

基于 SPA-BP 的河流水质评价模型

袁 伟¹, 闫小虎²

(1. 合肥工业大学土木与水利工程学院, 安徽 合肥 230009; 2. 长江水利委员会长江科学院, 湖北 武汉 430010)

摘要 将集对分析法(SPA)和 BP 神经网络结合起来建立基于 SPA-BP 的河流水质评价模型, 直接由样本联系度矩阵得出评价结果, 并给出评价样本与评价标准之间的各级隶属度值。模型应用表明, 该模型评价河流水质具有客观性和实用性, 评价结果较为直观、可靠。

关键词 河流水质评价; 集对分析; BP 神经网络; SPA-BP 模型

中图分类号: X824 文献标识码: A 文章编号: 1006-7647(2011)S1-0011-03

河流水质评价方法概括起来大致可以分为指标法和和不确定性方法两大类^[1], 由于水环境是一个有着随机性、模糊性和灰色性的不确定系统^[2], 所以采用不确定性评价方法对河流水质进行评价相对较为合理。不确定性评价方法有集对分析法(SPA)^[3-5]、人工神经网络^[6-7]和信息论等, 用集对分析法综合评价河流水质需要确定评价指标的权重, 采用不同的确定权重的方法(定权方法), 评价结果可能也会有所差别; 用 BP 神经网络评价河流水质, 往往只能得到 1 个评价等级值, 没有给出与其余等级的隶属关系。笔者尝试用集对分析法处理水环境系统的不确定性和随意性, 将集对分析法与 BP 神经网络结合起来建立基于 SPA-BP 的河流水质评价模型, 直接由输入矩阵得出评价结果, 避免因指标定权方法不同导致评价结果不准确, 并给出评价样本与评价标准之间的各级隶属度值。

1 基于 SPA-BP 的河流水质评价模型

1.1 集对分析理论

集对分析理论是由我国学者赵克勤^[3]首创、用于处理确定和不确定性问题的新方法。集对分析的基本思想: 在给定问题背景的情况下对所论的 2 个集合 A 和 B 进行同、异、反的定量比较分析, 计算联系度^[8]的一般形式见式(1):

$$\mu = \frac{S}{N} + \frac{F_1}{N}i_1 + \frac{F_2}{N}i_2 + \dots + \frac{F_m}{N}i_m + \frac{P_1}{N}j_1 + \frac{P_2}{N}j_2 + \dots + \frac{P_n}{N}j_n \tag{1}$$

式(1)可简化为

$$\mu = a + b_1i_1 + b_2i_2 + \dots + b_mi_m + c_1j_1 + c_2j_2 + \dots + c_nj_n \tag{2}$$

其中

$$a + \sum_{u=1}^m b_u + \sum_{v=1}^n c_v = 1$$

$$a = \frac{S}{N} \quad b_u = \frac{F_u}{N} \quad c_v = \frac{P_v}{N}$$

$$a \geq 0, b_u \geq 0, c_v \geq 0$$

$$u = 1, 2, \dots, m \quad v = 1, 2, \dots, n$$

式中: S 为集对所共有的特性个数; N 为每个集合所具有的总特性个数; F_u 为集对的任一差异性个数; P_v 为集对的任一对立性个数; i₁, i₂, ..., i_m 为差异度系数, 均在 [-1, 1] 中取值; j₁, j₂, ..., j_n 为对立度系数, 一般均取为 -1。

由于实际水体含有各种不同组分, 各种物质对水质的影响不同, 即使水质处于同一等级, 也会因指标的含量不同而有所差异, 为此需要进一步相对于分级标准作同一、差异、对立的集对分析^[9]。将待评价样本集与评价标准集看作 1 个集对, 假设有 Q 个评价指标, 评价标准分为 M 级, x_{kl} (k = 1, 2, ..., Q; l = 1, 2, ..., M) 为第 k 个评价指标第 l 类水质的分界点值。对于越小越优指标(如 COD), 将高于 x_{kM} 的质量浓度值统一化为 M 类; 对于越大越优指标(如 DO), 将低于 x_{kM} 的质量浓度值统一化为 M 类。同时将 I 类水质归为同一; M 类水质归为对立; 两者之间归为差异。越小越优或越大越优指标的联系度 μ_k 分别采用式(3)和(4)计算^[8]:

$$\mu_k = \begin{cases} 1 + 0i_1 + 0i_2 + \dots + 0i_{M-2} + 0j & 0 < x_{ki} \leq x_{k1} \\ \frac{x_{k2} - x_{ki}}{x_{k2} - x_{k1}} + \frac{x_{ki} - x_{k1}}{x_{k2} - x_{k1}} i_1 + 0i_2 + \dots + 0i_{M-2} + 0j & x_{k1} < x_{ki} \leq x_{k2} \\ 0 + 0i_1 \dots + 0i_{M-3} + \frac{x_{kM} - x_{ki}}{x_{kM} - x_{k(M-1)}} i_{M-2} + \frac{x_{ki} - x_{k(M-1)}}{x_{kM} - x_{k(M-1)}} j & x_{k(M-1)} < x_{ki} \leq x_{kM} \\ 0 + 0i_1 + 0i_2 + \dots + 0i_{M-2} + 1j & x_{ki} > x_{kM} \end{cases} \quad (3)$$

