



全国中文核心期刊  
中国科技核心期刊

# 环境工程学报

Chinese Journal of Environmental Engineering



第8卷 第10期

Vol.8 No.10

中国科学院  
生态环境研究中心 主办  
科学出版社 出版

10  
2014

## 目 次

### 水 污 染 防 治

镁铝水滑石及焙烧产物去除水中的天然有机物 ··· 张仙娥 肖寒 肖峰 王东升 方利平 岳娜娜 王继库 (4053)	
PVA-I 修饰介孔 TiO <sub>2</sub> 光催化剂对盐酸四环素的光降解 ··········· 赵雪辉 蒋彩云 王玉萍 (4060)	
SPG 膜微气泡曝气生物膜反应器运行性能——空气通量的影响 ········· 张明 刘春 张磊 张静 杨景亮 年永嘉 (4067)	
在线混凝-超滤中膜污染与絮体水动力学的关系 ··········· 宋婷婷 赵长伟 王涛 <b>栾兆坤</b> (4074)	
酸根离子对铝板电絮凝处理含镍废水的影响 ··········· 张峰振 杨波 张鸿 赵绪新 刘剑洪 (4081)	
核桃果皮炭粉对 3 种重金属离子吸附的影响 ··········· 刘晓红 吕俊芳 左伟伟 杨飞超 康妮娜 (4086)	
臭氧对饮用水中氯化消毒副产物生成的影响 ··········· 张锁娜 王海波 李肖肖 朱莹 胡春 (4091)	
混凝好氧颗粒污泥的絮凝剂选择 ··········· 易诚 湛含辉 程胜高 (4097)	
pH、温度及初始铈浓度对硫酸盐还原菌脱铈的影响 ··········· 陈炜婷 张鸿郭 陈永亨 罗定贵 吴启航 (4105)	
Ti/SnO <sub>2</sub> -Sb <sub>2</sub> O <sub>3</sub> /β-PbO <sub>2</sub> 阳极消毒处理医院污水 ··········· 周键 王三反 薛志强 宋小三 (4110)	
(B-NZVI)/CTMAB 的制备及对水中铬(VI)的去除性能与机理 ········· 彭秀达 刘红 周亚端 秦雄 冯涛 (4115)	
氨基功能化碳材料的水热制备及其对活性艳红 K-2BP 的吸附性能 ········· 谭腾飞 闫碌碌 廉优芬 漆新华 (4122)	
潜流人工湿地去除毒死蜱、双酚 A、4-壬基酚 ··········· 高士博 林培伦 郭宗楼 王维 王振午 (4129)	
活性炭纤维对水中典型除草剂的吸附行为 ··········· 倪若晨 陈思嘉 冯景伟 (4135)	
三峡库区城市污水处理厂工艺特征分析 ··········· 付国楷 喻晓琴 张春玲 张智 陆磊 (4141)	
电增强载铝活性炭纤维吸附氟离子性能 ··········· 胡欣琪 宋永会 张旭 刘瑞霞 曾萍 贾建丽 (4147)	
曝气扰动对泥水界面硝化-反硝化性能的影响 ··········· 沈叔云 何岩 黄民生 顾佩弟 饶应福 姚丽平 张一璠 (4153)	
硫酸盐还原对三氯乙烯生物降解的影响 ··········· 郭莹 崔康平 (4159)	
水平潜流人工湿地在不同水力负荷下的水力效率及可视化分析 ········· 牛瑞华 王宇晖 宋新山 (4163)	
温度对正渗透工艺性能的影响 ··········· 李洁 王军 白羽 侯得印 <b>栾兆坤</b> (4168)	
昆明市污水处理厂运行综合评价 ··········· 李波 吴光学 胡洪营 吴毅晖 郭昉 郭玉梅 (4175)	
正渗透复合膜的制备及表征 ··········· 宁静恒 赵俊 李玉平 曹宏斌 李海波 (4183)	
微电场-零价铁强化厌氧水解酸化性能 ··········· 王婧馨 崔康平 (4191)	
