

BP神经网络在长江水质COD预测中的应用

郭庆春^{1,2,3}, 郝源¹, 李雪^{2,3}, 杜北方¹, 张向阳¹

(1. 陕西广播电视大学, 陕西 西安 710119;

2. 中国科学院地球环境研究所黄土与第四纪国家重点实验室, 陕西 西安 710075;

3. 中国科学院大学, 北京 100049)

摘要:水质变化具有非线性、突变性,且含有噪声,传统线性预测模型不能全面反映其变化规律,预测精度低,误差大。针对水质变化规律复杂,影响因素间非线性程度高的问题,为了提高水质预测精度,将改进算法的BP神经网络引入化学需氧量(COD)预测预报领域,以pH、溶解氧(DO)、氨氮(NH₃-N)为输入向量,以COD为输出向量,建立了COD的预测模型并对效果进行检验。结果表明:检验样本中COD的预测值与实测值的线性相关系数为0.991。BP神经网络模型预测精度高,收敛速度快,具有良好的泛化能力,能较好地反映COD和影响因子的变化规律。

关键词:神经网络;水质;化学需氧量;溶解氧;氨氮

中图分类号: X522

文献标识码: A

文章编号: 1673-629X(2014)04-0235-04

doi:10.3969/j.issn.1673-629X.2014.04.060

Application of BP Neural Network in Predicting COD of Yangtze River

GUO Qing-chun^{1,2,3}, HAO Yuan¹, LI Xue^{2,3}, DU Bei-fang¹, ZHANG Xiang-yang¹

(1. Shaanxi Radio & TV University, Xi'an 710119, China;

2. State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710075, China;

3. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Water quality change is of nonlinear and dynamicity, it is a kind of complex time series data, therefore, the traditional linear prediction model cannot reflect the variation rule, and the prediction accuracy is low. For the problems of complex water quality change rule and high degree of nonlinear between factors, in order to improve the water quality prediction accuracy, introduce the BP neural network of improved algorithm into a model of COD, with pH, DO, NH₃-N as input and COD as output, the prediction model of COD is established and tested. The research results show the linear correlation coefficient of COD between forecasting and the monitoring in the test samples is 0.991. BP neural network has high forecast precision, fast convergence rate and the good generalization ability, which can better reflect the change rule between COD and impact factors.

Key words: neural network; water quality; COD; DO; NH₃-N

0 引言

化学需氧量(COD),指一定条件下水样被强氧化剂氧化时所需氧化剂的量,主要指示水体被还原性物质污染的程度,由于水样普遍地受到有机物污染,因此,化学需氧量可以作为有机物相对含量的指标之一,同时也是河流水质规划的一项重要指标。长江流域水质规划把COD作为最重要的指标,建立河流污染断面的COD预测模型,是河流水质规划的一项基础工作。

关于COD的预测研究已经有很多成果,如田智勇等^[1]利用数理统计的方法研究了COD和pH与氮素基质浓度之间的关系,结果表明COD和pH与基质NH₄⁺-N浓度之间呈良好的线性相关关系;官宝锐等^[2]分别建立了逐步回归分析、ARMA、基于时间序列分析的神经网络模型和基于回归分析的神经网络模型,通过比较选择了基于时间序列分析的神经网络模型作为对污水厂出水COD的预测模型,其平均预测精

收稿日期:2013-05-31

修回日期:2013-09-06

网络出版时间:2014-01-07

基金项目:国家重点基础研究发展规划项目(2010CB833406);国家自然科学基金资助项目(40975020,41075067);陕西省教育科学研究计划项目(12JK0123,12JK0414)

作者简介:郭庆春(1979-),男,博士研究生,讲师,研究方向为环境信息系统。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140107.1648.029.html>

