

基于可拓关联度比的三支决策模型及应用

李 敏^{1,2,3}, 杨亚锋^{1,2,3}, 雷 宇⁴, 李丽红^{1,2,3}

(1. 华北理工大学 理学院 河北 唐山 063210; 2. 河北省数据科学与应用重点实验室 河北 唐山 063210;
3. 唐山市数据科学重点实验室 河北 唐山 063210; 4. 华北理工大学 电气工程学院 河北 唐山 063210)

摘要: 三个域的构造和解释一直是三支决策研究重点之一,将可拓评价的基本方法与“三分而治”的决策思想进行融合,构建基于可拓关联度比的三支决策模型。分析关联函数的构造原理及数值特征,定义更具决策指导意义的可拓关联度比,论述其作为三支决策中评价函数的可能性与合理性;将归一化的关联度比作为评价函数,对每个等级的指标进行三个域划分,构造三支决策规则,实现动态决策与挖掘优化指标的目的。将模型应用于决策哈尔滨市水资源承载力,结果表明,哈尔滨市的水资源承载力处于Ⅱ级水平,城镇化率、人口自然增长率、Remanenko 潜在蒸发量、干燥度、人均生活用水量、GDP 总量、人均 GDP、GDP 增长率、森林覆盖率这些指标亟待优化,以实现水资源系统的可持续发展。

关键词: 可拓集; 关联度比; 三支决策; 可拓评价; 水资源承载力

中图分类号: TP181

文献标志码: A

文章编号: 1671-6841(2021)02-0072-10

DOI: 10.13705/j.issn.1671-6841.2020247

0 引言

三支决策是在粗糙集的基础上提出的,是一种处理不确定性决策的数学方法,姚一豫等^[1]基于各个领域三支决策的共性及特征,提出三支决策理论框架。自提出以来,三支决策备受学者关注,从不同角度进行深入研究。针对勾股模糊三支决策概率阈值难以确定的问题,文献[2-3]给出概率阈值确定方法和基于勾股模糊集评价的三支决策方法,同时结合直觉模糊集理论提出一种基于直觉模糊相似度的三支决策方法;薛占熬等^[4]通过研究粒度重要度的构造方法,提出一种新的基于粒度重要度的三支决策模型将证据理论引入三支决策中,分析延迟信任区间可能包含的可变语义;陈玉金等^[5]构建基于证据理论的确定和可变三支决策模型;岳文琦等^[6]在混合决策信息系统中,提出一种模糊效用三支决策模型;张春英等^[7]在深入研究 PS-粗糙集模型特点的基础上,提出一种基于 PS-粗糙集的动态三支决策算法。此外,三支决策模型在水资源承载力、垃圾邮件过滤、图像分割等领域^[8-11]得到广泛的应用。

“可拓集”从变换的角度研究对象具有某种性质的程度及其变化,并用关联函数进行定量化描述^[12]。随着可拓学研究的深入,关联函数成为重点。文献[13]对可拓集无量纲一维关联函数进行研究,为进一步研究多维关联函数奠定基础。杨春燕等^[14]介绍可拓集中初等关联函数构造的进展并进一步深入研究。在单评价特征基元可拓集基础上,杨春燕^[15]研究了多评价特征基元可拓集的构造方法及其关联函数的建立方法,为解决不相容问题提供了理论依据。胡宝清等^[16]从实际出发,研究区间论域上的可拓集及其关联函数。还有学者利用可拓集进行可拓评价,解决实际问题。安永林等^[17]为综合评价服役隧道结构健康状况,利用可拓法进行评估并有针对性地给出不同的策略。李倩倩等^[18]以山东半岛蓝色经济区为研究对象,基于熵权可拓模型对城镇化发展水平进行评价。郭嘉伟等^[19]基于熵权可拓物元模型对会宁县水资源承载力进行评价。文献[20-21]将可拓集方法应用于水资源承载力的评价。

三支决策模型大多以条件概率作为评价函数进行决策,而可拓集中关联函数表示事物具有或不具有某种性质的程度,可对同一域中的元素进行优劣排序,因此将可拓集应用于三支决策,不仅可以实现域的划分,

收稿日期:2020-08-10

基金项目:河北省数据科学与应用重点实验室项目(10120201);唐山市数据科学重点实验室项目(10120301)。

作者简介:李敏(1995—),女,硕士研究生,主要从事粒计算与三支决策研究,E-mail:2744997922@qq.com;通信作者:李丽红(1979—),教授,主要从事粗糙集和模糊集研究,E-mail:22687426@qq.com。

同时能够挖掘决策优化方向,进而制定系统状态提升策略。基于此,本文利用可拓关联函数的方法,定义关联度比,构建一种基于关联度比的三支决策模型,并对哈尔滨市水资源承载力系统给出优化方向与建议。

1 预备知识

1.1 三支决策

三支决策的提出为决策粗糙集的正域、负域、边界域赋予合理的语义解释,分别对应接受、拒绝和延迟决策,符合人类的决策思想。

贝叶斯理论是风险型决策方法之一,其思想是以期望损失最小为决策依据,选择合适的决策方案^[22]。三支决策是基于贝叶斯理论进行的决策,规则如下。

设状态空间 $\Theta = \{D_i, \neg D_i\}$, 表示对象 x 是否属于集合 D_i , 行动集 $A = \{a_p, a_B, a_N\}$ 表示对一个对象 x 进行分类决策时采取的行动。 a_p, a_B, a_N 分别表示对象确定属于、不一定属于和一定不属于集合的行动,不同状态下对应不同行动的代价函数如表1所示。

