

未确知测度 BP 神经网络模型 在黑河流域水质预测中的应用

李 慧, 周轶成

(甘肃省张掖水文水资源勘测局, 甘肃 张掖 734000)

摘要:水质预测是水环境规划、评价和管理的重要依据,对促进水资源可持续利用及生态发展具有重要意义。针对水质预测中各项因子的不确定性,基于未确知测度理论(unascertained measure, UM),采用改变网络初值的方法,对 BP 神经网络加以改进,并利用黑河流域莺落峡水文站 1998~2011 年的水质监测资料进行分析和预测。以挥发酚为参考序列,用灰色关联方法分析参考序列与其他因子的关联度,并最终确定 BP 网络的输入节点为 COD_{Mn}、DO、SO₄²⁻、Cr⁶⁺以及挥发酚,输出节点为挥发酚,从而建立 UM-BP 模型。分析结果表明,UM-BP 预测模型比标准的 BP 神经网络模型具有更高的预测精度。因此,该模型应用于黑河流域水质预测是可行的。

关键词:未确知测度;权值;BP 神经网络;水质预测;黑河流域

中图分类号:X830

文献标识码:A

文章编号:1000-0852(2014)03-0066-04

引言

科学合理的水质预测对水资源开发利用,实时了解河流水质的发展趋势,制定水环境监测和保护实施计划都有重要意义。水质预测方法主要有非点源污染、神经网络、支持向量回归机等^[1-3]。BP 网络是利用非线性可微分函数进行权值训练的多层网络,其结构简单,可塑性强,故在函数逼近、模式识别、信息分类及预测等领域得到广泛应用^[4]。但传统的 BP 算法易陷入局部极小点,存在收敛速度慢、预测误差较大等不足。为此,许多学者从权值方面对 BP 网络进行了改进,取得了许多成果^[5-8]。由于水质监测存在很多天然或人工的误差,得到的数据往往具有不精确性、不确定性,资料总体呈现着未确知性,未确知测度可以有效的解决评价或预测系统中许多不确定的问题,其权重并不凭经验而是根据属性的实测值来确定,选择过程较简便^[9]。因此,本文运用未确知测度理论得出的权重值作为 BP 网络的权值,建立 UM-BP 模型,并对黑河流域莺落峡水文站水质进行预测。

1 基于未确知测度的 BP 神经网络模型(UM-BP 模型)

1.1 流域概况

黑河流域是我国的第二大内陆河流域,位于 98°~102°E、37°50'~42°40'N,流经青甘蒙三省(区),流域面积 14.29×10⁴km²,出山口莺落峡以上为上游,是黑河流域的产流区和水源来源区;莺落峡与正义峡之间为中游,地处河西走廊,为平原盆地,是流域的耗水区和径流利用区;正义峡以下为下游,属于阿拉善高原区,大部分为荒漠、沙漠和戈壁,属极端干旱区。

黑河流域水质状况较好,但中游部分河段水质已有下降趋势。据 1998 年对洪水河、梨园河、黑河、山丹河共 589km 河段水质评价结果,属至类水的河段占评价河长的 85.4%,类、类水河长占 14.6%。因此,对黑河流域水环境合理的进行规划管理、对水污染综合治理已经成为亟待完成的重要任务之一,而水质预测则是其中一项不可或缺的基础工作。

莺落峡水文站是出山口的控制性水文站,位于黑河流域的中游地段,控制流域面积约 1×10⁴km²,是黑

河的主要产流区。因此研究莺落峡水文站的水质对整个黑河流域的分析监测有至关重要的作用。

1.2 未确知测度理论^[9-10]