基于 WASP 模型计算尾水回用河道污染负荷 ··········· 冯斌 李大鹏 周琦 马焱 李铮 (4196)	
浊度与微生物指标在污水再生处理系统中的去除特性及相关性 ········· 庞宇辰 席劲瑛 张彤 胡洪营 黄晶晶 宗祖胜 (4203)	
黑藻根际对沉积物中氨氧化细菌和古菌的影响 ··········· 李倩 田翠翠 肖邦定 (4209)	
可吸附生物反应墙修复地下水中 BTEX ··········· 孙本山 崔康平 洪天求 (4215)	
络合萃取技术分离制药废水中的金刚烷胺 ··········· 宋永会 魏健 马印臣 曾萍 (4221)	
湘江株洲段镉污染动态模拟与情景分析 ··········· 张青梅 刘湛 刘翔宇 向仁军 郑一 (4227)	
纳米零价铁处理冶炼废水 ··········· 刘一源 雷轰 王伟 滑熠龙 (4233)	
无机离子对石墨烯/TiO <sub>2</sub> 纳米管阵列光电极光催化性能影响 ········· 张清哲 辛言君 马东 包南 公丕成 (4239)	
电混凝去除水中铈污染物 ··········· 张家兴 王超 杨波 张宝锋 赵旭 (4244)	
透水坝原位净化山溪性污染河流 ··········· 董慧峪 王为东 强志民 (4249)	
零价铁与甲醇支持的生物-化学法去除富氧水中的硝酸盐氮 ··········· 黄国鑫 刘菲 黄园英 王晓宏 邹静 时满 (4254)	
固定化硝化细菌去除生活污水中的氨氮 ··········· 刘少敏 储磊 朱敬林 (4261)	
K <sup>+</sup> 、Ca <sup>2+</sup> 、Mg <sup>2+</sup> 对高盐肝素废水处理的影响 ··········· 杨红薇 陈佼 张建强 (4267)	
强化循环厌氧反应器处理印染废水的中试启动研究 ··········· 杨波 毕深涛 李方 田晴 (4273)	
高压脉冲放电等离子体对水中土霉素的降解 ··········· 张艳 孙亚兵 何东 边琳 陆素洁 赵泽华 荣少鹏 (4280)	
倒置 A/O 工艺参数变化对污水中医药类污染物去除的影响 ··········· 冯莉 卜龙利 白丽荣 高宁 赵佩 王文东 (4285)	
Fe/Cu/C 微电解深度处理钻井废水 ··········· 杨德敏 袁建梅 (4291)	
硅酸镁/PES 复合膜吸附材料的制备及其对亚甲基蓝的吸附 ··········· 汪清 毛丽莉 王海增 (4296)	
改性膨润土联合絮凝剂处理微污染水中有机物 ··········· 吴光锋 聂锦旭 谢伟楠 陈仕稳 范恒 (4301)	
CF-TiO <sub>2</sub> 的制备及其在模拟太阳光下的光催化性能 ··········· 邓琴 母康生 刘燕 陈忠毅 张延宗 (4307)	
人工湿地组合工艺对太湖三山岛农村生活污水的脱氮效果 ··········· 王学华 张娜 苏祥 沈耀良 陈骏 (4313)	
补水景观湖水生态系统构建对水质的影响及因子分析 ··········· 钟铮 张明珠 孟庆强 余帆洋 (4319)	
制药废水 UASB-HAR-ICEAS 组合处理系统效能建模及优化研究 ········· 姜涛 赵延华 王鸿程 (4326)	
废弃棉布制备活性炭纤维处理六价铬微污染水 ··········· 郑杰莹 赵泉林 郭云红 叶正芳 (4333)	
基于支持向量机的地下水质量综合评价 ··········· 司训练 张旭峰 (4340)	
木镁基碳质吸附剂的制备及其性能 ··········· 李志萍 刘千钧 冯卿明 孙斌 (4345)	

