

基于属性识别模型的湖泊富营养化评价

胡吉炜¹, 高军省^{1,2}

(1. 长江大学 地球环境与水资源学院, 湖北 武汉 430100; 2. 华北水利水电学院, 河南 郑州 450011)

摘要: 为合理地确定湖泊的富营养化状态,从属性识别理论出发,建立了湖泊富营养化评价的属性识别模型,并结合熵权法确定的权重系数对东昌湖的富营养化状态进行评价。评价结果为各湖区全年均属于富营养化级别,并且7月份的富营养化程度最高,12月份次之,4月份最低,这与集对分析法和模糊识别方法评价的结果一致。表明该模型的评价结果是合理的;同时该模型简便、易懂,可作为湖泊富营养化评价的一种可行方法。

关键词: 属性识别; 富营养化; 熵权; 湖泊

中图分类号: X824

文献标识码: A

文章编号: 1672-643X(2013)04-0127-04

Eutrophication assessment of lake based on attribute recognition model

HU Jiwei¹, GAO Junsheng^{1,2}

(1. College of Earth Environment and Water Resources, Yangtze University, Wuhan 430100, China;

2. North China University of Water Resources and Electric Power, Zhengzhou 450011, China)

Abstract: In order to determine the state of lake eutrophication, starting from the attribute recognition theory, the paper established the attribute recognition model for the eutrophication assessment of lake, and combined with the weight coefficients determined by entropy method to evaluate the state of eutrophication of Dongchang lake. The result is that the lake district throughout the year belongs to eutrophication level, and the highest degree of eutrophication is July, followed by December, the lowest degree of that is April. This result is consistent with the result evaluated by set pair analysis and fuzzy recognition method. The evaluation result of the model is reasonable. The model is simple, easy to be understand and can be used as a feasible way for the eutrophication assessment of lake.

Key words: attribute recognition; eutrophication; entropy weight; lake

湖泊富营养化(lake eutrophication)是指湖泊水体接纳过多的氮、磷等营养性物质,使藻类以及其他水生生物过量繁殖的现象。湖泊水体一旦进入富营养阶段,就会因藻类等水生植物过量繁殖而导致水质恶化,影响湖泊周围居民的生活和经济的发展^[1]。近些年来由于人口的快速增加和工农业快速发展,进入河流湖泊的污染物质急剧增多,人类赖以生存的淡水生态系统日益退化,特别是富营养化问题,已严重威胁社会经济的持续发展。为了防治湖泊水体的富营养化,不但要对湖泊水体的富营养化指标进行长期监测,而且要对湖泊水体所处的营养状态进行评价。只有湖泊富营养化程度评价准确,才能为湖泊富营养化的防治提供科学依据。目前评价湖库富营养化的方法有模糊评判法^[2]、可拓法^[3]、集对分析法^[4]等多种,但这些方法也存在一

些问题。如模糊评判法计算过程中信息丢失问题,可拓学法节域上限(第五级的上限)的合理确定问题,集对分析法中不确定系数的合理取值问题等。本文使用属性识别模型对湖泊富营养化进行评价,同时采用熵值法计算各指标权重,它能尽量消除各评价指标权重确定的人为因素影响,避免了权重确定的主观偏差,使权重的确定更客观,提高了评价模型的可靠度。

1 属性识别模型^[5]

1.1 属性空间矩阵

在评价对象空间 X 中取 n 个样本 $X_1, X_2, \dots, X_n$ 。对于每一个样本有 m 个评价指标 $I_1, I_2, \dots, I_m$ 。第 i 个样本 X_i 的第 j 个指标 I_j 的测量值为 x_{ij} ,因此,第 i 个样本 X_i 可以表示为一个 m 维向量,即 $X_i = (x_{i1},$

收稿日期:2013-01-12; 修回日期:2013-03-14

基金项目:国家水体污染控制与治理科技重大专项(2009ZX07210-006-3-2)

作者简介:胡吉炜(1988-),男,湖北武汉人,硕士研究生,主要从事水文学及水资源方面的研究。

通讯作者:高军省(1964-),男,陕西华阴人,教授,博士,主要从事水环境评价方面的研究。

$x_{i2}, \dots, x_{im}), 1 \leq i \leq n$ 。设 F 为 X 上某类属性空间, $(C_1, C_2, \dots, C_K)$ 为属性空间 F 的有序分割类, 并且满足 $C_1 > C_2 > \dots > C_K$ 或 $C_1 < C_2 < \dots < C_K$, 每一个评价指标均有其相应的评价等级分类标准 $\{z_{j1}, z_{j2}, \dots, z_{jk}\}, (j = 1, 2, \dots, m)$ 。写成分类标准矩阵为:

$$\begin{matrix} & C_1 & C_2 & \cdots & C_K \\ \begin{matrix} I_1 \\ I_1 \\ \vdots \\ I_m \end{matrix} & \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1K} \\ z_{21} & z_{22} & \cdots & z_{2K} \\ \vdots & \vdots & & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mK} \end{bmatrix} \end{matrix} \quad (1)$$

