

基于 GA-BP 模型的景观小水体富营养化评价方法

张维砚¹, 沈蓓雷¹, 童琰¹, 胡雪芹¹, 徐春燕¹, 由文辉^{1*}, 薄芳芳² (1.华东师范大学环境科学系, 上海市城市化生态过程与生态恢复重点实验室, 上海 200062; 2.上海市绿化管理指导站, 上海 200021)

摘要: 为了研究适用于景观小水体富营养化评价的方法, 以上海市 10 个公园景观水体为实验对象, 在初步了解了景观小水体水质和营养状况的基础上运用主成份分析法对各项水质指标进行评价因子筛选, 并对筛选出来的因子建立富营养化评价等级。运用 MATLAB 软件的 GA-BP 算法建立评价模型对 10 个公园 4 个季度共 40 组数据进行评价, 结果显示与营养度指数法结果总体一致。选用的 Chla, TP, TN, COD_{Mn}, SD 五个评价因子的数据都比较易得, 模型可推广性强。

关键词: 景观小水体; 富营养化评价; GA-BP 模型

中图分类号: X703.5 **文献标识码:** A **文章编号:** 1000-6923(2011)04-0674-06

Evaluation method of small landscape waters eutrophication based on the GA-BP model. ZHANG Wei-yan¹, SHEN Bei-lei¹, TONG Yan¹, HU Xue-qin¹, XU Chun-yan¹, YOU Wen-hui^{1*}, Bo Fang-fang² (1.Shanghai Key Laboratory of Urbanization and Ecological Restoration, College of Resources and Environment Science, East China Normal University, Shanghai 200062, China; 2.Shanghai City Greening Management Station, Shanghai 200021, China). *China Environmental Science*, 2011,31(4): 674-679

Abstract: In order to study evaluation methods on eutrophication of small landscape waters, we take 10 garden landscape waters of Shanghai into account. We use principle component analysis method to filter out the evaluation factors and determine the evaluation criteria. The GA-BP model of Matlab is used to evaluate totally 40 data sets, which contains the data of 10 gardens under four seasons. Our results demonstrate that it is consistent with that AHC-PCA. Five evaluation factors, including Chla, TP, TN, COD_{Mn} and SD, can be easily obtained, as a result, our model could be widely extended.

Key words: small landscape waters; eutrophication evaluation; GA-BP model

上海市目前城市景观水体水域面积已经超过 210 万 m²^[1]。因其普遍具有闭锁性、流动性慢、交换周期长的特点使得富营养化较易产生, 如果管理控制不善易引发水华^[2]。藻类大量繁殖严重阻碍了景观水体的生态系统服务功能, 因此应加强景观水体富营养化的研究。

水体富营养化评价方法主要有营养度指数法、灰色聚类法以及综合营养状态指数法等^[3-5]。水体富营养化是多因子共同作用的结果, 各因子间的相互关系具有非确定性、非线性、以及模糊性的特点, 因此上述富营养化评价方法虽然各有优势但是均存在一定的局限性。基于遗传算法(GA)和 BP 算法有机结合的 GA-BP 算法所建立的水体富营养化评价模型具有较好的非线性映射能力, 适合在水体富营养化评价中运用。

地理位置和规划施工等因素的差异, 容易造成各景观小水体富营养化诱发因素的不同, 监测指标较多也加大评价难度, 营养度指数法(AHP-PCA)能有效的减少评价指标, 同时也能筛选出对评价目标最有效、最合理的评价指标。因此将该方法作为经典方法加以比较, 不仅可以利用主成分分析法为模型的运算筛选合适的因子, 也可印证 GA-BP 模型的有效性。

传统的 BP 算法在水质分析中已得到一定的运用^[6]。但该算法在分析过程中存在容易陷入局部极小、收敛速度慢和引起振荡效应的局限性, 因此

收稿日期: 2010-08-28

基金项目: 上海市科技兴农重点攻关资助项目(沪农科攻字(2007)第 1-4 号)

* 责任作者, 教授, youwenhui1964@126.com

在水体富营养化等级评价上和实际情况有一定的偏差^[7].遗传算法(GA)是借鉴生物界自然选择和基因遗传学原理的优化搜索算法,具有良好的全局搜索性能,减少了陷入局部寻优的风险,鲁棒性强等特点,该方法目前已应用在水质模型的建立研究中^[8-9].GA-BP 模型将上述两点集合能较好的反映水体营养等级的实际情况,实用性较强.

