

基于遗传算法湖泊水质富营养化的投影寻踪分析

黄伟军, 方 崇, 张永祥

(广西水利电力职业技术学院, 广西南宁 530023)

摘要:根据湖泊富营养化特性,选择总磷(TP)、总氮(TN)、耗氧量(COD_{Mn})、透明度(SD)、生物量共5个指标作为评价因子,建立湖泊富营养化评价的投影寻踪分析模型,采用遗传算法对评价模型进行优化,并将该模型应用于我国8个湖泊富营养化程度的评价。研究表明,投影寻踪回归分析法,避免了传统评价方法由于主观原因造成的误差,评价结果合理可信,评价方法简单易行,为湖泊水体营养化状态的评价提供了新途径。

关键词:湖泊;富营养化;综合评价;投影寻踪;遗传算法

中图分类号:X824 **文献标志码:**A **文章编号:**1674-3075(2010)03-0104-05

近年来,由于人类活动的加剧、经济的快速发展,大量工业废水和生活污水以及农田耕作中的植物营养物质排入湖泊等水体后,导致我国湖泊生源要素严重富集,大大加速了湖泊水体的富营养化过程(秦伯强等,2006)。日趋严重的湖泊富营养化对区域经济发展、饮用水安全和生态环境构成了威胁,制约着社会和经济的可持续发展(邓大鹏等,2006)。为此,研究湖泊富营养化也是我国基础研究领域亟待解决的重大科学问题。

湖泊富营养化评价的目的是为了准确反映水环境的质量和污染状况,预测其发展趋势,为开展环境污染和综合治理及环境规划提供科学依据;同时,也是对湖泊进行有效管理的重要手段(张忠祥和钱易,1998)。多年来,国内外学者对湖泊富营养化评价进行了不少的研究,提出了许多富营养化评价的模型和方法(邬敏等,2009)。最早的富营养化评价是由Carlson等提出的卡森指数公式以及适用于我国若干湖泊的营养状态指数公式,这些方法的共同特点是在给出单项指标分类界限后,通过对单项指标的计算,确定水质的类别归属;但这类方法应用时经常会出现某个单项指标值接近或等于某个界限值,或按不同的单项指标值判定,则分属于不同水质类别的情况,使得该水源的水质类别难以确定。近

年来,随着新的人工智能理论和智能优化技术的引入,人们又相继提出了若干湖泊富营养化评价的新方法。诸如模糊综合评价法(龚绍琦等,2005)、灰色评价法(李祚泳和邓新民,1992)、集对分析(李凡修和辛焰,2000)、贝叶斯分析法(谢平等,2005)、物元分析评价法(饶清华和张江山,2006)、人工神经网络评价法(刘首文和冯尚友,1996)等多种湖泊富营养化综合评价模型。这些人工智能理论对湖泊富营养化评价理论的发展起到一定的促进作用,但在评价过程中尚存在明显不足。例如,模糊综合评判理论相对较完善和成熟,但在权重确定上存在人为主观因素;灰色系统评价方法能较好处理不确定性问题,并且能与传统的层次分析法进行有机结合,评价结果也是综合考虑各种影响因素的具体量值,但在评价指标体系方面,大多进行了归纳和简化,使得很多具体的问题被掩盖,有时并不能全面真实的反映出被评价湖泊水体的富营养化状况。尽管国内外湖泊富营养化的评价模型和方法很多,但因为湖泊的成因、地理位置、环境条件等多方面的差异以及受人类活动影响程度的不同,湖泊富营养化的类型和富营养化的进程均不一样,其评价标准也应该不尽相同(苏平和伊海龙,2006)。

本文综合考虑了影响湖泊富营养化的各种因素,选择总磷(TP)、总氮(TN)、耗氧量(COD_{Mn})、透明度(SD)、生物量共5个具有代表性的因素作为评判指标,采用投影寻踪回归理论对原始数据进行处理,提取原始样本的特征,将遗传算法引入到投影方向优化计算中,建立湖泊富营养化等级评价的数学模型,在此基础上,编制基于MATLAB语言相应的计算程序。利用该评价模型对湖泊水体富营养化进行综合评价,并通过实例予以验证。

收稿日期:2009-08-12

基金项目:国家自然科学基金项目(No. 50679047);广西高校优秀人才资助计划(RC2007052);广西水利水电科技计划项目(桂水科合字2005-2号)。