式中: z_{jk} 满足 $z_{j1} < z_{j2} < \dots < z_{jk}$ 或者 $z_{j1} > z_{j2} > \dots > z_{jk}$ 。

1.2 样本属性测度

第 i 个样本的第 j 个指标值 I_j 的测量值为 x_{ij} , 具有属性 C_k 的属性测度为:

$$\mu_{ijk} = \mu(x_{ij} \in C_k), \text{ 且 } \sum_{k=1}^K \mu_{ijk} = 1, (1 \leq k \leq K, 1 \leq i \leq n, 1 \leq j \leq m)。$$

不妨假定 $z_{j1} < z_{j2} < \dots < z_{jk}$, 则:

$$\begin{cases} \text{当 } x_{ij} \leq z_{j1} \text{ 时, 取 } \mu_{ij1} = 1, \mu_{ij2} = \dots = \mu_{ijk} = 0 \\ \text{当 } x_{ij} \geq z_{jk} \text{ 时, 取 } \mu_{ijk} = 1, \mu_{ij1} = \dots = \mu_{ijK-1} = 0 \\ \text{当 } z_{jl} \leq x_{ij} \leq z_{jl+1} \text{ 时, 取 } \mu_{ijl} = \left| \frac{x_{ij} - z_{jl+1}}{z_{jl} - z_{jl+1}} \right|, \\ \mu_{ijl+1} = \left| \frac{x_{ij} - z_{jl}}{z_{jl} - z_{jl+1}} \right|, \mu_{ijk} = 0 \\ (1 \leq k \leq l \text{ 或 } l+1 \leq k \leq K) \end{cases} \quad (2)$$

式中: $1 < l < K$ 。

按(2) 式计算得单指标属性测度矩阵:

$$(\mu_{ijk})_{m \times K} = \begin{bmatrix} \mu_{i11} & \mu_{i12} & \cdots & \mu_{i1K} \\ \mu_{i21} & \mu_{i22} & \cdots & \mu_{i2K} \\ \vdots & \vdots & & \vdots \\ \mu_{im1} & \mu_{im2} & \cdots & \mu_{imK} \end{bmatrix} \quad (3)$$

在知道第 i 个评价样本第 j 个指标的属性测度之后, 再计算第 i 个评价样本 x_{ij} 的属性测度 μ_{ik} 。由于各评价指标的重要性可能不同, 因此, 需要考虑指标权重: $(W_1, W_2, \dots, W_j), W_j \geq 0, \sum_{j=1}^m W_j = 1$ 。 W_j 的值可由专家或数据计算确定, 一般采用熵值法或比例测度法。

由单指标属性测度加权求和即可得到综合属性测度:

$$\mu_{ik} = \mu(x_i \in C_k) = \sum_{j=1}^m W_j \mu_{ijk} \quad (4)$$

式中: $1 \leq i \leq n, 1 \leq k \leq K$ 。

按式(4) 计算得到综合属性测度矩阵:

$$(\mu_{ik})_{n \times K} = \begin{bmatrix} \mu_{i1} & \mu_{i2} & \cdots & \mu_{iK} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2K} \\ \vdots & \vdots & & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nK} \end{bmatrix} \quad (5)$$

1.3 评判准则

通过上述方法, 计算得出各评价指标权重系数和各样本属性测度, 在此基础上, 就可以进行待测样本的属性识别和比较分析。按照置信度准则, 对置信度 λ (一般取 $0.5 < \lambda \leq 1$) 计算如下:

$$k_i = \min \left\{ k: \sum_{l=1}^k \mu_{xi}(C_l) \geq \lambda, 1 \leq k \leq K \right\} \quad (6)$$

当满足式(6) 时, 则认为 x_i 属于 C_{ki} 类, 即对样本作出了评价。

按照属性识别模型的评分准则, 如果样本水质类别相同, 计算:

$$q_{xi} = \sum_{l=1}^K n_l \mu_{xi}(C_l), n_l = k + 1 - i \quad (7)$$

则可根据 q_{xi} 的大小对 x_i 进行比较评价和排序。

2 权重的确定方法^[6]

