

基于改进神经网络的环境空气质量预测

蒲国林, 刘笃晋

(四川文理学院 智能制造学院, 四川 达州 635000)

摘要:为提高环境空气质量预测的精度,提出一种由改进人工蜂群算法和反向传播神经网络相结合的环境空气质量预测方法(KABC-BP)。对人工蜂群算法中雇佣蜂、跟随蜂的搜索空间提出一种随迭代次数递减的搜索公式,以随机初始化此改进人工蜂群算法的不同初始解作为不同组反向传播神经网络权值,以蜂群算法迭代代替人工神经网络的梯度下降修正迭代,以蜂群个体的对应权值下训练误差倒数作为适应度函数,该改进人工蜂群算法所求全局最优解就是所求反向神经网络最优权值。通过基于改进蜂群算法的反向传播神经网络算法、传统蜂群算法的反向传播神经网络算法(ABC-BP)及反向传播神经网络算法(BPNN)的环境空气质量预测的仿真实验表明,该算法的环境空气质量预测精度是最高的。

关键词:人工蜂群算法;迭代递减;反向传播神经网络;环境空气质量预测;误差函数;适应度函数

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2018)09-0181-04

doi:10.3969/j.issn.1673-629X.2018.09.037

Ambient Air Quality Prediction Based on Improved Neural Network

PU Guo-lin, LIU Du-jin

(School of Intelligent Manufacturing, Sichuan University of Arts and Science, Dazhou 635000, China)

Abstract: In order to improve the accuracy of ambient air quality prediction, we propose a method of ambient air quality prediction based on improved artificial bee colony algorithm and back propagation neural network (KABC-BP). In the artificial bee colony algorithm, we give a search formula with decreasing number of iterations for the search space of employed bees and onlookers. The different initial solutions of the improved artificial bee colony algorithm are randomly initialized as the weights of different groups of back propagation neural networks. The gradient iteration of artificial neural network is replaced by iterative algorithm of artificial bee colony algorithm. The reciprocal of training errors is used as fitness function under the corresponding weight of colony individuals. The global optimal solution of the improved artificial bee colony algorithm is the optimal weight of the back propagation neural network. The simulation of ambient air quality prediction on back propagation neural network algorithm based on improved bee colony algorithm, back propagation neural network algorithm based on the traditional bee colony algorithm (ABC-BP) and back propagation neural network algorithm (BPNN) shows that the method proposed is the highest in the prediction of ambient air quality.

Key words: artificial bee colony algorithm; iterative descending; back propagation neural network; ambient air quality prediction; error function; fitness function

0 引言

当前环境污染问题日益引起人们的关注,颗粒物、二氧化硫、氮氧化物等大气污染物对人体健康和生态环境都产生了很大影响^[1]。空气质量预测^[2-3]对于评价空气质量、管控环境污染、保持人类健康具有重要作用。传统的环境空气质量预测方法较多,如统计预测法、数值预测法等,由于影响环境空气质量的因素较为复杂^[4],大多具有非线性特性,这些传统方法在描述和

刻画非线性因素方面并不适合,因而这些方法对环境空气预测的准确度并不理想^[5]。而人工神经网络是现代神经科学发展的重要成果^[6],由于天然具有的自适应、自学习、自组织等优点,并具有良好的非线性表现能力^[7],许多研究者运用人工神经网络进行环境空气质量预测,取得了明显进步。其中最为典型的是运用反向传播神经网络进行环境空气质量预测^[8-10],但BP传播神经网络收敛时间较长,并且容易陷入局部最优,

收稿日期:2017-08-17

修回日期:2017-12-21

网络出版时间:2018-05-16

基金项目:国家自然科学基金(61152003);四川省教育重点科研项目(16ZA0353);四川省教育科研项目(17ZB0377,16ZB0360);四川文理学院2015年度特色培育一般项目(2015TP001Y)

作者简介:蒲国林(1971-),男,博士,教授,CCF会员(39278M),研究方向为智能计算、大数据。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20180515.1651.022.html>