通讯作者:方崇,1965年生,男,壮族,广西横县人,博士,研究方向为湖泊环境质量评价和系统分析。E-mail:fangchong2009@163.com

作者简介:黄伟军,1967年生,男,湖南娄底人,教授,博士生导师,研究方向为水资源与水环境。

1 基于遗传算法投影寻踪回归评价模型

投影寻踪回归是用来处理和分析高维数据的一种探索性数据分析的有效方法,其基本思想是利用计算机技术,把高维数据通过某种组合,投影到低维空间上,借助优化算法,寻找出能反映原高维数据结构或特征的投影,在低维空间上对数据结构进行分析,以达到研究和分析高维数据的目的(邬敏等,2009)。

用投影寻踪回归探索高维数据的结构或特征时,一般采用迭代模式。首先根据经验或猜想给定一个初始模型;其次把数据投影到低维空间上,找出数据与现有模型相差最大的投影,然后把上述投影中所包含的结构并在现有模型上,得到改进了的新模型;再从这个新模型出发,重复以上步骤,直到数据与模型在任何投影空间都没有明显的差别为止。

设等级评价表中各指标值的样本集为 $\{x_{ij} | i = 1, 2, \dots, n; j = 1, 2, \dots, m\}$,其中 x_{ij} 是第 i 个样本的第 j 个指标值,则建立投影寻踪评价模型的步骤如下(付强等,2002;Collins & McGown,1974)。

1)数据预处理。即对数据进行归一化处理,消除各指标值的量纲和统一各指标值的变化范围。

对于越大越优的指标:

$$x_{ij}^* = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}} \quad (1)$$

对于越小越优的指标:

$$x_{ij}^* = \frac{x_{j\max} - x_{ij}}{x_{j\max} - x_{j\min}} \quad (2)$$

其中: $x_{j\max}$ 、 $x_{j\min}$ 分别为第 j 个指标的最大值和最小值, x_{ij}^* 为指标值归一化处理后的序列。

2)构造投影指标函数。把 m 维数据综合成某一方向的一维投影值,并要求投影值散布特征为局部投影点尽可能密集,整体上各个点团之间尽可能散开。用 m 维的单位向量 $a(a_1, a_2, \dots, a_m)$ 表示某一投影方向,样本在该方向上的投影值可表示为:

$$Z_j = \sum_{i=1}^m a_j x_{ij}^*, i = 1, 2, \dots, n \quad (3)$$

设 $S(a)$ 为投影值的标准差, $D(a)$ 为投影值的局部密度,则投影指标函数可以表示为:

$$Q(a) = S(a) \cdot D(a) \quad (4)$$

其中:

$$S(a) = \sqrt{\frac{\sum_{i=1}^n (Z_i - \bar{Z}_j)^2}{(n-1)}} \quad (5)$$

\bar{Z}_j 为投影方向上的平均值。

$$D(a) = \sum_{i=1}^n \sum_{k=1}^n (R - r_{ik}), r_{ik} = |Z_i - Z_k| \quad (6)$$

R 为局部密度的窗口半径,可取 $r_{\max} + \frac{m}{2} \leq R \leq 2m$,函数 f 为一单位跃阶函数, $R - r_{ik} \geq 0$ 时,其值为1, $R - r_{ik} \leq 0$ 时,其值为0。

3)估计最佳投影方向。 $Q(a)$ 值与投影方向 a 有关,当 $Q(a)$ 取最大值时的 a 方向是最能反映数据结构特征的方向,所以可以通过求解投影指标函数优化问题来估计最佳投影方向,即求解:

$$\begin{cases} \max Q(a) \\ \text{s. t. } \|a\| = 1 \end{cases} \quad (7)$$

遗传算法是模拟生物在自然环境中的遗传和进化过程而形成的一种自适应全局优化概率搜索算法,可以用来求解最佳投影方向。限于篇幅,本文略去了该算法的详细步骤,基于实数编码加速遗传算法优化的实现过程参考金菊良和杨晓华(2000)。

4)等级评价。得到近似最佳投影方向后,计算各经验等级样本点的投影值,根据各经验等级及其对应的投影值建立基于投影值的等级评价方法,并将待评价样本进行归一化处理计算其投影值,利用投影值等级评价方法得到评价样本所属等级。