目前国内外关于评价指标权重系数的确定方法有数十种之多, 根据计算权重系数时原始数据的来源以及计算过程的不同, 这些方法大致可分为 3 类: 主观赋值法、客观赋值法和主客观综合集成赋权法。

评价指标的权重通常采用主观赋权法确定。这样做虽然能比较充分地反映领域专家的知识 and 经验, 能够给出评价指标实际意义上的重要性系数, 但是容易因个人的主观因素而造成偏差。在本文权重系数的确定过程中, 为了避免人为主观因素的干扰, 使用熵值法, 从实测数据指标入手, 充分利用数据自身信息, 客观地确定出指标权重。

2.1 实测数据的标准化

设有 n 各评价对象, 每个评价对象有 m 个评价指标, 形成的实测数据矩阵 X , 采用极差变换对原始数据进行处理, 则标准化矩阵 Y 为:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \quad Y = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix} \quad (8)$$

2.2 熵值

根据熵的定义,第 j 个指标的熵为:

$$H_j = \frac{1}{\ln n} \sum_{i=1}^n f_{ij} \ln f_{ij} \tag{9}$$

其中: $f_{ij} = y_{ij} / \sum_{i=1}^n y_{ij}$,当 $f_{ij} = 0$ 时,令 $f_{ij} \ln f_{ij} = 0$ 。

2.3 熵权

第 j 个评价指标的熵权为:

$$w_{ej} = \frac{1 - H_j}{m - \sum_{j=1}^m H_j} \tag{10}$$

3 实例分析

3.1 基础数据

东昌湖位于聊城市的城区西南部,环绕古城四周,总面积为 4.2 km²,容积在 0.1 亿 m³ 以上,水深 2~3 m。引黄河水为源,常年不竭,是我国长江以北最大的城市人工湖泊。近年来,随着聊城市经济的快速发展,东昌湖出现了一定程度的富营养化。因此,需要对东昌湖的富营养化状态进行评价,从而为湖区污染控制及生态环境管理提供依据。东昌湖水质监测数据^[7]如表 1 所示。

表 1 东昌湖富营养化监测数据 mg/m ³ , mg/L, m				
评价 指标	月份	西北湖区 (A)	西南湖区 (B)	东南湖区 (C)
Chla	4	21.25	22.02	22.54
	7	110.07	118.63	120.85
	12	45.53	57.26	56.87
TP	4	180	161	173
	7	234	218	226
	12	196	179	185
TN	4	1980	1760	1840
	7	3630	3350	3420
	12	2320	2070	2190
COD _{Mn}	4	14.10	13.48	13.36
	7	17.05	16.78	16.54
	12	14.26	14.54	13.94
SD	4	1.21	1.17	1.11
	7	0.70	0.53	0.59
	12	1.33	1.25	1.31

3.2 评价结果

参照我国湖泊富营养化水质分级标准,以东昌湖 4 月份富营养化监测数据为例进行评价。

3.2.1 评价样本空间矩阵与属性空间矩阵

评价样本空间矩阵:

	A	B	C
Chla	21.25	22.02	22.54
TP	180	161	173
TN	1980	1760	1840
COD _{Mn}	14.10	13.48	13.36
SD	1.21	1.17	1.11

属性空间矩阵:

	I 类	II 类	III 类	IV 类	V 类	VI 类
Chla	1	2	4	10	65	160
TP	2.5	5	25	50	200	600
TN	30	50	300	500	2000	6000
COD _{Mn}	0.3	0.4	2	4	10	25
SD	10	5	1.5	1	0.4	0.3

3.2.2 评价指标的权重 利用公式(8)、(9)、(10)计算,得到各评价指标的权重见表 2。

表 2 各评价指标权重值					
评价指标	Chla	TP	TN	COD _{Mn}	SD
权重	0.195	0.193	0.203	0.206	0.202

3.2.3 属性测度 利用公式(2)~(5),计算出样本的综合属性测度矩阵:

$$(\mu_{ik})_{6 \times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.1174 & 0.1336 & 0.1579 \\ 0.2689 & 0.3043 & 0.2518 \\ 0.5575 & 0.5144 & 0.5442 \\ 0.0562 & 0.4770 & 0.0461 \end{bmatrix}$$

