

# RBF 神经网络在平顶山市地表水评价中的应用

李磊, 孙卉, 翟秋敏<sup>△</sup>, 郭志永 (河南大学资源与环境科学研究所, 河南开封 475001)

**摘要** 为准确和客观地评价地表水环境质量状况, 运用 MATLAB 软件中的神经网络工具箱, 结合 K 均值聚类方法建立径向基函数网络, 对平顶山市 2004 年市控 5 个地表水断面进行了环境质量评价。在评价前根据平顶山市的实际情况对训练样本范围进行更改, 将训练和测试样本进行归一化处理, 同时利用 RAND 函数对训练样本进行插值保证神经网络充分学习。结果发现, K 均值聚类法能快速准确地确定网络中心, 用建立的径向基函数网络进行地表水质量评价, 其评价结果与单因子方法的评价结果一致, 并且具有计算速度快、量化评价结果便于同类水质间互相比对的优点。

**关键词** RBF 神经网络; 平顶山市; 地表水; 环境质量评价

**中图分类号** X824 **文献标识码** A **文章编号** 0517-6611(2008)26-11514-03

## Application of RBF Neural Network in the Surface Water Assessment in Pingdingshan City

LI Lei et al (Institute of Natural Resources and Environmental Science Research, Henan University, Kaifeng, Henan 475001)

**Abstract** In order to assess the surface water environment quality accurately and objectively, neural network toolbox of MATLAB was used combining with K-means method. The surface water quality of five sections in Pingdingshan City in 2004 was assessed. Due to the actual situation, the scope of training samples were changed before assessment. The training and test samples were normalized. The RAND function was used to construct enough training samples in order to keep the network full learning. The result showed that using K-means method could determine the network center fast and accurately, and the result was the same with that by the single factor method. The RBF network could compute fast and quantify the result, which was advantageous for the comparison of same kind of water quality.

**Key words** RBF neural network; Pingdingshan City; Surface water; Environmental quality assessment

为了准确地评价地表水环境质量, 我国根据水域使用目的和环境保护要求颁布了《地表水环境质量标准》, 以化学需氧量、氨氮、溶解氧等 30 项评价指标为依据将其划分为五类水体。目前, 国内外地表水环境质量评价方法主要有: 指数评价法、模糊评价法、灰色评价法、人工神经网络评价法等。其中人工神经网络法 (Artificial Neural Network, 简称 ANN) 是一种具有分布并行处理、自适应学习, 对于处理非线性问题十分有效的新方法体系, 如今已被广泛地运用到水环境质量评价当中<sup>[1-4]</sup>。但环境质量评价使用的大都是 BP 网络, 由于 BP 网络自身训练速度慢、容易陷入局部极小点等局限性, 使得在网络设计中需要反复试凑和训练。而径向基函数网络与 BP 网络相比, 具有学习速度较快的优点, 并且网络的函数逼近能力、模式识别与分类能力都优于后者。李泽应等使用径向神经网络对长江芜湖段水域进行评价<sup>[5]</sup>, 表明运用径向基函数网络进行地表水环境质量评价是可行的。

笔者参阅了相关的技术资料<sup>[6-8]</sup>, 尝试将评价标准范围进行针对性的修改, 利用 RAND 函数对评价标准插入足够的训练样本, 并将训练和测试样本进行归一化处理。运用 MATLAB 神经网络工具箱的径向基函数网络结合 K 均值聚类法对平顶山市 2004 年市控 5 个地表水断面进行环境质量评价, 取得了较为理想的评价结果。

## 1 径向神经网络简介

**1.1 径向神经网络结构** 径向基函数 (Radial Basis Function, 简称 RBF) 网络是由 Moody 和 Darken 于 20 世纪 80 年代末提出的一种神经网络, 网络中的神经元是分层排列的, 每个神经元只与前一层神经元相连, 共分为 3 层 (图 1)。它是由 3 部分组成的前馈式网络: 第 1 部分是输入层, 由信号源节点组成; 第 2 部分是隐含层, 一般采用高斯函数作为其节点的基函数; 第 3 部分是单输出层, 神经元采用线性传递函数。

