

基于最大熵原理的水环境模糊优化评价模型

王 栋¹, 朱元生²

(1. 南京大学城市与资源学系, 江苏 南京 210093 2. 河海大学水资源环境学院, 江苏 南京 210098)

摘要: 基于最大熵原理, 应用工程模糊集理论中的相对隶属度等概念, 综合考虑水环境评价的随机性和模糊性, 定义了不同的加权广义权距离来表征待评价水样与水质标准的差异, 建立了两个水环境相对隶属度模糊优化评价模型. 我国 12 个有代表性的湖库营养化评价实例表明: 所建模型的评价结果基本一致, 与调查结果相符, 对比单一考虑模糊性的模型, 所建两个模型评价结果的描述要细致一些, 且具有较小的 Shannon 熵, 即此次评价的不确定性小.

关键词: 最大熵原理, 相对隶属度, 模糊优化, 营养化, 水环境评价

中图分类号: X824 **文献标识码:** A **文章编号:** 1000-198X(2002)06-0056-05

熵是一个极其重要的物理量, 但又以抽象隐晦著称. 最大熵原理(Principle of Maximum Entropy, POME)成功解决了广为存在的不适定问题(ill-posed problem), 信息科学由此取得了长足进步. POME 现已在诸多研究领域得到广泛应用, 例如, 在水文水资源科学中^[1]. 随着我国经济的快速发展, 水环境评价问题日益引起广泛关注. 本文综合考虑水环境评价的随机性和模糊性, 基于最大熵原理, 应用工程模糊集理论中的相对隶属度等概念, 定义了不同的加权广义权距离来表征待评价水样与水质标准的差异, 以建立水环境相对隶属度模糊优化评价模型, 评价实例验证所建模型的有效性、准确性和可靠性.

1 基本概念和理论

1.1 信息熵及最大熵原理

在一个孤立系统中, 任何变化不可能导致熵的总值减少. 这是热力学第二定律的熵表述. 缘于此, 热力学第二定律亦被称为熵恒增定律. 1948 年, C. Shannon 创立了信息论, 他把通讯过程中信源讯号的平均信息量称为熵, 实现了信息熵的实际应用.

定义 将离散无记忆信源 X 的信息量定义为概率矢量的函数 $P = (p_1, p_2, \dots, p_n)$ 的函数:

$$H(X) = - \sum_{i=1}^n p_i \log_2 p_i = H(P) \quad (1)$$

$H(P)$ 称为信息熵的熵函数, 具有熵的全部基本性质(非负性、对称性、扩展性等).

熵恒增定律在信息熵领域则叫做最大熵原理. 1957 年, E. T. Jaynes 首次明确提出了 POME, 由此开创了 POME 发展之先河^[2, 3]. 根据 POME, 可以认为: 在不适定问题的所有的可行(可能)解中, 应选其中熵值最大的一个解. 熵最大意味着对因为数据不足而作的人为假定(人为添加信息)最小, 从而所获得的解最合乎自然, 偏差最小.

1.2 工程模糊集理论中的相对隶属度、隶属函数

定义 设论域 U 上的一个模糊子集/模糊概念 \tilde{A} , 分别赋给 \tilde{A} 处于共维差异的中介过渡段的两个极点以 0 和 1 的数, 构成 $[0, 1]$ 闭区间数的连续统. 在该连续统的数轴上建立参考系, 使其中的任两个点定为参考系坐标上的两极, 赋给参考系的两极以 0 和 1 的数, 并构成参考系 $[0, 1]$ 数轴上的参考连续统. 对于任意 $\mu \in U$, 在参考连续统上指定了一个数 $\mu_{\tilde{A}}(u)$ 称为 u 对 \tilde{A} 的相对隶属度, 下述映射称为 \tilde{A} 的相对隶属函数^[4]:

$$\mu_{\tilde{A}} : U \rightarrow [0, 1] \quad u \mapsto \mu_{\tilde{A}}(u) \in [0, 1] \quad (2)$$

应用模糊集合理论时,根据实际问题的不同性质和要求,常常只需确定相对隶属度、隶属函数,计算相对隶属度、隶属函数远比确定绝对隶属度、隶属函数容易,更为重要的是,可以在基本理论与概念上消除在应用领域长期存在的所谓确定隶属度、隶属函数中的‘主观任意性’的困惑,这种困惑在很大程度上影响着模糊集合论的应用,尤其在工程领域的应用。

