

群智能优化算法优化支持向量机的方法及应用

崔丽洁 程焕新 刘军亮 张远绪

(青岛科技大学 自动化与电子工程学院 青岛 266042)

摘要: 水质评价过程具有多变量、非线性、不确定等特点,传统的粒子群算法训练神经网络的水质评价模型收敛速度慢、泛化性能差。为了克服传统模型的缺点,提出了利用动态多种群粒子群算法训练支持向量机的模型,并利用多种群粒子群算法优化支持向量机结构参数。该模型结合了粒子群算法的搜索性能以及支持向量机的高效性、强鲁棒性等优点,提高了模型的泛化能力。通过对新疆某流域站点的水文数据进行仿真,结果得出该方法的相对误差为2.74%,远低于传统粒子群算法4.21%的相对误差,由此证明该模型的应用效率及精度得到提高,适用于日常水质评价工作。

关键词: 水质评价;动态多种群粒子群算法;支持向量机

中图分类号: TN98 **文献标识码:** A **国家标准学科分类代码:** 510.8010

Method and application of swarm intelligence optimization algorithm for support vector machine optimization

Cui Lijie Cheng Huanxin Liu Junliang Zhang Yuanxu

(College of Automation and Electrical Engineering, Qingdao University of Science and Technology, Qingdao 266042, China)

Abstract: The process of water quality assessment is multivariable, nonlinear and uncertain. The traditional particle swarm optimization training neural network water quality evaluation model has slow convergence speed and poor generalization performance. In order to overcome the shortcomings, this paper proposes a new model to use dynamic multigroup particle swarm optimization algorithm to train support vector machine, in which the DMPSO is used to optimize the parameters of the fuzzy neural network model. The model combines the search performance of PSO algorithm, the efficiency and robustness of SVM, which can improve the generalization ability of the model. Through the simulation experiment on the hydrological data, the results show that the relative error of this model is 2.74%, which is much lower than the 4.21% relative error of the traditional particle swarm optimization. It is proved that the efficiency and accuracy of the model are improved, and is suitable for the daily water quality evaluation.

Keywords: water quality evaluation; dynamic multigroup particle swarm optimization algorithm; support vector machine

0 引言

当今环境污染问题是人类面临的一大严重问题,尽管环保治理力度不断加强,环境污染问题仍然是制约和困扰我国可持续发展的一大障碍。水质评价是水污染治理环节中关键的一步,通过定量分析水质状况的影响因子以对环境的质量进行综合的评判。近年来,水质评价的预测理论和方法成了水污染治理的一项重点,各种新的理论和方法不断涌现,如灰色聚类分析法^[1]、人工神经网络法^[2]、层次分析法^[3]等。灰色聚类分析法能直观、全面地反映实际结果,计算简便,但对于较复杂的数据集计算比较困难。人

工神经网络应用最为广泛,但人工神经网络过分强调克服学习错误而自适应性与泛化性能不强,评价效果也不太满意。层次分析法对于多目标、多准则的评价系统应用效果较好,分析结果清晰、明确、简单,但使用层次分析法缺乏创造性,评价结果定性成分较多,不易令人信服。

近年来,机器学习与人工智能成为当下研究热点,基于统计学习理论的支持向量机(support vector machine, SVM)方法也开始广泛用于智能预测中。支持向量机作为一种机器学习方法,不仅能从已知事实的分析中学习总结规律,还可以利用学习到的经验对未来的现象做出正确的预测与判断^[4]。支持向量机算法参数的优化直接影响算法的有

效性,为了使支持向量机回归算法在水质评价中的应用效率更高,本文采用动态多种群粒子群优化算法(dynamic multigroup particle swarm optimization algorithm, DMPSO)来优化 SVM 算法的参数,以提高 SVM 算法预测精度和鲁棒性。