度为 85%,取得了满意的预测结果;宓云翥等^[3]分别运用 LM-BP 神经网络、最小二乘法、支持向量机的相关性模型进行 COD 预测,结果表明 2 种机器学习方法的预测能力明显优于最小二乘法,能够得到满意的预测精度。水质浓度预测过程可认为是一种复杂非线性函数关系的逼近过程,要对化学需氧量进行比较精确的预测,应该利用能够捕捉非线性变化规律的预测手段。神经网络具有自组织、自学习能力、非并行结构、并行处理信息等特点,特别是神经网络具有逼近任意连续函数和非线性映射的能力,并可进行高维非线性的精确映射,已在很多领域得到应用^[4-10],并取得理想效果。如李国庆等^[4]应用 BP 人工神经网络方法计算分层注,以有效厚度、渗透率、砂岩厚度和沉积相影响系数等 16 个影响因素作为模型输入参数,单层吸水量作为模型输出。实例计算结果表明 BP 人工神经网络法计算分层注水量与实测值的最大误差为 6.51%,平均误差为 3.21%,准确性较好,说明 BP 人工神经网络方法在分层注水量计算方面具有较好的应用前景。在脑机接口(BCI)中,脑电信号(EEG)的特征提取和分类识别可以通过多层前馈神经网络的大量学习来实现,但是基于误差反向传播的 BP 人工神经网络标准算法收敛速度慢,在训练中效率不高,分类正确率也很有限,陈悦等^[5]针对这些问题,提出使用一种快速稳定的 Levenberg-Marquardt 算法来代替 BP 算法进行神经网络的学习训练,并利用 BCI 2008 竞赛的 Graz 数据集 B 进行了对左右手想象运动脑电信号分类的 Matlab 仿真实验。该方法使得脑电信号分类的正确率达到 87.1%,比 BP 算法的正确率 78.2% 要高,并且具有更好的收敛性,该算法为脑电信号的分类提供了有效的手段。彭勇等^[10]针对公路客货运量预测的问题,对现有的常用预测方法进行研究,提出了改进 BP 人工神经网络预测模型。该模型首先采用动态陡度因子改变激励函数的陡峭程度,改善激励函数的响应特征,得到更好的非线性表达能力;其次利用附加动量因子,通过将以前的经验进行积累,有效地降低了人工神经网络对误差曲面局部细节的敏感性,很好地控制了网络陷于局部最小的缺陷;再次采取变学习率学习算法,先给一个较大初值,随着学习过程的进行,学习率不断减小,网络趋于稳定。改进的 BP 算法既可以找到更优解,又可以缩短训练时间。结合某地区的公路运量相关数据,对改进 BP 人工神经网络预测模型进行了验证。实验结果表明该模型的相对误差和迭代次数都取得了较大的改善,对公路客货运量预测很有效。针对长江 COD 时间序列变化的非线性问题,应用 BP 神经网络来预测长江干流江苏南京林山断面的 COD,为长江干流的水质浓度预测提供新的思路和方法。

1 数 据

水质数据来自中国环境保护部数据中心,时间范围为 2004 年~2011 年,资料为江苏南京林山断面逐周水质数据,水质好坏的变化主要因为是水体中氨氮($\text{NH}_3\text{-N}$)、溶解氧(DO)、化学需氧量(COD)、pH 变化造成的,所以采用自动监测站四项指标:pH、DO、 $\text{NH}_3\text{-N}$ 、COD。为有效地加快训练速度,提高灵敏性并能够有效地避开 Sigmoid 函数的饱和区,普遍地要求输入向量的值在 0~1 之间。所以,需要对输入向量采取预处理的办法。为了确保构建的神经网络模型具有一定的外推能力,最佳的方式是预处理后的值在 0.2~0.8 之间,把 0.8~1.0 和 0~0.2 的区间值留出来,用来预测。以下是数据预处理法:

$$f(x) = 0.6 * (x - \min(x)) / (\max(x) - \min(x)) + 0.2$$

其中, $f(x)$ 是水质浓度预处理后的数据; x 是水质浓度原始数据; $\min(x)$ 和 $\max(x)$ 分别是江苏南京林山断面原始数据变化范围的最小值和最大值。

2 构建人工神经网络模型

人工神经网络(Artificial Neural Network, ANN)是为了模仿人类大脑工作原理而建立的计算理论与技术方法。神经生物学家是采用自下而上的方法从单个神经到神经网络对刺激响应进行研究,而神经生理学家是采用自上而下的方式对大脑的行为和认识功能进行研究的,在对大脑的一些基本功能有了一定认识的基础上,计算机科学家、工程技术人员、数学家,尝试采用数学方程的方法描述大脑的工作原理。从 1943 年 Pitts 和 McCulloch 提出 M-P 神经网络模型到现在,ANN 已经取得了在技术、理论、应用方面的很多成果,应用范围涉及了认识科学、神经生理科学、心理科学、数理科学、计算机科学、信息科学、环境科学、管理科学等学科。