2 基于 POME 的水环境相对隶属度模糊优化评价模型

2.1 建模前期资料基本处理

设水环境评价标准分级数为 c ,评价指标有 m 项,各指标分级标准浓度值为 y_{ih} ,则有水环境评价分级标准浓度矩阵 $Y = [y_{ih}]_{m \times c}$ 。现有 n 个待评价水样,每一水样有 m 项评价指标的实测值 x_{ij} ,则有待评价水样实测浓度矩阵 $X = [x_{ij}]_{m \times n}$ 。若规定 1 级评价标准中评价指标 i 的标准浓度 y_{i1} 对于某评价概念 \tilde{A} 的相对隶属度 $s_{i1} = 0$; c 级评价标准中评价指标 i 的标准浓度 y_{ic} 对于该评价概念的相对隶属度 $s_{ic} = 1$,则可根据式 (3) 确定 h 级评价标准中指标 i 的标准浓度 y_{ih} 的相对隶属度

$$s_{ih} = \frac{y_{ih} - y_{i1}}{y_{ic} - y_{i1}} \quad (3)$$

从而将水环境分级标准浓度矩阵 Y 变换为分级标准浓度相对隶属度矩阵 $\tilde{S} = [s_{ih}]_{m \times c}$ 。

将指标值越大评价等级越大的指标实测浓度值 x_{ij} 转化为相应的相对隶属度

$$r_{ij} = \begin{cases} 1 & x_{ij} > y_{ic} \\ \frac{x_{ij} - y_{i1}}{y_{ic} - y_{i1}} & y_{i1} \leq x_{ij} \leq y_{ic} \\ 0 & x_{ij} < y_{i1} \end{cases} \quad (4)$$

相反地,对于指标值越小评价等级越大的指标实测浓度值 x_{ij} 而言,以式 (5) 进行转化

$$r_{ij} = \begin{cases} 1 & x_{ij} < y_{ic} \\ \frac{x_{ij} - y_{i1}}{y_{ic} - y_{i1}} & y_{i1} \geq x_{ij} \geq y_{ic} \\ 0 & x_{ij} > y_{i1} \end{cases} \quad (5)$$

从而将待评价水体实测浓度矩阵 X 变换为水体实测浓度相对隶属度矩阵 $\tilde{R} = [r_{ij}]_{m \times n}$ 。

此外,各指标对模糊概念 \tilde{A} 的影响也不同,以指标权向量表示各指标自身的影响权重

$$v = (v_1, v_2, \dots, v_m) \quad \sum_{i=1}^m v_i = 1 \quad (6)$$

综合考虑超标权重和评价指标权重,可建立 n 个样本 m 项指标的综合权重矩阵 $A = [v_i v_{ij}]_{m \times n}$ 。将之按列归一化,得到环境指标综合权重矩阵 $W = [w_{ij}]_{m \times n}$ 。矩阵 W 既考虑了指标的超标权重,又考虑了评价指标自身权重,综合表达了各项指标对模糊概念 \tilde{A} 的作用大小或影响程度。

待评价水体样本以不同的相对隶属度隶属于评价各级标准。设 n 个样本对于 c 级水的相对隶属度矩阵为 $U = [u_{hj}]_{c \times n}$ 。模糊矩阵 \tilde{U} 的约束条件为

$$\sum_{h=1}^c u_{hj} = 1 \quad u_{hj} \geq 0, \quad j = 1, 2, \dots, n \quad (7)$$

显然,这样的矩阵有无数个。进行水环境评价的目的就是确定满足式 (7) 的最优模糊矩阵 \tilde{U} 。

2.2 模糊优化评价模型 I

由于水环境评价具有随机性和模糊性,因而 $[u_{hj}]_{c \times n}$ 的确定也具有随机性和模糊性等不确定性。其中,随机性包括了至少两方面的含义:其一是应用相对隶属度的概念进行评价过程中的随机性,其二就是实测指标观测过程中的观测误差的随机性。

现将 u_{hj} 视为第 j 个待评价水样属于第 h 级水的‘概率’,以 Shannon 信息熵来描述并比较水环境评价中的不确定性。

$$H_j = - \sum_{h=1}^c u_{hj} \ln u_{hj} \quad (8)$$

定义 以加权广义权距离 $^1D_{hj}$ 表示第 j 个待评价水样与第 h 级水的差异

$$^1D_{hj} = u_{hj}^1 d_{hj} = u_{hj} \left\{ \sum_{i=1}^m [w_{ij}(r_{ij} - s_{ih})]^p \right\}^{\frac{1}{p}} \tag{9}$$

式中 p 为距离参数, $p=1$ 为海明距离, $p=2$ 为欧式距离.