表1 损失代价函数

Table 1 Loss cost function

类别	D_i	$\neg D_i$
a_p	λ_{pp}	λ_{pn}
a_B	λ_{Bp}	λ_{Bn}
a_N	λ_{Np}	λ_{Nn}

$\lambda_{pp}, \lambda_{Bp}, \lambda_{Np}$ 分别表示 x 本属于 D_i 而采取行动 a_p, a_B, a_N 时的损失; $\lambda_{pn}, \lambda_{Bn}, \lambda_{Nn}$ 分别表示 x 本不属于 D_i 而采取行动 a_p, a_B, a_N 时的损失。 $P(D_i | [x])$ 是条件概率,对于特定 x , 采取一个决策行动时的期望代价为

$$\begin{cases} T(a_p | [x]) = \lambda_{pp}P(D_i | [x]) + \lambda_{pn}P(\neg D_i | [x]), \\ T(a_B | [x]) = \lambda_{Bp}P(D_i | [x]) + \lambda_{Bn}P(\neg D_i | [x]), \\ T(a_N | [x]) = \lambda_{Np}P(D_i | [x]) + \lambda_{Nn}P(\neg D_i | [x]), \end{cases}$$

式中:条件概率 $P(D_i | [x]) = \frac{|D_i \cap [x]|}{|[x]|}$ 。根据贝叶斯风险决策过程,最小代价决策规则为

- P) 如果 $T(a_p | [x]) \leq T(a_B | [x])$ 且 $T(a_p | [x]) \leq T(a_N | [x])$ 成立,则 $x \in POS(X)$,
- B) 如果 $T(a_B | [x]) \leq T(a_p | [x])$ 且 $T(a_B | [x]) \leq T(a_N | [x])$ 成立,则 $x \in BND(X)$,
- N) 如果 $T(a_N | [x]) \leq T(a_p | [x])$ 且 $T(a_N | [x]) \leq T(a_B | [x])$ 成立,则 $x \in NEG(X)$ 。

由于 $P(D_i | [x]) + P(\neg D_i | [x]) = 1$, 且 $\lambda_{pp} \leq \lambda_{Bp} \leq \lambda_{Np}, \lambda_{Nn} \leq \lambda_{Bn} \leq \lambda_{pn}$, 则决策规则可简化为

- P) 如果 $P(D_i | [x]) \geq \alpha$ 成立,则 $x \in POS(X)$,
- B) 如果 $\beta < P(D_i | [x]) < \alpha$ 成立,则 $x \in BND(X)$,
- N) 如果 $P(D_i | [x]) \leq \beta$ 成立,则 $x \in NEG(X)$,

其中: $\alpha = \frac{(\lambda_{pn} - \lambda_{Bn})}{(\lambda_{pn} - \lambda_{Bn}) + (\lambda_{Bp} - \lambda_{pp})}$; $\beta = \frac{(\lambda_{Bn} - \lambda_{Nn})}{(\lambda_{Bn} - \lambda_{Nn}) + (\lambda_{Np} - \lambda_{Bp})}$ 。

若 $x \in POS(X)$, 则表示接受 x ; 若 $x \in BND(X)$, 则表示对 x 延迟决策; 若 $x \in NEG(X)$, 则表示拒绝 x 。

1.2 可拓集与可拓关联函数

定义1^[23](可拓集合) 设 U 为论域, u 为 U 中任一元素, k 是 U 在实数 \mathbf{R} 的一个映射, $T(T_u, T_k, T_u)$ 是给定的变换, 称

$$\tilde{E}(T) = \{(u, y, y') \mid u \in T_u U, y = k(u) \in \mathbf{R}, y' = T_k k(T_u u) \in \mathbf{R}\}, \quad (1)$$

为论域 U 上的一个可拓集合, 其中 $y = k(u)$ 是可拓集 $\tilde{E}(T)$ 的关联函数, $y' = T_k k(T_u u)$ 为可拓集 $\tilde{E}(T)$ 的可拓函数。

定义2^[13](正域为有限区间的关联函数) 若正域 $X_0 = [a, b]$ 为有限区间, 节域 $X (X \neq \mathbf{R})$ 为实数域上任何区间且 $X_0 \subset X$, 非端点 $x_0 \in X_0$, 记 X_0 和 X 的公共端点为 x_0 (若无公共端点, 则为空), 则对任意实数 $x \neq x_0$, 则 x 关于 X_0 和 X 且在 X_0 的非中点 x_0 取得最大值的一般关联函数为

$$k(x) = \begin{cases} \rho(x, x_0, X_0)/D(x, X_0, X) - 1, \rho(x, X) \geq \rho(x, X_0), & x \notin X_0, \\ \rho(x, x_0, X_0)/D(x, X_0, X), & \text{其他,} \end{cases} \quad (2)$$

其中: $D(x, X_0, X)$ 为 x 关于 X_0 和 X 的一般位置值。

定义 3^[13](正域为无限区间的关联函数) 设无限区间 X_0 为正域, 节域 $X (X \neq \mathbf{R})$ 为实数域上任何区间且 $X_0 \subset X$, 非端点 $x_0 \in X_0$, 则对实数域上的任意点 x 关于无限区间套 X_0 和 X 且在点 x_0 取最大值的关联函数为

$$k(x) = \begin{cases} \rho(x, x_0, X_0)/D(x, X_0, X) - 1, \rho(x, X) \geq \rho(x, X_0), & x \notin X_0, \\ \rho(x, x_0, X_0)/D(x, X_0, X), & \text{其他,} \end{cases} \quad (3)$$