本研究应用 GA-BP 模型对上海市城市景观水体进行水体富营养化分析的同时也选择营养度指数法(AHP-PCA)作为经典方法与其比较,旨在为景观水体富营养化治理和跟踪监测提供相关理论依据.

1 材料与方法

1.1 材料来源

本文根据上海市城市景观水体分布区域、水面大小等基本特征,在市内选取 10 个典型景观水体作为研究对象,总体上代表了上海城市城市景观水体的主要类型^[10].选取的研究对象为中山公园、漕溪公园、世纪公园、人民公园、大牌坊绿地、曹杨公园、上海植物园、杨浦公园、滨江森林公园陆家嘴中心绿地(图 1).测定项目包括:水温、pH 值、化学需氧量(COD_{Mn})、透明度(SD)、总氮(TN)、总磷(TP)、溶解氧(DO)、叶绿素(Chla),按国家标准方法进行测定^[11].

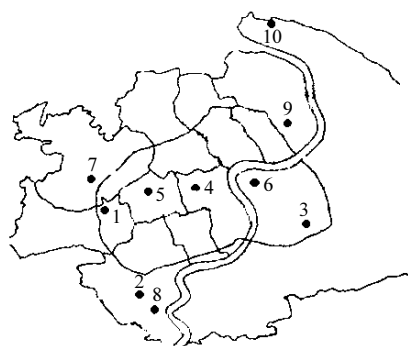


图 1 上海市公园样点分布示意

Fig.1 The garden sampling sites in Shanghai

1-大牌坊绿地,2-曹杨公园,3-漕溪公园,4-人民公园,5-陆家嘴中心绿地,6-中山公园,7-杨浦公园,8-上海植物园,9-滨江森林公园,10-世纪公园

为了研究上述城市景观水体营养化程度变

化情况,在 2008 年 7 月至 2009 年 3 月时期内按季度进行采样,共采样 4 次.

1.2 主成分分析法筛选变量

用 SPSS18.0 对所有采集的指标(变量)进行主成分分析,找到对总体贡献最小的主分量的特征值所对应的特征向量中权数最大的分量,删除该分量所对应的变量.重复操作后(直至最后一个特征值不是很小为止)剩下的变量不仅基本上保留了原来多个变量的信息,而且减少了评价因子,有助于模型构建的简单化^[11-13].经过筛选,本文模型运用的水质指标为叶绿素(chla)、总氮(TN)、总磷(TP)、高锰酸盐指数(COD_{Mn})以及透明度(SD).

1.3 GA-BP 神经网络原理

1.3.1 遗传算法基本原理 遗传算法(GA)基本原理是从“初始种群”开始,通过交叉和变异等遗传操作产生新的染色体,通过一定的“适应度值”来选择染色体,得到新的种群;此步骤反复循环若干代后,算法最终收敛于适应度值最佳的染色体,得到本次运算的最优解^[14].运用遗传算法可以全局寻优和进行网络权重优化的特点对 BP 算法的连接权值进行全局寻优.

1.3.2 BP 网络基本原理 BP 网络由输入层、隐含层和输出层三部分组成.BP 算法的学习过程由正向传播和反向传播两部分组成.在正向传播过程中,输入模式从输入层经过隐层神经元的处理,传向输出层.根据最陡坡降法原理,若测试输出与期望输出误差过大,则转入反向传播,此时误差信号从输出层向输入层传播并逐步调整各层连接权值,以使误差不断减小,直到达到精度要求^[15].

