

常用水质综合评价方法探讨

王凡, 周扬, 刘伟, 刘灵, 何国富* (华东师范大学资源与环境科学学院, 上海 200062)

摘要 介绍目前运用较多的3种水质综合评价方法,即主成分分析评价法、模糊模式识别评价法、BP神经网络评价法的结构原理和算法。根据建立的不同水质综合评价模型,采用MATLAB语言编写相应的评价程序进行实例评价,试图找出3种方法在水质综合评价中的优缺点。

关键词 主成分分析;模糊综合评价法;BP神经网络法;水质;综合评价

中图分类号 X824 **文献标识码** A **文章编号** 0517-6611(2009)31-15394-03

Research on Common Methods of Comprehensive Water Quality Evaluation

WANG Fan et al (College of Resources and Environmental Sciences, East China Normal University, Shanghai 200062)

Abstract The structure principle and algorithm of three water quality comprehensive evaluation methods (principal component analysis, fuzzy model recognition and BP neural network) widely used were introduced. Based on these theories, comprehensive water quality evaluation models were established and used to evaluate the water quality. The aim was to find out the advantages and disadvantages of these methods in the water quality evaluation.

Key words Principal component analysis; Fuzzy comprehensive assessment; BP neural network; Water quality; Comprehensive evaluation

水环境质量评价是环境质量评价中的重要组成部分,是通过一定的数理方法和其他手段对水环境要素的优劣进行定量描述的过程。通过水环境质量评价可以了解环境质量的过去、现在和将来发展趋势及其变化规律,从而为水环境质量的科学管理和规划提供科学依据^[1]。

水质综合评价实际上是一种依据水质标准进行水质等级划分的典型的模式识别问题。国内外用于水质综合评价的方法主要有简单指数法、分级加权平均法、综合污染指数法、模糊模式识别法、物元分析评价法、主成分分析法、层次分析法、灰色聚类法、BP神经网络法等^[2-5]。笔者就目前运用较多、评价结果较准确的3种方法,即主成分分析法、BP神经网络法和模糊综合评价法进行探讨,试图找出其在水质综合评价中的优缺点,旨在为水质评价方法的改进和应用研究积累资料。

1 水质综合评价方法

1.1 主成分分析评价法 主成分分析是一种数学变换方法,它把给定的一组相关变量通过线性变换转换成另一组不相关的变量,这些新的变量按照方差依次递减的顺序排列。在数学变换中保持变量的总方差不变,使第一个变量具有最大方差,称为第一主成分,第二个变量的方差次大,并和第一个变量不相关,称为第二主成分,依此类推,最后一个主成分方差最小,且与此前的主成分都不相关^[6]。根据主成分分析的基本原理,可以把主成分分析的计算步骤归纳如下^[7-9]。

(1)为排除量纲和数量级的不同,首先要对原始数据进行标准化处理, $x_{ij}^* = (x_{ij} - \bar{x}_i) / \sigma_i$ ($i = 1, 2, \dots, p; j = 1, 2, \dots, p$);式中, x_{ij} 为第 i 个指标第 j 监测点的原始数据; \bar{x}_i 和 σ_i 分别为第 i 个指标的样本均值和标准差。

(2)根据标准化数据表 $(x_{ij}^*)_{n \times p}$, 计算相关系数矩阵 $R = (r_{ij})_{n \times n}$; 其中, $r_{ij} = \frac{1}{n} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) / \sigma_i \sigma_j$ 。

(3)计算特征值与特征向量,根据特征方程 $|\lambda I - R| =$

0, 求出特征值 λ_i ($i = 1, 2, \dots, p$), 并使其按大小顺序排列, 然后分别求出对应于特征值 λ_i 的特征向量 e_i ($i = 1, 2, \dots, p$), 要求 $\|e_i\| = 1, \sum_{j=1}^p$ 即表示向量 e_i 的第 j 个分量。

(4)计算主成分 z_i 贡献率 $\lambda_i / \sum_{k=1}^p \lambda_k$ 和累计贡献率 $\sum_{k=1}^i \lambda_i / \sum_{k=1}^p \lambda_k$, 一般取累计贡献率大于 80% 的特征值 $\lambda_1, \lambda_2, \dots, \lambda_m$ 所对应的第 1、第 2、...、第 m 个主成分。

