

基于改进投影寻踪的海洋生态环境综合评价

李彦苍*, 周书敬

(河北工程大学土木工程学院, 邯郸 056038)

摘要: 为了克服现有的海洋环境评价中存在的主观性强、不易处理高维数据的缺陷, 提出了基于改进投影寻踪模型的海洋环境评价新方法。该方法利用改进蚁群算法实现了投影寻踪技术, 将方案的多维评价指标值投影为一维投影数据, 并据投影值大小对样本进行综合评价。工程应用实例表明, 该模型易于决策, 具有很强的客观性、适用性和可操作性, 为海洋生态环境评价提供了新的技术工具。

关键词: 海洋生态环境学指标; 综合评价; 投影寻踪模型; 蚁群算法; 改进

文章编号: 1000-0933(2009)10-5736-05 中图分类号: Q178, X171, X820.2 文献标识码: A

Comprehensive assessment of marine ecological indexes based on improved projection pursuit method

LI Yan-Cang*, ZHOU Shu-Jing

College of Civil Engineering, Hebei University of Engineering, Handan 056038, China

Acta Ecologica Sinica, 2009, 29(10): 5736 ~ 5740.

Abstract: To overcome the shortcomings of the marine ecological indexes comprehensive assessment methods used today, an improved projection pursuit (PP) model based on an adaptive ant colony algorithm was presented. The improved ant colony algorithm was employed to realize the projection pursuit. And the values of the evaluation index were synthesized into an one-dimension projection value which was used to finish the assessment. Engineering practice shows that the method can realize the assessment simply, objectively and feasibly. This work has significance for the marine ecological indexes comprehensive assessment.

Key Words: marine ecological indexes; comprehensive assessment; projection pursuit model; ant colony algorithm; improvement

21 世纪以来,随着“海洋世纪”的到来,全球海洋环境,特别是近海海域环境持续恶化。如何有效地保护海洋环境已经成为世界各沿海国家所共同面临的问题。欲有效地解决海洋的生态与环境问题,需要通过海洋环境监测,快速准确地获取相关的海洋环境数据,并采取恰当的方法对这些海洋环境数据进行评价,通过构建“数字海洋”,最终获得对海洋环境保护、海洋资源开发和可持续发展有指导意义的科学依据和决策支持。作为保障国家安全、合理开发利用海洋资源、促进海洋经济可持续发展、保护海洋环境和减灾防灾不可或缺的基础性工作,海洋环境评价日益得到国际社会和沿海各国的高度关注。如澳大利亚利用区域项目“生态系统健康监测计划”进行了定量化的生态健康综合评价的尝试^[1]。欧盟和欧洲保护北大西洋海洋环境的奥斯陆-巴黎公约组织在实施海洋环境监测与评价项目的同时,首先推出了一系列完整的监测与评价技术指南(导则),并在实际工作中不断修订和完善^[1]。

基金项目: 河北省教育厅自然科学研究指导计划资助项目(Z2003404)

收稿日期: 2008-06-08; 修订日期: 2009-03-02

* 通讯作者 Corresponding author. E-mail: liyancang@163.com

但目前国际上的评价方法尚不完善成熟,缺少较为有效的算法及模型。同时,针对如何合理地将方案的多维评价指标转换成一维评价指标这一瓶颈问题的研究还很少。近年来,国内外学者提出的模糊综合优选模型、灰色关联分析模型、人工神经网络优选模型、单因子标准指数法等方模型对综合评价起到了积极作用^[2~4],但由于海洋环境综合评价涉及众多评价指标、各指标的量纲不尽相同和各指标的权重较难确定,实际的评价中所采用的专家打分的方法存在主观性强、不易处理高维数据的缺陷。

为此,本文引入直接利用样本本身指标数据的改进投影寻踪(projection pursuit, PP)模型和改进蚁群算法进行方案优选,应用实例表明,该方法具有人为干扰小、结果稳定等优点,是一种值得推荐的综合评价方法。

1 投影寻踪模型简介

作为一种分析和处理非正态高维数据的降维分析方法,投影寻踪法(projection pursuit, PP)最先是由 Friedman 于 1974 年提出的一种新型数理统计分析方法^[5],其基本思想是利用计算机技术,把高维数据(尤其是高维非正态数据),投影到低维(1~3 维)子空间上,并通过优化投影指标函数,寻找出能反映原高维数据结构或特征的投影向量,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。