1 动态多种群粒子群优化算法

粒子群算法(partical swarm optimization, PSO)是一种用来处理非线性优化问题的智能算法,也是一种进化算法^[5]。粒子模拟鸟类捕食行为,通过追寻当前时刻的最优粒子来不断更新自己的速度及位移以在解空间中寻找最优解。其速度和位移更新公式如下:

$$V_{id}^{k+1} = \omega V_{id}^k + c_1 r_1 (P_{id}^k - X_{id}^k) + c_2 r_2 (P_{gd}^k - X_{id}^k) \quad (1)$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \quad (2)$$

式中: V_{id}^k 表示第 i 个粒子在第 k 次迭代飞行速度矢量的第 d 维分量的速度; X_{id}^k 表示第 i 个粒子在第 k 次迭代飞行速度矢量的第 d 维分量的位置; k 为当前迭代次数; $1 \leq i \leq M, 1 \leq d \leq D$; ω 为惯性权重; c_1, c_2 为学习因子; r_1, r_2 为 $(0, 1)$ 之间均匀分布的随机数。

为了提高算法的搜索能力,有学者在标准 PSO 算法的基础上对惯性权重 ω 进行动态调整,即

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \frac{t}{T_{\max}} \quad (3)$$

式中: ω_{\max} 为初始惯性权重; ω_{\min} 为进化至最大代数时的惯性权重; t, T_{\max} 分别为当前迭代次数和最大迭代次数。

标准 PSO 算法在实际应用中,由于粒子的趋向同一化性,使得粒子在前期的收敛速度非常快,极易陷入局部极值,出现早熟收敛现象。因此,本文提出动态多种群粒子群算法,该算法能有效避免局部最优现象,提高全局搜索能力。

1.1 动态多种群算法种群划分策略

动态多种群粒子群优化算法引入了物种的概念,按照各粒子适应度值的大小将粒子群划分为不同物种。各物种在各自的解空间中分别对最优值进行搜索。种群划分步骤如下^[6]:

1) 计算所有粒子的适应度值。设粒子群 \mathbf{X} 共有 m 个粒子,即 $\mathbf{X} = [x_1, x_2, \dots, x_m]$ 。用 $f(x)$ 表示粒子的适应度值,并对其做降序排序。假设 $f(x_1) \geq f(x_2) \geq \dots \geq f(x_m)$, 则:

$$F(x) = [f(x_1), f(x_2), \dots, f(x_m)] \quad (4)$$

2) 确定物种半径 r_s 。物种半径的选择决定了物种的多少,合理地选择物种半径是物种划分成功的关键。当物种半径的取值比较大时会导致划分的物种比较少,一些重要的个体可能会因此被忽视甚至淘汰;当物种半径较小时则划分的物种较多,从而增大计算量,使算法过于复杂^[7]。本文对物种半径的选择进行自适应调整,在保证物种多样性的前提下,在种群进化前期选取较大的物种半径,对整个粒

子群进行粗分,提高算法的全局搜索能力;在进化过程中随着粒子的更新逐渐减小物种半径,从而对粒子群进行细分,提高算法在极值点附近的局部搜索能力。物种半径调整公式为:

$$r_s(t) = [1 - e^{-D(t)}] r_s(0) \quad (5)$$

$$D(t) = \frac{1}{k} \sum_{i=1}^k D_i(t) \quad (6)$$

$$D_i(t) = \frac{1}{m_i} \sum_{j=1}^m \sqrt{\sum_{d=1}^n (x_{jd} - x_d)^2} \quad (7)$$

式中: $D(t)$ 为种群的多样性指数; $D_i(t)$ 为第 i 个物种的多样性指数; $r_s(0)$ 为物种半径的初始值; k 为物种个数; x_{jd} 为第 j 个粒子位移的第 d 个分量; x_d 为物种内各粒子位移的第 d 个分量的平均值。