2 实验方法及数据处理

2.1 富营养化指标测试

本文水体富营养化评价指标以金相灿、屠清瑛的《湖泊富营养化调查规范》^[16]中富营养化状态指标与水质参数的关系一表中的数值为依据.

2.2 GA-BP 神经网络运行流程

GA-BP 神经网络模型运行流程以张德丰等人^[17]研究结果为参考,该神经网络运行流程

见图 2.

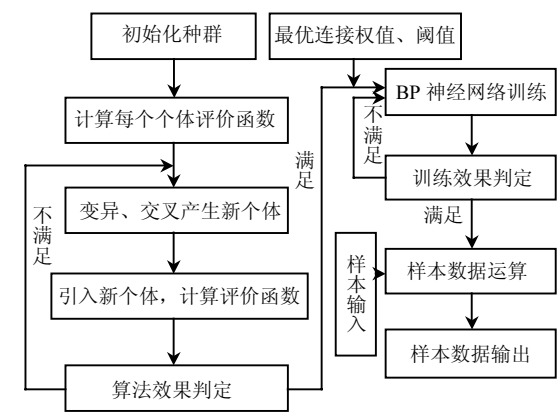


图 2 GA-BP 神经网络模型运行流程

Fig.2 The routine of GA-BP neural network model

2.3 学习样本生成

GA-BP 神经网络的学习样本分为训练样本和测试样本.本研究根据《湖泊富营养化调查规范》^[16]中富营养化状态指标与水质参数的关系表来确定研究指标的等级划分,并作为模型学习样本的数据来源.由于公园小水体水深一般较浅,如按原表透明度等级进行评价,将无法准确反映水体实际的营养化程度,因此在透明度各个等级数值的确定上根据采样数值进行了修正(表 1).

学习样本采用 MALAB 软件的 Linspace 函数均匀线性内插生成,每一级富营养化等级中每一项水质指标均生成 100 组学习样本,其中随机抽取 80 组为训练样本,其余 20 组作为测试样本.

表 1 景观小水体富营养化评价参数标准

Table 1 Landscape water eutrophication evaluation parameter criteria

等级	叶绿素(mg/m ³)	TN(mg/L)	TP(mg/L)	COD(mg/L)	SD(cm)
贫	≤1.6	≤0.079	≤0.0046	≤0.48	≥90
贫-中	≤4.1	≤0.160	≤0.0100	≤0.96	≥80
中	≤10.0	≤0.310	≤0.0230	≤1.80	≥60
中-轻富	≤26.0	≤0.650	≤0.0500	≤3.60	≥50
轻富	≤64.0	≤1.20	≤0.1100	≤7.10	≥40
中富	≤160.0	≤2.30	≤0.2500	≤14.0	≥30
重富	≤400.0	≤4.60	≤0.5550	≤27.0	≥20
超富	>400.0	>4.60	>0.5550	>27.0	<20

2.4 GA-BP 模型参数设定

本文中 GA-BP 模型输入神经元共有 5 个 (TN、TP、DO、NH₄-N、COD_{Mn}),隐含层神经元数量根据经验进行调试,本文最终确定隐含层神经元个数为 25.遗传算法的种群规模为 100,交叉概率为 0.8,变异概率设定为 0.1.BP 网络的学习速率为 0.05,最大训练次数为 1000 次,训练精度为 10⁻⁶.

2.5 数据处理

对 GA-BP 神经网络模型进行训练测试后,将上海市城市景观水体水质指标数据作为实验样本引入模型进行数据运算.实验样本共 40 组,数据来源是 2008 年 7 月至 2009 年 3 月对上海市 10 个具典型性的城市景观水体 4 个季度水质监测数据,中将每个水体每次水质监测作为独立评价样点,以期更好的反映富营养化变化趋势.为了模型更好的运行,运算前需对各项数据进行归一化处理.所有模型数据处理、分析基于 MATLAB 软件(Version 2010a)运行完成.

