决策风险最小。决策粗糙集模型利用 2 个状态集和 3 个行动集来描述决策过程。状态集 $\Omega = \{X, X^c\}$ 中的元素分别表示某事件属于 X 和不属于 X ,行动集 $A = \{a_P, a_B, a_N\}$ 中的元素分别表示将一个对象 x 分给正区域、边界域和负区域。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 $\lambda_{NN}, \lambda_{BN}, \lambda_{PN}$ 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。因此,根据式(2),采取 a_P, a_B 和 a_N 3 种行动下的期望损失可分别表示为:

$$\begin{aligned} R_P &= R(a_P | [x]) = \lambda_{PP} \cdot P(X|[x]) + \lambda_{PN} \cdot P(X^c|[x]) \\ R_B &= R(a_B | [x]) = \lambda_{BP} \cdot P(X|[x]) + \lambda_{BN} \cdot P(X^c|[x]) \\ R_N &= R(a_N | [x]) = \lambda_{NP} \cdot P(X|[x]) + \lambda_{NN} \cdot P(X^c|[x]) \end{aligned} \quad (6)$$

根据贝叶斯决策准则,需要选取期望损失最小的行动集

作为最佳行动方案。于是得到如下 3 条决策规则:

- (P) 如果 $R_P \leq R_B$ 且 $R_P \leq R_N$, 则选择 $x \in POS(X)$;
- (B) 如果 $R_B \leq R_P$ 且 $R_B \leq R_N$, 则选择 $x \in BND(X)$;
- (N) 如果 $R_N \leq R_P$ 且 $R_N \leq R_B$, 则选择 $x \in NEG(X)$ 。

根据 $P(X|[x]) + P(X^c|[x]) = 1$, 上述规则只与 $P(X|[x])$ 和相关的损失函数有关。此外,考虑到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,则可得出一个合理的假设,为 $0 \leq \lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}$, $0 \leq \lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}$ 。根据上述条件,决策规则可重新定义为:

- (P) 如果 $P(X|[x]) \geq \alpha$ 且 $P(X|[x]) \geq \gamma$, 则选择 $x \in POS(X)$;
- (B) 如果 $P(X|[x]) \leq \alpha$ 且 $P(X|[x]) \geq \beta$, 则选择 $x \in BND(X)$;
- (N) 如果 $P(X|[x]) \leq \beta$ 且 $P(X|[x]) \leq \gamma$, 则选择 $x \in NEG(X)$ 。

其中, α, β, γ 分别为:

$$\begin{aligned} \alpha &= \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} \\ \beta &= \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} \\ \gamma &= \frac{\lambda_{PN} - \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})} \end{aligned} \quad (9)$$

由规则(B)可知, $\alpha > \beta$ 。据此,可以得到 $0 \leq \beta < \gamma < \alpha \leq 1$ 。这样,决策规则可重写如下:

- (P1) 如果 $P(X|[x]) \geq \alpha$, 则选择 $x \in POS(X)$;
- (B1) 如果 $\beta < P(X|[x]) < \alpha$, 则选择 $x \in BND(X)$;
- (N1) 如果 $P(X|[x]) \leq \beta$, 则选择 $x \in NEG(X)$ 。

3 模糊数及模糊运算

为了解决不确定环境下的问题, Zadeh 在 1965 年提出了模糊集的概念:给定论域 U 上的一个模糊集 A , 对任何 $x \in U$, 都有一个数 $\mu_A(x) \in [0, 1]$ 与之对应, 则称 $\mu_A(x)$ 为 $x \in A$ 的隶属度, 从而一般称 $\mu_A(x)$ 为 A 的隶属函数。

定义 1 设 $A \in F(U)$, $\forall \eta \in [0, 1]$, 称 $A_\eta = \{x | x \in U, \mu_A(x) \geq \eta\}$ 为 A 的 η -截集。称 $A_\eta = \{x | x \in U, \mu_A(x) > \eta\}$ 为 A 的强 η -截集。

若模糊集定义在实数 R 上, 则模糊集的隶属函数称为模糊分布。在模糊数学中, 常见的模糊分布有三角形模糊分布、梯形模糊分布、正态模糊分布等。区间值和三角形模糊分布是梯形模糊分布的特例, 本文主要介绍梯形模糊数。