$$\mu_k = \begin{cases} 1 + 0i_1 + 0i_2 + \dots + 0i_{M-2} + 0j & x_{ki} > x_{k1} \\ \frac{x_{ki} - x_{k1}}{x_{k2} - x_{k1}} + \frac{x_{k2} - x_{ki}}{x_{k2} - x_{k1}} i_1 + 0i_2 + \dots + 0i_{M-2} + 0j & x_{k2} < x_{ki} \leq x_{k1} \\ 0 + 0i_1 \dots + 0i_{M-3} + \frac{x_{ki} - x_{k(M-1)}}{x_{kM} - x_{k(M-1)}} i_{M-2} + \frac{x_{kM} - x_{ki}}{x_{kM} - x_{k(M-1)}} j & x_{kM} < x_{ki} \leq x_{k(M-1)} \\ 1 + 0i_1 + 0i_2 + \dots + 0i_{M-2} + 0j & 0 < x_{ki} \leq x_{kM} \end{cases} \quad (4)$$

式中: x_{ki} ($k=1, 2, \dots, Q$) 为第 k 个评价指标的实际质量浓度值; 差异度系数 i_m ($m=1, 2, \dots, M-2$) 和对立度系数 j 仅起标记作用。

1.2 样本联系度矩阵和标准指标联系度矩阵的建立

设有 P 个待评价样本, Q 个评价指标, 第 p 个样本的联系度矩阵为

$$U = (\mu_{kt})_{Q \times T} \quad (5)$$

式中: $t=1, 2, \dots, T$ ($T=MP$); μ_{kt} 为第 p 个样本第 k 个评价指标第 l 级分类标准的联系度, 由式(3)或式(4)计算得到。

对于标准指标联系度矩阵

$$Y = (y_{kl})_{Q \times M} \quad (6)$$

式中: y_{kl} 为第 k 个评价指标第 l 级分类标准的联系度。

y_{kl} 由第 k 个评价指标第 l 级分类标准区间内的任意值代入式(3)或式(4)计算得到, 输出值为 Q 个指标同一等级的平均值。根据这种方法, 可以生成任意多的标准指标联系度样本。在本次研究中, 采用取特殊值的方法生成训练样本输入矩阵和输出目标矩阵。另外, 作为对训练样本的补充^[7], 笔者将文献[10]中元台大桥断面 2007 年 1—5 月份的实测数据代入式(3)或式(4)计算得到这 5 个月的联系度矩阵, 同时为得到 1—5 月份的输出目标矩阵, 采用文献[10]的各指标权重计算得到这 5 个月的隶属度矩阵, 并将这 5 个补充样本作为训练样本一同输入网络, 对网络进行训练。

定义待评价样本 p 与标准等级 l 的隶属度矩阵

$$S = (s_{lt})_{\lambda \times T} \quad (7)$$

对第 p 个评价样本的隶属度矩阵作归一化处理, 得到归一化隶属度矩阵为

$$S = \frac{s_{pl}}{\sum_{l=1}^M s_{pl}} \quad (8)$$

采用最大隶属度原则确定待评价样本的水质等级。

1.3 确定 BP 神经网络的结构与参数

选用 3 层 BP 神经网络, 利用 LM 优化算法构建神经网络, 经过试算确定隐含层的激励函数为 LOGSIG, 输出层为 PURELIN。以大沙河干流水质评价^[10]为例, 共设置 5 项评价指标: DO, $\text{NH}_3\text{-N}$, COD_{Cr} , BOD 和 COD_{Mn} , 因此输入层神经元个数为 5 个; 取河流水质评价等级为 5 级, 分别为 I 类、II 类、III 类、IV 类和 V 类, 所以输出层神经元个数为 5 个; 隐含层神经元个数等于 $\sqrt{p+q} + t^{[11]}$, p 为输入层神经元个数, q 为输出神经元个数, t 为系数, $t \in [1, 10]$, 据此公式反复检验和试算, 确定隐含层个数为 10 个, 从而构成 5-10-5 的 BP 神经网络。

1.4 训练样本的生成和 SPA-BP 模型的建立

采用标准指标联系度矩阵和目标隶属度矩阵以及补充的训练样本对 BP 神经网络进行训练, 设置迭代次数为 20 次, 学习效率为 0.1, 目标误差为 1×10^{-9} , 满足目标误差时结束训练, 此时便建立了具有预测功能的 SPA-BP 模型, 代入样本联系度矩阵 U 即可得到输出目标隶属度矩阵 S 。