$[u_{hj}]_{c \times n}$ 的确定应使全体待评价水样与评价标准中各级水的加权广义距离之和最小, 即

$$\begin{aligned} \min_{u_{hj}} \quad & ^1D = \sum_{j=1}^n \sum_{h=1}^c u_{hj} \left\{ \sum_{i=1}^m [w_{ij}(r_{ij} - s_{ij})]^p \right\}^{\frac{1}{p}} \\ \text{s.t.} \quad & \sum_{h=1}^c u_{hj} = 1 \quad u_{hj} \geq 0 \quad j = 1, 2, \dots, n \end{aligned} \tag{10}$$

同时, 根据 POME $[u_{hj}]_{c \times n}$ 的确定还应该使得 Shannon 信息熵最大, 即

$$\begin{aligned} \max_{u_{hj}} \quad & H = \sum_{j=1}^n \left(- \sum_{h=1}^c u_{hj} \ln u_{hj} \right) \\ \text{s.t.} \quad & \sum_{h=1}^c u_{hj} = 1 \quad u_{hj} \geq 0 \quad j = 1, 2, \dots, n \end{aligned} \tag{11}$$

这是一个双目标规划问题. 为寻求这类问题的非劣解, 可应用加权法、约束法、混合法等^[5]. 现应用加权法, 为此, 构造如下单目标规划

$$\begin{aligned} \min_{u_{hj}} \quad & Y = ^1D + \frac{1}{\eta_1} H = \sum_{j=1}^n \sum_{h=1}^c \left\{ u_{hj} \left\{ \sum_{i=1}^m [w_{ij}(r_{ij} - s_{ih})]^p \right\}^{\frac{1}{p}} + \frac{1}{\eta_1} u_{hj} \ln u_{hj} \right\} \\ \text{s.t.} \quad & \sum_{h=1}^c u_{hj} = 1 \quad u_{hj} \geq 0 \quad j = 1, 2, \dots, n \end{aligned} \tag{12}$$

式中 η_1 为加权因子, 其最优解应满足 Kuhn-Tucker 条件.

最后解得

$$u_{hj} = \frac{\exp \left\{ - \eta_1 \left\{ \sum_{i=1}^m [w_{ij}(r_{ij} - s_{ih})]^p \right\}^{\frac{1}{p}} \right\}}{\sum_{h=1}^c \exp \left\{ - \eta_1 \left\{ \sum_{i=1}^m [w_{ij}(r_{ij} - s_{ih})]^p \right\}^{\frac{1}{p}} \right\}} \tag{13}$$

此即基于 POME 的水环境相对隶属度模糊优化评价模型 I.

2.3 模糊优化评价模型 II

定义 以加权广义权距离 $^2D_{hj}$ 来表示第 j 个待评价水样与第 h 级水的差异

$$^2D_{hj} = u_{hj}^2 d_{hj} = u_{hj} \left[\sum_{i=1}^m (w_{ij} |r_{ij} - s_{ih}|) \right] \tag{14}$$

同样, 将双目标规划问题构造为如下单目标规划

$$\begin{aligned} \min_{u_{hj}} \quad & Y = ^2D + \frac{1}{\eta_2} H = \sum_{j=1}^n \sum_{k=1}^c \left\{ u_{hj} \left[\sum_{i=1}^m (w_{ij} |r_{ij} - s_{ih}|) \right] + \frac{1}{\eta_2} u_{hj} \ln u_{hj} \right\} \\ \text{s.t.} \quad & \sum_{h=1}^c u_{hj} = 1 \quad u_{hj} \geq 0 \quad j = 1, 2, \dots, n \end{aligned} \tag{15}$$

式中 η_2 为加权因子.

最后可以得到

$$u_{hj} = \frac{\exp \left[- \eta_2 \sum_{i=1}^m (w_{ij} |r_{ij} - s_{ih}|) \right]}{\sum_{h=1}^c \exp \left[- \eta_2 \sum_{i=1}^m (w_{ij} |r_{ij} - s_{ih}|) \right]} \tag{16}$$

此即基于 POME 的水环境相对隶属度模糊优化评价模型 II. 对比式(16)和式(13), 实际上, 这里取的是绝对值距离.