(5)计算主成分负荷: $l_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} e_{ij}$ ($i, j = 1, 2, \dots, p$)。

1.2 模糊模式识别评价法 1965年,查德提出模糊集合的概念,并把“模糊(Fuzzy)”一词引入技术文献当中。水质评价涉及到许多模糊概念,1991年陈守煜等提出了模糊模式识别理论模型并将其应用于水质评价,该法利用模糊数学的理论和方法解决模式识别问题。随着模糊理论的不断完善和发展,改进的模糊模式识别模型相继出现,其主要的计算步骤如下^[10-14]。

(1)标准矩阵 Y 和样本矩阵 X 规范化。对于 BOD、COD 等数值愈大污染愈严重的指标,采用下列 2 式进行 Y, X 矩阵元素的变换:

$$S_{ih} = \begin{cases} 1 & y_{ih} = y_{ic} \\ (y_{ih} - y_{iu}) / (y_{ic} - y_{iu}) & y_u < y_{ih} < y_{ic} \\ 0 & y_{ih} = y_{iu} \end{cases} \quad (1)$$

$$r_{ij} = \begin{cases} 1 & x_{ij} \geq y_{ic} \\ (x_{ij} - y_{iu}) / (y_{ic} - y_{iu}) & y_u < x_{ij} < y_{ic} \\ 0 & x_{ij} \leq y_{iu} \end{cases} \quad (2)$$

对于 DO 等数值愈大污染程度愈轻的指标,采用下列变换公式:

$$S_{ih} = \begin{cases} 1 & y_{ih} = y_{ic} \\ (y_{ih} - y_{iu}) / (y_{ic} - y_{iu}) & y_{ic} < y_{ih} < y_{iu} \\ 0 & y_{ih} = y_{iu} \end{cases} \quad (3)$$

$$r_{ij} = \begin{cases} 1 & x_{ij} \geq y_{iu} \\ (x_{ij} - y_{iu}) / (y_{ic} - y_{iu}) & y_{ic} < x_{ij} < y_{iu} \\ 0 & x_{ij} \leq y_{ic} \end{cases} \quad (4)$$

作者简介 王凡(1985-),女,河南南阳人,硕士研究生,研究方向:环境污染治理。*通讯作者。

收稿日期 2009-06-29

将矩阵 X, Y 变为相应的实测指标相对隶属度矩阵 R 和指标标准相对隶属度矩阵 S 。

(2) 计算样本 j 对于 h 级水质标准的最优相对隶属度 μ_{hj} , 公式为:

$$\mu_{hj} = \begin{cases} 1 & h = \alpha_{\min} \text{ 或 } h = \alpha_{\max} \\ \frac{1}{\sum_{k=\alpha_{\min}}^{\alpha_{\max}} \left[\frac{\sum_{i=1}^m (w_{ij} |r_{ij} - s_{ik}|)^p}{\sum_{i=1}^m (w_{ij} |r_{ij} - s_{ik}|)^p} \right]^{2/p}} & \alpha_{\min} < h < \alpha_{\max} \\ 0 & h < \alpha_{\min} \text{ 或 } h > \alpha_{\max} \end{cases} \quad (5)$$

式中, 指标权向量 $w = \sum_{j=1}^n r_{ij} / \sum_{j=1}^n \sum_{i=1}^m r_{ij}$ 。

(3) 根据最大隶属度原则确定水质等级: $J(j) = (\mu_{hj})_H = \sum_{h=1}^c h \mu_{hj}$ 。

1.3 BP 神经网络评价法 神经网络(ANN)是以模仿人脑神经系统的组织方式来构成新型的信息处理系统, 具有很强的非线性映射能力和自适应、自组织、自学习能力, 能实现对任意复杂函数的映射。1986年, D E Rumelhart 等提出多层网络学习的误差反传播算法(Back Propagation, 简称BP)。人们也常把按误差反传递算法训练的多层前馈网络直接称为BP网络。

其中, BP神经网络是目前发展比较成熟的一种人工神经网络, 由一个输入层、一个输出层及多个隐含层组成, 每个层可以有若干个节点, 常见为3层BP网络, BP神经网络的工作原理如图1所示^[15-18]。