其中: $D(x, X_0, X)$ 为 x 关于区间套 X_0 和 \mathbf{R} 的一般位置值。

2 基于可拓关联度比的三支决策模型的构建

2.1 评价函数的构造

三支决策是在概率粗糙集的基础上提出的, 为粗糙集的三个域提供合理语义解释, 三支决策中评价函数的选取直接影响决策结果, 因此评价函数的选取至关重要。可拓关联度是可拓评价与决策的主要工具, 其大小关系着决策的结果。因此, 关联函数是一种重要而有效的评价函数。由于关联度值大部分为负, 而且取绝对值后大于 1, 评价函数会影响决策, 因此结合模糊集中隶属函数的思想, 定义可拓关联度比如下。

定义 4 设 $K_j(V_i)$ 为评价指标 j 关于等级 i 的关联度, 则评价指标 j 关于等级 i 的关联度比为

$$P_i(K_j) = \begin{cases} \frac{|K_j(X_i)| + |K_j(X_{\min})|}{|K_j(X_{\max})| + |K_j(X_{\min})|}, & K_j(X_{\min}) < 0, \\ \frac{K_j(X_i)}{K_j(X_{\max})}, & K_j(X_{\min}) > 0, \end{cases} \quad (4)$$

其中: $\frac{K_j(X_{\min})}{K_j(X_{\max})} \leq P_i(K_j) \leq 1; K_j(X_{\max}) = \max_{1 \leq i \leq n} \{K_j(X_i)\}; K_j(X_{\min}) = \min_{1 \leq i \leq n} \{K_j(X_i)\}$ 。

由于关联度比的取值范围为 $(0, 2)$, 所以需要关联度比进行归一化处理。由关联度比的定义可知, 指标关于等级的关联度比和非关联度比的和为 1。故归一化的关联度比适合作为三支决策的评价函数来进行决策。

2.2 决策规则

将表 1 的代价函数作为损失, 将关联度比作为评价函数构建三支决策模型。

对于特定对象 j , 采取一个决策行动时的期望代价为

$$\begin{cases} T(a_p | j) = \lambda_{PP} \times P_i(K_j) + \lambda_{PN} \times \overline{P_i(K_j)}, \\ T(a_B | j) = \lambda_{BP} \times P_i(K_j) + \lambda_{BN} \times \overline{P_i(K_j)}, \\ T(a_N | j) = \lambda_{NP} \times P_i(K_j) + \lambda_{NN} \times \overline{P_i(K_j)}. \end{cases} \quad (5)$$

由于 $P_i(K_j) + \overline{P_i(K_j)} = 1$, 且 $\lambda_{PP} \leq \lambda_{BP} \leq \lambda_{NP}, \lambda_{NN} \leq \lambda_{BN} \leq \lambda_{PN}$, 则式(5)可改写为

$$\begin{cases} T(a_p | j) = (\lambda_{PP} - \lambda_{PN}) \times P_i(K_j) + \lambda_{PN}, \\ T(a_B | j) = (\lambda_{BP} - \lambda_{BN}) \times P_i(K_j) + \lambda_{BN}, \\ T(a_N | j) = (\lambda_{NP} - \lambda_{NN}) \times P_i(K_j) + \lambda_{NN}. \end{cases} \quad (6)$$

从式(6)可知, 代价目标函数 $T(a_p | j)$ 、 $T(a_B | j)$ 、 $T(a_N | j)$ 的几何意义是以 $P_i(K_j)$ 为变量, 相关代价函数组合为斜率和截距的线性函数, 因此三支分类的阈值 α, β 可以看作是上述三个代价目标线性函数之间交集的横坐标 $P_i(K_j)$ 的取值。显然 α, β 可以看作是 $T(a_p | j)$ 与 $T(a_B | j)$ 、 $T(a_B | j)$ 与 $T(a_N | j)$ 交点的 $P_i(K_j)$ 处取值, $T(a_p | j)$ 与 $T(a_N | j)$ 的交点记为 γ , 如图 1 所示, 图中的 T_p, T_B, T_N 分别代表式(6)中的 $T(a_p | j)$ 、 $T(a_B | j)$ 、 $T(a_N | j)$, T_* 表示 T_p, T_B, T_N 中的任意量。

根据贝叶斯风险决策过程, 最小代价决策规则如下。

P) 如果 $T(a_p | j) \leq T(a_B | j)$ 且 $T(a_p | j) \leq T(a_N | j)$ 成立, 则 $j \in POS(X)$;

B) 如果 $T(a_B | j) \leq T(a_p | j)$ 且 $T(a_B | j) \leq T(a_N | j)$ 成立, 则 $j \in BND(X)$;

N) 如果 $T(a_N | j) \leq T(a_p | j)$ 且 $T(a_N | j) \leq T(a_B | j)$ 成立, 则 $j \in NEG(X)$ 。

由于 $0 \leq \beta < \gamma < \alpha \leq 1$, 则决策规则可简化为

P) 如果 $P_i(K_j) \geq \alpha$ 成立, 则 $j \in POS(X)$;

B) 如果 $\beta < P_i(K_j) < \alpha$ 成立, 则 $j \in BND(X)$;

N) 如果 $P_i(K_j) \leq \beta$ 成立, 则 $j \in NEG(X)$,

$$\text{其中: } \alpha = \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})};$$