综合评价的实质是降维数。由于评价指标较多,寻找高维空间上的点映射到低维空间上的最佳投影方向,需优化多个参数,传统方法同时优化多个参数相当困难,而 PP 方法能成功地克服此点。PP 技术是一种直接由样本数据驱动的探索性数据分析方法,特别适用于分析和处理非线性、非正态高维数据。投影寻踪模型在遥感图像解译、光(色)谱数据分析、环境质量评价、环境毒理学分析等诸多领域已得到成功应用^[6,7]。

2 基于改进投影寻踪方法的海洋生态环境综合评价模型

欧洲保护北大西洋海洋环境的奥斯陆-巴黎公约(OSPAR)组织指出^[8],海洋环境监测和评价需涵盖至少 3 个层面的重复测定与评价:海洋环境各介质(包括水、沉积物和生物体)的质量和海洋环境的综合质量;自然变化及人为活动向海洋输入的、可能会对海洋环境质量产生影响的物质和能量;人类活动所产生的环境效应。所以海洋环境综合评价涉及技术、经济、环境、社会等国计民生的各个方面,是一个典型的多目标、多层次、多属性决策问题。现今多由决策者和专家凭经验确定,因此存在较大的主观性和人为干扰因素。为此,有必要采用投影寻踪方法这一有效的降维技术对方案进行优选。

(1) 评价指标规范化处理

设评价指标值的样本集为: $\{x^*(i, j), i=1 \sim n, j=1 \sim p\}$,其中, $x^*(i, j)$ 为第 i 个样本第 j 个指标值。 n, p 分别为样本的个数和指标的数目。为消除各指标值的量纲差异并统一各指标值的变化范围,对各指标值进行归一化处理。

对效益型指标,采用:

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

对成本型指标,采用:

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

式中, $x(i, j) i=1, 2, \dots, n; j=1, 2, \dots, p$ 为经归一化处理后的评价指标集; $x^*(i, j)$ 为第 i 个样本的第 j 项指标值; $x_{\max}(i, j)$ 和 $x_{\min}(i, j)$ 分别为第 j 项评价指标值的最大值和最小值。

(2) 构造投影指标函数 $Q(a)$

PP 法实质是寻找最能充分表现数据特征的最优投影方向。设 $a = a\{a(1), a(2), \dots, a(p)\}$ 为 p 维单位向量,PP 法就是把经归一化处理的 p 维评价指标集 $x(i, j)$ 投影到 $\bar{a}$ 上。得到对应的一维线性空间投影值 $z(i)$:

$$z(i) = \sum_{j=1}^p a(j)x(i, j), i = 1, 2, \dots, n \quad (3)$$

式中, $a(j) j=1, 2, \dots, p$ 为投影方向向量,是单位长度向量。

可根据 $\{z(i), i=1 \sim n\}$ 的一维散布图进行分类。确定综合评价指标 $z(i)$ 的关键是找到能反映高维数据特征结构的最优投影方向 $\bar{a}$ 。此问题转化为求有约束的投影指标函数 $Q(a)$ 的极值问题。 $Q(a)$ 的约束条件是 $z(i)$ 的散布特征应满足局部投影点分类,即内部尽可能密集;类间尽可能散开。因此, $Q(a)$ 可表达为:

$$Q(a) = S_z D_z \quad (4)$$

式中, S_z 为类间散开度,为投影值 $z(i)$ 的标准差; D_z 为类内密集度,即投影值 $z(i)$ 的局部密度。两者的表达式分别如式(5)、式(6)所示:

$$S_z = \left\{ \sum_{i=1}^n [z(i) - \bar{z}]^2 / (n-1) \right\} \quad (5)$$

$$D_z = \sum_{i=1}^n \sum_{j=1}^p (R - r_{ij}) \cdot I(R - r_{ij}) \quad (6)$$

式中, $\bar{z}$ 为序列 $\{z(i), i=1 \sim n\}$ 的平均值; R 为局部密度的窗口半径,建议取值为 $r_{\max} + \frac{p}{2} \leq R \leq 2p^{[9]}$ 。 r_{ij} 表示样本之间距离, $r_{ij} = |z(i) - z(j)|$; 符号函数 $I(R - r_{ij})$ 为单位阶跃函数,当 $R \geq r_{ij}$ 时,函数值取 1,否则取 0。