2 应用实例

2.1 评价因子及评价标准的选取

选择参数时要视研究区域的具体情况而定,尽可能选取从不同方面、不同角度真实客观地反映湖泊水质富营养化程度的因子(邱林等,2008)。本文参考李祥泳等(2004)的分级标准和原始数据进行计算。根据富营养化的特性,选择总磷(TP)、总氮(TN)、耗氧量(COD_{Mn})、透明度(SD)和生物量作为评价因子。湖泊富营养化状态综合评价等级为5级,即:I级(极贫营养)、II级(贫营养)、III级(中营养)、IV级(富营养)、V级(重富营养)。富营养化评价分级标准如表1所示。

2.2 湖泊富营养化状态投影寻踪评价模型的建立

表1中的每个样本有总磷、总氮、耗氧量、透明度和生物量共5个评价指标。对湖泊富营养化评价等级的标准样本建立投影寻踪评价模型,即以5个水质特征参数来综合判定湖泊富营养化状态,为了使得投影函数 $Q(a)$ 为最大,必需对以上5个参数同时进行优化,实质上属于多维参数寻优的问题。本文采用遗传算法进行参数优化选,在RAGA优化过程中选定参数:父代初始种群规模 $N = 400$,交叉

表1 水质富营养化评价因子与分级标准

Tab.1 Water quality eutrophication evaluation factor and grading standards

营养状态级别	评价指标					投影值
	TP/ $\text{mg} \cdot \text{m}^{-3}$	TN/ $\text{mg} \cdot \text{m}^{-3}$	COD _{Mn} / $\text{mg} \cdot \text{L}^{-1}$	生物量/ $10^4 \text{个} \cdot \text{L}^{-1}$	SD/ m	
I级(极贫营养)	1	20	0.09	4	37.00	0.000
II级(贫营养)	4	60	0.36	15	12.00	0.010
III级(中营养)	23	310	1.80	50	2.40	0.076
IV级(富营养)	110	1 200	7.10	100	0.55	0.438
V级(重富营养)	660	4 600	27.10	1 000	0.17	2.136

概率 $P_c = 0.8$, 变异概率 $P_m = 0.8$, 优秀个体数目选定为 20 个, $\alpha = 0.05$, 加速次数为 15, 迭代次数为 50。遗传算法的迭代过程见图 1。

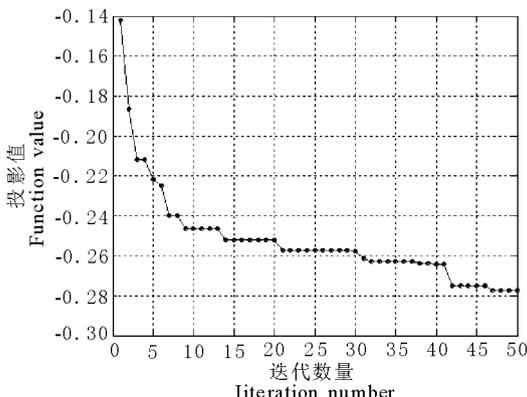


图1 遗传算法的迭代过程

Fig.1 A iterative process diagram of simulated annealing algorithms

在回归分析过程中,首先用式(1)、(2)对样本数据进行归一化处理,在 MATLAB 环境下运行程序后,得到最佳投影方向: $a^* = (0.712, 0.939, 0.782, 0.780, -0.041)$, 把最佳投影方向代入式(3)即可求得等级标准样本投影特征值。将待评价湖泊样本的投影特征值与等级标准样本的投影特征值相比较,根据投影值的大小,即可确定待评价湖泊富营养化等级。

表2 湖泊富营养化指标值及评价结果

Tab.2 The indicators value and evaluation results of lake eutrophication

湖泊名称	评价指标					投影值	评价等级	SVM评价等级
	TP/ $\text{mg} \cdot \text{m}^{-3}$	TN/ $\text{mg} \cdot \text{m}^{-3}$	COD _{Mn} / $\text{mg} \cdot \text{L}^{-1}$	生物量/ $10^4 \text{个} \cdot \text{L}^{-1}$	SD/ m			
青海湖	20	220	1.40	14.60	4.50	0.048	中营养	中营养
抚仙湖	20	210	1.61	19.00	7.03	0.052	中营养	中营养
洪泽湖	100	460	5.50	11.50	0.30	0.240	富营养	富营养
巢湖	30	1 670	8.26	25.30	0.25	0.457	富营养	富营养
滇池	20	230	10.13	189.20	0.50	0.198	富营养	富营养
武汉东湖	105	2 000	10.70	1 913.70	0.40	0.852	富营养	重富营养
杭州西湖	130	2 160	10.32	6 920.00	0.35	1.460	重富营养	重富营养
草海	177	1 860	51.00	7 060.00	0.41	2.090	重富营养	重富营养