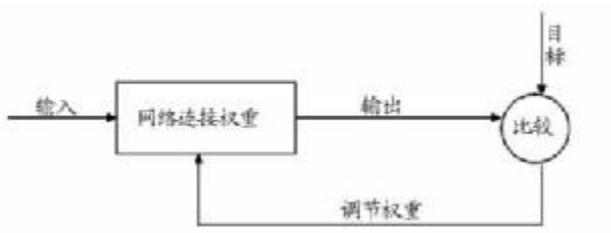


图1 BP神经网络工作示意

Fig. 1 The working of BP neural network

经过反复学习直到样本输入输出间的总误差达到设定的精度要求, 样本总误差为 $E = \frac{1}{2} \sum_{k=1}^n (y^k - c^k)^2$ 。式中, y^k 为网络期望输出, c^k 为网络实际输出; m 为学习样本数。

利用BP网络分析水质, 首先对水质评价分级标准和待评价河流的水质监测数据进行标准化处理。由于各监测指标具有不同的量纲, 且类型不同, 故指标间具有不可公度性, 难以进行直接比较。因此, 在综合评价前必须把这些分指标按某种规则归一化到某一无量纲区间^[15]。由于BP网络中sigmoid函数的输出范围在0~1, 所以也把数据归一到[0, 1]。具体的标准化处理如下^[16]。

对于成本型因子(如TP、AS等), 其数值越大表示水质越差, 可用下式进行标准化:

$$Z_{ij} = \begin{cases} 1 & X_{ij} > S_{ij} \\ X_{ij}/S_{ij} & 0 \leq X_{ij} \leq S_{ij} \end{cases} \quad (6)$$

式中, X_{ij} 为实测值, S_{ik} 为水质评价标准的第 k 级(水质最差的一级)。

对于效益型因子(如DO), 数值越大表示水质越好, 可用下式进行标准化:

$$Z_{ij} = \begin{cases} 0 & X_{ij} > S_{i1} \\ 1 - X_{ij}/s_{i1} & 0 \leq X_{ij} \leq S_{i1} \end{cases} \quad (7)$$

式中, X_{ij} 为实测值; S_{i1} 对应于水质评价标准的第一级(水质最好的一级)。

2 实例评价

2.1 监测数据与评价指标 选取长江下游某河流监测断面2004年1、3、5、7、9、11月监测数据月平均值, 以高锰酸盐指数(COD_{Mn})、氨氮(NH₃-N)、总磷(TP)、总氮(TN)4项为评价指标, 依据《地表水环境质量标准》(GB3838-2002), 分别采用主成分分析法、模糊模式识别法和BP神经网络法对其进行评价, 3种模型计算程序均采用MATLAB语言编写。具体监测数据及水质标准见表1、2。

表1 某河流监测断面2004年水质数据
Table 1 The data of the water quality monitoring section in 2004

监测月份 Monitoring month	水温//℃ Water temperature	pH 值 pH value	COD _{Mn} mg/L	NH ₃ -N mg/L	TP mg/L	TN mg/L
1	8.0	7.21	2.8	0.302	0.065	0.43
3	13.5	7.38	2.9	0.289	0.061	0.42
5	24.6	7.08	3.0	0.315	0.072	0.48
7	28.6	7.15	2.5	0.298	0.072	0.49
9	15.3	7.12	2.2	0.264	0.065	0.43
11	7.6	7.22	2.8	0.297	0.068	0.45

表2 地表水分级评价标准

Table 2 The grading evaluation criteria of surface water quality
mg/L

级别 Grade	COD _{Mn}	NH ₃ -N	TP	TN
I	2	0.15	0.02	0.2
II	4	0.50	0.10	0.5
III	6	1.00	0.20	1.0
IV	10	1.50	0.30	1.5
V	15	2.00	0.40	2.0

2.2 评价结果分析 以高锰酸盐指数(COD_{Mn})、氨氮(NH₃-N)、总磷(TP)、总氮(TN)4项为评价指标, 依据《地表水环境质量标准》(GB3838-2002), 分别采用主成分分析法、模糊模式识别法和BP神经网络法对某河流水质进行评价, 评价结果如表3所示。