定义 2 设论域 U 上的模糊数为 M , 如果 M 的隶属度函数 $\mu_M: U \rightarrow [0, 1]$ 表示为:

$$\mu_M = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & b \leq x < c \\ \frac{d-x}{d-c}, & c \leq x < d \\ 0, & x \geq d \end{cases} \quad (11)$$

则称 $M = (a, b, c, d)$ 为梯形模糊数。

梯形模糊数的几何表示如图 1 所示。

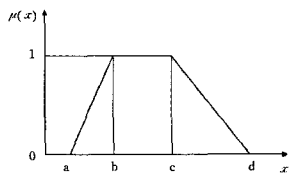


图1 梯形模糊数的几何表示

一个梯形模糊数可以表示为 $M = (r_1, r_2, r_3, r_4)$, 其中 r_1, r_2, r_3, r_4 为实数, $r_1 \leq r_2 \leq r_3 \leq r_4$ 。特别地, 若 $r_1 = r_2 = r_3 = r_4, M$ 为一个实数; 若 $r_1 = r_2 < r_3 = r_4, M$ 为一个区间值; 若 $r_1 < r_2 = r_3 < r_4, M$ 为一个三角模糊数。

定义3 若 $M_1 = (r_1, r_2, r_3, r_4), M_2 = (s_1, s_2, s_3, s_4)$ 是两个梯形模糊数, 且参数是非负实数, 则其模糊运算定义如下:

1) 模糊数加法

$$M_1 + M_2 = (r_1 + s_1, r_2 + s_2, r_3 + s_3, r_4 + s_4)$$

2) 模糊数减法

$$M_1 - M_2 = (r_1 - s_4, r_2 - s_3, r_3 - s_2, r_4 - s_1)$$

3) 模糊数乘法

$$M_1 \times M_2 = (r_1 s_1, r_2 s_2, r_3 s_3, r_4 s_4)$$

4) 模糊数除法

$$M_1 \div M_2 = (r_1/s_4, r_2/s_3, r_3/s_2, r_4/s_1)$$

其中 s_1, s_2, s_3, s_4 不为 0。

5) 模糊数倒数

$$M_1^{-1} = (r_4^{-1}, r_3^{-1}, r_2^{-1}, r_1^{-1})$$

其中 r_1, r_2, r_3, r_4 不为 0。

4 损失函数的模糊分布

在实际决策问题中, 风险很难确定, 用单值来刻画它是不合理的。针对这个问题, 刘盾博士等在文献[12]中讨论了用区间值来刻画损失函数, 使得决策模型具有更好的适应性。本文从更一般性的角度出发, 假定风险值是个模糊数, 而其模糊分布通过模糊统计方法^[12]得出。

在进行专家咨询时, 专家们往往会给出一些模糊量, 例如三值判断: 最低可能值、最可能值、最高可能值; 或者二值区间判断。

由多名专家分别给出一组损失函数 $\lambda_{PP} = [\lambda_{PP}^L, \lambda_{PP}^U]$, $\lambda_{BP} = [\lambda_{BP}^L, \lambda_{BP}^U]$, $\lambda_{NP} = [\lambda_{NP}^L, \lambda_{NP}^U]$, $\lambda_{NN} = [\lambda_{NN}^L, \lambda_{NN}^U]$, $\lambda_{BN} = [\lambda_{BN}^L, \lambda_{BN}^U]$, $\lambda_{PN} = [\lambda_{PN}^L, \lambda_{PN}^U]$ 的区间值, 考虑到模型的实际语义性, 每组区间值需要满足:

$$\begin{aligned} 0 &\leq \lambda_{PP}^L \leq \lambda_{PP}^U < \lambda_{BP}^L \leq \lambda_{BP}^U < \lambda_{NP}^L \leq \lambda_{NP}^U \\ 0 &\leq \lambda_{NN}^L \leq \lambda_{NN}^U < \lambda_{BN}^L \leq \lambda_{BN}^U < \lambda_{PN}^L \leq \lambda_{PN}^U \end{aligned} \quad (12)$$

对每个损失函数 $\lambda_{..}$ 进行模糊统计, 得出其模糊分布。模糊统计方法如下:

1) 假定在一次统计中, 对于 λ_{NP} 的值, n 位专家给出了 n 个区间值 $\lambda_{NP} = [\lambda_{NP}^L, \lambda_{NP}^U]$ 。