2.3 实例分析

本文参考李祚泳等(2004)的分级标准和实例进行分析。对我国青海湖、抚仙湖、洪泽湖、巢湖、滇池、武汉东湖、杭州西湖和草海共8个湖泊进行富营养化等级评价。选取5项富营养化评价因子的监测数据作为待评价样本值,对水质评价样本建立综合分类评价的投影寻踪评价模型,其计算结果列于表2。

从湖泊样本投影特征值的大小来看,投影特征值越大,说明湖泊营养化程度越严重,水质越差。由表2可知,杭州西湖、草海的投影特征值分别为1.460、2.090,与分级标准中V级的投影特征值2.136接近,因此,可以判定杭州西湖和草海的富营养化程度等级为V级,即属重富营养化状态,说明水体污染最严重;同理,可以判定洪泽湖、巢湖、滇池、武汉东湖的富营养化程度等级IV级,属富营养化状态;青海湖、抚仙湖的富营养化程度等级III级,属中营养化状态,水质最优。

杭州西湖受工业废水和生活污水排放污染影响较大,近年来,尽管西湖水质经过清淤和换水,但局部区域受有机污染较重,因而其富营养化依然严重。草海位于贵州省威宁县,草海周边土壤流失严重。3万多居民的每天约2万t未经处理的生活污水直接

排放到草海,加上草海周边农民耕地每年施用约 10t 化肥、农药等污染物质的大量入侵,致使草海遭受污染,水质富营养化成分增大(李忠将,2005)。青海湖是中国最大的内陆湖泊,湖水来源主要依赖地表径流和湖面降水补给,入湖的河流众多,多达 40 余条河流,湖泊净化能力强。抚仙湖水的补给除靠雨季四周沟溪汇集外,湖岸周围还有大量的地下水涌出,湖水清澈纯净。

由表 2 可知,基于遗传算法投影寻踪分析模型的评价结果与支持向量机方法(SVM)的评价结果、水质状况基本相符,从而验证了本文方法的准确性。2 种方法的评价结果不同的是武汉东湖,从原始数据来看,武汉东湖除了生物量指标超过 V 级外,其余的 4 项指标值均属 IV 级范围内,说明本方法所得结果更切合实际情况。

3 结论

1) 本文成功地将投影寻踪评价模型应用于湖泊富营养化等级评价,并采用实码加速遗传算法对评价模型进行优化,并利用该模型对我国 8 个湖泊进行了实例检验。结果表明,利用投影寻踪进行湖泊富营养化等级评价是切实可行的。

2) 投影寻踪评价模型将指标体系的高维数据投影到一维子空间上,借助加速遗传算法,建立投影寻踪评价模型,多次运算,寻找最佳投影方向,形成评价指标值,按大小进行排序。与传统评价方法比较,该方法可以排除与数据结构和特征无关的,或关系很小的变量的干扰,而且避免了评价过程中人为主观因素,评价结果客观公正、精度较高,易于推广应用,具有较大的实用性。

3) 加速遗传算法模拟了生物遗传进化过程,是一种多点随机优化技术,对实际问题搜索空间的大小变化适应能力、计算速度及全局优化特性有很大改进,所得结果是令人满意的。将加速遗传算法应用于投影寻踪评价模型中的优化问题,可以有效地找到最佳投影方向,获得最佳投影值,为投影寻踪优化问题的解决提供了一条新的思路。但是,加速遗传算法的诸多运行参考选取方法尚缺乏理论指导,这是今后研究的方向。

4) 将投影寻踪回归用于湖泊富营养化等级评价,目前尚处于探索性阶段,还需要进一步研究和完善。此外,由于我国地域宽广,水质差别较大,影响湖泊水质的因素越来越多,水体富营养化发生原因是多方面的,在水体富营养化日益严重的今天,尽管