3 结果及分析

3.1 上海市 10 个城市景观水体水质分析

上海市 10 个城市景观水体水质状况见图 3.图 3 显示,TN 浓度最高值出现在 2009 年冬季的大牌坊绿地,最低值出现在 2008 年秋季的世纪公园.研究水体 TN 浓度都有一定波动,大牌坊绿地波动相对较大而漕溪公园和世纪公园 TN 浓度波动相对较小.研究水体 TP 总体上波动不是很大,大牌坊绿地 TP 数值也呈现较大的波动.漕溪公园、杨浦公园和世纪公园 TP 数值较为稳定.夏季的叶绿素含量总体上高于其余 3 个季节.大牌坊绿地和中山公园夏季以及滨江森林公园的叶绿素含量都超过了 100mg/m³.4 个季度的透明度数值波动较小.春季是研究水体透明度总体上最高的时期,同时春季是研究水体叶绿素 a 含量最低的时期,显示叶绿素 a 同水体透明度有一定的负相关性.研究水体 2009 年 3 月份的 COD_{Mn} 数值波动不大,除漕溪公园和陆家嘴中心绿地外其余 8 个城市景观水体水质总体变化不大.陆家嘴中心绿地 2008 年 11 月 COD_{Mn} 数值最高,徐家

汇公园、杨浦公园和世纪公园的 COD_{Mn} 数据较 为接近。

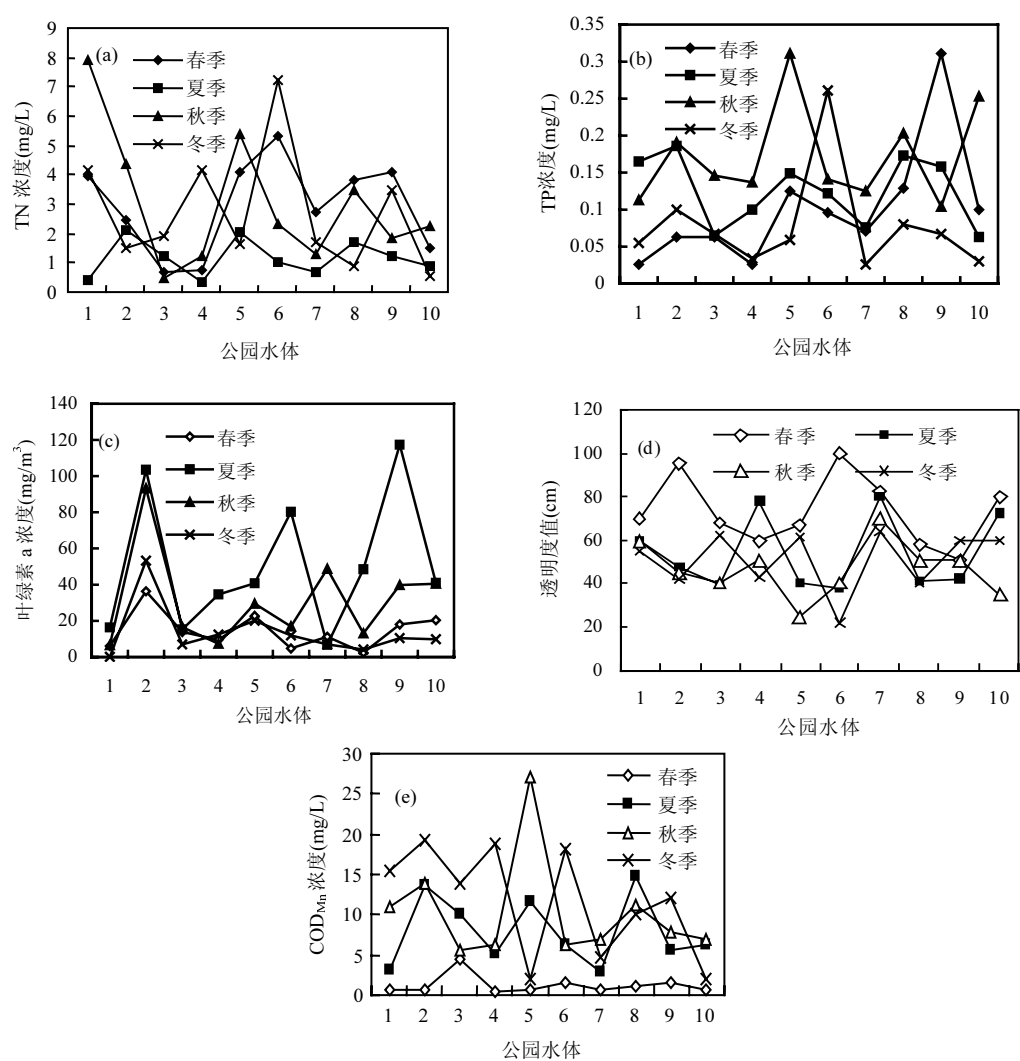


图3 上海市 10 个景观水体 TN、TP、叶绿素 a、透明度及 COD_{Mn} 指标数据

Fig.3 Data of TN ,TP, Chla , SD and COD_{Mn} of 10 landscape waters in Shanghai

1-大牌坊绿地,2-曹杨公园,3-漕溪公园,4-人民公园,5-陆家嘴中心绿地,6-中山公园,7-杨浦公园,8-上海植物园,