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

2.3 模型步骤

步骤 1 确定待评物元、经典域和节域。将给定数据集的指标根据以往研究进行分级, 并确定每个等级具体取值范围。经典域是每个等级中每个指标具体取值范围, 节域表示所有等级中每个指标的取值范围, 待评物元表示需要评价的物元, 根据分级确定经典域和节域。

$$\text{经典域表示为 } D_t = (C_j, V_{oj}) = \begin{bmatrix} c_1 & (a_{o1}, b_{o1}) \\ c_2 & (a_{o2}, b_{o2}) \\ \vdots & \vdots \\ c_n & (a_{on}, b_{on}) \end{bmatrix},$$

式中: C_j 表示 c_1, c_2, \dots, c_n 这 n 个指标的集合; V_{oj} 表示 $(a_{oj}, b_{oj}), j = 1, 2, \dots, n$ 的集合; 而 (a_{oj}, b_{oj}) 为等级 t 中样本关于第 j 个指标的取值范围。

$$\text{节域表示为 } D_p = (C_j, V_{pj}) = \begin{bmatrix} c_1 & (a_{p1}, b_{p1}) \\ c_2 & (a_{p2}, b_{p2}) \\ \vdots & \vdots \\ c_n & (a_{pn}, b_{pn}) \end{bmatrix},$$

式中: C_j 表示 c_1, c_2, \dots, c_n 这 n 个指标的集合; V_{pj} 表示 $(a_{pj}, b_{pj}), j = 1, 2, \dots, n$ 的集合, 而 (a_{pj}, b_{pj}) 为所有等级中样本关于第 j 个指标的取值范围。

$$\text{待评物元可以将影响因素的作用参数用物元表示 } D_o = (C_j, V_i) = \begin{bmatrix} c_1 & x^*(i, 1) \\ c_2 & x^*(i, 2) \\ \vdots & \vdots \\ c_n & x^*(i, n) \end{bmatrix},$$

式中: C_j 表示 c_1, c_2, \dots, c_n 这 n 个指标的集合; V_i 表示 $x^*(i, j), j = 1, 2, \dots, n$ 的集合; $x^*(i, j)$ 为待评样本 i 关于第 j 个指标的取值。

步骤 2 根据关联度公式计算各指标的关联度。

$$K_j(x_i) = \begin{cases} \frac{-\rho(x^*(i, j), V_{oj})}{|V_{oj}|}, & x^*(i, j) \in V_{oj} \\ \frac{\rho(x^*(i, j), V_{oj})}{\rho(x^*(i, j), V_{pj}) - \rho(x^*(i, j), V_{oj})}, & x^*(i, j) \notin V_{oj} \end{cases}, \quad (7)$$

式中: $K_j(x_i)$ 表示第 i 个样本、第 j 个指标的关联度, 各个指标值到各评价等级范围值的距离为

$$\rho(x^*(i, j), V_{oj}) = \left| x^*(i, j) - \frac{(a_{oj} + b_{oj})}{2} \right| - \frac{(b_{oj} - a_{oj})}{2},$$

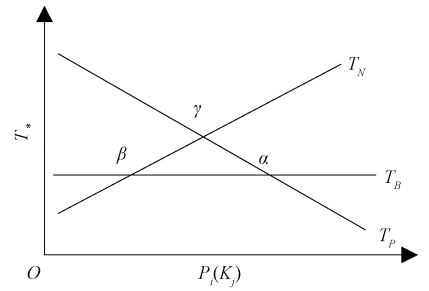


图 1 T_p, T_B, T_N 之间的关系示意图

Figure 1 Schematic diagram of the relationship among T_p, T_B, T_N

$$\rho(x^*(i,j), V_{pj}) = |x^*(i,j) - \frac{(a_{pj} + b_{pj})}{2}| - \frac{(b_{pj} - a_{pj})}{2}.$$

步骤 3 由可拓关联度比公式(4)计算每个指标关于每个等级的关联度比,并做归一化处理。

步骤 4 由熵权可拓评价法计算城市所属等级。

1) 由于各指标的计量单位并不统一,而且正向指标和负向指标表示的含义不同,正向指标数值越高越好,负向指标数值越低越好,因此,对于正向、负向指标采用不同的运算进行数据标准化处理。同时为使后面计算的指标熵值有意义,即标准化后的指标值不为 0,故标准化时后面要加 1,取值范围为[1,2]。

$$\text{正向指标 } x'(i,j) = \frac{x^*(i,j) - x_{\min}(i,j)}{x_{\max}(i,j) - x_{\min}(i,j)} + 1, \text{ 负向指标 } x'(i,j) = \frac{x_{\max}(i,j) - x^*(i,j)}{x_{\max}(i,j) - x_{\min}(i,j)} + 1,$$

式中: $x^*(i,j)$ 表示第 i 个样本第 j 个评价指标值; $x_{\max}(i,j) = \max\{x(1,j), x(2,j), \dots, x(n,j)\}$, 即所有样本中第 j 个评价指标的最大值; $x_{\min}(i,j) = \min\{x(1,j), x(2,j), \dots, x(n,j)\}$, 即所有样本中第 j 个评价指标的最小值。

2) 由熵权法确定属性的权重。第 j 项指标下第 i 个样本值占该指标的比重为 $P_{ij} = \frac{x'(i,j)}{\sum_{i=1}^n x'(i,j)}$; 计算第 j