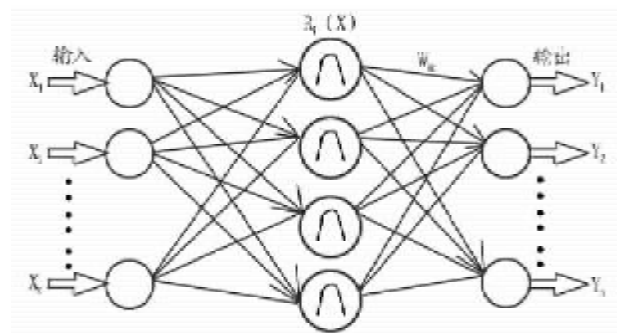


图 1 RBF 神经网络结构

Fig. 1 Structure of RBF neural network

**1.2 径向神经网络特点** RBF 神经网络模拟了人脑中局部调整、相互覆盖接受域的神经网络结构, 因此具有较强的局部逼近特性。它对任何非线性连续映射能够用任意精度逼近<sup>[9]</sup>, 同时具有良好的泛化能力、计算量小、学习速度快等特点。由于 RBF 神经网络强大的函数逼近能力, 已被广泛应用于分级分类、系统辨识和参数估计等方面。

**1.3 径向神经网络设计与仿真** RBF 神经网络设计的过程实际上就是选取基函数中心的过程, 如果由于基函数中心选取不当而构造出来的 RBF 网络, 其性能一般不能令人满意。目前选择 RBF 网络中心的方法主要有聚类算法和随机选择中心点等, 在基函数中心确定之后, 根据输入的扩展常数来确定隐含层的宽度; 由于输出单元为线性单元, 输出层的权值和阈值通过最小二乘法求出。

由于 RBF 网络在设计的过程中通过训练样本对网络的中心、宽度以及输出权值和阈值进行了设定, 所以 RBF 网络在建立的同时也完成了对网络的训练, 最后利用 SIM 函数对设计好的网络进行仿真。

## 2 运用 RBF 神经网络评价地表水质量

**2.1 数据的选取与预处理** 根据平顶山市实际情况, 选取化学需氧量、氨氮、氟化物等 7 项对断面水质影响较大的监测项目作为评价因子, 同时选取国家环保总局颁布的《地表

作者简介 李磊 (1982-), 男, 河南平顶山人, 硕士研究生, 研究方向: 环境变化与管理。\* 通讯作者。

收稿日期 2008-06-30

水环境质量标准》(表 1)中各评价因子一至五类水质标准作为训练输入,相应的评价等级作为训练输出。选取平顶山市

2004 年市控 5 个监测断面(简称 A、B、C、D、E 断面)的监测数据(表 2)作为测试样本进行质量评价。

表 1 地表水环境质量标准(部分)

评价等级	水质标准	化学需氧量	氨氮	生化需氧量	高锰酸盐指数	石油类	挥发酚	氟化物
Evaluation level	Standard of water quality	Chemical oxygen demand	Ammonia nitrogen	Biochemical oxygen demand	Permanganate index	Petroleum oil	Volatile phenol	Fluoride
1	一类 Type 1	15	0.15	3	2	0.05	0.002	1.0
2	二类 Type 2	15	0.50	3	4	0.05	0.002	1.0
3	三类 Type 3	20	1.00	4	6	0.05	0.005	1.0
4	四类 Type 4	30	1.50	6	10	0.50	0.010	1.5
5	五类 Type 5	40	2.00	10	15	1.00	0.100	1.5

表 2 平顶山市 2004 年市控各断面水质监测结果

断面	化学需氧量	氨氮	生化需氧量	高锰酸盐指数	石油类	挥发酚	氟化物
Section	Chemical oxygen demand	Ammonia nitrogen	Biochemical oxygen demand	Permanganate index	Petroleum oil	Volatile phenol	Fluoride
A 断面 Section A	6	0.29	1.33	2.10	0.01	0.001	0.46
B 断面 Section B	7	0.49	1.65	2.42	0.01	0.001	0.44
C 断面 Section C	16	7.50	5.90	5.30	0.01	0.001	0.60
D 断面 Section D	88	39.00	37.00	26.90	0.12	0.100	1.91
E 断面 Section E	6	0.12	1.50	2.65	0.01	0.001	0.60

为使建立的 RBF 网络具有比较准确的评价结果,将作为训练输入的国家标准进行调整,同时设置相应的训练输出。经调整后的训练数据具有较高的代表性,使神经网络得到充分训练和学习,确保对测试数据具有较高的仿真程度。

**2.1.1 训练样本范围设置。**国家环保总局颁布的《地表水环境质量标准》中仅规定了一至五类水质标准,对于超出这五类的水体并没有划定明确的范围。这使得训练数据无法让神经网络学习到超出五类的水质样本,在遇到测试样本水质超出五类标准的情况时,网络将会给出误差较大的评价结果。

为了避免因神经网络学习不充分而导致误差较大的情况出现,人为增加了评价级别为六的劣五类水质标准即化学需氧量 200 mg/L,氨氮 50 mg/L,生化需氧量 100 mg/L,高锰酸盐指数 100 mg/L,石油类 15 mg/L,挥发酚 10 mg/L,氟化物 20 mg/L,各评价因子取值范围均远远超出五类标准,以确保新生成的训练样本包含多种水质状况。