因此许多研究者对此进行了改进^[11-12],并取得了积极进步。但由于各地的环境空气实际情况并不一样,除主要受各地环境空气质量和当地大气污染物排放的影响外,还受气候变化和气象因素变化的影响。研究表明^[13-14],在一个地区环境污染物排放物相对稳定的情况下,气象因素对当地环境空气质量的变化将产生重要影响。因而研究具有非线性特性的人工神经网络对环境空气质量预测具有重要意义,目前研究的重点主要在人工神经网络的网络结构调整、误差函数构造、权值和阈值学习算法的选取^[15]等方面。

针对人工神经网络易陷入局部最优、收敛速度慢等问题,文中应用群体智能算法中应用较为广泛的人工蜂群算法^[16-17]进行改进,将改进后的人工蜂群算法和反向传播神经网络相结合以预测环境空气质量,并通过实验进行验证。

1 误差反向传播神经网络学习原理

人工神经网络实际上是用大量的计算单元(神经元)构成的一个具有非线性处理功能的网络系统,并且具有一定的记忆、学习和计算等人脑神经系统具有的智能处理能力。人工神经网络的神经元是人工神经网络的基本处理单元,一般具有多输入/单输出的特点,其基本结构如图 1 所示。

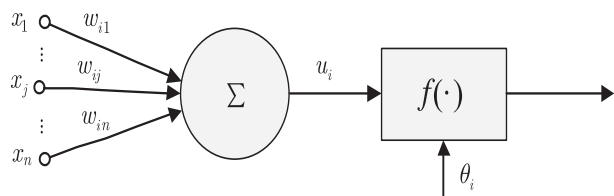


图 1 典型的神经元模型

图中, $x_j (j = 1, 2, \dots, n)$ 是神经元 i 的输入信号; w_{ij} 是神经网络权值; u_i 是输入信号计算后的输出; θ_i 为阈值(偏差),用 b_i 表示;神经元 i 的输出函数公式如下:

$$y_i = f\left(\sum_j w_{ij}x_j + b_i\right) \quad (1)$$

人工神经网络的学习就是一个根据一定方法或规则不断调整修改权值和阈值的过程,调整神经网络权值和阈值的学习方法较多,其中主要是有监督学习的神经网络应用较为广泛,主要原理是在神经网络训练过程中输入/输出是成对提出,根据神经网络实际输出和期望输出之间误差的方向和大小按一定规则调整。研究表明,含一个隐含层的 3 层 BP 神经网络,只要有足够多的神经元,就能以任意精度逼近有界区域上的任意连续函数。

误差反射传播神经网络是一种多层前馈神经网络,具有良好的非线性映射、泛化和容错能力,训练目

标函数采用的是实际输出和期望输出的平方和误差,其误差函数如下:

$$E_D = \frac{1}{2} \sum_{d \in D} (t_0 - D_0)^2 \quad (2)$$

其中, D 表示训练的数据集; t_0 和 D_0 分别表示 BPNN 模型训练中的目标输出和实际输出,但其采用的是基于梯度下降的 BP 算法,故存在收敛速度慢、容易陷入局部极小值等缺陷,这些缺陷严重影响了 BP 网络的应用。

2 改进蜂群算法优化神经网络的学习算法

2.1 基本人工蜂群算法

人工蜂群算法以随机初始化方式产生 HS 个蜜源的蜂群,第 i 个蜜源对应算法一个可行解为 EX_i ,每个蜜源对应可行解维度为 D ,令 $j \in \{1, 2, \dots, D\}$,第 i 个蜜源随机产生的位置公式如下:

$$EX_{i,j} = EX_{\min,j} + \beta(EX_{\max,j} - EX_{\min,j}) \quad (3)$$

其中, $EX_{i,j}$ 表示随机产生的第 i 个解第 j 维分量位置的解; $EX_{\min,j}$ 和 $EX_{\max,j}$ 分别表示第 i 个蜜源对应可行解第 j 维分量的下限和上限; $\beta \in (0, 1)$ 是一个随机数。