项指标的熵值 $e_j = -\frac{1}{\ln(n)} \sum_{i=1}^n P_{ij} \ln(P_{ij})$; 计算信息熵冗余度 $d_j = \frac{1 - e_j}{m - \sum_{j=1}^m e_j}$; 各项指标的权重为 $w_j = \frac{d_j}{\sum_{j=1}^m d_j}$, n

表示样本数, m 表示指标数。

3) 确定所属等级。城市关于众指标的综合关联度用公式(8)计算

$$K_i(d_0) = \sum_{j=1}^m w_j K_i(x_j), \quad (8)$$

若

$$K_{i_0}(d_0) = \max K_i(x_i), \quad (9)$$

则待评样本的等级为 i_0 。式中: $K_{i_0}(d_0)$ 表示待评样本的关联度; $\max K_i(x_i)$ 表示待评样本关于所有等级关联度的最大值。

步骤 5 结合数据集特点,经过多次实验,并根据不同阈值划分后每个域中元素个数情况,可得选取阈值 $\alpha = 0.28, \beta = 0.22$ 比较合适,因此选取 $\alpha = 0.28, \beta = 0.22$ 作为阈值。由基于关联度比的三支决策的决策规则,对每个等级的指标进行域的划分,得到基于每个等级的三个域划分。

步骤 6 分析域的划分结果,对城市水资源承载力系统给出优化方向与建议。

2.4 算法分析

步骤 1 中确定待评物元时,用到两层 for 循环,一层步长为指标数 n , 一层步长为样本数 m , 最终得到每一待评样本关于所有指标的取值,其时间复杂度为 $O(mn)$; 步骤 2、3 中计算各指标的关联度与关联度比时,也用到步长分别为 m 和 n 的两层 for 循环,其时间复杂度也均为 $O(mn)$; 步骤 4 中仍通过步长为 m 和 n 的两层 for 循环计算出正、负项指标及属性权重,其时间复杂度也为 $O(mn)$; 因此最终得到的时间复杂度为 $O(mn)$ 。

3 案例分析

选取黑龙江省 13 市水资源承载力数据,包括水土资源和承载力两个指标体系,共 19 个指标,利用本文模型对哈尔滨市水资源承载力进行评价,并给出优化建议。

步骤 1 根据黑龙江省 13 市的水资源承载力数据,结合以往研究和专家经验,将指标承载力等级设定为 4 级,即 I(强)、II(一般)、III(弱)、IV(很弱)。各等级的经典域和节域为

$$D_t = \begin{bmatrix} c_1 & [0,0.1] & [0.1,0.15] & [0.15,0.6] & [0.6,1.5] \\ c_2 & [0,0.02] & [0.02,0.05] & [0.05,0.09] & [0.09,0.19] \\ c_3 & [40,70] & [70,100] & [100,130] & [130,180] \\ c_4 & [0.1,0.3] & [0.3,0.5] & [0.5,0.7] & [0.7,0.9] \\ c_5 & [-0.4,-0.019] & [-0.019,-0.008] & [-0.008,-0.006] & [-0.006,0] \\ c_6 & [0,0.2] & [0.2,0.3] & [0.3,0.4] & [0.4,0.5] \\ c_7 & [38,52] & [52,57] & [57,64] & [64,68] \\ c_8 & [0.05,0.09] & [0.09,0.11] & [0.11,0.13] & [0.13,0.15] \\ c_9 & [66,110] & [110,130] & [130,150] & [150,170] \\ c_{10} & [100,500] & [500,1500] & [1500,5000] & [5000,6500] \\ c_{11} & [2.1万,2.5万] & [2.5万,3.5万] & [3.5万,5.5万] & [5.5万,10万] \\ c_{12} & [0.02,0.04] & [0.04,0.05] & [0.05,0.06] & [0.06,0.08] \\ c_{13} & [0.8,1] & [0.6,0.8] & [0.4,0.6] & [0.2,0.4] \\ c_{14} & [21,28] & [18,21] & [14,18] & [6,14] \\ c_{15} & [0.6万,1.3万] & [0.3万,0.6万] & [0.1万,0.3万] & [0.05万,0.1万] \\ c_{16} & [700,800] & [600,700] & [500,600] & [200,500] \\ c_{17} & [0.59,0.62] & [0.54,0.59] & [0.48,0.54] & [0.3,0.48] \\ c_{18} & [0.7,0.9] & [0.6,0.7] & [0.4,0.6] & [0.1,0.4] \\ c_{19} & [0.9,0.92] & [0.88,0.9] & [0.86,0.88] & [0.83,0.86] \end{bmatrix},$$

$$D_p = \begin{bmatrix} c_1 & [0,1.5] \\ c_2 & [0,0.19] \\ c_3 & [40,180] \\ c_4 & [0.1,0.9] \\ c_5 & [-0.04,0] \\ c_6 & [0,0.5] \\ c_7 & [38,68] \\ c_8 & [0.05,0.15] \\ c_9 & [66,170] \\ c_{10} & [100,6500] \\ c_{11} & [2.1万,10万] \\ c_{12} & [0.02,0.08] \\ c_{13} & [0.2,1] \\ c_{14} & [6,28] \\ c_{15} & [0.05万,1.3万] \\ c_{16} & [200,800] \\ c_{17} & [0.3,0.62] \\ c_{18} & [0.1,0.9] \\ c_{19} & [0.83,0.92] \end{bmatrix},$$