3 应用实例

在湖库藻类繁殖的高峰期, 我国 12 个代表性湖库的 Chl-a, TP, TN, COD, SD 等评价指标的实测浓度和评

价标准参见文献^[6]. 由式(3)求得水环境分级标准浓度相对隶属度矩阵 \tilde{S} , 由式(4)(5)可求得待评价水体实测浓度相对隶属度矩阵 \tilde{R} . 文献[5]对水体营化评价指标的权重进行了分析研究. 本文亦采用这一权重, 可以求得环境指标综合权重矩阵 W .

由式(13)将有关参数 \tilde{S}, \tilde{R}, W 代入, 取 $p = 2$, 最后得到基于 POME 的水环境相对隶属度模糊优化评价模型 I:

$$\tilde{U}_1^T = \begin{matrix} & \begin{matrix} \text{贫} & \text{贫中} & \text{中} & \text{中富} & \text{富} & \text{重富} \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{matrix} & \begin{bmatrix} 0 & 0.1470 & 0.5918 & 0.2118 & 0.0487 & 0.0007 \\ 0 & 0.0050 & 0.7875 & 0.1815 & 0.0239 & 0.0021 \\ 0 & 0.0003 & 0.4777 & 0.4499 & 0.0717 & 0.0004 \\ 0 & 0.0001 & 0.3403 & 0.5928 & 0.0664 & 0.0003 \\ 0 & 0.0002 & 0.1866 & 0.4860 & 0.3264 & 0.0007 \\ 0 & 0.0003 & 0.0710 & 0.1610 & 0.7668 & 0.0009 \\ 0 & 0.0006 & 0.0132 & 0.0258 & 0.9595 & 0.0008 \\ 0 & 0.0014 & 0.0078 & 0.0176 & 0.9613 & 0.0118 \\ 0 & 0.0007 & 0.0065 & 0.0133 & 0.9775 & 0.0019 \\ 0 & 0 & 0 & 0 & 0.0002 & 0.9998 \\ 0 & 0 & 0 & 0 & 0.0012 & 0.9988 \\ 0 & 0 & 0 & 0 & 0.0003 & 0.9997 \end{bmatrix} \end{matrix} \quad (17)$$

由式(22)可得基于 POME 的水环境相对隶属度模糊优化评价模型 II:

$$\tilde{U}_2^T = \begin{matrix} & \begin{matrix} \text{贫} & \text{贫中} & \text{中} & \text{中富} & \text{富} & \text{重富} \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{matrix} & \begin{bmatrix} 0 & 0.1671 & 0.4902 & 0.2770 & 0.0655 & 0.0002 \\ 0 & 0.0274 & 0.6622 & 0.2875 & 0.0225 & 0.0004 \\ 0 & 0.0028 & 0.4936 & 0.4552 & 0.0484 & 0 \\ 0 & 0.0020 & 0.3234 & 0.6315 & 0.0430 & 0.0001 \\ 0 & 0.0017 & 0.1849 & 0.5747 & 0.2385 & 0.0002 \\ 0 & 0.0011 & 0.0752 & 0.2250 & 0.6982 & 0.0005 \\ 0 & 0.0004 & 0.0105 & 0.0321 & 0.9576 & 0.0003 \\ 0 & 0.0002 & 0.0024 & 0.0076 & 0.9849 & 0.0049 \\ 0 & 0.0002 & 0.0038 & 0.0116 & 0.9893 & 0.0005 \\ 0 & 0 & 0 & 0 & 0.0001 & 0.9999 \\ 0 & 0 & 0 & 0 & 0.0005 & 0.9995 \\ 0 & 0 & 0 & 0 & 0.0001 & 0.9999 \end{bmatrix} \end{matrix} \quad (18)$$

由表 1 可见, 对于我国 12 个有代表性的湖库, 本文所建基于 POME 的水环境相对隶属度模糊优化评价模型 I 和模型 II 的评价结果基本一致(仅水样 3 稍有差别), 与调查结果相符. 表中还列出了单一考虑模糊性的文[7]模型的评价结果, 对照可见: 对于水样 2, 3, 4 和水样 6, 本文两个模型的评价结果的描述要细致一些.