## 大气污染防治

水泥预分解炉内流场及温度场的数值模拟 .....	蔡鑫	马永亮(4349)
镁基-海水法船舶废气脱硫实船实验 .....	唐晓佳 李铁 刘全 陈晨 李叶叶	朱益民(4355)
商业钛白粉的脱硝性能 .....	王登辉 惠世恩 张庚	尚桐(4361)
西安夏季 PM <sub>2.5</sub> 中碳组分与水溶性无机离子特征分析 .....	代志光 张承中 李勇 周变红 马文静 韩婧 李文韬	(4366)

## 固体废物处置

纤维素酶、半纤维素酶降解膨化玉米秸秆工艺优化 .....	邵丽杰 寇巍 曹焱鑫 武崇辉 刘齐	张大雷(4373)
羽毛蛋白海绵材料对亚甲基蓝的吸附去除 .....	刘畅 高品 曹张军 薛昱	周美华(4379)
餐厨垃圾两相厌氧发酵产甲烷相恶臭排放规律 .....	黄丽丽 张妍 商细彬 韩萌 李伟芳 宁晓宇 王元刚 卢志强	(4386)
以芦苇灰为硅源制备三胺改性 SBA-15 及其对溶液中 Cd <sup>2+</sup> 的吸附 .....	秦睿 王力 张增强 杨亚提 李荣华	孟昭福(4393)
嗜酸硫杆菌和黑曲霉对电镀污泥重金属浸出效果 .....	毕文龙 崔雨琪 方迪 崔春红	周立祥(4402)
利用钢渣制备多孔吸声材料的研究 .....	李鹏 郭占成 孙朋	郭茂盛(4409)
内源过氧化氢(H <sub>2</sub> O <sub>2</sub> )对 <i>Stenotrophomonas maltophilia</i> DHHJ 降解羽毛 .....	蒋晶君 吴斌 曹张军	周美华(4415)
微波法制备工业污泥含碳吸附剂及其除铬性能 .....	陈建发 陈艺敏 黄慧珍	刘福权(4419)
应用混料实验设计制备生猪养殖发酵床复合菌剂 .....	钟仁方	吴祖芳(4427)
污泥-秸秆活性炭制备过程热化学分析 .....	曹群 李炳堂	舒威(4433)
聚丙烯酸钠作为调理剂对好氧堆肥的保氮影响 .....	杨文卿 陈潇川 郭梓波 卓倩 肖荔人	陈庆华(4439)
蔗渣发酵生产 L-乳酸 .....	李市场 刘雅 刘胜男	王延凯(4445)
利用废液晶屏玻璃基板制备发泡材料工艺优化 .....	李龙珠 唐惠东	孙媛媛(4451)
水泥窑共处置过程中水泥生料对砷的吸附/冷凝特性 .....	丛璟 闫大海 李丽 蒋旭光 周英男 吕任生 王琪	(4456)
钢渣掺量对陶瓷地砖性能的影响 .....	王维 王可祯 董翰琼 王标	赵云超(4463)

## 土壤污染防治

Tween 系列表面活性剂对白腐菌降解氯丹的影响 .....	肖鹏飞 孙丽军 李玉文	(4468)
生物炭对邻苯二甲酸二甲酯在土壤中自然降解和吸附行为的影响 .....	周震峰 张海光 王茜	(4474)
外源营养物质对土壤中六氯苯厌氧降解效能的影响 .....	王琪 刘辉 姜林 姚珏君	(4480)
环境友好型淋洗剂对重金属污染土壤的修复效果 .....	吴烈善 吕宏虹 苏翠翠 曾东梅 莫小荣	(4486)