2) 为了确定区间中某个值 μ_0 在分布中的隶属度, 可以对 μ_0 做统计处理。得出隶属频率 $f = m/n$, 其中 n 表示随机选择的样本总数, m 为区间样本盖住 μ_0 的频数。

3) 为了得到 λ_{NP} 的模糊分布函数, 采用“方框图法”先找出 n 个区间值中的最小值 $\mu_{\min}$ 和最大值 $\mu_{\max}$ 。于是, 以 $\mu_{\min}$ 为起点, $\mu_{\max}$ 为终点, 以 $(\mu_{\max} - \mu_{\min})/k$ 为长度 ($k \in N^*$, 依据具体数据进行取值), 作 k 个区间的划分。计算每个区间的隶属频率。

4) 以实数 x 为横坐标、相对频率为纵坐标, 绘出 λ_{NP} 的模糊分布曲线, 假定结果如图 2 所示。

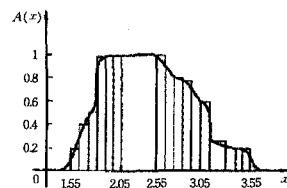


图2 损失函数的模糊分布曲线

5) 对该模糊分布左右两边的曲线进行直线拟合, 得到一个梯形分布函数, 如图 3 所示。

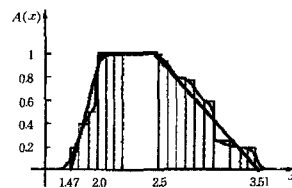


图3 损失函数的梯形模糊分布

5 模糊量 α, β 的分布

本节主要介绍通过模糊运算得出阈值 α, β 模糊分布的方法。与刘盾等^[12]的阈值 α, β 上、下确界相比, 本文可以得出更为紧凑的上、下确界。

通过模糊统计方法^[13], 可以得出损失函数的梯形分布 (若得到区间分布或者三角形分布, 则拓展写成梯形分布形式)。

设损失函数 $\lambda_{..}$ 的模糊表示为 $M_{..}$, 损失函数可分别表示如下:

$$M_{PP} = (r_{11}, r_{12}, r_{13}, r_{14})$$

$$M_{BP} = (r_{21}, r_{22}, r_{23}, r_{24})$$

$$M_{NP} = (r_{31}, r_{32}, r_{33}, r_{34})$$

$$M_{NN} = (s_{11}, s_{12}, s_{13}, s_{14})$$

$$M_{BN} = (s_{21}, s_{22}, s_{23}, s_{24})$$

$$M_{PN} = (s_{31}, s_{32}, s_{33}, s_{34})$$

在决策粗糙集模型中, 有:

$$\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} \quad (13)$$

$$\beta = \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}$$

因为分子和分母中包含同一个模糊数, 故不能直接对 α, β 进行模糊运算。先对其取倒数得:

$$\frac{1}{\alpha} = 1 + \frac{\lambda_{BP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{BN}} \quad (14)$$

$$\frac{1}{\beta} = 1 + \frac{\lambda_{NP} - \lambda_{BP}}{\lambda_{BN} - \lambda_{NN}}$$

将模糊数代入 $1/\alpha, 1/\beta$ 中进行模糊运算, 得到 $1/\alpha, 1/\beta$ 的梯形分布, 结果如下:

$$\begin{aligned} \frac{1}{\alpha} &= (1 + \frac{r_{21} - r_{14}}{s_{34} - s_{21}}, 1 + \frac{r_{22} - r_{13}}{s_{33} - s_{22}}, 1 + \frac{r_{23} - r_{12}}{s_{32} - s_{23}}, 1 + \frac{r_{24} - r_{11}}{s_{31} - s_{24}}) \\ \frac{1}{\beta} &= (1 + \frac{r_{31} - r_{24}}{s_{24} - s_{11}}, 1 + \frac{r_{32} - r_{23}}{s_{23} - s_{12}}, 1 + \frac{r_{33} - r_{22}}{s_{22} - s_{13}}, 1 + \frac{r_{34} - r_{21}}{s_{21} - s_{14}}) \end{aligned} \quad (15)$$