人工神经网络是由大量简单元件—神经元(节点)广泛相互连接而成的具有非局域性、非定常性、非线性和非凸性的复杂网络系统,具有并行分布的信息处理结构和自适应性的脑模式的信息处理的本质与能力,它可以通过“自学习”掌握大量的知识,完成特定的工作。其中应用较多的是具有非线性传递函数神经元构成的利用误差反向传播算法当做学习算法的前馈网络(Back error Propagation, BP)。图 1 为研究化学需氧量预测采用的神经网络结构图。

典型的 BP 网络模型是人工神经网络中使用最广泛和最具代表性的一类模型。从结构上来说,BP 网络是多层网络模型,分为输出层、隐含层、输入层,各层之间实行全连接。BP 网络模型实现了多层学习的构想,

假如给定 BP 网络一个输入模式,它由输入层神经元传到隐含层神经元,然后经过隐含层神经元逐层处理后再发到输出层神经元,由输出层神经元处理后产生一个输出模式,这时完成了一个逐层状态更新过程,被称为前向传播,如果输出结果与期望输出模式有误差,不能够满足需求,则进入误差反向传播,将误差值沿连接通路逐层反向传播并修正各层连接阈值和权值,当各个训练模式都满足需求时,那么学习结束。

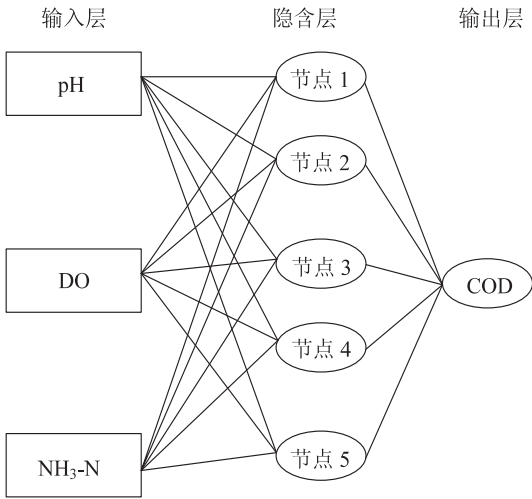


图 1 BP 神经网络结构图

在地表水化学需氧量的测定中,目前应用最普遍的是酸性高锰酸钾氧化法。高锰酸盐指数是以高锰酸钾溶液为氧化剂测得的化学耗氧量,其影响因素很多,而且物理化学生物过程复杂,其中,氨氮是水体中的主要耗氧污染物,水体中溶解的氧量称之为溶解氧量(DO),可用来指示水体洁净的程度。由于受实测资料的限制,选 pH、DO、NH₃-N 的浓度作为输入向量,COD 为输出向量,由此神经网络的输入层神经元为 3 个,输出层神经元为 1 个。为了有效地避开训练时出现的“过拟合”问题,确保足够高的泛化能力和网络性能,决定隐含层神经元数量的基本原则是:在确保预测精度要求的情况下,利用尽可能紧凑的结构,就是取尽量少的隐含层神经元数量。因此在水质浓度预测中采用的方法是分别构建了隐含层数量从 1~30 的 BP 网络,最后依据预测效果,决定了隐含层神经元数量是 5。在设置初始权值和阈值等参数时,都从 0.1~0.9 取值分别作了实验,当分别取值是 0.4 和 0.5 时计算速度较快,效果最好。转换函数经常用到的是两种:线性函数、Sigmoid 函数,研究化学需氧量预测时取尽了这二种函数的搭配组合方式,测试了不同组合方式对网络性能的影响结果,实验结果表明了当隐层-输出层为线性函数、输入-隐层为 tansig 函数时,结果最佳。

标准的 BP 算法是基于梯度下降法,通过计算目标函数对神经网络阈值和权值的梯度进行修正。改进算法多数是在标准梯度下降法的基础上发展起来的,

它们只是用到目标函数对阈值和权值的一阶导数(或者梯度)信息。标准的 BP 算法虽然为训练网络提供了简单而有效的方法,但是由于训练过程中学习速率为一较小的常数,因而存在收敛速度慢和局部极小问题。为解决这些问题,研究人员提出很多改进的算法。

附加动量法是在利用 BP 神经网络修正其阈值和权值时,不仅仅兼顾了误差在梯度上的影响,而且兼顾了在误差曲面上变化规律的作用,它许可忽略网络上的微小变化特征。这个方法就是在反向传播法的基础上,在每一个阈值和权值的变化基础上加上了一项正比前次阈值和权值变化量的值,并且依据反向传播法来产生新的阈值和权值的变化,这样的话就降低了网络性能对参数调节的敏感性,有效地制止了局部极小的缺点,但是收敛速度相对慢,4 500 迭代次数的训练并把目标误差增大为 0.06 时,还是难以找出最优的网络阈值和权值。