## 环境生物技术

利用 MFC 阳极室暗发酵培养沼泽红假单胞菌 .....	王兴祖 程翔 孙德智 任以伟 徐贵华	(4492)
全细胞脂肪酶(菌)对 SBR 系统中微生物群落演替的影响 .....	冷欢 周莉 刘家书 李谥 赵勇 江正兵 宋慧婷	(4497)
一株光合细菌的分离鉴定及其产电特性 .....	吴义诚 肖勇	赵峰(4503)
废水中硝氮和 COD 浓度对 AD-MFC 脱氮产电性能的影响 .....	张吉强 郑平 何薪飞 季军远 张萌 陈慧 谢作甫	(4508)
双室污泥微生物燃料电池产电的影响因素 .....	连静 孙永军 李鹤超 刘俊新 肖本益	(4515)
MFC-MBR 耦合系统运行效果及产电性能 .....	李俐频 张军 左薇 李慧	(4521)
厌氧流化床微生物燃料电池及其串并联性能 .....	宫本月 刘新民 郭庆杰	(4527)
华北某油田石油降解菌的筛选及降解特性 .....	张帆 郝春博 王丽华 盛益之 王鹤立	(4533)
碳毡和碳布 2 种生物阴极材料微生物燃料电池产电性能 .....	刘松山 陈亮 黄满红 郑宇	(4540)

## 环境监测

基于主成分和粒子群优化支持向量机的水质评价模型 .....	王成杰 张森	(4545)
基于小波变换的 ARIMA 模型在水质预测中的应用研究 .....	石子泊 邹志红	(4550)

## 工程实例

畜禽养殖污染防治典型工程的综合效益评价 .....	张昌爱 徐振 李彦 张玉凤 林海涛 王江涛	(4555)
---------------------------	-----------------------	--------

# 基于主成分和粒子群优化支持向量机的水质评价模型

王成杰 张 森

(河海大学理学院,南京 210098)

**摘 要** 水质的评价是治理水污染必不可少的工作。为了准确、快速地对水质进行评价,利用主成分分析法从水质监测常见的多个物化指标提取出主成分,然后将主成分作为支持向量分类机的输入,利用历史数据进行水质评价训练并用粒子群算法优化参数,构造出水质评价模型,将从物化指标中得出的主成分代入此模型即可得到水质类别。最后,选取水质监测点实测数据进行试验,结果表明,模型的水质评价结果准确且稳定。

**关键词** 主成分分析 粒子群优化算法 支持向量机 水质评价

中图分类号 X824 文献标识码 A 文章编号 1673-9108(2014)10-4545-05

## Water quality evaluation mode based on principal component analysis and support vector machine optimized by PSO

Wang Chengjie Zhang Sen

(College of Science, Hohai University, Nanjing 210098, China)

**Abstract** Water quality evaluation is essential to control the water pollution. We extract the main components from several physical and chemical indicators in water quality monitoring by the use of principal component analysis, and then put them into the support vector machine to evaluate the water quality with the help of historical statistics, and optimize the parameters by using the particle swarm algorithm, as a result of which the mode is built. By plugging the principal components from the indicators to the built mode, the categories of the water quality will be obtained. At last, it's proved that the results of the water quality are correct and stability by conducting the experiment on the real samples from the monitoring station.

**Key words** principal component analysis; particle swarm algorithm; support vector machine; water quality evaluation

水资源事关人类生产生活的方方面面,然而随着我国工业化、城镇化发展的不断推进,水质污染状况不断加剧。水质评价是防治水污染问题的重要工作,许多专家学者利用不同的方法和理论对此进行了研究和探讨。水质综合指数法和不确定性理论被先后用于水质评价工作。考虑到水环境系统的复杂性和动态性,一些智能方法逐渐被引入水质评价方法的研究中,向娜<sup>[1]</sup>即用人工蜂群算法优化的神经网络进行水质评价。

支持向量机理论不仅适用于复杂问题的机器学习,还能够克服传统智能算法在过学习、小样本学习、非线性问题等多方面的缺点。因此,支持向量机理论适合于水环境等类似系统的处理。郑一华<sup>[2]</sup>和徐红敏<sup>[3]</sup>的研究都表明,支持向量机方法在水质评价中的合理性和可行性。在实际水质监测活动中