表 1 我国 12 个代表性湖库营养化程度评价结果

Table 1 Evaluated results of eutrophication of 12 representative lakes and reservoirs in China

水样序号	湖库名称	用水障碍调查结果	本文模型 I	本文模型 II	文[7]模型
1	邛海	无	中营养	中营养	中营养
2	洱海	无	中营养(偏中富)	中营养(偏中富)	中营养
3	博斯腾湖	无	中营养(偏中富)	中富营养(偏中)	中营养(偏中富)
4	于桥水库	局部出现水花, 对供水有影响	中富营养(偏中)	中富营养(偏中)	中富营养
5	磁湖	局部出现水花, 影响观瞻	中富营养(偏富)	中富营养(偏富)	中富营养(偏富)
6	巢湖	大面积水花, 夏季供水障碍	富营养(偏中富)	富营养(偏中富)	富营养
7	甘棠湖	出现水花, 影响观瞻	富营养	富营养	富营养
8	蘑菇湖	局部出现水花, 大量死鱼现象	富营养	富营养	富营养
9	杭州西湖	出现水花, 严重影响观瞻	富营养	富营养	富营养
10	南京玄武湖	全湖出现水花, 大量死鱼	重富营养	重富营养	重富营养
11	武汉墨水湖	全湖出现水花, 严重影响养殖	重富营养	重富营养	重富营养
12	广州东山湖	全湖出现水花, 严重影响观瞻	重富营养	重富营养	重富营养

由表2可见:在这次评价的12个水样之中,除了4、5、6水样之外,对于其他9个水样,本文所建的两个基于POME的模糊优化模型都比单一考虑模糊性的文[7]模型具有较小的Shannon熵,即依照本文所建两个模型进行这次评价的不确定性小,可靠性较高。

表2 各评价模型 Shannon 熵计算结果

Table 2 Shannon entropy of each evaluation model

水样序号	湖库名称	文[7]模型	本文模型 I	本文模型 II	水样序号	湖库名称	文[7]模型	本文模型 I	本文模型 II
1	邗海	1.3985	1.0730	1.1852	7	甘棠湖	0.2964	0.2017	0.2030
2	洱海	0.8292	0.6261	0.8186	8	蘑菇湖	0.5683	0.2091	0.0943
3	博斯腾湖	0.9128	0.9071	0.8706	9	杭州西湖	0.4186	0.1300	0.0944
4	于桥水库	0.6508	0.8607	0.8040	10	南京玄武湖	0.2542	0.0018	0.0008
5	磁湖	0.6631	1.0361	0.9851	11	武汉墨水湖	0.4804	0.0094	0.0043
6	巢湖	0.5798	0.6942	0.7921	12	广州东山湖	0.1742	0.0030	0.0008

参考文献:

- [1] 王栋, 朱元生. 最大熵原理在水文水资源科学中的应用[J]. 水科学进展, 2001, 12(3): 424~430.
 [2] Jaynes E T. Information theory and statistical mechanics[J]. Phys Rev, 1957, 106: 620~630.
 [3] Jaynes E T. Information theory and statistical mechanics[J]. Phys Rev, 1957, 108: 171~190.
 [4] 陈守煜. 工程模糊集理论与应用[M]. 北京: 国防工业出版社, 1998. 1~21.
 [5] 马振华. 现代应用数学手册: 运筹学与最优化理论卷[M]. 北京: 清华大学出版社, 1998. 307~370.
 [6] 舒金华. 我国湖泊富营养化程度评价方法的探讨[J]. 环境污染与防治, 1990(5): 2~7.
 [7] 陈守煜, 熊德琪. 湖泊富营养化评价模糊集理论与模式[J]. 湖泊科学, 1992(2): 144~152.

POME-based fuzzy optimal evaluation model of water environment

WANG Dong¹, ZHU Yuan-sheng²

(1. Dept. of Urban and Resources Sciences, Nanjing Univ., Nanjing 210093, China;
 2. College of Water Resources and Environment, Hohai Univ., Nanjing 210098, China)

Abstract: Based on the Principle of Maximum Entropy (POME) and the concept of relative membership degree in the theory of fuzzy set in engineering, and in consideration of the randomness and fuzziness of water resources evaluation, the difference is defined between water quality standard and the qualities of a sample evaluated and represented with different generalized weighted distances, and two fuzzy optimal models are developed for evaluation of the relative membership degree of water environment. The eutrophication evaluation of 12 representative lakes and reservoirs in China shows that the evaluated results by the two models are in good agreement, and consistent with investigations; compared with the results from the model only considering fuzziness, the results evaluated by the present models are sophisticated, and a small Shannon entropy is included, indicating less uncertainty of the evaluated results.

Key words: Principle of Maximum Entropy; relative membership degree; fuzzy optimization; eutrophication; water environment evaluation