**2.1.2 样本插值及归一化处理。**由于训练数据范围更改后的样本数目仍然偏少,而径向基函数网络则需要较多的输入输出数据来训练网络。有文献指出,利用 MATLAB 自带的 RAND 函数在各级评价标准中间按随机均匀分布方式内插训练数据的方法是可行的<sup>[10]</sup>。所以针对训练样本过少的情况,在一至五类水质标准中,每 2 个相邻标准之间内插生成 10 组训练样本,为了使网络对劣五类水质更充分地学习,在五类与劣五类水质标准间内插生成 20 组训练样本,并按照同样的方法对相应的评价等级进行插值操作。

由于神经网络训练样本输入的每一维代表一个特征,当神经网络的输入是多维时,要识别的模式就有多个特征。当这些特征的数据相差很大,如几个数量级时,就需要进行归一化,使其变成相同数量级,以防某些数值低的特征被淹没。所以还要用 PREMNMX 函数对训练输入和输出数据进行归一化处理,使数据变成 -1 ~ 1 范围内的数值,以提高网络泛

化能力。

**2.2 RBF 网络的建立与仿真** 在确定基函数中心的问题上,由于随机选择中心点法在输入样本有一定的冗余时,精度就会下降,不能保证所设计网络的最优性能,所以选择不需要数据矢量的先验知识,并且运算速度较快的 K 均值聚类法来确定径向基函数网络的中心。

根据评价等级的设置将归一化后的训练输入用 K 均值聚类法划分为 6 个聚类,以各类中样本到对应中心的距离的总和最小的一次聚类分析结果来初始化 RBF 网络基函数的中心,很大程度上解决了 K 均值聚类算法易陷入局部最优的问题,从而为 RBF 网络提供了良好的隐含层节点基函数中心。

用 NEWFF 函数建立层数为 2,网络输入维数为 7,网络输出维数为 1 的 RBF 神经网络。其隐含层神经元个数为 6 个,神经元传递函数为 RADBAS,加权函数为 DIST,输入函数为 NETPROD;输出层神经元个数为 1 个,神经元传递函数为纯线性函数 PURELIN,加权函数为 DOTPROD,输入函数为 NETSUM。扩展常数经试验定为 0.2,输出层的权值和阈值通过最小二乘法求出。

将进行过归一化处理的水质标准作为训练输入,对应的评价等级作为训练输出。在网络建立后,对 5 个断面的监测数据用 TRAMNMX 函数同样进行归一化处理作为测试输入,用 SIM 函数仿真来获取测试输出。最后结果用 POSTMNMX 函数进行反归一化求出真正的评价等级。评价等级对应数值 < 1 的为 一类水质、1 ~ 2 的为 二类水质、2 ~ 3 的为 三类水质、3 ~ 4 的为 四类水质、4 ~ 5 的为 五类水质、5 ~ 6 的为 劣五类水质,从而对各断面水质进行整体评价。

### 3 评价结果与讨论

依据国家环保总局颁布的《地表水环境质量标准》,将 2004 年平顶山市市控 5 个断面主要影响指标的单因子评价

等级列出(表3)。

从表3可以看出 A、B、E 3 个断面水质整体情况较好,C、D 断面水质较差。其中 E 断面仅高锰酸盐指数一项评价因子为二类标准,其他 6 项因子均为一类标准;A、B 断面氨氮、高锰酸盐指数为二类标准,其他 5 项评价因子等级均为一类标准;C 断面石油类、挥发酚等 3 项评价因子均为一类标准,仅氨氮 1 项超五类标准,水质比 D 断面稍好;D 断面化学需氧量、氨氮、氟化物等 5 项评价因子均超五类标准,水质最差。

根据单因子评价方法的水质类别判定原则:在每个断面中选择污染最为严重的因子的水质类别代表该断面的水质类别。可以得出 A、B、E 断面为二类水质,C、D 断面为劣五类水质。

表3 2004 年各断面水质单因子评价等级

Table 3 Single factor assessment level of all sections in 2004

监测项目 Monitoring item	断面 Section				
	A	B	C	D	E
化学需氧量 Chemical oxygen demand	I类	I类	III类	劣V类	I类
氨氮 Ammonia nitrogen	II类	II类	劣V类	劣V类	I类
生化需氧量 Biochemical oxygen demand	I类	I类	IV类	劣V类	I类
高锰酸盐指数 Permanganate index	II类	II类	III类	劣V类	II类
石油类 Petroleum oil	I类	I类	I类	IV类	I类
挥发酚 Volatile phenol	I类	I类	I类	V类	I类
氟化物 Fluoride	I类	I类	I类	劣V类	I类