其中: D_t 矩阵表示经典域矩阵; D_p 矩阵表示节域矩阵。

步骤2 由关联度公式(7)及关联度比公式(4)计算哈尔滨市各指标关于每个等级的关联度及关联度比,见表2。

表2 哈尔滨市各指标关于每个等级的关联度及关联度比

Table 2 Correlation degree and correlation degree ratio of each index in Harbin

评价指标	评价函数	I	II	III	IV
供水模数/($10^4 \cdot \text{m}^3 \cdot \text{km}^{-2}$)	关联度	-0.484 6	-0.465 5	-0.198 3	0.198 3
	关联度比	0.295 0	0.289 2	0.207 9	0.207 9
水资源开发利用/%	关联度	-0.305 5	0.476 3	-0.285 8	-0.603 2
	关联度比	0.222 5	0.264 3	0.217 7	0.295 4
人口密度/(人 $\cdot \text{km}^{-2}$)	关联度	-0.999 4	-0.999 1	-0.998 6	0.001 4
	关联度比	0.285 7	0.285 7	0.285 6	0.143 1
城镇化率/%	关联度	-0.327 0	0.055 0	-0.027 5	-0.351 7
	关联度比	0.313 1	0.187 6	0.174 9	0.324 4
人口自然增长率/%	关联度	-0.611 6	-0.077 5	0.310 2	-0.157 5
	关联度比	0.339 5	0.191 3	0.255 8	0.213 5
第一产业比重/%	关联度	0.415 0	-0.415 0	-0.610 0	-0.707 5
	关联度比	0.225 5	0.225 5	0.264 7	0.284 3
Remanenکو 潜在蒸发量	关联度	-0.542 3	-0.334 2	0.474 8	-0.312 2
	关联度比	0.283 0	0.228 7	0.265 4	0.222 9
干燥度	关联度	-0.907 2	-0.860 8	-0.721 6	0.278 4
	关联度比	0.283 6	0.276 4	0.254 6	0.185 3
人均生活用水量/($\text{m}^3 \cdot \text{人}^{-1}$)	关联度	-0.460 0	-0.190 0	0.380 0	-0.276 8
	关联度比	0.292 4	0.206 6	0.266 9	0.234 1
GDP 总量/亿元	关联度	-0.959 5	-0.951 4	-0.838 1	0.161 9
	关联度比	0.284 3	0.283 1	0.266 4	0.166 2
人均 GDP/(万元 $\cdot \text{人}^{-1}$)	关联度	-0.540 3	-0.469 5	-0.233 8	0.233 8
	关联度比	0.297 0	0.277 5	0.212 7	0.212 7
GDP 增长率/%	关联度	-0.675 0	-0.566 7	-0.350 0	0.350 0
	关联度比	0.290 8	0.267 5	0.220 8	0.220 8
地表水资源比例/%	关联度	0.338 8	-0.338 8	-0.669 4	-0.779 6
	关联度比	0.213 2	0.213 2	0.276 3	0.297 3
产水模数/($10^4 \cdot \text{m}^3 \cdot \text{km}^{-2}$)	关联度	0.114 1	-0.114 1	-0.379 9	-0.557 1
	关联度比	0.197 8	0.197 8	0.276 1	0.328 3
人均水资源量/($10^4 \text{ m}^3 \cdot \text{人}^{-1}$)	关联度	-0.870 6	-0.715 4	0.105 8	-0.229 2
	关联度比	0.322 2	0.293 5	0.180 7	0.203 5
降雨量/mm	关联度	-0.438 4	-0.298 0	-0.064 0	0.064 0
	关联度比	0.334 9	0.281 3	0.191 9	0.191 9
第三产业比重/%	关联度	0.196 4	-0.803 6	-0.926 4	-0.957 9
	关联度比	0.171 9	0.262 3	0.280 6	0.285 3
森林覆盖率/%	关联度	-0.402 0	-0.282 4	0.294 0	-0.140 8
	关联度比	0.294 8	0.251 0	0.255 2	0.199 0
污水处理能力/%	关联度	0.500 0	-0.500 0	-0.750 0	-0.833 3
	关联度比	0.225 4	0.225 4	0.267 6	0.281 7

步骤3 由公式(8)~(9)计算可得,目前哈尔滨市水资源承载力处于II级。

步骤4 由决策规则得每个等级指标三个域的划分为

$$\begin{aligned}
 \text{I 级: } & \begin{cases} POS = \{1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 15, 16, 18\} \\ BND = \{2, 6, 19\} \\ NEG = \{13, 14, 17\} \end{cases}; \quad \text{II 级: } \begin{cases} POS = \{1, 3, 10, 15, 16\} \\ BND = \{2, 6, 7, 8, 11, 12, 17, 18, 19\}; \\ NEG = \{4, 5, 9, 13, 14\} \end{cases} \\
 \text{III 级: } & \begin{cases} POS = \{3, 17\} \\ BND = \{5, 6, 7, 8, 9, 10, 12, 13, 14, 18, 19\}; \\ NEG = \{1, 2, 4, 11, 15, 16\} \end{cases}; \quad \text{IV 级: } \begin{cases} POS = \{2, 4, 6, 13, 14, 17, 19\} \\ BND = \{7, 9, 12\} \\ NEG = \{1, 3, 5, 8, 10, 11, 15, 16, 18\} \end{cases}。
 \end{aligned}$$

步骤5 给出指标优化方向与建议。由熵权可拓评价法可知哈尔滨市目前水资源承载力处于Ⅱ级水平,根据每个等级指标三个域的划分以及水资源承载力现状可知,哈尔滨市需要优化指标有城镇化率、人口自然增长率、Remanenko 潜在蒸发量、干燥度、人均生活用水量、GDP 总量、人均 GDP、GDP 增长率、森林覆盖率。而在这些指标中,城镇化率、人口自然增长率的关联度比的值较大,在Ⅰ级正域中所处层次较高,所以应先优化这些指标,这为政府决策和管理提供依据。