9-滨江森林公园,10-世纪公园

3.2 GA-BP 神经网络模型水质评价结果

3.2.1 GA-BP 神经网络模型训练误差 GA-BP 神经网络模型的训练误差见图 4.由图 4 可知,训练后的 GA-BP 神经网络误差除第 92 组样本误差较大外,其余各组样本误差范围总体在 $4\times10^6\sim4\times10^{-6}$ 之间.说明 GA-BP 神经网络引入训练样本训练后,预测精度较高,可用性较好.

3.2.2 GA-BP 神经网络模型水质评价结果 运用训练好的 GA-BP 神经网络对上海市 10 个城市景观水体水质进行评价,结果见图 5.纵坐标代表本文采用的 8 个等级富营养化指标数值.为了更好的反映该方法的可用性,本文选取营养度指数法作为比较方法,运用该方法对 10 个城市景观水体水质进行评价,结果见表 2.

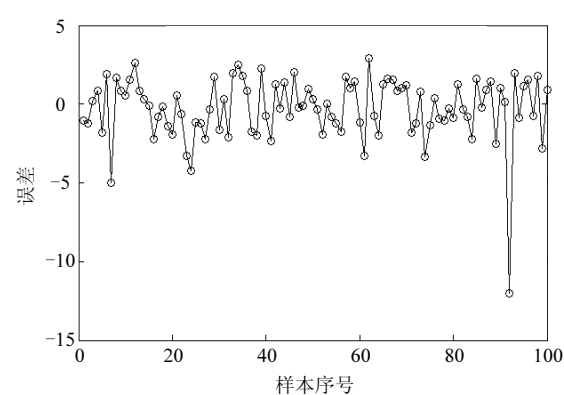


图 4 GA-BP 神经网络模型的训练误差结果

Fig.4 The training errors of the GA-BP neural network model

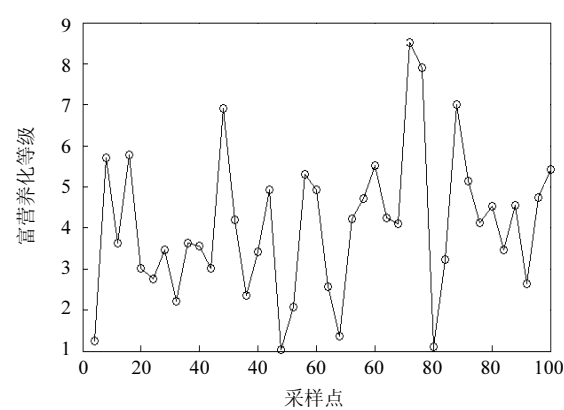


图 5 GA-BP 神经网络对上海市 10 个景观水体水质评价结果

Fig.5 The evaluation results of water quality of 10 greenbelts in Shanghai based on GA-BP neural network