往往会监测水体的多个物化指标,从而使得一个样本具有较大维数,虽然支持向量机能够处理高维样本,但是过高的维数会使得训练时间过长。主成分分析是多元统计分析中一种能够较好地保留样本全部信息的降维技术,也常用于水质评价<sup>[4]</sup>在内的综合评价方法中。此外,支持向量机的参数选择是影响其性能的重要因素,由于缺乏足够的理论指导参数选择,常通过大量的实验来搜索最优参数。粒子群算法具有实现容易、精度高、收敛快等优点,在工程计算上得到了广泛的应用,邵信光等<sup>[5]</sup>提出粒子

基金项目:国家级创新训练计划项目(201205XCX096)

收稿日期:2013-10-06; 修订日期:2013-11-02

作者简介:王成杰(1993—),男,本科生,主要研究方向为水质参数测试。E-mail:washingtoni@yeah.net

群算法用于参数选择能够得到较优的解。

本研究运用主成分分析从水质监测的多个物化指标中提取出主成分,将历史数据作为训练样本代入支持向量机中,选用径向基函数作为核函数并用粒子群算法优化参数,得到了水质评价模型。最后,以实测水质数据作为实例对其水质进行了评价,得到了较高的准确率,表明本文模型具有一定实用价值。

## 1 相关理论基础

### 1.1 主成分分析

主成分分析通过计算协方差或相关系数矩阵来尽量保留原始变量信息并使得新变量各自含有的信息不重叠。运用相关系数进行主成分分析的一般计算过程如下<sup>[6]</sup>:

计算相关系数矩阵

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (1)$$

式中: $r_{ij}(i, j = 1, 2, \dots, p)$ 为原变量 $x_i$ 与 $x_j$ 的相关系数, $r_{ij} = r_{ji}$ ,计算公式为:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (2)$$

解特征方程 $|\lambda I - R| = 0$ (常用雅可比法),求出特征值,并使其按大小顺序排列 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ,分别求出对应于特征值的特征向量 $e_i(i = 1, 2, \dots, p)$ ,要求 $e_i = 1$ ,其中 $\sum_{j=1}^p e_{ij}^2 = 1$ 表示向量 $e_{ij}$ 的第 $j$ 个分量。

式子 $\lambda_i / \sum_{k=1}^p \lambda_k(i = 1, 2, \dots, p)$ 的值称为第 $i$ 主成分的贡献率,式子 $\sum_{k=1}^i \lambda_k / \sum_{k=1}^p \lambda_k(i = 1, 2, \dots, p)$ 称为第 $i$ 主成分的累计贡献率一般取累计贡献率达85%~95%的特征值所对应的前 $m(m \leq p)$ 个主成分作为新指标。

### 1.2 支持向量机

支持向量机算法是从计算线性可分情况下的最优分类面提出的。高维空间中线性判别函数的一般形式为 $g(x) = \omega x + b$ ,分类面方程是 $\omega x + b = 0$ ,将判别函数进行归一化,使得两类样本都满足

$|g(x)| \geq 1$ 。要求分类面对所有样本都正确分类,即要求满足下式:

$$y_i(\omega x_i + b) \geq 1 \quad i = 1, 2, \dots, n \quad (3)$$

式中: $x_i$ 为输入向量, $y_i$ 为分类标签,使等号成立的那些样本就叫做支持向量(support vectors)。此时分类间隔为 $2/\omega$ ,寻找最优分类面即求解以下规划问题:

$$\min \frac{\omega^2}{2} \quad s. t. \quad y_i(\omega x_i + b) \geq 1 \quad i = 1, 2, \dots, n \quad (4)$$

当样本不能被线性分类器分开,为构造最优超平面,引入非负松弛变量集合 $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ ,这样分类超平面的最优化问题变为

$$\min \frac{\omega^2}{2} + C \sum_{i=1}^n \xi_i \quad s. t. \quad y_i(\omega x_i + b) \geq 1 - \xi_i \quad \xi_i \geq 0, i = 1, 2, \dots, n \quad (5)$$