弹性 BP 算法(RPROP)的基本原理是阈值和权值更新值的直接修改,它与基于学习速率的算法相反,弹性 BP 算法引入了 Resilient(有弹性的)更新值的概念进行直接修改权步的大小值,所以修改结果不会因为不可预见的梯度性能而变模糊。因为学习规律简单、清楚,与开始的反向传播算法相比较,在计算上只有很少的耗费,除了快速之外,弹性 BP 算法的主要优点是对很多问题不需要参数的选择就可以得到最优或者接近最优收敛时间。

自适应学习速率法利用自适应调整学习速率可以有效地减少学习时间,标准的 BP 算法收敛速度慢,其主要原因就是学习速率选择不当,如果学习速率选得太小,收敛就会很慢;如果学习速率选取得太大,就会修正过头,引起发散。所以出现了自适应调整的改进算法,一般地调整学习速率 η 的标准法则是:检查权值的修正值是否有效地降低了误差函数,如果降低了误差函数,那么表明选取的学习速率值过小,可以对其增加一个数值;如果不是这样,而是产生了过调,那么就应该减小学习速率的数值。

共轭梯度法就是采用共轭梯度向量来确定共轭方向的算法。在共轭梯度法中,顺着共轭方向进行行搜索,收敛速度会比自适应学习速率法快的多,在自适应学习速率法的训练算法中,是采用学习速率决定阈值和权值更新的步长,而在共轭梯度算法中权值和阈值的步长各自经过反复调整,顺着共轭梯度利用行搜索来决定阈值和权值步长以减少在行中的完成功能。共轭梯度法一般不用计算或者存储二阶导数信息就具有了二阶导数方法的性能,与拟牛顿法相比较,它的计算代价非常低,所以对于大规模的问题非常有用。

拟牛顿算法(BFGS)比共轭梯度法收敛速度快的

多,但是在前馈反向网络中海森矩阵计算耗时、复杂,而拟牛顿法是基于导数信息的,尝试通过迭代次数逐渐构建对海森矩阵的近似,尽管拟牛顿算法一般只要很少的迭代次数就能够收敛,但它每步迭代的内存和计算量需求大于共轭梯度法。由于需要存储近似海森矩阵,它相当于网络中阈值和权值的个数,通常对于大型网络来说,建议采用弹性 BP 算法或共轭梯度法,而训练小型网络时建议采用拟牛顿算法效果会较好。

由于拟牛顿算法每一步迭代时的计算量和存储量都大于共轭梯度法,因此需要有更少计算量的正切逼近。一步割线 (OSS) 算法,即 One Step Secant (OSS) 方法,就是拟牛顿法和共轭梯度法的折衷方法。该方法不需要存储整个海森矩阵,假如每次迭代时上一次的海森矩阵是正定的,并且在计算新的搜索方向的时候,没有用到海森矩阵的逆,一步割线算法每步需要的计算量和内存都小于拟牛顿算法,但是稍微大于共轭梯度法。

Levenberg-Marquardt 优化方法可以依据网络训练误差变化趋势,自动控制网络训练参数 μ ,能够让网络不断更新适宜的训练方法。如果 μ 相对大的时候,网络训练过程主要根据梯度下降算法,如果 μ 相对小的时候,网络训练过程主要根据 Gauss-Newton 法;利用 Levenberg-Marquardt 优化方法,可以让学习时间更短,在实验中效果比较理想,但对于相对复杂的难题,该方法需要很大的内存。

传统的标准 BP 算法存在收敛速度慢和局部极小问题,为解决这些问题,研究中将上述方法做了对比分析,最后确定训练方式为 trainlm。该算法期望在不计算 Hessian 矩阵的情况下获得高阶的训练速度,由于牛顿法在误差极小点附近通常能够收敛的更快、更准确,因此算法的目的就是尽快转换为牛顿法。该方法收敛速度较快,拟合精度高,预防出现“过度训练”,训练的网络稳定性会优于其他函数,仅仅 68 次的训练就可以让网络收敛,找到最优的网络阈值和权值。采用 LM 优化方法,能够让学习时间缩短,在 COD 预测中效果很理想。

3 结果分析

采用江苏南京林山断面 2004 ~ 2010 年的逐周水质指标资料作为神经网络模型的训练样本,进行训练,建立了 COD 预测模型。为了验证预测模型的正确性,采用训练好的神经网络模型,利用调整后的阈值和权值,然后采用 2011 年的数据作为检验样本进行预测预报检验,图 2 为预报结果。