由表3可知, 3种方法评价结果基本一致, 可见这3种方法均适用于水质综合评价。通过3种评价方法结果显示, 该河流监测断面水质除3月份外各个月份水质均属II级, 3月份水质介于I类与II类之间。另外, 该监测断面已位于下游, 所以整体来看该河流水质常年处于污染较轻状态, 这与流域内无重大工业污染源, 主要为农业面源污染有关。

通过3种评价方法比较可得出, 采用人工神经网络模型进行水环境质量评价只需对分类标准进行训练, 就能自动获得水质参数间的合理规则, 自动调节连接权矩, 无需人为干预, 因此评价结果具有客观性; 模糊综合评价法由于考虑到了水质分级界限的模糊性和污染因素的权重, 因而能够比较客观地反映水体的水质现状, 但受权重因素制约, 可操作性

相比较差;主成分分析法能够在保证原始数据信息损失最小的基础上,经过线性变化和舍弃部分信息,以少数的综合变量取代原有的多维变量,简化了数据结构,可以在损失较少信息的情况下减少评价指标,抓住主要矛盾进行分析评价。3种方法各有优缺点,因此在实际中进行水质评价时,至少应选择2种方法进行评价,并综合分析比较后最终确定水质等级。

表3 水质数据评价结果

Table 3 The evaluation results of water quality

监测月份 Monitoring month	主成分分析法 Principal component analysis	模糊综合评价法 Fuzzy comprehensive assessment	BP神经网络法 BP neural network
1	II	II	II
3	II	I、II	II
5	II	II	II
7	II	II	II
9	II	II	II
11	II	II	II

3 结语

(1)人工神经网络模型进行水环境质量评价运算速度快,评价结果准确,需要在构造BP人工神经网络时,不断试验和尝试,以满足实际应用的要求;模糊综合评价法由于考虑到了水质分级界限的模糊性和污染因素的权重,因而能够比较客观地反映水体的水质现状,但可操作性相比较差;主成分分析法能够在损失较少信息的情况下减少评价指标,抓住主要矛盾进行分析评价,简化了数据结构,但在实际应用中还应综合考虑各个评价指标,以免信息损失。

(2)在实际中进行水质评价时,至少应选择2种方法进行评价,并综合分析比较后最终确定水质等级。同时,建立具体评价模型时,应考虑把不同的方法结合起来,根据实际建立更适用的评价模型,如采用模糊数学或灰色关联度方法处理数据,然后应用BP人工神经网络方法进行评价的方法,在实际中已得到了初步探究和应用。

(3)随着新学科和新理论的创立,水环境质量评价也会有新的突破。尤其是计算机技术越来越广泛地用于环境科学领域,与之相关的新的环境评价方法也在不断地得到开发和应用,如地理信息系统、化学计量学在水环境质量评价中的初试体现了这一新的发展趋势,它们的研究应用也将对目前环境质量评价方法和体系给以革新、补充与完善。