4 结论

本文将可拓集与三支决策结合建立基于关联度比的三支决策模型并应用于决策哈尔滨市水资源承载力。第一,基于关联度比构建三支决策模型,为三个域的构造提供合理的解释,丰富三支决策和可拓评价理论;第二,模型可以在每一级上划分三个域,定量给出评级改变时各指标的变化,并根据关联度比的值,可以量化同一域中属性的重要程度,使属性分出层次,既实现动态决策,又细化决策;第三,将模型应用于哈尔滨市水资源承载力评价中可知,若希望承载力由原来的Ⅱ级提升为Ⅰ级,则应优化城镇化率、人口自然增长率等指标,这为决策与管理提供理论依据。在阈值变化时,关于每个等级三个域如何变化,以及建立自适应的阈值计算方法将成为今后研究的重点。

参考文献:

- [1] 姚一豫,祁建军,魏玲. 基于三支决策的形式概念分析、粗糙集与粒计算[J]. 西北大学学报(自然科学版), 2018, 48(4):477-487.
YAO Y Y, QI J J, WEI L. Formal concept analysis, rough set and granular computing based on three-way decision[J]. Journal of northwestern university (natural science edition), 2018, 48(4):477-487.
- [2] 刘久兵,王天行,周献中,等. 基于勾股模糊集评价的三支决策方法[J]. 模式识别与人工智能, 2019, 32(12): 1080-1092.
LIU J B, WANG T X, ZHOU X Z, et al. Three-way decisions method based on evaluations of pythagorean fuzzy sets[J]. Pattern recognition and artificial intelligence, 2019, 32(12): 1080-1092.
- [3] 刘久兵,周献中,李华雄,等. 基于直觉模糊相似度的直觉模糊三支决策方法[J]. 系统工程理论与实践, 2019, 39(6): 1550-1564.
LIU J B, ZHOU X Z, LI H X, et al. An intuitionistic fuzzy three-way decision method based on intuitionistic fuzzy similarity degrees[J]. Systems engineering-theory & practice, 2019, 39(6): 1550-1564.
- [4] 薛占熬,韩丹杰,吕敏杰,等. 一种新的基于粒度重要度的三支决策模型[J]. 计算机科学, 2019, 46(2): 236-241.
XUE Z A, HAN D J, LU M J, et al. New three-way decisions model based on granularity importance degree[J]. Computer science, 2019, 46(2): 236-241.
- [5] 陈玉金,李续武,邢瑞康. 基于证据理论的三支决策模型[J]. 计算机科学, 2018, 45(6): 241-246, 283.
CHEN Y J, LI X W, XING R K. Three-way decisions model based on evidence theory[J]. Computer science, 2018, 45(6): 241-246, 283.
- [6] 岳玉琦,张楠,童向荣,等. 混合决策信息系统的模糊效用三支决策模型[J]. 郑州大学学报(理学版), 2020, 52(1): 24-32.
YUE W Q, ZHANG N, TONG X R, et al. Fuzzy utility three-way decisions model in hybrid decision information systems[J]. Journal of Zhengzhou university (natural science edition), 2020, 52(1): 24-32.
- [7] 张春英,乔鹏,王立亚,等. 一种基于 PS-粗糙集的动态三支决策算法[J]. 郑州大学学报(理学版), 2019, 51(2): 72-77.
ZHANG C Y, QIAN P, WANG L Y, et al. A dynamic three-way decision algorithm based on PS-rough set[J]. Journal of Zhengzhou university (natural science edition), 2019, 51(2): 72-77.
- [8] 孙国营,孙新杰,霍兴赢,等. 基于三支决策改进 TOPSIS 的水资源承载力评价[J]. 节水灌溉, 2019(12): 77-81.
SUN G Y, SUN X J, HUO X Y, et al. Evaluation of water resources carrying capacity based on three-branch decision making improved TOPSIS method[J]. Water saving irrigation, 2019(12): 77-81.
- [9] 谢秦,张清华,王国胤. 基于相似度量的自适应三支垃圾邮件过滤器[J]. 计算机研究与发展, 2019, 56(11): 2410-2423.
XIE Q, ZHANG Q H, WANG G Y. An adaptive three-way spam filter with similarity measure[J]. Journal of computer