其中, $C$ 是一个常数,称为惩罚因子,表征了对错误的惩罚程度。这样得到的最优解决定的最优超平面称为广义最优超平面。

该优化问题可通过求解Lagrange函数的鞍点得到,本问题的Lagrange函数为:

$$L(\omega, b, a) = \frac{1}{2}(\omega \cdot \omega) - \sum_{i=1}^n \alpha_i \{ [(x \cdot \omega) + b] y_i - 1 \} \quad (6)$$

式中: $\alpha_i$ 为Lagrange乘子。

在鞍点上,解必须满足:

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad \frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0 \quad (7)$$

又由Kuhn-Tucker条件可知,最优超平面的充要条件是分类超平面满足:

$$\alpha_i \{ [(x \cdot \omega) + b] y_i - 1 \} = 0 \quad (8)$$

将(7)式代入(6)式中,利用对偶原理,Lagrange函数可转化为求解如下泛函优化问题:

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad s. t. \quad \begin{cases} 0 \leq \alpha_i \leq C & i = 1, 2, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (9)$$

对于非线性问题,通过非线性变换 $\varphi(x)$ 转化为某个高维空间中的线性问题,在高维空间求最优分

类面,即采用核函数  $K(x_i, x_j)$ ,使得  $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ ,把优化问题中的点积运算都用核函数运算代替,得到一般的支持向量机算法<sup>[7]</sup>:

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s. t.} \quad & \begin{cases} 0 \leq \alpha_i \leq C & i = 1, 2, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \end{aligned} \quad (10)$$

具体算法步骤如下:

Step1 设已知训练集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in (X \times Y)^n$ ,其中  $x_i \in X = R^n, y_i \in Y = \{+1, -1\}, i = 1, 2, \dots, n$

Step2 构造并求解最优化问题式(10),得到最优解  $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)^T$

Step3 选择  $\alpha^*$  的一个分类  $0 \leq \alpha_i^* \leq C$ ,并据此计算  $b^* = y_j - \sum_{i=1}^n y_i \alpha_i^* (x_i \cdot x_j)$ 。

Step4 由此计算求得决策函数如下

$$f(x) = \text{sign} \left[ \sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x) + b^* \right] \quad (11)$$

### 1.3 粒子群算法

粒子群优化算法 (particle swarm optimization, PSO) 源于对鸟群捕食行为的研究。在算法中,一只鸟为一个“粒子”,解群相当于一个鸟群,从一地到另一地的迁徙相当于解群的进化,“好消息”相当于解群每代进化中的最优解,食物源相当于全局最优解。粒子群算法的优势在于简单容易实现,同时有着深刻的智能背景,即适合科学研究,也适合工程应用。

基本粒子群优化算法的流程<sup>[8]</sup>如下:

Step1 初始化粒子群中每个粒子的位置和速度,记粒子群中有  $N$  个粒子的初始位置为  $X(1)$  初始速度为  $V(1), t = 1$ ;

Step2 对  $X(t)$  中的第  $i (i = 1, 2, \dots, N)$  个粒子  $pop(t)_i$  计算适应值  $f_i = f(x), x = X(t)_i$ ,若适应值为该粒子历史最大值,记  $pbest_i = x$ 。若适应值为粒子群 (或第  $i$  个粒子的邻近粒子) 历史最大值,记  $gbest_i = x$ ;

Step3 若满足终止条件,则算法停止;否则,依次更新粒子群所有粒子的速度和位置:

$$\begin{aligned} V(t+1)_i = & V(t) + c_1 r_1 (pbest - X(t)_i) \\ & + c_2 r_2 (gbest - X(t)_i) \end{aligned} \quad (12)$$

$$X(t+1)_i = X(t)_i + V(t+1)_i \quad (13)$$

式中  $r_1, r_2$  是  $[0, 1]$  上的随机数,  $c_1, c_2$  是加速系数 (可以调整)。由此得到下一代粒子,令  $t = t + 1$ ,返回 Step2。