然后对 $1/\alpha, 1/\beta$ 进行模糊数的倒数运算, 可以得到 α, β 的模糊分布。

$$\alpha = (1/(1 + \frac{r_{24} - r_{11}}{s_{31} - s_{24}}), 1/(1 + \frac{r_{23} - r_{12}}{s_{32} - s_{23}}), 1/(1 + \frac{r_{22} - r_{13}}{s_{33} - s_{22}}), 1/(1 + \frac{r_{21} - r_{14}}{s_{34} - s_{21}}))$$

$$\beta = (1/(1 + \frac{r_{34} - r_{21}}{s_{21} - s_{14}}), 1/(1 + \frac{r_{33} - r_{22}}{s_{22} - s_{13}}), 1/(1 + \frac{r_{32} - r_{23}}{s_{23} - s_{12}}), 1/(1 + \frac{r_{31} - r_{24}}{s_{24} - s_{11}}))$$

(16)

依据式(12),可以得出 α, β 分布中每个参数的值在 $[0, 1]$ 内。

接下来讨论一种特殊的情形:如果梯形模糊数 $(r_{*1}, r_{*2}, r_{*3}, r_{*4}), (s_{*1}, s_{*2}, s_{*3}, s_{*4})$ 同时满足 $r_{*1} = r_{*2} = \lambda_{*P}^L, r_{*3} = r_{*4} = \lambda_{*P}^U, s_{*1} = s_{*2} = \lambda_{*N}^L$ 和 $s_{*3} = s_{*4} = \lambda_{*N}^U$, 则损失函数可以表示为区间值。各损失函数分布可以简化表示为:

$\lambda_{PP} = [\lambda_{PP}^L, \lambda_{PP}^U], \lambda_{BP} = [\lambda_{BP}^L, \lambda_{BP}^U], \lambda_{NP} = [\lambda_{NP}^L, \lambda_{NP}^U], \lambda_{NN} = [\lambda_{NN}^L, \lambda_{NN}^U], \lambda_{BN} = [\lambda_{BN}^L, \lambda_{BN}^U], \lambda_{PN} = [\lambda_{PN}^L, \lambda_{PN}^U]$, 且需要满足式(12)。

刘盾博士等^[12]给出 α, β 的上、下确界为:

$$\alpha \in [\alpha_L, \alpha_U] = [\frac{\lambda_{PN}^L - \lambda_{BN}^L}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)}, \min(\frac{\lambda_{PN}^U - \lambda_{BN}^U}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)}, 1)]$$

$$\beta \in [\beta_L, \beta_U] = [\frac{\lambda_{BN}^L - \lambda_{NN}^L}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}, \min(\frac{\lambda_{BN}^U - \lambda_{NN}^U}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)}, 1)]$$

(17)

在式(17)的上、下确界值计算中,分子和分母中的同一个损失函数使用了不同的值代入,如 α_L 的损失函数 λ_{PN} , 分母代入 λ_{PN}^L , 而分子代入 λ_{PN}^U 。这不够合理,对此,按照式(16)进行运算,得到另一组上、下确界值:

$$\alpha \in [\alpha_L', \alpha_U'] = [1/(1 + \frac{r_{24} - r_{11}}{s_{31} - s_{24}}), 1/(1 + \frac{r_{21} - r_{14}}{s_{34} - s_{21}})]$$

$$= [\frac{\lambda_{PN}^L - \lambda_{BN}^L}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)}, \frac{\lambda_{PN}^U - \lambda_{BN}^U}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)}]$$

$$\beta \in [\beta_L', \beta_U'] = [1/(1 + \frac{r_{34} - r_{21}}{s_{21} - s_{14}}), 1/(1 + \frac{r_{31} - r_{24}}{s_{24} - s_{11}})]$$

$$= [\frac{\lambda_{BN}^L - \lambda_{NN}^L}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}, \frac{\lambda_{BN}^U - \lambda_{NN}^U}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)}]$$

(18)