未确知信息是不同于模糊信息、随机信息和灰色信息的一种新的不确定信息。未确知测度理论是未确知数学理论的重要成果,目前该理论已被运用到多个领域。

1.2.1 分类标准矩阵的建立

设评价对象空间 $Z=\{z_1, z_2, \dots, z_n\}$, 对于每个评价的对象 $z_i(i=1, 2, \dots, n)$ 有 m 个单项评价指标空间 $X=\{x_1, x_2, \dots, x_m\}$, 则 Z_i 可以表示为 m 维向量, $Z_i=\{x_{i1}, x_{i2}, \dots, x_{im}\}$, 其中 x_{ij} 表示研究对象 Z_i 关于评价指标 x_j 的实测值。对每个 x_{ij} 有 p 个评价等级。设 $c_k(k=1, 2, \dots, p)$ 表示第 k 个评价等级, 则 k 级好于 $k+1$ 级, 记作 $c_k > c_{k+1}$ 。设评价空间 $U=\{c_1, c_2, \dots, c_p\}$, 若满足 $c_1 > c_2 > \dots > c_p$ 或 $c_1 < c_2 < \dots < c_p$ 则称 $\{c_1, c_2, \dots, c_p\}$ 是评价空间 U 的一个有序分割类。

1.2.2 单指标未确知测度的构造

若 $\mu_{ijk}=\mu(x_{ijk} \in c_k)$ 表示监测值 x_{ij} 属于第 k 个评价等级 c_k 的程度, 则 μ 满足:

$$0 \leq \mu(x_{ij} \in c_k) \leq 1 \quad (1)$$

$$\mu(x_{ij} \in U) = 1 \quad (2)$$

$$\mu\left(x_{ij} \in \bigcup_{q=1}^k c_q\right) = \sum_{q=1}^k \mu(x_{ij} \in c_q) \quad (3)$$

其中, (3) 式的 $q=1, 2, \dots, k$, 上式分别为“非负有限性”, “可加性”, “归一性”, 称满足 (1)、(2)、(3) 的 μ 为未确知测度, 简称确度, 称矩阵

$$(\mu_{ijk})_{m \times p} = \begin{pmatrix} \mu_{i11} & \mu_{i12} & \dots & \mu_{i1p} \\ \mu_{i21} & \mu_{i22} & \dots & \mu_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{im1} & \mu_{im2} & \dots & \mu_{imp} \end{pmatrix} \quad (4)$$

为单指标测度评价矩阵。

1.2.3 指标区分权重

将待评价对象 x_{ij} 所属等级区分开的程度在所有 m 项评价指标中占的比例, 称分类权重, 记为 w_{ij} 。利用 μ_{ijk} 所确定的信息熵和分类权重的计算公式, 得到评价指标的分类权重向量 $w_i=(w_{i1}, w_{i2}, \dots, w_{ip})$ 。其中 $(i=1, 2, \dots, n; j=1, 2, \dots, m)$ 。

$$v_{ij} = 1 + \frac{1}{\lg m} \sum_{k=1}^m \mu_{ijk} \lg \mu_{ijk} \quad (\text{其中 } \sum_{j=1}^m w_{ij} = 1; 0 \leq w_{ij} \leq 1) \quad (5)$$

$$w_{ij} = v_{ij} / \sum_{j=1}^m v_{ij} \quad (6)$$

式中: m 为评价级别的数目; v_{ij} 为监测值 x_{ij} 提供的信息量; μ_{ijk} 为单指标测度。

若式中某个 $\mu_{ijk}=0$, 则观测值 x_{ij} 不属于第 k 个等级。所以, 在上式求和计算时取 $\mu_{ijk} \lg \mu_{ijk}=0$ 。

1.3 建立 UM-BP 模型

将未确知测度理论与 BP 神经网络结合可以在一定程度上克服 BP 算法收敛速度慢, 易陷入局部极值的不足, 提高 BP 网络的适应能力。基于未确知测度的 BP 神经网络模型的具体算法见图 1。