3) 物种的划分。选取粒子群中适应度值最大的 n 个粒子作为物种的核 s_i^* , $i = 1, 2, 3, \dots, n$, 放入集合 \mathbf{S} 中,形成物种集合。以物种的核 s_i^* 为圆心,物种半径 r_s 为半径形成圆,将圆内所有粒子放入集合 \mathbf{S}_i 中,形成以 s_i^* 为中心的物种 \mathbf{S}_i , 即: $s_i^* \in \mathbf{S}_i$, 若 $\forall y \in \mathbf{S}_i, \exists f(s_i^*) \geq f(y)$, 则 s_i^* 是物种 \mathbf{S}_i 的核。若集合 \mathbf{X} 中存在某个粒子不在任意物种的核所形成的圆中,则将该粒子作为新的物种的核放入集合 \mathbf{S} 中;若集合 \mathbf{S} 中存在两物种的核之间的距离小于物种半径时,保留适应值更优的核所在的物种,淘汰适应值较差的核所在的物种。当集合 \mathbf{X} 为空集时物种生成完毕。

4) 物种规模调整。经步骤 1)~3) 形成的物种其规模大小可能会相差很大,从而影响算法的性能。因此本文将物种规模限定在一定的范围内 ($m_{\min} \leq m \leq m_{\max}$)。当物种规模超过 m_{\max} 时,需要对物种内个体进行淘汰,淘汰规则如下^[8]: 分别计算该物种内所有粒子个体的当前适应度值与个体最优适应值之和,和小的个体淘汰;当物种规模低于 m_{\min} 时,则在该物种内随机生成一定数量的粒子,使其规模达到 m_{\min} 。

1.2 DMPSO 算法流程

物种生成完毕后,各物种根据式(1)~(3)进行进化,每个核都是各自物种内的最优粒子,物种内其他粒子通过追寻核在各自的区域内进行搜索,从而收敛到每个区域内的最优点附近。动态粒子群优化算法的流程如下:

Step 1: 初始化粒子种群;

Step 2: 计算粒子适应度;

Step 3: 选取物种的核;

Step 4: 根据种群的多样性指数计算物种半径;

Step 5: 种群划分生成物种,并对物种规模进行调整;

Step 6: 根据式(1)、(2)更新粒子;

Step 7: 物种半径的自适应调整;

Step 8: 判断是否满足终止条件,若是,停止更新,种群进化完成;否则重复 Step 2~Step 8。

1.3 DMPSO 算法性能测试

为了进一步验证 DMPSO 算法的性能,本文采用

Ackley 标准函数对其寻优过程进行仿真:

$$y = -c_1 \exp\left(-0.2 \sqrt{\frac{1}{n} \sum_{j=1}^n x_j^2}\right) - \exp\left(\frac{1}{n} \sum_{j=1}^n \cos(2\pi x_j)\right) + c_1 + e \quad (8)$$

利用该函数测试 DMP SO 算法寻优过程,测试结果如图 1 所示。

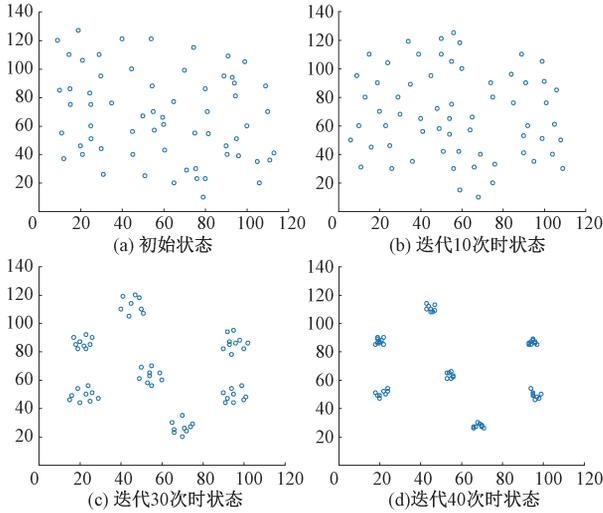


图 1 DMP SO 寻优过程

由图 1 可知,随着粒子的进化,各种群逐渐收敛于不同的最优值附近,说明该方法是有有效的。

2 基于 DMP SO 的支持向量机回归估计

2.1 支持向量机回归估计

支持向量机作为一种基于统计学习理论的算法,其原理是寻找一个最优分类线,在保证将样本准确无误分开的前提下,使得分类间隔最大化^[9]。在处理非线性回归问题时,将该问题通过某种非线性映射,由低维空间映射到高维空间,根据对偶原理,该问题便可转换为高维空间中的线性回归问题。因此在处理非线性、高维数问题过程中,支持向量机有较好的应用效果^[10]。