同时,根据式(12),有:

$$\alpha_U' < 1, \beta_U' < 1 \quad (19)$$

将两个上、下确界值进行比较,明显地,有:

$$\frac{\lambda_{PN}^L - \lambda_{BN}^L}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)} = \alpha_L'$$

$$\geq \alpha_L = \frac{\lambda_{PN}^L - \lambda_{BN}^L}{(\lambda_{PN}^L - \lambda_{BN}^L) + (\lambda_{BP}^L - \lambda_{PP}^L)}$$

$$\frac{\lambda_{PN}^U - \lambda_{BN}^U}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)} = \alpha_U'$$

$$\leq \alpha_U = \frac{\lambda_{PN}^U - \lambda_{BN}^U}{(\lambda_{PN}^U - \lambda_{BN}^U) + (\lambda_{BP}^U - \lambda_{PP}^U)}$$

$$\frac{\lambda_{BN}^L - \lambda_{NN}^L}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)} = \beta_L'$$

$$\geq \beta_L = \frac{\lambda_{BN}^L - \lambda_{NN}^L}{(\lambda_{BN}^L - \lambda_{NN}^L) + (\lambda_{NP}^L - \lambda_{BP}^L)}$$

$$\frac{\lambda_{BN}^U - \lambda_{NN}^U}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)} = \beta_U'$$

$$\leq \beta_U = \frac{\lambda_{BN}^U - \lambda_{NN}^U}{(\lambda_{BN}^U - \lambda_{NN}^U) + (\lambda_{NP}^U - \lambda_{BP}^U)} \quad (20)$$

式(19)和式(20)说明,本文得出比文献[12]更加紧凑的 α, β 上、下确界值,为进一步研究提供了更加可靠的基础。

下面,通过简单数学公式证明 $\alpha \in [\alpha_L', \alpha_U']$ 。

证明:将 $\frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}$ 表示成函数 $f(x) = \frac{x}{x+c}, (\lambda_{PN} - \lambda_{BN})$ 对应自变量 $x, x \in [X_L, X_U] = [\lambda_{PN}^L - \lambda_{BN}^L, \lambda_{PN}^U - \lambda_{BN}^U]; (\lambda_{BP} - \lambda_{PP})$ 对应常数 $c, c > 0$ 。

对 $f(x)$ 求导可得: $f'(x) = \frac{c}{(x+c)^2} > 0, f(x)$ 是个递增函数,代入对应 $(\lambda_{PN} - \lambda_{BN})$ 和 $(\lambda_{BP} - \lambda_{PP})$ 值可得 $\alpha \in [\alpha_L', \alpha_U']$ 。

6 模糊决策

从上节的计算中,得出了阈值 α, β 的梯形模糊分布,本节首先介绍模糊数的一般比较规则,然后介绍两种利用阈值 α, β 的梯形模糊分布来生成决策规则的方法。

在决策粗糙集模型中, α, β 是单值的,且 $\alpha > \beta$ 。通过对比条件概率 $P(X|[x]) = |x \cap X|/|x|$ 与 α, β 的大小关系,做出三支决策。

使用模糊数进行决策,一般有两个方向:一是将模糊数 α, β 和实数 $P(X|[x])$ 进行比较,据此做出决策;二是分析 α, β 的模糊分布,分解出不同风险倾向的决策,如乐观决策准则、悲观决策准则^[12]。

将模糊数 α, β 和实数 $P(X|[x])$ 进行比较。模糊数间的比较方法有很多,常见的思路是模糊量转化为实数,进而利用实数的自然序关系导出模糊量的序关系。这类指标主要有 Yager 的指标^[14,15]、Adamo 的指标^[16]、Chang 的指标^[17]、Campos-munoz 指标^[18]。其基本思路如下:

作一个映射 $F: \Omega \rightarrow R$ (实数域),从而将每一个模糊量 A_i 转化为实数 $F(A_i)$, 通过比较 $F(A_i), F(A_j)$ 得到 A_i, A_j 之间的序关系为:

$$A_i \geq A_j \Leftrightarrow F(A_i) \geq F(A_j)$$

本文使用模糊满意度^[14]来比较两个模糊数的大小。对于梯形模糊数 $a = (a_1, a_2, a_3, a_4)$, 将其任一 α 水平截集的左右端点分别记为 $a_\alpha^L, a_\alpha^U, b_\alpha^L, b_\alpha^U$, 则有 $a_\alpha^L = a_2\alpha + a_1(1-\alpha), a_\alpha^U = a_3\alpha + a_4(1-\alpha)$ 。