2 SPA-BP 模型的应用

为验证本文所建模型的可行性, 利用文献[10]中大沙河干流 3 个监测断面之一元台大桥断面 2007 年 6—12 月份的水质监测资料进行网络训练, 令 $Q=5$, 满足误差要求后结束训练, 建立具有预测功能的 SPA-BP 模型。将 6—12 月的实测样本联系度矩阵输入训练好的 SPA-BP, 得到 7 个月的隶属度矩

阵。由于采用随机初始化权重和阈值,每次训练、输入得到的结果不完全相同,为提高评价结果的可靠性,采用20次训练、输入的结果平均值作为评价结果,得到待评样本的归一化隶属度矩阵,见表1。

表1 评价结果对比

月份	归一化隶属度					评价结果	
	I类	II类	III类	IV类	V类	本文评价结果	文献[10]模糊综合评价结果
6	0.1045	0.1914	0.1303	0.1349	0.4389	V	V
7	0.2994	0.1645	0.2172	0.0785	0.2404	I	V
8	0.0604	0.1623	0.2101	0.1280	0.4392	V	V
9	0.5868	0.2008	0.0717	0.0703	0.0703	I	I
10	0.6553	0.1292	0.0718	0.0718	0.0718	I	I
11	0.4926	0.1320	0.0690	0.1107	0.1956	I	I
12	0.4964	0.0978	0.2518	0.0770	0.0770	I	I

从评价结果对比(表1)来看,本文与文献[10]的评价结果基本一致,但7月份的评价等级出现较大偏差,主要与单个指标权重过大(表2)和计算权重的方法不同有关。各评价指标的评价结果见表2,评价结果由指标NH₃-N决定,其他4个指标的权重和为0.53,对确定待评价样本的水质等级影响不明显,类似于单因子评价法^[12](取该断面的最高水质类别作为评价结果,该方法评价河流水质存在不足);同时根据文献[10]介绍的权重计算方法,每个实测样本采用的指标权重矩阵皆不同,而训练样本对神经网络训练之后,输入与输出向量之间存在的对应关系就被确定了,对6—12月的水质预测采用统一的这种对应关系也是造成结果出现较大偏差的原因之一,这就充分说明评价指标权重的确定直接影响到评价结果的准确性。

表2 7月份实测联系度矩阵

评价指标	联系度矩阵					指标权重 ^[10]	单因子评价法 ^[12]
	I类	II类	III类	IV类	V类		
DO	0.47	0.53	0	0	0	0.24	II
NH ₃ -N	0	0	0	0	1	0.47	V
COD _{Cr}	0	0	0.88	0.12	0	0.15	III
BOD	1	0	0	0	0	0.06	I
COD _{Mn}	0.35	0.65	0	0	0	0.08	II

3 结 论

- a. 笔者运用集对分析法(SPA)处理水环境系统的不确定性和模糊性,并且结合BP神经网络建立了基于SPA-BP的河流水质评价模型。
- b. 本文与文献[10]的评价结果基本一致,从而证明了笔者建立的新模型用来评价河流水质是可行的。
- c. 单独用BP神经网络评价水质等级,只能得

到1个评价等级值;将BP神经网络与SPA结合建立SPA-BP模型,能够给出评价样本与评价等级之间的各级隶属度值,评价结果较为直观、全面。

参考文献:

[1] 梁德华,蒋火华.河流水质综合评价方法的统一和改进[J].中国环境监测,2002,18(2):63-66.

[2] 刘国东,丁晶.水环境中不确定性方法的研究现状与展望[J].环境科学进展,1996,4(4):46-51.

[3] 赵克勤.集对分析及初步应用[M].杭州:浙江科学技术出版社,2000.

[4] 童英伟,刘志斌,常欢.集对分析法在河流水质评价中的应用[J].安全与环境学报,2008,8(6):84-86.

[5] 王文圣,金菊良,丁晶,等.水资源系统评价新方法:集对评价法[J].中国科学,2009,39(9):1529-1534.

[6] 李雪,刘长发,朱学慧,等.基于BP人工神经网络的海水水质综合评价[J].海洋通报,2010,29(2):225-230.

[7] 陈守煜,李亚伟.基于模糊人工神经网络识别的水质评价模型[J].水科学进展,2005,16(1):88-91.

[8] WANG Dong, ZHU Yuan-sheng, ZHAO Ke-qin. Research and application of model based on set pair analysis and fuzzy met theory for evaluation of water eutrophication[J]. Hydrology, 2004, 24(3):9-13.

[9] LIU Hui, GONG Shi-liang. The set pair analysis method to environmental quality evaluation of groundwater[J]. Shanghai Geology, 2000(2):21-23.

[10] 杨佳佳.大沙河干流水质评价与容量研究[D].大连:辽宁师范大学,2009.

[11] 张立明.人工神经网络的模型及其应用[M].上海:复旦大学出版社,1993:55-70.

[12] 安乐生.地表水水质评价方法和水质预测模型的综合研究[D].青岛:青岛大学,2009.

(收稿日期:2011-01-07 编辑:骆超)