参考文献

- [1] 王梦. 环境质量评价中几种方法的比较[J]. 渤海大学学报:自然科学版,2008,29(1):34-37.
- [2] 李奇珍,何俊仕. 水质综合方法探讨[J]. 人民黄河,2007,29(1):50-52.
- [3] 薛巧英. 水环境质量评价方法的比较分析[J]. 环境保护科学,2004,30(4):64-66.
- [4] 夏军. 区域水环境及生态环境质量评价:多级关联评估理论与应用[M]. 武汉:武汉水利电力大学出版社,1999.
- [5] 兰文辉,安海燕. 环境水质评价方法的分析与探讨[J]. 干旱环境监测,2002,16(3):167.
- [6] 盛周君,孙世群,王京城,等. 基于主成分分析的河流水环境质量评价研究[J]. 环境科学与管理,2007,32(12):172-175.
- [7] 丁春,盛周君. 基于主成分分析法的南淝河水质综合评价[J]. 安徽农业科学,2007,35(35):11583-11584.
- [8] 高卫东. 基于主成分分析的矿区地下水水质评价[J]. 节水灌溉,2009,4(1):58-60.
- [9] 尹元荣,海米提,依米提,等. 主成分分析法在城市河流水质评价中的应用[J]. 干旱区研究,2008,25(4):497-501.
- [10] 陈守煜. 复杂水资源系统优化模糊识别理论与应用[M]. 长春:吉林大学出版社,2002.
- [11] 陈守煜. 工程水文水资源系统模糊集分析理论与实践[M]. 大连:大连理工大学出版社,1998.
- [12] 席北斗,于会斌,郭旭晶,等. 基于模糊权元理论的地下水水质评价模型构建及应用[J]. 环境工程学报,2009,3(2):381-384.
- [13] 刘章军,叶燎原. 模糊概率模型及其在地下水水质评价中的应用[J]. 应用基础与工程科学学报,2007,15(3):289-293.
- [14] 刘章军,叶燎原. 模糊概率模型及其在地下水水质评价中的应用[J]. 应用基础与工程科学学报,2007,15(3):289-293.
- [15] 杜富芝,傅瓦利,杜小红,等. 基于BP神经网络的三峡库区小流域水质评价[J]. 节水灌溉,2009,1(1):8-12.
- [16] 杨浩,吴晗名,万飏,等. 汉江水质综合评价的BP网络方法[J]. 武汉大学学报:工学版,2004,37(1):51-54.
- [17] 楼文高,王延政. 基于BP网络的水质综合评价模型及其应用[J]. 环境污染治理技术与设备,2003,4(8):23-27.
- [18] Statsoft. Statistica neural networks[M]. Tulsa:Statsoft, Inc.,1999.
- [14] 单保庆,杜国祯,刘振恒. 不同养分条件下和不同生境类型中根茎草本黄帚囊吾的克隆生长[J]. 植物生态学报,2000,24(1):46-51.
- [15] 孙海群,康晓燕. 鹅绒委陵菜在不同养分条件下的克隆生长研究[J]. 草业与畜牧,2008(10):18-19,39.
- [16] 李根前,黄宝龙,唐德瑞,等. 毛乌素沙地中国沙棘无性系生长格局与生物量分配[J]. 西北农林科技大学学报:自然科学版,2001,29(2):51-55.
- [17] 王长庭,龙瑞军,丁路明. 高寒草甸不同海拔梯度下多年生黄帚囊吾的克隆生长特征[J]. 西北植物学报,2004,24(10):1805-1809.
- [18] LOVETT D L. Population dynamics and local specialization in a clonal perennial (*Rununculus repens*) III. Responses to light and nutrient supply[J]. Journal of Ecology,1987,75:555-568.
- [19] SACKVILLE H N R. Variation and adaptation in wild population of white clover (*Trifolium repens*) in East Anglia[M]. Cambridge:Cambridge University Press,1982.
- [20] DONG M, KROON H. Plasticities in morphology and biomass location in *Cynodon dactylon*, a grass species forming stolons and rhizomes[J]. Oikos,1994,70:99-106.
- [21] 罗学刚,董鸣. 匍匐茎草本蛇莓克隆构型对土壤养分的可塑性反应[J]. 生态学报,2001,21(12):1957-1963.
- [22] 陈尚,王刚,李自珍. 白三叶分枝格局的研究[J]. 草业科学,1995,12(2):35-39.
- [23] 李静,马小凡,郭平,等. 施肥及施加植物激素对羊草克隆构型的影响[J]. 草业科学,2005,23(1):18-21.
- [24] 张玉芬,张大勇. 克隆植物的无性与有性繁殖对策[J]. 植物生态学报,2006,30(1):174-183.
- [25] WILLIAMS G C. Natural selection, the cost of reproduction and refinement of lacks of principle[J]. American Naturalist, 1966,100:687-690.
- [26] STEARMS S C. The Evolution of Life Histories[M]. Oxford:Oxford University Press,1992.
- [27] WESTLEY L C. The effect of inflorescence bud removal on tuber production in *Helianthus tuberosus* L. (Asteraceae)[J]. Ecology,1993,74:2136-2144.
- [28] 张运春,杜晓军,张桥英,等. 克隆乔木黄牛奶树两种繁殖方式的适合度分析[J]. 应用生态学报,2005,16(9):1687-1692.
- [29] XU L M, HUANG J Y, LIU X Q, et al. Cloning of *Brassica napus* EIN3 gene and its expression induced by *sclerotinia sclerotiorum*[J]. Agricultural Science & Technology,2009,10(2):33-36.

(上接第15344页)