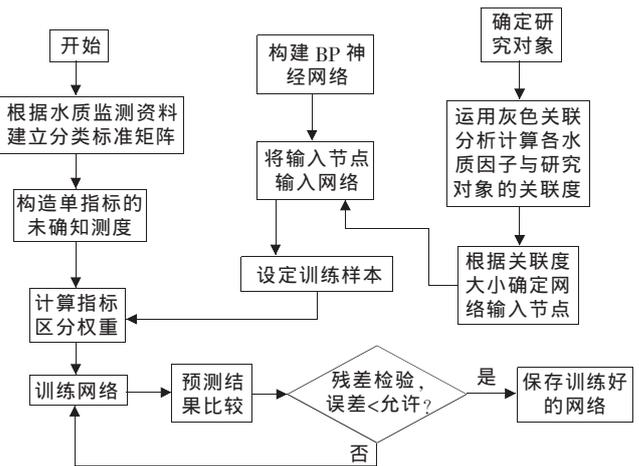


图 1 UM-BP 模型算法流程图

Fig.1 Algorithm flow chart of UM-BP model

2 实例应用

2.1 数据样本

本文选取甘肃省水环境监测中心张掖分中心黑河流域莺落峡水文站 1998~2011 年的水质监测资料, 其中 1998~2008 年水质监测资料为训练样本; 2009~2011 年水质监测资料作为检验样本。选择 7 个水质因子数据建立模型, 挥发酚是反映黑河流域水污染情况的一个重要指标, 而莺落峡水文站的水质监测资料中污染因子历年资料最全的也是挥发酚。因此本研究确定挥发酚为研究对象。文章首先采用拉依达准则方法^[11]剔除异常的原始数据, 经检验, 该水质监测资料的数据无异常。

为了加快网络学习的收敛速度, 需要对原始的输入、输出样本做一些处理, 即对样本系列进行归一化:

$$y_{ij} = \left(x_{ij} - \min_{1 \leq i \leq N} x_{ij} \right) / \left(\max_{1 \leq i \leq N} x_{ij} - \min_{1 \leq i \leq N} x_{ij} \right) \quad (7)$$

2.2 确定 BP 网络输入节点

经过灰色关联分析^[12]计算, 7 个水质因子与挥发

酚的关联度分别为: COD_{mn} (0.7128) > DO (0.5959) > SO_4^{2-} (0.5634) > Cr^{6+} (0.5579) > pH (0.5222) > NH_3-N (0.5138) > Cl^- (0.4495)。最终选择关联度高的 COD_{mn} 、 DO 、 SO_4^{2-} 、 Cr^{6+} 以及挥发酚本身作为 BP 网络的输入节点。

2.3 UM-BP 模型的应用

通过公式(5)和(6)计算分别得到 COD_{mn} 、 DO 、 SO_4^{2-} 、 Cr^{6+} 和挥发酚分类权重向量 w_i 。见表 1。(表 1 中每列分别表示这 5 个水质因子的分类权重向量)

表 1 莺落峡水文站 1998~2008 年水质指标权重

Table 1 Water quality index weight for the Yingluoxia station from 1998 to 2008

年度	COD_{mn}	DO	SO_4^{2-}	Cr^{6+}	挥发酚
1998	0.200	0.200	0.200	0.200	0.200
1999	0.211	0.211	0.211	0.156	0.211
2000	0.207	0.207	0.207	0.173	0.207
2001	0.218	0.218	0.218	0.130	0.218
2002	0.213	0.213	0.213	0.147	0.213
2003	0.137	0.240	0.240	0.143	0.240
2004	0.211	0.211	0.211	0.156	0.211
2005	0.207	0.207	0.207	0.173	0.207
2006	0.203	0.203	0.203	0.188	0.203
2007	0.203	0.203	0.203	0.188	0.203
2008	0.200	0.200	0.200	0.200	0.200

本文采用 3 层的 BP 网络建立挥发酚浓度的非线性映射关系模型进行水质预测, 输入样本为 5 维的输入向量, 输入层应有 5 个神经元, 则隐含层有 11 个神经元。网络只有 1 组输出向量, 则输出层有 1 个神经元。网络结构为 $5 \times 11 \times 1$ 的结构, 按照 BP 网络的一般设计原则, 中间层神经元的传递函数采用 $tansig$ 型函数, 此时输出已被归一化到区间 $[0, 1]$ 中, 因此输出层神经元的传递函数可以采用 $purelin$ 型函数。训练函数采用 $trainlm$ 型函数。确定网络的输入节点为 COD_{mn} 、 DO 、 SO_4^{2-} 、 Cr^{6+} 以及挥发酚 11 年的监测值, 输出节点为挥发酚。学习速率为 0.01。网络的权值则由表 1 计算的指标区分权重组合而成。