(3) 优化投影指标函数

当被评价的对象海洋生态学各指标值的样本集给定时,投影指标函数 $Q(a)$ 只随投影方向 a 的变化而变化。不同的投影方向反应不同的数据特征,最佳投影方向就是最大可能暴露高维数据某类特征结构的投影方向。因此可通过求解投影指标函数最大化问题来估计最佳投影方向,即:

$$\text{Max: } Q(a) = S_z O_z \quad (7)$$

$$\text{s. t. } \sum_{j=1}^p a^2(j) = 1 \quad (8)$$

这是一个以 $\{a(j) | j=1 \sim p\}$ 为优化变量的复杂非线性优化问题。需用现代优化算法处理,本文采用蚁群算法进行求解。蚁群优化算法,简称蚁群算法 (Ant Colony Optimization, ACO), 是 1991 年由意大利学者 Dorigo, M. V. Mahiezzo 和 A. Colomi 提出的一类较为有效的求解组合优化问题的新型算法。本文采用文献的基于信息熵的改进蚁群算法求解^[10]。

(4) 评价

将 PP 模型中投影指标函数 $Q(a)$ 求最大作为目标函数,各个指标的投影 $a(j)$ 作为优化变量,运行步骤(3),可求得最佳投影方向 $a^*(i)$,将求得的最佳投影方向 $a^*(j)$ 代入式(3),可得各评价样本点的投影值 $z^*(i)$ 。将 $z^*(i)$ 与 $z^*(j)$ 进行比较,二者越接近,表明样本 i 与 j 越倾向于同一类。按 $z^*(i)$ 的值从大到小排序,则可将样本从优到劣进行排序,从而实现对各样本的评价。

3 应用实例

为了便于比较说明模型的有效性,采用本文提出的模型,对文献^[8]中我国大连湾生态环境进行了综合评价,测点的选取和数据同文献^[8],评价指标及 6 个测点处的数据见表 1。

该问题待评价的样本数为 6 个,每个样本有 17 个评价指标,即属于 17 维数据。对此问题建立 PP 模型,然后将样本指标值按式(1)~(2)进行归一化处理后,依次代入式(3)~(6),即得投影指标函数,然后用改进蚁群算法优化由式(7)和(8)所给定的问题,蚁群算法的参数为: $\alpha = 1.5$, $\beta = 4.0$, $\rho = 0.6$, $Q = 50$,停机准则为所得的信息熵的值小于等于 0.001,得出的最佳投影方向 $a^* = (0.10076, 0.14587, 0.1721, 0.1692, 0.10786, 0.10755, 0.1094, 0.1215, 0.1094, 0.1188, 0.1133, 0.1058, 0.1523, 0.2090, 0.1203, 0.1124, 0.1672)$,将 a^* 代入(3)式,求得各样本的投影值 $z^*(j) = (2.1563, 1.9651, 2.8927, 2.8651, 3.1622, 1.2329)$,比较投影值,便可得出测点 5 处海洋环境最优,其次为 3 和 1,近岸海域水质较差,湾内中心区域水质较好,与实际情况基本相符。评价结果与文献^[8]所采用的综合指数评价法的结果相吻合。同时,投影方向 a^* 中各分量大小,实质上反映了各指标对海洋生态评价的影响程度,由其可知,该问题的 17 个评价指标中,重要的是生物学指标,即水体交换能力较差,因工业废水造成的营养物质在湾内集结导致的水体的富营养化

是大连湾所面临的主要问题。所以,应采取控制污染物排放、调整产业结构等措施保护该海域的生态环境。

表 1 大连湾生态环境综合评价指标值
Table 1 Comprehensive assessment of marine ecological indexes in Dalian bay

项目 Item	1#测点 No.	2#测点 No.	3#测点 No.	4#测点 No.	5#测点 No.	6#测点 No.
	1 spot	2 spot	3 spot	4 spot	5 spot	6 spot
海水水质现状统计结果						
Statistic result of quality						
status of the sea water						
悬浮物 suspended matter(mg/L)	6	7	7.7	6.4	6.3	6.1
石油类 petroleum(mg/L)	0.0205	0.036	0.0205	0.018	0.0095	0.0105
无机氮 inorganic nitrogen(mg/L)	0.4795	0.465	0.4065	0.04685	0.195	0.06625
活性磷酸盐 activated phosphate(mg/L)	0.0119	0.2515	0.02515	0.01	0.0077	0.0021
溶解氧 dissolved oxygen(mg/L)	10.2	9.85	9.76	9.81	9.78	9.92
盐度 salinity of sea water	31.5	31.3	31.3	31.5	31.4	31.6
海域沉积物质量						
Quality of marine						
sediment						
Cu (μg/L)	32.1	45.7	36.2	27.8	22.6	32.3
Pb (μg/L)	39.5	42.1	37.1	39.8	42.5	47.8
Zn (ug/L)	103.5	87.5	92.3	83	121.5	78.5
Cd (ug/L)	0.61	0.87	0.96	1.32	0.81	0.77
有机碳 organic carbon(ug/L)	2400	3300	2100	2600	1700	1600
硫化物 sulphide(ug/L)	874	550	969	1349	448	535
海洋生物学调查						
Biological						
investigation						
in monitoring spots						
浮游植物多样性指数 Shannon-Weaver indexes of phytoplankton in monitoring spots	2.46	2.24	1.99	2.42	2.48	1.47
浮游动物多样性指数 Shannon-Weaver indexes of zooplankton in monitoring spots	3.16	2.68	2.42	2.33	2.72	1.98
底栖生物多样性指数 Shannon-Weaver indexes of zoobenthos in monitoring spots	0.996	1.192	2.570	3.348	3.1699	3.736
细菌总数 Gross quantity of bacteria in monitoring spots(ind./L)	23000	17000	14000	20000	16000	13000
生物残毒 Cd Analysis result of organism samples	0.234	0.205	0.147	0.356	0.127	0.109