1—贫,2—贫中,3—中,4—中轻富,5—轻富,6—中富,7—重富,8—超富;图中每个数据点数值代表该水体该次富营养化评价的具体营养化程度

图 5 数据显示,10 个景观水体 4 次水体富营养化数据集中在 2~5 范围内,说明 10 个水体富营养化在贫-中等级和轻富等级之间,说明上述 10 个水体的水体富营养化程度在较理想范围内.个别公园在研究周期内富营养化程度较高,出现了营养状态的波动,也反映城市景观小水体富营养化程度易受周遭环境影响的特点.

表 2 数据显示,10 个城市景观水体研究期间水体营养状态较为稳定,多为中营养和轻富营养,重富营养出现次数仅为两次,显示良好的水体营养状态.GA-BP 模型运算所得结果与此结果类似,

两种方法对 40 个样点的分析结果均显示研究目标水体的营养化程度在中营养化和轻富营养状态之间.

表 2 营养度指数法对上海市 10 个景观水体富营养化评价结果

Table 2 The result of water quality of 10 greenbelts in Shanghai based on AHC-PCA

采样点编号	春季	夏季	秋季	冬季
1	中	轻富	中	中
2	轻富	重富	中富	中富
3	中	中	轻富	轻富
4	中	轻富	轻富	中
5	中	中	中富	轻富
6	中	中富	轻富	轻富
7	中	贫	中	轻富
8	贫	中富	中富	中
9	中	轻富	重富	轻富
10	中	轻富	轻富	轻富

注:1—大牌坊绿地,2—曹杨公园,3—漕溪公园,4—人民公园,5—陆家嘴中心绿地,6—中山公园,7—杨浦公园,8—上海植物园,9—滨江森林公园,10—世纪公园

将 GA-BP 模型计算得出的数值与营养状态指数法的评价结果比较表明,该模型对于上述 10 个公园水体的富营养化程度评价结果与营养状态指数法总体相近,说明该模型能有效的对公园景观水体进行富营养化等级分析.

4 讨论

本文对于景观水体富营养化等级的划分是以相崎宇弘和郁根森的 2 个标准为依据,并结合《湖泊富营养化调查规范》的湖泊富营养化等级划分标准以位于上海市域的 10 个城市景观小水体 4 个季度份的水质调查数据为基础确定的,目的是为了为了更好的建立能准确分析景观小水体富营养化等级的模型.

景观小水体由于水体面积小,水体营养化程度较易受到外界干扰,波动较大,因此对各个小水体进行分时评价,能更好地反映小水体本身的营养状态,并提供科学合理的控制依据.

目前常见的水体富营养化评价方法多以大型湖泊作为研究对象,在建立评价指标和分析标

准也多以大型湖泊作为参照系,造成很多方法在景观小水体富营养化评价方面使用难度加大.由于城市景观小水体自身水质波动加大,造成富营养化爆发速度较大型湖泊更快,因此对水体富营养化分级的细化可更好把握水体营养化状态现状,对今后水体富营养化治理的应急响应机制建立上提供依据.

与营养度指数法相比,本文所建立的模型的等级更为细化,对于景观小水体的富营养化等级有更准确的评价.模型对样本进行评价之前,运用主成分分析法对各项评价因子进行筛选,使模型能在不影响评价结果的前提下更快速的对富营养化等级进行评价.通过大量的训练样本和测试样本的训练计算,使之在对未知水样进行分析时能更有效的评价富营养化等级.

本文所选方法结果与营养度指数法统一度较好,所出现的偏差也在合理范围内,本方法的优势在于通过微机对数据进行分析,排除了人为运算产生的误差,同时仅需将分析指标的数值输入已训练好的模型就可得出结果,为营养度法相比更快速,减少了富营养化发生时的响应时间.