2.4 预测结果比较

基于标准的 BP 神经网络模型和 UM-BP 模型, 应用相同的样本, 对网络进行训练, 预测结果见表 2。

由表 2 可知, UM-BP 组合模型预测的相对误差明显小于标准的 BP 网络预测, 标准 BP 神经网络的相对误差平均值为 13.46%, 在利用模型组合后, 2009 年水质预测的相对误差从 14.30% 下降到 3.80%, 2010 年从

表 2 BP 网络预测结果

Table 2 The forecast results for BP network

年份	实测值	BP	BP	UM-BP	UM-BP
		预测值	相对误差	预测值	相对误差
2009	0.001	0.0008570	14.30%	0.0009620	3.80%
2010	0.001	0.0009162	8.38%	0.0009768	2.32%
2011	0.001	0.0008230	17.70%	0.0009438	5.62%

8.38% 下降到 2.32%, 2011 年从 17.70% 下降到 5.62%, 预测精度分别提高了 12.30%、6.60%、14.70%。UM-BP 模型的预测值比标准的 BP 网络的更接近真实值 0.001, 故前者的预测精度比较高, 可以满足水质预测的要求。经残差检验^[13], 标准的 BP 网络得出的 $\bar{\Delta}=0.1346$, 属于四级精度标准, UM-BP 模型的 $\bar{\Delta}=0.0042 \leq 0.01$ 属于一级精度标准, 检验后的结果表明, UM-BP 模型较标准的 BP 网络模型预测精度高。

3 结语

本文尝试运用未确知测度理论得出的权重值作为权值应用于 BP 神经网络中, 建立了 UM-BP 模型, 并选取甘肃省水环境监测中心张掖分中心黑河流域莺落峡水文站 1998~2011 年的水质监测资料, 对莺落峡最近 3 年的水质资料进行了预测。经过灰色关联分析计算, 原始资料的 7 个水质因子中只有 COD_{mn} 、 DO 、 SO_4^{2-} 、 Cr^{6+} 与挥发酚的关联度最高, 因此选择 COD_{mn} 、 DO 、 SO_4^{2-} 、 Cr^{6+} 以及挥发酚本身为网络的输入节点, 输出节点为挥发酚。在利用模型组合后, 2009 年水质预测的相对误差从 14.3% 下降到 3.8%, 2010 年从 8.38% 下降到 2.32%, 2011 年从 17.7% 下降到 5.62%, 预测精度分别提高了 12.30%、6.60%、14.70%。经残差检验, UM-BP 模型得出的平均相对误差属于一级精度标准, 而标准的 BP 网络仅属于四级精度标准, UM-BP 模型较标准的 BP 网络模型预测精度高。说明此预测模型应用于黑河流域水质是可行的。

参考文献:

- [1] 陈丁江, 吕军, 沈晔娜, 等. 非点源污染河流水质的人工神经网络模拟[J]. 水利学报, 2007, 38(12): 1519-1525. (CHEN Dingjiang, LV Jun, SHEN Yena, et al. ANN approach for modeling and prediction of water quality in non-point sources polluted river [J]. Journal of Hydraulic Engineering, 2007, 38(12): 1519-1525. (in Chinese))
- [2] 武国正, 徐宗学, 李畅游. 支持向量回归机在水质预测中的应用与验证[J]. 中国农村水利水电, 2011, 30(9): 1888-1895. (WU Guozheng, XU Zongxue, LI Changyou. Water quality prediction and validation by using support vector regression machines [J]. China Rural Water