对湖泊富营养化评价的研究已经取得了一定的成果,但其评价指标的研究还有待于进一步深入。到目前为止,对于富营养化发生的各项指标还没有一个被广泛接受的严格量化的界定,还没有一套较成熟的理论能够在实际水体中用来预测富营养化发生。因此,系统深入地开展水体富营养化发生原因的研究,正确评价湖泊富营养化,对于有效地开展湖泊水体富营养化综合治理与防治具有重大的理论意义和实践价值。

参考文献:

- 邓大鹏,刘刚,李学德,等. 2006. 湖泊富营养化综合评价的坡度加权评分法[J]. 环境科学学报, 26(8): 1 386 - 1 392.
- 付强,金菊良,梁川. 2002. 基于实码加速遗传算法的投影寻踪分类模型在水稻灌溉制度优化中的应用[J]. 水利学报, (10): 39 - 45.
- 龚绍琦,黄家柱,李云梅,等. 2005. 基于 GIS 下的太湖水质富营养化模糊综合评价[J]. 环境科学, 26(5): 34 - 37.
- 金菊良,杨晓华. 2000. 基于实数编码的加速遗传算法[J]. 四川大学学报:工程科学版, 32(4): 20 - 24.
- 李凡修,辛焰. 2000. 集对分析用于湖泊富营养化评价研究[J]. 重庆环境科学, 22(6): 10 - 11.
- 李祚泳,邓新民. 1992. 湖泊富营养化的灰色层次决策法综合评价[J]. 上海环境科学, 11(3): 17 - 21.
- 李祚泳,丁晶,彭荔红. 2004. 环境质量评价原理与方法[M]. 北京:化学工业出版社.
- 刘首文,冯尚友. 1996. 人工神经网络在湖泊富营养化评价中的应用研究[J]. 上海环境科学, 15(1): 11 - 14.
- 秦伯强,杨柳燕,陈非洲,等. 2006. 湖泊富营养化发生机制与控制技术及其应用[J]. 科学通报, 51(16): 1 857 - 1 866.
- 邱林,冯晓波,冯丽云,等. 2008. 集对分析法在湖泊水质富营养化评价中的应用[J]. 人民长江, 39(5): 52 - 54.
- 饶清华,张江山. 2006. 熵权物元分析模型在湖泊富营养化评价中的应用[J]. 环境工程, 24(6): 80 - 82.
- 苏平,伊海龙. 2006. 多层前向人工神经网络在东部平原湖泊富营养化评价中的应用[J]. 系统仿真技术, 2(3): 166 - 170.
- 邬敏,李祚泳,刘智勇,等. 2009. 一种改进的集对分析法在湖泊富营养化评价中的应用[J]. 水资源保护, 25(2): 5 - 9.
- 谢平,李德,陈广才,等. 2005. 基于贝叶斯公式的湖泊富营养化随机评价方法及其验证[J]. 长江流域资源与环境, 14(2): 224 - 228.
- 张忠祥,钱易. 1998. 城市可持续发展与水污染防治对策[M]. 北京:中国建筑工业出版社.

Collins K, McGown A. 1974. The Form and Function of Microfab-
ric Features in a Variety of Nature Solis [J]. Geotech-
nique, 24(2): 223 - 254.

(责任编辑 万月华)

Projection Pursuit and its Application in Comprehensive Evaluation of the Eutrophication State in Lake

HUANG Wei-jun, FANG Chong, ZHANG Yong-xiang

(Guangxi Technology College of Water Conservancy and Electric Power, Nanning 530023, China)

Abstract: Based on characteristics of lake eutrophication, total phosphorus (TP), total nitrogen (TN), oxygen consumption (COD_{Mn}), transparency (SD), microbial population numbers, etc. 5 evaluation indicators was chosen for the evaluation factors. Projection Pursuit analysis model of evaluation of the lake eutrophication was established, The evaluation model was optimized using Genetic Algorithm, and the model was applied to analysis of eight lakes. The evaluation results are accurate, this method is simple, and some errors were eliminated because of the subjective factors for traditional methods. Projection Pursuit is a new evaluation method of the lake eutrophication, it has a wide application prospect.

Key words: Lake; Eutrophication; Comprehensive evaluation; Projection pursuit; Genetic algorithm