定义4 对两个模糊数 a, b (它们中至少一个有不是实数), 它们的隶属函数分别为 $\mu_a(x)$ 和 $\mu_b(x)$, $a \leq b$ 的模糊满意度定义为:

$$P(a \leq b) = \frac{\int_0^1 (\max(0, b_\alpha^U - a_\alpha^L) - \max(0, b_\alpha^L - a_\alpha^U)) d\alpha}{\int_0^1 (a_\alpha^U - a_\alpha^L) d\alpha - \int_0^1 (b_\alpha^U - b_\alpha^L) d\alpha} \quad (21)$$

满意度有如下性质:

- 1) $P(a \leq b) + P(b \leq a) = 1$ 。
- 2) 若 $b_\alpha^U \geq a_\alpha^U$, 则 $P(a \leq b) = 1$; 反之不成立。若 $b_\alpha^L \leq a_\alpha^L$, 则 $P(a \leq b) = 0$; 反之不成立。
- 3) $0 \leq P(a \leq b) \leq 1, 0 \leq P(b \leq a) \leq 1$ 。
- 4) $P(a \leq a) = 1/2$ 。

排序规则如下:若 $P(a < b) > 0.5$, 则 $a < b$;

若 $P(a < b) \leq 0.5$, 则 $a > b$ 。

使用式(21)比较 $P(X|[x]) = |[x] \cap X|/|[x]|$ 与模糊数 α, β 的大小关系, 做出三支决策, 决策规则如下。

(P2) 如果 $P(X|[x]) \geq \alpha$, 则选择 $x \in POS(X)$;

(B2) 如果 $\beta < P(X|[x]) < \alpha$, 则选择 $x \in BND(X)$; (22)

(N2) 如果 $P(X|[x]) \leq \beta$, 则选择 $x \in NEG(X)$ 。

可以看出, 决策规则(P2-N2)仍然是基于单值的决策规则, 未充分利用模糊数 α, β 的模糊特性。前文已得出 α, β 的梯形模糊分布。据此, 可以考虑取 α, β 的 1-截集, 左右端点分别记为 $\alpha_1^L, \alpha_1^R, \beta_1^L, \beta_1^R$, 参照文献[12], 分别得出风险偏好者以及风险厌恶者的决策阈值。具体应用过程参见下一节的案例分析。

7 案例分析

下面利用一个石油投资的例子来阐明基于模糊数风险最小化的拓展决策粗糙集模型的应用过程。把模糊数阈值转化为单值的过程比较简单, 下面主要利用阈值的模糊特性来得到不同风险倾向的决策阈值。

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, 将它们主要分为 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^[20]。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^[21]。

根据模糊数决策粗糙集的思想, 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 $\Omega = \{X, X^c\}$ 表示某油田有油和无油两种状态, 行动集 $A = \{a_P, a_B, a_N\}$ 表示投资、延迟决策和不投资 3 种行动。在群体决策中, 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。考虑到专家思维模糊性和评估过程中的不确定性情况, 最终每个专家给出的各个损失函数是由区间值来表示的, 每组区间值需要满足式(12)。假定对于某一损失函数, 40 位专家给出的区间如表 1 所列。

表 1 40 位专家给出 λ_{PP} 的区间损失函数

1.8~2.5	1.7~3.0	1.7~2.8	1.8~2.5	1.6~3.0
1.6~2.8	1.5~2.5	1.8~3.5	1.4~2.5	1.8~3.3
1.6~2.5	1.8~3.0	1.6~3.0	2.0~3.0	2.0~3.0
1.8~3.1	1.5~2.8	1.8~3.5	1.8~3.0	1.7~3.0
1.7~2.5	1.5~2.5	1.8~2.9	1.5~2.4	1.5~3.0
1.7~3.0	1.8~3.0	1.8~2.5	1.7~2.6	1.6~2.8
1.9~2.8	1.5~3.0	1.5~2.6	1.7~2.5	1.5~3.4
1.8~3.6	1.8~3.0	1.6~3.2	1.5~2.5	1.8~2.8

为了确定区间中某个值 μ_0 在分布中的隶属度, 采用“方框图法”。根据表 1 可知, 最小数据是 1.4, 最大数据是 3.6。于是, 以 1.35 为起点、3.65 为终点, 以 0.1 为长度, 做 23 个区间的划分。数据如表 2 所列。