假设训练样本集为 $\{(x_i, y_i), i = 1, 2, \dots, l\}$, x_i, y_i 分别为输入输出值, l 为样本数。为了确保对偶变量的稀疏性,本文采用 ϵ 不敏感损失函数:

$$|y - f(x)|_{\epsilon} = \begin{cases} 0, & |y - f(x)| \leq \epsilon \\ y - f(x) - \epsilon, & |y - f(x)| > \epsilon \end{cases} \quad (9)$$

式中: $f(x)$ 为回归函数; y 为目标值。当两者之间距离小于 ϵ 时,损失为 0。

假设样本集为非线性,为了在高维空间中进行线性回归,采用映射函数为 $\phi(x)$,则回归估计函数为:

$$f(x) = w \cdot \phi(x) + b \quad (10)$$

式中: w 为权值向量; b 为偏置。最优化问题为:

$$\begin{aligned} \min_{w, b, \xi} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (11) \\ \text{s.t.} & \begin{cases} y_i - w \cdot \phi(x_i) - b \leq \epsilon + \xi_i \\ w \cdot \phi(x_i) + b - y_i \leq \epsilon + \xi_i^*, i = 1, 2, \dots, l \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \end{aligned}$$

式中: $C > 0$ 为惩罚系数, C 越大,表示对超出 ϵ 的惩罚越大。 ξ_i 与 ξ_i^* 为松弛因子。

得到对偶最优化问题:

$$\begin{aligned} \max_{\alpha, \alpha^*} & \left\{ L_D = -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \right\} \quad (12) \\ \text{s.t.} & \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases} \end{aligned}$$

式中: $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ 为核函数。为了提高算法的抗干扰能力,本文选径向基核函数 $K(x_i, x) = \exp(-\|x - x_i\|^2 / \sigma^2)$, x_i 为核函数中心, σ 为函数宽度。

最终得到回归估计函数为:

$$f(x) = \sum_{x_i \in SV} (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (13)$$

$$\text{其中, } b = \frac{1}{N_{NSV}} \left\{ \sum_{0 < \alpha_i < C} [y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_j, x_i) - \epsilon] + \sum_{0 < \alpha_i^* < C} [y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_j, x_i) + \epsilon] \right\}$$

2.2 基于 DMP SO 的支持向量机参数优化

支持向量机参数的选择对算法的影响很大,合理地选择参数不仅能提高算法精度、减小误差,还能有效改善局部收敛问题。SVM 中待优化的参数有不敏感性参数 ϵ 、惩罚系数 C 以及核函数中的宽度参数 σ 。为了进一步提高算法学习能力,本文将前文提出的动态粒子群算法用于支持向量机的参数优化,形成 DMP SO-SVM 算法。

DMP SO-SVM 算法流程:

- 1) 种群初始化,将 SVM 的学习参数 C, ϵ, σ 设置为粒子种群的初始状态。
 - 2) 计算适应度,将实际值与目标值之间的均方误差设置为粒子的适应度:
- $$MES = \frac{1}{M} \sum_{i=1}^M (y_i - Y_i)^2 \quad (14)$$
- 3) 利用 DMP SO 算法对种群进化。
 - 4) 判断是否满足最大迭代次数或者达到最小误差。若满足则终止寻优,否则,返回步骤 2)。
 - 5) 粒子寻优结束,确定 SVM 算法的参数。
 - 6) 利用训练样本对 SVM 回归算法进行训练,得到回归函数式。