- research and development, 2019, 56(11): 2410-2423.
- [10] 武慧琼, 张素兰, 张继福, 等. 一种基于三支决策的花卉图像分类[J]. 小型微型计算机系统, 2019, 40(7): 1558-1563.
WU H Q, ZHANG S L, ZHANG J F, et al. Classification method of flower images based on three-way decisions[J]. Journal of Chinese computer systems, 2019, 40(7): 1558-1563.
- [11] 杨岚心, 苏健民, 欧长坤. 基于三支决策的图像过渡区提取与分割方法[J]. 黑龙江大学自然科学学报, 2019, 36(2): 219-226.
YANG L X, SU J M, OU C K. Method of image transition region extraction and segmentation based on three-way decision[J]. Journal of natural science of Heilongjiang university, 2019, 36(2): 219-226.
- [12] 杨春燕, 蔡文. 可拓学与矛盾问题智能化处理[J]. 科技导报, 2014, 32(36): 15-20.
YANG C Y, CAI W. Extensics and intelligent processing of contradictory problems[J]. Science & technology review, 2014, 32(36): 15-20.
- [13] 李桥兴, 杨春燕. 可拓集无量纲一维关联函数[J]. 系统工程, 2014, 32(11): 154-158.
LI Q X, YANG C Y. Dimensionless one-dimensional dependent function in extension set[J]. Systems engineering, 2014, 32(11): 154-158.
- [14] 杨春燕, 蔡文. 可拓集中关联函数的研究进展[J]. 广东工业大学学报, 2012, 29(2): 7-14.
YANG C Y, CAI W. Recent research progress in dependent functions in extension sets[J]. Journal of Guangdong university of technology, 2012, 29(2): 7-14.
- [15] 杨春燕. 多评价特征基元可拓集研究[J]. 数学的实践与认识, 2005, 35(9): 203-208.
YANG C Y. Study on the basic-element extension set of multi evaluating characteristics[J]. Mathematics in practice and theory, 2005, 35(9): 203-208.
- [16] 胡宝清, 何娟娟. 区间论域上的区间可拓集及其关联函数[J]. 广东工业大学学报, 2001, 18(1): 48-54.
HU B Q, HE J J. Interval extension set and correlation function on the interval[J]. Journal of Guangdong university of technology, 2001, 18(1): 48-54.
- [17] 安永林, 李佳豪, 赵丹, 等. 可拓法评估运营隧道衬砌结构健康实例[J]. 铁道科学与工程学报, 2020, 17(2): 422-428.
AN Y L, LI J H, ZHAO D, et al. Comprehensive extension assessment on tunnel structure health[J]. Journal of railway science and engineering, 2020, 17(2): 422-428.
- [18] 李倩倩, 董会忠. 基于熵权可拓模型的城镇化发展水平评价: 以山东半岛蓝色经济区为例[J]. 山东理工大学学报(自然科学版), 2020, 34(2): 58-63.
LI Q Q, DONG H Z. Evaluation on the development level of urbanization based on entropy-weighted extension model: an example of Shandong Peninsula Blue Economic Zone[J]. Journal of Shandong university of technology (natural science edition), 2020, 34(2): 58-63.
- [19] 郭嘉伟, 张军, 陈彦. 基于熵权可拓物元模型的会宁县水土资源承载力评价[J]. 甘肃农业大学学报, 2018, 53(6): 195-203.
GUO J W, ZHANG J, CHEN Y. Water and soil resources matching patterns and carrying capacity in Huining County[J]. Journal of Gansu agricultural university, 2018, 53(6): 195-203.
- [20] 屈豪. 基于物元可拓模型的流域水环境承载力研究[D]. 天津: 河北工业大学, 2017.
QU H. Study on water environment carrying capacity of watershed based on matter-element extension model[D]. Tianjin: Hebei University of Technology, 2017.
- [21] 谷红梅, 贾丽, 蒋晓辉, 等. 基于熵权物元可拓法的黑河中游水资源承载力评价[J]. 灌溉排水学报, 2016, 35(6): 87-92.
GU H M, JIA L, JIANG X H, et al. Evaluation of water resources bearing capacity based on entropy-weight and matter-element assessment methods in midstream of Heihe river[J]. Journal of irrigation and drainage, 2016, 35(6): 87-92.
- [22] 仇国芳, 王小宁. 基于三支决策的医院分级诊疗决策研究[J]. 河南师范大学学报(自然科学版), 2018, 46(3): 106-111.
QIU G F, WANG X N. Research on the decision-making of hospitals based on three decisions[J]. Journal of Henan normal university (natural science edition), 2018, 46(3): 106-111.
- [23] 屈云龙. 基于可拓理论的公路工程项目管理成熟度评价研究[D]. 西安: 西安建筑科技大学, 2017.
QU Y L. Research on maturity evaluation of highway project management based on the theory of extension set[D]. Xi'an: Xi'an University of Architecture and Technology, 2017.

Three-way Decision Model Based on Extension Correlation Ratio and Its Application

LI Min^{1,2,3}, YANG Yafeng^{1,2,3}, LEI Yu⁴, LI Lihong^{1,2,3}

- (1. *College of Science, North China University of Science and Technology, Tangshan, Hebei 063210, China;*
2. *Hebei Key Laboratory of Data Science and Application, Tangshan 063210, China;*
3. *Tangshan Key Laboratory of Data Science, Tangshan 063210, China;*
4. *College of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, China)*

Abstract: The construction and interpretation of the three domains had always been the research focuses of the three-way decision. The basic method of extension evaluation was integrated with the decision-making thought of “three parts and rule”, and the three decision-making model based on the extension correlation ratio were constructed. The construction principle and numerical characteristics of the correlation function were analyzed; the extension correlation ratio with more decision-making guiding significance was defined; and its possibility and rationality as an evaluation function in three decision-making were discussed. The normalized correlation ratio was taken as an evaluation function; and the indexes of each grade were divided into three domains; and three decision rules were constructed; the purpose of dynamic decision and mining optimized indexes were achieved. The model was applied to the decision-making process of water resources carrying capacity in Harbin. The results showed that Harbin’s water resources carrying capacity was in Grade II. The indexes of urbanization rate, natural population growth rate, Remanenکو potential evaporation, dryness, per capital domestic water consumption, GDP total, per capital GDP, GDP growth rate and forest coverage rate were urgent to optimize to realize the sustainable development of water resources system.

Key words: extension set; correlation ratio; three-way decision; extension evaluation; water resources carrying capacity

(责任编辑:王浩毅 方惠敏)