表 2 各损失函数值的分布频数

序号	分组值	频数	相对频数
1	1.35~1.45	1	0.025
2	1.45~1.55	10	0.250
3	1.55~1.65	16	0.400
4	1.65~1.75	23	0.575
5	1.75~1.85	37	0.925
6	1.85~1.95	38	0.950
7	1.95~2.05	40	1
8	2.05~2.15	40	1
9	2.15~2.25	40	1
10	2.25~2.35	40	1
11	2.35~2.45	40	1
12	2.45~2.55	39	0.975
13	2.55~2.65	29	0.725
14	2.65~2.75	27	0.675
15	2.75~2.85	27	0.675
16	2.85~2.95	21	0.525
17	2.95~3.05	20	0.500
18	3.05~3.15	7	0.175
19	3.15~3.25	6	0.500
20	3.25~3.35	5	0.125
21	3.35~3.45	4	0.100
22	3.45~3.55	3	0.075
23	3.55~3.65	1	0.025

以损失值为横坐标、相对频数为纵坐标, 绘出 λ_{PP} 的模糊分布曲线, 如图 4 所示。

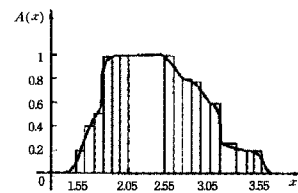


图 4 损失函数的模糊分布曲线

对模糊分布左右两边的曲线进行最小二乘法直线拟合, 得到一个梯形分布函数, 如图 5 所示。

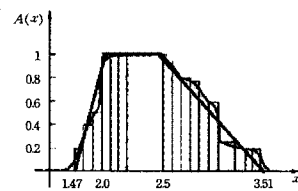


图 5 损失函数的梯形模糊分布

如此, 得到损失函数 λ_{PP} 的梯形模糊分布, 可表示为 $(1.47, 2.0, 2.5, 3.51)$ 。对其他损失函数做同样处理, 得到各损失函数的梯形分布如下。

$$M_{PP} = (1.47, 2.0, 2.5, 3.51)$$

$$M_{BP} = (3.6, 4.3, 4.9, 5.82)$$

$$M_{NP} = (7.2, 7.8, 8.4, 9.53)$$

$$M_{NN} = (0, 0.5, 0.8, 1.2)$$

$$M_{BN} = (1.4, 2.0, 2.6, 4.0)$$

$$M_{PN} = (4.6, 5.7, 6.3, 8.2)$$

将上述模糊量代入式(16), 得参数 α, β 的梯形模糊数如下:

$$\alpha = (0.121, 0.517, 0.705, 0.989)$$

$$\beta = (0.033, 0.226, 0.420, 0.734)$$

考虑置信度为 1 的情况, 取 α, β 梯形模糊数的 1-截集, 左

(下转第 75 页)

- [11] Min F, Zhu W. Granular association rules for multi-valued data [C]//Proceedings of the 2013 Canadian Conference on Electrical and Computer Engineering, 2013
- [12] Lin T Y. Granular computing on binary relations I: data mining and neighborhood systems [J]. Rough sets in knowledge discovery, 1998, 1: 107-121
- [13] Zhu W, Wang F Y. Reduction and axiomization of covering generalized rough sets [J]. Information sciences, 2003, 152: 217-230
- [14] Yao Y, Deng X. A granular computing paradigm for concept

learning [M]. Emerging Paradigms in Machine Learning. Springer Berlin Heidelberg, 2013: 307-326

- [15] 苗夺谦. 粒计算: 过去, 现在与展望[M]. 北京: 科学出版社, 2007
- [16] Min F, Zhu W, He X. Grale; Granular association rules[OL]. <http://grc.fjzs.edu.cn/~fmin/grale/>, 2013
- [17] Pawlak Z. Rough sets [J]. International Journal of Computer & Information Sciences, 1982, 11(5): 341-356
- [18] Skowron A, Stepaniuk J. Approximation of relations [M]. Rough Sets, Fuzzy Sets and Knowledge Discovery. London: Springer, 1994: 161-166

(上接第 54 页)