实验结果表明:训练样本中 COD 的拟合值的平均相对误差为 4.1%,通过显著性水平 0.05 的 t 检验,检

验样本中 COD 的拟合值与实测值的线性相关系数为 0.989;检验样本中 COD 的预测值与实测值的平均相对误差仅仅为 3.4%,COD 的预测值与实测值的线性相关系数为 0.991。说明模型拟合的效果均较好,预测值的相对误差较小,预测值与实测值的线性相关系数较高,此模型可以满足预测 COD 的需要。

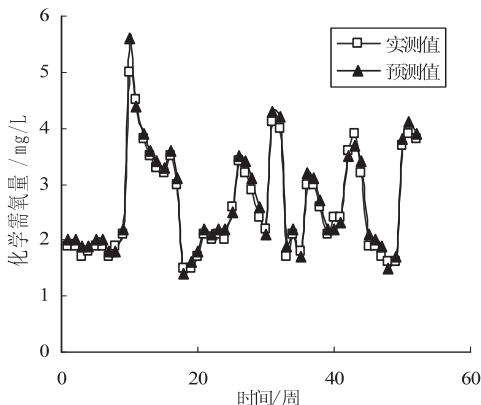


图 2 2011 年 COD 预测值和实测值的对比图

4 结束语

采用 BP 网络的改进算法,建立了基于改进型 BP 神经网络的 COD 预测模型。利用 2004 年 ~ 2012 年的数据作为训练样本,并对 2011 年的 COD 进行了预测,通过模型建立和研究可以得出以下结论:

(1) 采用江苏南京林山断面 2004 ~ 2010 年 COD 的数据作为训练样本,训练人工神经网络,调整、修正神经网络中各权值和阈值,并用 2011 年的 COD 数据作为独立样本进行预测,证实模型的合理性。用 Matlab 实现改进 BP 神经网络模型,具有计算精确、使用方便的优点,分析得出的 BP 神经网络预测模型,可以不断添加历史数据,进一步提高预测能力,提高网络的可信度。

(2) 从实验结果看,经训练样本训练后得到的神经网络模型不仅对训练样本具有较好的拟合能力,而且对测试样本也具有很高的模拟能力,说明所建立的模型具有较高的预测精度,该神经网络的泛化能力很强,能够取得良好的效果。由此可见,运用 BP 神经网络预测 COD 的方案是可行的,而且采用神经网络理论进行 COD 预测,具有明显的实用性。

(3) BP 网络模型能够很好地捕捉 COD 和影响因素的变化规律,不用设计复杂的数学方程,仅仅利用神经元之间的相互影响可以实现网络的信息过程处理,并且能够取得理想的预测准确度。利用人工神经网络模型可避免以往为寻找水质数学模型而消耗的大量人力、物力和财力。亦可大大改善因各种随机污染或模型误差而造成的预测精度低的问题,可在河流水质的

(下转第 242 页)

激活过程,就由议程来处理这些产生冲突的激活过程,指定这些过程的执行顺序。

(4)工作内存。
工作内存从已有的数据库中提取有效的并需要处理的气象数据信息,用于生成当前需要处理的气象灾害大数据预警信息数据对象,为逻辑推理模块提供匹配事实。该模块的实现较为简单,这里不做具体描述。

3 结束语

文中针对目前气象信息处理的弊端,根据软件工程的理论,将业务逻辑和逻辑推理功能、数据存储、数据处理等进行有效的剥离,设计了一套针对大数据量的气象信息灾害预警系统。系统给业务人员提供了创建规则功能,然后利用规则引擎技术对大数据信息进行逻辑推理,得出业务结果,有效提高了气象大数据业务处理的效率。同时,由于实现了业务规则和程序代码的分离,降低了系统的复杂性,提高了代码的可维护性和重用率。

参考文献:

[1] Sottara D, Mello P, Proctor M. A configurable Rete-OO engine for reasoning with different types of imperfect information[J]. IEEE transactions on knowledge and data engineering, 2010, 22(5):1-14.
[2] Doocy S E A. Implementing cash for work programmes in post-tsunami Aceh: Experiences and lessons learned[J]. Disas-

ters, 2006, 30(3):277-296.
[3] Meng Jin, Jiang Limin, Li Qianmu, et al. Research on risk correlation based on the module coupling[J]. Acta armamentarii, 2011, 32(2):211-216.
[4] Xiao Ding, Zhong Xiaolan. Improving rete algorithm to enhance performance of rule engine systems[C]//Proc of 2010 international conference on computer design and applications. Qinhuangdao:[s. n.], 2010:572-575.
[5] 张 彪. 基于 Rete 算法的数据库通知引擎技术研究[D]. 上海:上海海事大学, 2004.
[6] 鲍金玲. 基于规则引擎技术的 Rete 算法的研究[J]. 科技信息, 2008(32):90-90.
[7] 规则引擎研究(一)-Reta 算法[EB/OL]. 2010. <http://hi.baidu.com/dburu/blog/item/d34bb68b00130a779f2fb47e.html>.
[8] 彭 磊. 规则引擎原理分析[J]. 福建电脑, 2006(9):42-42.
[9] 数据海量增长引发高级分析之争[EB/OL]. 2010. <http://tech.sina.com.cn/b/2010-04-21/09454088878.shtml>.
[10] Chen Yeguo, Nong Mengsong. Causality analysis on a severe chilling icy rain and snow freezing disaster weather event in early 2008 in Guangxi[J]. Journal of meteorological research and application, 2008, 29(2):12-14.
[11] Huang Honghui, Lin Bomin, Wang Da. Discussion on the dic-tion of weather forecasting services[J]. Journal of meteorological research and application, 2010, 31(2):273-278.
[12] 刘 伟. Java 规则引擎-Drools 的介绍及应用[J]. 微计算机应用, 2005, 26(6):717-721.

(上接第 238 页)

智能化建模、流域规划、预测决策、污染控制等方面发挥积极作用。

参考文献:


[1] 田智勇,李 冬,张 杰. 厌氧氨氧化过程中 COD 及 Ph 与基质浓度之间的关系[J]. 环境科学, 2009, 30(11):3342-3346.
[2] 官宝锐,徐得潜. 污水厂出水 COD 预测模型的对比分析[J]. 环境保护科学, 2011(2):65-67.
[3] 宓云蜚,王晓萍,金 鑫. 基于机器学习的水质 COD 预测方法[J]. 浙江大学学报(工学版), 2008, 42(5):790-793.
[4] 李国庆,赵恩远,任帅勤,等. 基于 BP 神经网络方法的分层注水量计算[J]. 科学技术与工程, 2012, 12(10):2425-2427.
[5] 陈 悦,张少白. LM 算法在神经网络脑电信号分类中的研

究[J]. 计算机技术与发展, 2013, 23(2):119-122.
[6] 张凯歌,缪 毅,雷建坤,等. 结合小波插值与 K-means 方法提取彩色图像文字[J]. 计算机技术与发展, 2013, 23(3):31-33.
[7] 白 燕,马光思. 基于灰色径向基神经网络模型的流量预测与分析[J]. 计算机工程与科学, 2008, 30(10):122-124.
[8] 罗 莉,罗 强,何鸿君. 前馈多层神经网络 BP 算法与可靠性增长模型[J]. 计算机工程与科学, 2001, 23(3):55-58.
[9] 李 翔,朱全银. Adaboost 算法改进 BP 神经网络预测研究[J]. 计算机工程与科学, 2013, 35(8):96-102.
[10] 彭 勇,陈俞强,严文杰. 基于改进 BP 网络模型的公路流量预测[J]. 计算机技术与发展, 2012, 22(8):111-113.

BP神经网络在长江水质COD预测中的应用

作者：郭庆春，郝源，李雪，杜北方，张向阳，GUO Qing-chun，HAO Yuan，LI Xue，
DU Bei-fang，ZHANG Xiang-yang

作者单位：郭庆春, GUO Qing-chun(陕西广播电视大学, 陕西 西安710119; 中国科学院 地球环境研究所黄土与第四纪国家重点实验室, 陕西 西安710075; 中国科学院大学, 北京100049)，郝源, 杜北方, 张向阳, HAO Yuan, DU Bei-fang, ZHANG Xiang-yang(陕西广播电视大学, 陕西 西安, 710119)，李雪, LI Xue(中国科学院 地球环境研究所黄土与第四纪国家重点实验室，陕西 西安710075; 中国科学院大学，北京100049)

刊名：计算机技术与发展 

英文刊名：Computer Technology and Development

年，卷(期)：2014(4)

本文链接：http://d.wanfangdata.com.cn/Periodical_wjfz201404060.aspx