3.2.4 评价与评分结果 取置信度 $\lambda = 0.7$,则根据(6)式进行判断,其东昌湖富营养化评价结果见表 3。

表 3 东昌湖富营养化评价结果			
分区	4 月	7 月	12 月
A	富营养(V)	富营养(V)	富营养(V)
B	富营养(V)	富营养(V)	富营养(V)
C	富营养(V)	富营养(V)	富营养(V)

为了比较不同月份各湖泊水质富营养化的程度,运用(7)式计算分值,其结果见表 4。分值越大,表明湖泊富营养化程度越严重。

由表 4 可知,东昌湖富营养化程度 7 月最高,12 月次之,4 月最低。

表 4 东昌湖富营养化评价的评分结果			
分区	4 月	7 月	12 月
A	4.552	5.217	4.728
B	4.476	5.161	4.692
C	4.478	5.186	4.723

3.3 结果分析

3.3.1 λ 值对评价结果的影响 置信度 λ 不同的取值对评价结果的影响见表 5。

置信度 λ 的取值结合样本综合属性测度矩阵,可以得出当 $0.5 < \lambda \leq 0.7$ 时,东昌湖富营养化评

价结果为各湖区全年均属于富营养化级别;当 $0.7 < \lambda \leq 1$,4 月份和 12 月份东昌湖各区均属于富营养化级别,7 月份属于重富营养化级别。说明 λ 的取值对评价结果的细致程度有一定的影响。

3.3.2 不同评价方法的比较 采用集对分析法与模糊识别方法的评价结果见表 6。结果为各湖区全年均属于富营养化级别,并且 7 月份的富营养化程度最高,12 月份次之,4 月份最低。与属性识别方法评价结果一致。表明该模型在湖泊富营养化评价应用中具有可行性和合理性。

表 5 λ 不同取值的东昌湖富营养化评价对评价结果															
分区	$\lambda = 0.6$			$\lambda = 0.7$			$\lambda = 0.75$			$\lambda = 0.8$			$\lambda = 0.9$		
	4 月	7 月	12 月	4 月	7 月	12 月	4 月	7 月	12 月	4 月	7 月	12 月	4 月	7 月	12 月
A	V	V	V	V	V	V	V	Ⅵ	V	V	Ⅵ	V	V	Ⅵ	V
B	V	V	V	V	V	V	V	Ⅵ	V	V	Ⅵ	V	V	Ⅵ	V
C	V	V	V	V	V	V	V	Ⅵ	V	V	Ⅵ	V	V	Ⅵ	V

表 6 湖泊富营养化综合不同方法评价结果比较									
地区	属性识别			集对分析			模糊识别 ^[7]		
	4 月	7 月	12 月	4 月	7 月	12 月	4 月	7 月	12 月
A	4.552	5.217	4.728	4.550	5.187	4.729	4.703	5.154	4.715
B	4.476	5.161	4.692	4.474	5.123	4.707	4.689	5.197	4.738
C	4.478	5.186	4.723	4.477	5.152	4.736	4.710	5.213	4.753

4 结 语

本文将属性识别模型应用于湖泊富营养化评价中,可通过评分的方法对各湖区富营养化程度进行测评,进而可根据分值的大小对湖区富营养化程度进行比较;同时采用熵权来确定权重,避免了主观因素对权重值的影响,有效地解决了水质的优劣和所属问题。实例评价结果表明,属性识别法与集对分析法、模糊识别法的评价结果一致,且属性识别方法计算简便、实用,是一种有效的湖泊富营养化程度综合评价方法。

参考文献:

[1] 金根东. 我国湖泊富营养化研究现状[J]. 现代农业科

技,2008(16):334-336.
[2] 李进,陈益滨,师伟,等. 模糊综合评价法在地下水水质评价中的应用[J]. 地下水,2006,28(2):4-5+22.
[3] 高军省,吕小凡. 湖泊富营养化评价的可拓学方法及应用[J]. 长江大学学报(自然科学版),2010,7(1):33-37.
[4] 高军省. 湖泊富营养化综合评价的五元联系数法[J]. 人民长江,2010,41(21):81-84.
[5] 郭凤台,王瑞京,孙红. 属性识别模型在黑龙洞泉域地下水水质评价中的应用[J]. 节水灌溉,2008(11):43-45.
[6] 张先起,梁川,刘慧卿. 基于熵权的属性识别模型在地下水水质综合评价中的应用[J]. 四川大学学报(工程科学版),2005,37(3):28-31.
[7] 陈为国,许文杰,张晓平,等. 湖泊水体富营养化评价与可持续发展研究[J]. 节水灌溉,2008,(6):47-49+52.