3 DMP SO-SVM 算法在水质评价中的应用

3.1 水质评价方法

水质评价即通过检测当前水体的水质状况、流域内和

水体中含有的污染物的迁移特性来预测水质发生变化的趋势^[11]。常用的水质评价的方法有指数评价法、模糊评价法和基于人工神经网络的水质评价法^[12]。在水质评价中,根据影响水质状况的因素选取以下6项指标作为评价参数:氨氮、溶解氧、化学需氧量、高锰酸盐指数、总磷、总氮^[13]。各项指标值与水质等级对应关系如表1所示。

表1 地表水环境质量指标

分类	I类	II类	III类	IV类	V类
氨氮	0.15	0.5	1.0	1.5	2.0
溶解氧	7.5	6.0	5.0	3.0	2.0
化学需氧量	15	15	20	30	40
高锰酸盐指数	2.0	4.0	6.0	10	15
总磷	0.02	0.1	0.2	0.3	0.4
总氮	0.2	0.5	1.0	1.0	2.0

3.2 水质评价模型构建

将DMP-PSO-SVM算法应用到新疆某河流的水质评价中建立回归预测模型。首先,将水质评价的6项指标作为模型的输入变量,模型的输出为水质评价结果。

模型的输出结果规定在(1,5)之间,分别对应水质评价的I~V级,水质等级评价对应关系如表2所示。

表2 水质等级评价对应关系

输出值	[0,1.5]	[1.5,2.5]	[2.5,3.5]	[3.5,4.5]	[4.5,5]
评价等级	I类	II类	III类	IV类	V类

3.3 仿真结果分析

本文选取新疆某河流2012—2016年水文数据进行分析,从中抽取200组数据作为样本,其中150组数据作为训练样本,其他50组数据作为测试样本。设置粒子群规模 $N=40$,最大迭代次数 $T=200$,训练精度MSE设置为 10^{-3} 。分别利用本文提的方法(DMP-PSO-SVM)和文献^[14]提到的PSO-TSFNN以及PSO-BP方法进行仿真训练,不同模型的测试结果如图2~4以及表3所示。