将 RBF 神经网络对监测数据仿真后的评价结果与单因子评价结果相比较(表4)。经过训练后的 RBF 网络对测试样本具有较高的仿真程度,其评价结果与单因子评价结果一致。同属二类水质的 A、B、E 断面从数值上比较(从优到劣):E 断面 > A 断面 > B 断面;同属劣五类水质的 C、D 断面从数值上比较(从优到劣):C 断面 > D 断面。

#### 4 结论

(1)通过试验表明,基于 K 均值聚类方法的 RBF 神经网络

(上接第 11511 页)

的数量较少,浮游动物的种类数目也较少,这些往往是污染的象征。一些多污带种类如浮游球衣藻(*Sphaerotilus natans*)、铜绿微囊藻(*Microcystis aeruginosa*)、小颤藻(*Oscillatoria tenuis*)、绿色裸藻(*Euglena viridis*)、镰形纤维藻(*Ankistrodesmus faicatus*)等<sup>[3]</sup>,成为了文昌桥段抚河中的优势种类。

络用于地表水环境质量评价是可行的,与常规方法相比较,具有学习速度较快,仿真结果量化的优点。可以较准确地反映出在同一类水质情况下不同断面水质的优劣程度。

表4 不同方法水质评价结果比较

Table 4 Comparison of water quality assessment result by different methods

断面 Section	单因子评价	RBF 网络
	Single-factor evaluation	RBF network
A 断面 Section A	II类	1.468 2(II类)
B 断面 Section B	II类	1.518 9(II类)
C 断面 Section C	劣五类	5.193 6(劣五类)
D 断面 Section D	劣五类	5.548 7(劣五类)
E 断面 Section E	II类	1.351 7(II类)

(2)RBF 网络的训练样本需要根据实际情况进行预处理,对于超五类水质标准的断面要人为设定劣五类标准来扩大训练样本的范围。并且训练样本要通过 RAND 函数随机插值来保证网络的充分学习,同时对训练和测试样本进行归一化处理,加快网络收敛速度。

(3)对预处理样本进行 K 均值聚类,能快速而有效地找到隐含层的中心位置,从而避免了因随机选取中心导致网络性能下降的情况。通过扩展常数的调节就能迅速找出性能较好的网络,并且评价结果客观、准确。

#### 参考文献

- [1] 张文艺. 基于人工神经网络的地面水环境质量评价模型[J]. 环境保护科学, 2000, 26(4): 40-42.
- [2] 蒋火华, 梁德华, 吴贞丽. 河流水环境质量综合评价方法比较研究[J]. 干旱环境监测, 2000, 14(3): 141.
- [3] 冯利华, 章明卓. 基于 ANN 的环境质量评价[J]. 四川环境, 2002, 21(3): 43.
- [4] 刘连芳, 魏砾宏, 李爱民, 等. 神经网络模型在辽河水质量评价中的应用[J]. 城市环境与城市生态, 2003, 16(6): 251-253.
- [5] 李泽应, 曹箐箐, 陈俊, 等. RBF 神经网络及其在水质评价中的应用[J]. 测控技术, 2006, 25(7): 69-70.
- [6] 刘宏友, 李莉, 彭峰. MATLAB6 基础及应用[M]. 重庆: 重庆大学出版社, 2002.
- [7] SIMON HAYKIN. 神经网络原理[M]. 2 版. 叶世伟, 等, 译. 北京: 机械工业出版社, 2004.
- [8] 许东, 吴铮. 基于 MATLAB6. X 的系统分析与设计——神经网络[M]. 西安: 西安电子科技大学出版社, 2002.
- [9] 蔡原将文, 谷菽隆嗣, 山口亨. 人工神经网络与模糊信号处理[M]. 马炫, 译. 北京: 科学出版社, 2003: 38.
- [10] 罗定贵, 王学军, 郭青. 基于 MATLAB 实现的 ANN 方法在地下水水质评价中的应用[J]. 北京大学学报: 自然科学版, 2004, 40(2): 299.

建议有关部门加强抚河的水质保护工作。

#### 参考文献

- [1] 高柏, 卫忠元, 徐卫东. 抚河流域水污染现状及可持续发展对策[J]. 中国农村水利水电, 2005(11): 49-53.
- [2] 韩茂生. 淡水浮游生物图谱[M]. 北京: 农业出版社, 1986: 170.
- [3] 刘洋. 常见浮游生物图谱(二)[J]. 科学养鱼, 2007(11): 85-86.