将粒子群优化算法用于支持向量机的参数优化时,将支持向量机和核函数的参数分别作为粒子的位置坐标。适应度函数为一次支持向量机训练的正确分类率,即计算粒子的适应度为对样本进行一次训练,得到的训练集分类正确率即为适应度值。给定初始解及算法中相关参数,即构造出用于支持向量机参数优化的粒子群算法。

## 2 水质评价模型的建立

### 2.1 提取主成分

把监测数据实测的多个物化指标作为原始变量。考虑到数据标准化过程会消除样本包含的信息从而对后面支持向量机的训练产生影响,本文不对原始数据进行标准化。同时考虑到不同变量间量纲和取值范围的差异,利用其相关系数矩阵计算主成分。最后按照贡献率大于 85% 的原则选取出主成分。

### 2.2 构造支持向量分类机

将提取出的主成分作为输入变量,以国家地表水环境质量标准<sup>[9]</sup> 判别出的水质类别作为分类标签,构成样本集。

核函数的选择对支持向量机的性能至关重要。研究表明,在没有关于问题的先验知识时,选择径向基核函数比较好。这种核函数训练而成的模型具有比基于其它核函数的模型更好的总体性能。因此,选用径向基函数作为模型中支持向量机的核函数,其表达式为:

$$K(x, x') = \exp\left(-\frac{x - x'}{\sigma^2}\right) \quad (14)$$

另一影响支持向量机的性能的重要因素是支持向量机的参数。将上述构造的样本集分为训练集和测试集后,将训练集代入支持向量机中训练并用粒子群算法优化参数,得到最大分类正确率下的参数。本文选取全局粒子群算法,设置算法的参数  $c_1 = 1.5, c_2 = 1.7$ ,最大迭代次数为 200,种群数量为 20。

结合获得的支持向量机参数和上文的径向基核函数最终得到支持向量分类机水质评价模型。将测试集输入支持向量分类机模型中,根据输出结果可以检验支持向量机的性能。同样地,将相应的物化指标提取出的主成分输入支持向量分类机模型中,

即可得到与国家地表水环境质量标准相应的水质的评价结果。

### 3 水质评价实例与结果分析

获得香港环境署公布的水质监测数据进行水质评价,选取3个不同水域监测点数据,每个监测点共144组数据。选取国家地表水环境质量标准中的水质测量指标 DO、COD、BOD<sub>5</sub>、NH<sub>3</sub>-N、TP、氟化物、砷及铜、锌等5种重金属元素共12种水质指标的数据。

采用交叉验证的方式评价支持向量机的性能。其原理是将训练样本分成容量相同的 $k$ 个子集,并对模型训练 $k$ 次。在第 $i$ ( $i=1,2,\dots,k$ )次,要用除了第 $i$ 个子集的所有子集训练模型,再用得到的模型对第 $i$ 个子集计算误差。以 $k$ 次误差的平均数值作为模型分类性能的近似数值。此类方法一般能够得到稳定且较为全面的结果。本文选取一个样本作为一个子集。

在交叉验证中,由于需要对模型进行多次训练,其最优模型参数可能会有不同,为了统一模型参数以更客观评判模型性能,参数优化时将全部样本进行参数优化的结果作为本模型的参数。

分别对3个监测站应用本模型,同时应用网格搜索法(grid search,GS)和遗传算法(genetic algorithm,GA)优化参数,利用matlab软件编程计算,得到表1结果。表1中结果为分类正确率,括号内为正确例数与总数比。