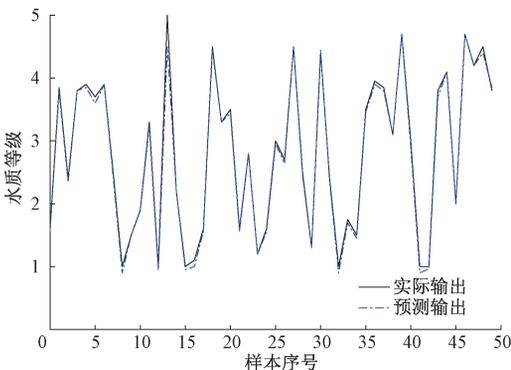


图2 DMP-PSO-SVM模型测试结果

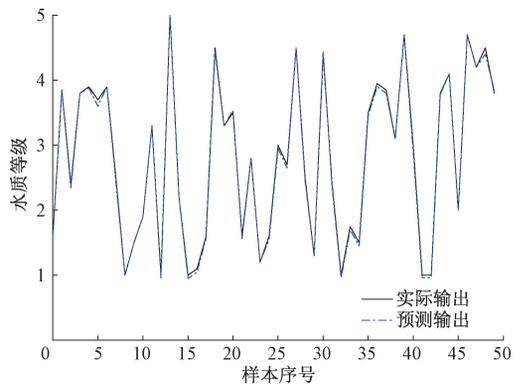


图3 PSO-TSFNN模型测试结果

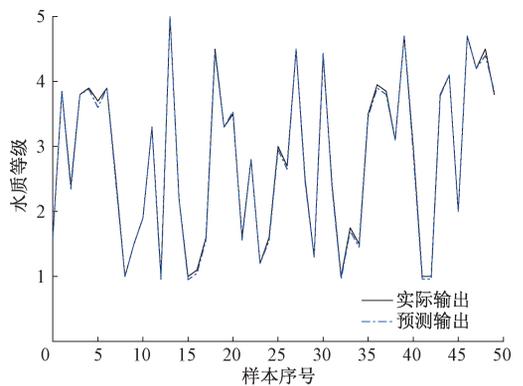


图4 PSO-BP模型测试结果

表3 测试结果

样本序号	DMP-PSO-SVM	PSO-TSFNN	PSO-BP	实际输出
5	3.651	3.613	3.821	3.749
10	1.879	1.878	1.789	1.880
15	0.989	0.982	0.957	1.015
20	3.562	3.511	3.502	3.558
25	2.989	2.903	2.887	3.030
30	4.542	4.672	4.652	4.532
35	3.723	3.684	3.711	3.695
40	2.874	2.883	2.910	2.881
45	2.213	2.187	2.114	2.225
最大相对误差/%	2.614	4.191	5.714	

由图2~4可以看出,DMP-PSO-SVM的最大相对误差为2.614%,PSO-TSFNN算法的最大相对误差为4.191%,PSO-BP的最大相对误差为5.714%,因此可以得出基于动态多种群粒子群算法支持向量机模型预测结果最好,精度最高。

4 结 论

水质评价是污水处理及水产养殖等过程中至关重要的一个环节,对水质状况作出准确有效的评价则对评价模型的性能有着更加严格的要求^[15],动态多种群粒子群算法克服了标准粒子群算法易陷入局部最优的问题,搜索性能更优,再结合支持向量机的学习推广能力,应用于水质评价等级的划分及质量预测中是十分有效的。但是在实际应用中由于环境的复杂性,受温度以及地域环境的影响,测量结果的准确性会受到一定程度影响,这个问题需要进一步解决。

参考文献

- [1] 贺北方,王效宇,贺晓菊,等.基于灰色聚类决策的水质评价方法[J].郑州大学学报(工学版),2002(1):10-13.
- [2] 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.
- [3] SAATY T L, BENNETT J P. A theory of analytical hierarchies applied to political candidacy[J]. Systems Research & Behavioral Science, 2010, 22(4):237-245.
- [4] 唐亮贵,程代杰.基于小波的支持向量机预测模型及应用[J].计算机科学,2006,33(3):202-204.
- [5] 满春涛.粒子群算法研究及其在过程控制系统稳态优化中的应用[D].哈尔滨:哈尔滨理工大学,2009.
- [6] XU S, ZHAO Y, WANG M, et al. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis-NIR spectroscopy[J]. Geoderma, 2018, 310:29-43.

- [7] 李丹.粒子群优化算法及其应用研究[D].沈阳:东北大学,2007.
- [8] 王宁,谢敏,邓佳梁,等.基于支持向量机回归组合模型的中长期降温负荷预测[J].电力系统保护与控制,2016,44(3):92-97.
- [9] 李茜,张建辉,林兰钰,等.水环境质量评价方法综述[J].现代农业科技,2011(19):285-287,290.
- [10] 隆重,朱则祥,胡冬梅,等.基于改进支持向量机的绝缘子污秽度在线监测[J].电子测量技术,2017,40(11):208-212.
- [11] LIU D, ZOU Z. Water quality evaluation based on improved fuzzy matter-element method[J]. Journal of Environmental Sciences, 2012, 24(7):1210-1216.
- [12] 梅恒荣,殷礼胜,刘冬梅,等.改进粒子群算法优化的SVM模拟电路故障诊断[J].电子测量与仪器学报,2017,31(8):1239-1246.
- [13] 彭越兮,徐蔚鸿,陈沅涛,等.改进量子粒子群算法的模糊神经网络水质评价[J].计算机工程与应用,2018,54(11):211-216.
- [14] 王坤,张立广,王琪.基于改进型粒子群算法的扩张状态观测器[J].国外电子测量技术,2017,36(8):66-70.
- [15] 薛青松.基于改进粒子群算法的模糊神经网络研究[D].重庆:重庆大学,2008.

作者简介

崔丽洁,硕士研究生,主要研究方向为人工智能与大数据。

E-mail:2533960770@qq.com