本文训练样本和测试样本的数据是以 8 个富营养化等级标准为依据,运用 MATLAB 软件的 Linspace 函数通过线性均匀内插生成,解决了该模型训练样本和测试样本过少的问题.运用该模型对水体富营养化进行评价具有直观、快速、结果稳定的优点,因此本研究对于景观小水体富营养化评价具有一定的指导意义.

5 结论

5.1 GA-BP 模型可以运用到城市景观水体富营养化程度的评价中.该模型通过训练样本和测试样本校正后输入相应分析因子即可进行水体富营养化程度分析.

5.2 主成分分析可有效减少模型运算因子.在模型运算时减少了次要指标对模型计算结果的影响,缩短了模型的运行时间.

5.3 本模型在景观水体富营养化评价中以城市小水体作为研究对象,加入其他类型的景观水体分析数据可进一步提高模型的通用性.

参考文献:

- [1] 汪松年.上海市景观水体水质调研及分析 [J]. 中国水利, 2004, 11:40-42.
- [2] 李 畅,秦华鹏, Soon-Thiam Khu,等.中水回用对社区景观水体叶绿素 a 变化的影响 [J]. 中国环境科学, 2010,30(10):1338-1343.
- [3] 王明翠,刘雪芹,张建辉.湖泊富营养化评价方法及分级标准 [J]. 中国环境监测, 2002,18(5):47-49.
- [4] 朱庆峰,廖秀丽,陈新庚,等.用灰色聚类法对荔湾湖水水质富营养化程度的评价 [J]. 中国环境监测, 2004,20(2):47-50.
- [5] 许秋瑾,郑丙辉,朱延中,等.三峡水库支流营养状态评价方法 [J]. 中国环境科学, 2010,30(4):453-457.
- [6] 刘国东,黄川友,丁 晶.水质综合评价的人工神经网络模型 [J]. 中国环境科学, 1998,18(6):514-517.
- [7] 原 清.遗传算法和神经网络在导弹测试设备故障诊断中的应用研究 [J]. 测试技术学报, 2006,16(z1):702-706.
- [8] 李 炜,邝 鹏.基于遗传优化的模糊神经网络在管道泄露检测中的应用研究 [J]. 科学技术与工程, 2008,8(13):3490-3494, 3499.
- [9] 罗固源,郑剑锋,许晓毅,等.基于遗传算法的次级河流回水段水质模型多参数识别 [J]. 中国环境科学, 2009,29(9):962-966.
- [10] 薄芳芳,杨 虹,左 倬,等.上海公园水体夏季浮游植物群落与环境因子的关系 [J]. 生态学杂志, 2009,28(7):1259-1265.
- [11] 金相灿.中国湖泊富营养化 [M]. 北京:中国环境科学出版社,1990.
- [12] Martin J L, Hanke A R, LeGresley L M. Long term phytoplankton monitoring, including harmful algal blooms, in the Bay of Fundy, eastern Canada [J]. Journal of Sea Research, 2009, 61(12):76-83.
- [13] 于秀林,任雪松.多元统计分析 [M]. 北京:中国统计出版社, 1999.
- [14] 张丽丽.基于遗传算法的区域水资源系统多目标优化方法研究 [D]. 辽宁工程技术大学, 2005.
- [15] 魏本征,田质广.基于遗传优化 BP 网络的汽轮发电机故障诊断 [J]. 控制工程, 2007,14(S0):175-177,195.
- [16] 金相灿,屠清瑛.湖泊富营养化调查规范 [M]. 北京:中国环境科学出版社, 1987.
- [17] 张德丰. MATLAB 神经网络应用设计 [M]. 北京:机械工业出版社, 2009.

致谢: 本文在数据分析过程中得到了郑晓宇学姐的帮助和指导,在此表示衷心感谢.

作者简介: 张维砚(1986-),男,福建厦门人,硕士研究生,主要研究方向为水域生态学.