右端点分别记为 $\alpha_l^I = 0.517$, $\alpha_r^I = 0.705$, $\beta_l^I = 0.226$, $\beta_r^I = 0.420$ 。据此, 可以得出 1-置信度下的风险偏好者以及风险厌恶者的决策阈值; 同时也给出 α 、 β 的阈值下界和阈值上界。根据不同风险倾向者的决策阈值, 可以分别得出风险偏好者以及风险厌恶者的决策规则。更一般地, 可以取 α 、 β 的 η 截集, $\eta \in (0, 1]$ 。得到置信度 η 下的不同风险倾向者的决策阈值, 从而得出相应的决策规则。可以看出, 在损失函数不确定的条件下, 本文通过一系列模糊运算, 将这些不确定的信息保留到模糊决策阈值 α 、 β 中, 在信息损失最少的情况下帮助决策者在实际决策中做出合理的决策。

结束语 本文从决策粗糙集出发, 利用模糊数来刻画损失函数, 首先提出用模糊量来处理损失函数不确定性特征的问题; 其次, 通过模糊统计得到损失函数的梯形分布; 再次, 介绍通过模糊运算得出 α 、 β 模糊分布的方法; 然后, 介绍一种通过比较模糊数 α 、 β 与条件概率大小而生成决策规则的方法; 最后, 用一个石油投资的例子来说明模型的应用过程。然而, 本文对引入模糊量进行粗糙集决策的研究尚在初级阶段, 其相关数学性质以及如何充分利用模糊数的分布来进一步分析决策过程有待进一步挖掘。

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. Rough Sets; Theoretical Aspects of Reasoning about Data[M]. Boston: Kluwer Academic Publishers Press, 1991: 90-166
- [3] Yao Y Y. Three-way decision: an interpretation of rules in rough set theory[J]. LNAI, 2009(5589): 642-649
- [4] Yao Y Y. Three-way decisions with probabilistic rough sets[J]. Information Sciences, 2010, 180: 341-353
- [5] Yao Y. The superiority of three-way decisions in probabilistic rough set models[J]. Information Sciences, 2011, 181(6): 1080-1096
- [6] Pawlak Z, Wong S K M, Ziarko W. Rough sets: probabilistic versus deterministic approach[J]. Inter. Journal of Man-Machine Studies, 1988, 29: 81-95

- [7] Yao Y Y, Wong S K M. A decision theoretic framework for approximation concepts[J]. Inter. Journal of Man-machine Studies, 1992, 37: 793-809
- [8] Yao Y Y, Wong S K M. A decision theoretic framework for approximating concepts[J]. International Journal of Man-machine Studies, 1992, 37(6): 793-809
- [9] Ziarko W. Variable precision rough set model [J]. Journal of computer and system sciences, 1993, 46(1): 39-59
- [10] Slezak D. Rough sets and Bayes factor[J]. LNCS Transactions on Rough Sets, 2005, 111: 202-229
- [11] Slezak D, Ziarko W. The investigation of the Bayesian rough set model[J]. International Journal of Approximate Reasoning, 2005, 40: 81-91
- [12] 刘盾, 李天瑞, 李华雄. 区间决策粗糙集[J]. 计算机科学, 2012, 39(7): 178-181, 215
- [13] 谢季坚, 刘承平. 模糊数学方法及其应用[M]. 武汉: 华中科技大学出版社, 2000: 29-32
- [14] Liu X. Measuring the satisfaction of constraints in fuzzy linear programming[J]. Fuzzy Sets and Systems, 2001, 122(2): 263-275
- [15] Yager R R. On choosing between fuzzy subsets[J]. Kybernetes, 1980, 9(2): 151-154
- [16] Yager R R. A procedure for ordering fuzzy subsets of the unit interval[J]. Information Sciences, 1981, 24(2): 143-161
- [17] Adamo J M. Fuzzy decision trees[J]. Fuzzy sets and systems, 1980, 4(3): 207-219
- [18] Chang W. Ranking of fuzzy utilities with triangular membership functions[C]//Proceedings of International Conference on Policy Analysis and Systems, 1981: 272
- [19] de Campos Ibáñez L M, Muñoz A G. A subjective approach for ranking fuzzy numbers[J]. Fuzzy sets and systems, 1989, 29(2): 145-153
- [20] Xie G, Yue W, Wang S, et al. Dynamic risk management in petroleum project investment based on a variable precision rough set model[J]. Technological Forecasting and Social Change, 2010, 77(6): 891-901
- [21] Yusgiantoro P, Hsiao F S T. Production-sharing contracts and decision-making in oil production: The case of Indonesia[J]. Energy economics, 1993, 15(4): 245-256