4 结论

海洋环境评价具有重要的理论及现实意义。影响海洋环境的变量多、作用因子复杂,对海洋环境质量作出客观的评价,需要有一个科学的客观方法。现有的方法存在主观性强、不易处理高维数据的缺陷,本文引入基于投影寻踪模型的方案优选新方法,并采用改进蚁群算法实现了投影寻踪操作。应用实例证明,因为直接采取各样本的原始数据进行分析,该模型信息量不会丢失,同时,避免了人为赋权的主观干扰,克服了传统方法的不足。通过此方法,还可知道各评价指标对综合评价结果的影响程度,即各指标的贡献率的大小,以便于针对关键因素采取专项措施,提高了决策的可行性。该研究为海洋生态环境评价及其他领域的评价问题提供了一较为有效的方法。PP 模型中局部密度的窗口半径的取值问题仍需进一步研究。

References:

- [1] Jo Fodena, Stuart I. Rogersb and Andrew P. Jonesa. A critical review of approaches to aquatic environmental assessment. Marine Pollution Bulletin, 2008, 56(11): 1825 — 1833.
- [2] Yang W H, Wang L F Wang K, *et al.* Application of modified osculating value method in evaluation of water environmental quality. Journal of Water Resources and Water Engineering, 2005, 16(2): 69 — 74.
- [3] Fu H, Sun Y L, Sun L, *et al.* The application of grey relational analysis in assessment of marine environmental quality. Transactions of Oceanology and Limnology, 2007, 3:127 — 131.
- [4] Chen L, Zhu W D. The environmental quality assessment of neural network algorithm trained by particle swarm optimization. Acta Ecologica

Sinica, 2008, 28(3) : 1072 — 1079.

- [5] Friedman J H, Turkey J W A. Projection pursuit algorithm for exploratory data analysis. IEEE Trans on Computer, 1974, 23(9) : 881 — 890.
- [6] Webb-Robertson Bobbie-Jo M, Jarman Kristin H, D Harvey Scott, *et al.* An improved optimization algorithm and a Bayes factor termination criterion for sequential projection pursuit. Chemometrics and Intelligent Laboratory Systems, 2005, 77(1/2) : 149 — 160.
- [7] Vasil L, Stefan T. Multivariate statistical approaches as applied to environmental physics studies. Central European Journal of Physics, 2006, 4(2) : 277 — 298.
- [8] Hu J. Research on ecological environment comprehensive assessment methods for coast marine Area. Dalian University of Technology, 2007.
- [9] Michael D S. Statistical modeling of high-dimensional nonlinear systems: a projection pursuit solution. Atlanta: Georgia Institute of Technology, 2005.
- [10] Li Y C, Li W Q. Adaptive Ant Colony Optimization Algorithm Based on Information Entropy: Foundation and application. Fundamenta Informaticae, 2007, 77(3) : 229 — 242.

参考文献:

- [2] 杨文海, 王丽芳, 王坤, 等. 改进密切值法在水环境质量评价中的应用. 水资源与水工程学报, 2005, 16(2) : 69 ~ 74.
- [3] 付会, 孙英兰, 孙磊, 等. 灰色关联分析法在海洋环境质量评价中的应用. 海洋湖沼通报, 2007, 3 : 127 ~ 131.
- [4] 陈莉, 朱卫东. 微粒群优化神经网络及其在环境评价中的应用. 生态学报, 2008, 28(3) : 1072 ~ 1079.
- [8] 胡婕. 沿岸海域生态环境质量综合评价方法研究. 大连理工大学, 2007.