由以上结果可以得出,原始数据共有12个变量,按照累计贡献率达85%的标准选取出6或7个主成分,降维效果比较明显;采用粒子群算法相比采

用网格搜索法和遗传算法优化参数,支持向量机能得到更好的正确率,验证了粒子群算法用于支持向量机参数优化是优于网格搜索法和遗传算法的,另外,在计算中得到相同结果所需的运行时间粒子群算法比遗传算法少很多,这与理论上粒子群算法的收敛速度较快是吻合的,也表明粒子群算法更适用于实际应用;无论是将原始数据,还是提取的主成分作为输入变量,支持向量机的水质评价模型均有着很高的准确率,再次验证了支持向量机理论在水质评价中的有效性,同时也表明主成分分析方法的引入是有效的,即在降低输入样本维度的情况下并未对模型的准确性产生较大影响;对于不同的水域,本模型均得到类似的结果,除了表明模型的有效性之外,也说明模型具有较好的稳定性和适应性,可以用于不同水质状况的水质评价。

### 4 结论

本研究将主成分分析法、粒子群优化算法与支持向量机理论结合起来构建了水质评价模型。用实测水质数据进行了交叉检验,3个实例均得到准确率较高的水质评价结果,表明了主成分分析法和粒子群优化算法引入的有效性和模型的适应性。分析可知,由于粒子群算法优化参数具有较强的搜索能力且收敛速度较快,主成分分析法则能够消除冗余信息,减少计算量,所以模型能够得到稳定的结果。因此,基于主成分和粒子群优化支持向量机建立的水质评价模型应用于不同水域能够得到较准确、稳定的水质评价结果的同时,也减少了计算工作,在实际应用中具有一定的价值。

表1 不同模型的水质评价结果

Table 1 Water quality evaluation results based on different models

河溪	原始数据作为输入变量			提取的主成分作为输入变量			主成分个数
	GS	GA	PSO	GS	GA	PSO	
1	68% (98/144)	75% (109/144)	82% (118/144)	62% (89/144)	69% (100/144)	83% (120/144)	7
2	73% (105/144)	78% (112/144)	81% (117/144)	76% (110/144)	79% (110/144)	79% (114/144)	6
3	67% (97/144)	72% (103/144)	83% (120/144)	67% (97/144)	77% (97/144)	80% (115/144)	7

## 参 考 文 献

- [1] 向娜. 基于神经网络和人工蜂群算法的水质评价和预测研究. 广州:华南理工大学硕士学位论文,2012  
Xiang Na. Exploration for water quality assessment and prediction based on neural networks and artificial bee colony algorithm. Guangzhou; Master Dissertation of South China University of Technology,2012(in Chinese)
- [2] 郑一华. 基于支持向量机的水质评价和预测研究. 南京: 河海大学硕士学位论文,2006  
Zheng Yihua. Study of water quality assessment and parameter prediction based on Support Vector Machine. Nanjing; Master Dissertation of Hohai University,2006(in Chinese)
- [3] 徐红敏. 基于支持向量机理论的水环境质量预测与评价方法研究. 长春:吉林大学博士学位论文,2007  
Xu Hongmin. Study on the prediction and assessment methods of water environment quality based on support vector machines theory. Changchun; Doctor Dissertation of Jilin University,2007(in Chinese)
- [4] 万金保,何华燕,曾海燕,等. 主成分分析法在鄱阳湖水  
质评价中的应用. 南昌大学学报(工科版),2010,32(2):113-117  
Wan Jinbao, He Huayan, Zeng Haiyan, et al. Application of principle component analysis in evaluating water quality of Poyang Lake. Journal of Nanchang University (Engineering & Technology),2010,32(2):113-117(in Chinese)
- [5] 邵信光,杨慧中,陈刚. 基于粒子群优化算法的支持向量机参数选择及应用. 控制理论与应用,2006,23(5):740-743,748  
Shao Xinguang, Yang Huizhong, Chen Gang. Parameters selection and application of support vector machines based on particle swarm optimization algorithm. Control Theory & Applications,2006,23(5):740-743,748(in Chinese)
- [6] 余锦华,扬维权. 多元统计分析与应用. 广州:中山大学出版社,2005
- [7] 张燕平,张铃. 机器学习理论与算法. 北京:科学出版社,2012
- [8] 雷秀娟. 群智能优化算法及其应用. 北京:科学出版社,2012
- [9] GB3838-2002,地表水环境质量标准