- and Hydropower, 2011,30(9):1888-1895.(in Chinese))
- [3] 熊伟,程加堂,徐绍坤. 证据理论融合蚁群神经网络的水质预测方法[J]. 人民黄河, 2012,34 (5):76-81. (XIONG Wei, CHENG Jiatang, XV Shaokun. Water quality prediction based on evidence theory combines with ant colony algorithm-neural network[J]. Yellow River, 2012,34(5):76-81.(in Chinese))
- [4] 郭晶,孙伟娟. 神经网络理论与 MATLAB7 实现[M]. 北京:电子工业出版社, 2005. (GUO Jing, SUN Weijuan. MATLAB Application[M]. Beijing: Publishing House of Electronics Industry, 2005. (in Chinese))
- [5] 雷斌,陶海龙,徐晓光. 基于改进粒子群优化算法的灰色神经网络的铁路货运量预测 [J]. 计算机应用, 2012,32 (10):2948-2951. (LEI Bin, TAO Hailong, XU Xiaoguang. Railway freight volume prediction based on grey neural network with improved particle swarm optimization [J]. Journal of Computer Applications, 2012,32(10):2948-2951. (in Chinese))
- [6] 王改革,郭立红,段红,等. 基于萤火虫算法优化 BP 神经网络的目标威胁估计[J]. 吉林大学学报(工学版), 2012,42(3):34-37. (WANG Gaige, GUO Lihong, DUAN Hong. Target threat assessment using glowworm swarm optimization and BP neural network [J]. Journal of Jilin University (Engineering and Technology), 2012,42(3):34-37. (in Chinese))
- [7] Nie S, Zhong Y. Application of BP neural network model based on improved DNA genetic algorithm to tool wear monitoring [J]. Advanced Science Letters, 2012,(6):756-759.
- [8] Tang N, Zhang D J. Application of a load forecasting model based on improved grey neural network in the smart grid [J]. Energy Procedia, 2011,(12):180-184.
- [9] 刘开第,李万庆,庞彦军. 未确知集[J]. 数学的实践与认识, 2006,36 (10):197-204. (LIU Kaidi, LI Wanqing, PANG Yanjun. Unascertained set[J]. Mathematics in Practice and Theory, 2006,36(10):197-204. (in Chinese))
- [10] 杜泽生. 基于粗糙集与未确知测度理论的突出危险性评价模型[J]. 煤矿安全, 2012,43(10):7-10. (DU Zesheng. Outburst risk evaluation model based on rough set and unascertained measure theory [J]. The Coal Mine Safety, 2012,43(10):7-10. (in Chinese))
- [11] 张敏,袁辉. 拉依达(PauTa)准则与异常值剔除[J]. 郑州工业大学学报, 1997,18 (1):84-88. (ZHANG Min, YUAN Hui. The PauTa criterion and rejecting the abnormal value [J]. Journal of Zhengzhou University of Technology, 1997,18 (1):84-88. (in Chinese))
- [12] 谭鹏,曹平. 基于灰色关联支持向量机的地表沉降预测[J]. 中南大学学报(自然科学版), 2012,43(2):632-637. (TAN Peng, CAO Ping. Predicting surface settlement of tunnel using grey relational-support vector machine [J]. Journal of Central South University (Science and Technology), 2012,43(2):632-637.(in Chinese))
- [13] 房明惠,张国义,赵丽,等. 基于灰色模型的合肥市用水量预测[J]. 西安建筑科技大学学报(自然科学版), 2009,41(6):862-866. (FANG Minghui, ZHANG Guoyi, ZHAO Li, et al. Forecast of water consumption of Hefei based on GM (1,1) model [J]. Journal of Xi'an University of Architecture and Technology (Natural Science Edition), 2009,41(6):862-866. (in Chinese))

Unascertained Measure BP Neural Network Model for Water Quality Prediction in Heihe River Basin

LI Hui, ZHOU Yicheng

(Zhangye Hydrology and Water Resources Survey Bureau of Gansu Province, Zhangye 734000, China)

Abstract: Water quality prediction is an important basis of water environment planning, evaluation and management, and has important significance to promote the sustainable utilization of water resources and ecological development. According to the uncertainty of various factors in water quality prediction, based on unascertained measure (UM) theory, the method of changing the initial value of the network to improve BP neural network. And analysis and forecasting were made with the quality data from the Yingluoxia Station in the Heihe River Basin during 1998-2011. To analyze the correlation degree of reference sequence and other factors with gray correlation method (volatile phenol as the reference sequence), and determine the input nodes of BP network which includes COD_{mn}, DO, SO₄²⁻, Cr⁶⁺ and volatile phenol, the output nodes is volatile phenol, thereby establish UM-BP model. The analysis results show that UM-BP prediction model has a higher prediction accuracy than standard BP neural network model. Therefore, the model used in water quality prediction in the Heihe River Basin is feasible.

Key words: unascertained measure; weights; BP neural network; water quality prediction; Heihe River Basin