在人工蜂群算法的不断迭代搜索过程中,雇佣蜂、跟随蜂采用同一个搜索公式寻找蜜源对应可行解,最终目标是寻找到全局最优解。侦察蜂用式 3 随机搜索全局最优解,其中雇佣蜂、跟随蜂用式 4 搜索最优解。

$$EV_{i,j} = EX_{i,j} + SR_{i,j}(EX_{i,j} - EX_{i,k}) \quad (4)$$

其中, $EX_{i,j}$ 表示当前位置可行解; $SR_{i,j} \in [-1, 1]$ 是一个随机数,用来控制蜜蜂的搜索空间大小; $EX_{i,k}$ 表示不同于 $EX_{i,j}$ 的随机解, $k \in \{1, 2, \dots, HS\}$; $EV_{i,j}$ 表示在当前位置 $EX_{i,j}$ 后产生的下一个可行解位置。

2.2 人工蜂群算法的改进

研究发现,在人工蜂群算法雇佣蜂和跟随蜂的不断迭代寻优过程中,在搜索早期为了保证尽快得到全局最优解,需要较大的搜索空间,在算法搜索后期的全局最优解附近需要较小的搜索空间。但传统人工蜂群算法中,对于控制参数 $SR_{i,j}$ 是一个不加控制的随机产生,随机产生有利于多样性的寻优,但其搜索空间大小不加控制是不利于算法的全局最优解的寻求,因而文中对式 4 的搜索空间控制参数 $SR_{i,j}$ 进行改进,用一个递减函数 ε 来控制空间搜索参数 $SR_{i,j}$,其函数如下所示:

$$\varepsilon = (SR_{\max} - SR_{\min} - w_1) \times \frac{1}{e^{1+w_2 \cdot ct/t_{\max}}} \quad (5)$$

其中, SR_{\max} 和 SR_{\min} 分别取 0.9 和 0.4; $w_1 \in (0, 0.5)$, $w_2 \in [2\pi, 4\pi]$; ct 表示当前蜂群的迭代次数; t_{\max} 表示最大迭代次数。

根据式 5,改进后雇佣蜂及跟随蜂搜索公式如下:

$$EV_{i,j} = EX_{i,j} + \varepsilon SR_{i,j}(EX_{i,j} - EX_{i,k}) \quad (6)$$

式 5 是一个递减函数,随着迭代次数的增加,在随机搜索的基础上限制搜索空间逐渐缩小,加快了全局最优解的求解过程。

2.3 基于改进蜂群神经网络算法的环境空气质量预测

本项目用人工蜂群算法优化 BP 神经网络,主要是用人工蜂群算法的迭代来代替 BP 神经网络算法中的梯度修正。由于人工蜂群算法的搜索过程主要是蜂群在不同维度上的速度和位置的改变,而神经网络学习过程主要是权值和阈值的不断更新,因此蜂群中每只蜜蜂的位置向量对应 BP 神经网络迭代中的权值和阈值。蜂群中每只蜜蜂的维数由神经网络中起连接作用的权值的数量和阈值个数所决定,用神经网络输出误差作为适应度函数,神经网络的误差表示适应度值,误差越小表明蜜蜂在搜索空间中的性能越好。蜜蜂在其搜索空间内不断移动搜索使网络输出层的误差最小,改变蜜蜂的速度即更新网络的权值。每次迭代过程中产生误差最小的蜜蜂为目前全局最优位置的蜜蜂。当训练过程不断重复直到达到满足要求的误差或者超过预先设定的迭代次数时,迭代终止,算法结束,此时得到的权值集合即为最终结果。

要实现蜂群算法优化神经网络对环境空气质量的预测,蜂群的适应度函数的选择至关重要。文中所用的神经网络模型属于 MLP,训练算法是采用 BP 方法进行训练,故称为 BP 神经网络,并采用三层 MLP 结构,实现人工蜂群算法优化环境空气质量预测的具体算法如下:

- (1)随机初始化文中提出的改进人工蜂群算法(KABC),随机产生 HS 个蜜源,每个蜜源对应一组神经网络权值参数,并设定权值参数的取值范围。
- (2)定义每个蜂群个体的适应度函数为该个体对应的一组权值下误差值的倒数,并根据适应度大小对蜂群所有个体按从大到小排序。
- (3)选择适应度值较大的一组位置指派给雇佣蜂,适应度值居中的一组指派为跟随蜂,最小的一组为侦察蜂。
- (4)雇佣蜂在当前迭代位置附近根据式 4 搜索候选蜜源,若搜索到的蜜源适应度值高于当前蜜源适应度值,则抛弃当前蜜源,否则保持不变。
- (5)跟随蜂在雇佣蜂所求蜜源中根据轮盘赌策略选择所要跟随的蜜源。
- (6)若经多次迭代更新后,蜜源适应度值仍然没有更新,对应的雇佣蜂降为侦察蜂,按式 3 随机搜索蜜源。
- (7)将三种蜜蜂的搜索结果适应度值进行排序比

较,若不能满足终止条件,则转向步骤 3,若满足条件则适应度最大的蜜源即是全局最优解。

(8)所得的全局最优解对应的一组神经网络权值即为所求。

3 实验结果与分析

实验数据来自四川省达州市环境局官网发布的日空气质量数据(数据地址为:<http://www.tianqihoubao.com/aqi/dazhou.html>),时间段为 2016 年最近 6 个月数据和对应的 2015 年历史数据。实验中输入的是输入时段的空气质量浓度数据,包括 PM2.5、PM10、NO₂、SO₂、CO、O₃,还有历史天气数据,包括天气状况、气温、风力、风向。这些空气浓度数据和天气数据都经过相应的预处理归一化再进行神经网络训练。实验硬件平台 CPU 为 2.4 GHz,内存 4 GB,Win7 旗舰版操作系统, MATLAB2014b 实验环境,在预测时输入当前时段空气浓度数据和预测时段天气预报的天气数据,输出的是预测时段的空气质量数据。采用基于反向传播神经网络(BPNN)、基于人工蜂群算法的反向传播神经网络(ABC-BP)、基于改进蜂群算法的反向传播神经网络(BPKABC-BP)进行实验,结果如图 2 和图 3 所示。采用均方误差(MSE)、平均绝对百分比误差(MAPE)、平均浓度准确率(PJB)三个指标对三种算法性能进行定量比较,结果如表 1 和表 2 所示。

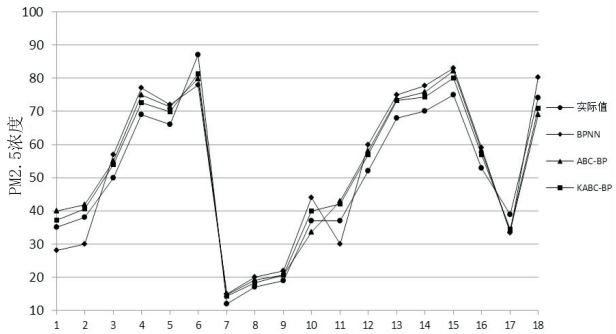


图 2 三种算法预测值与实际值对比
(2016 年 11 月达州 PM2.5)

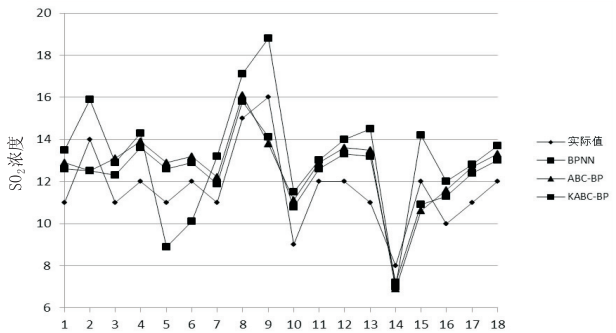


图 3 三种算法预测值与实际值对比
(2016 年 11 月达州 SO₂)

从图 2 和图 3 可以看出,在 BPNN、ABC-BP、

KABC-BP 三种算法中,KABC-BP 算法与实际数据最为接近,基于 ABC-BP 算法的预测数据和实际数据较为接近,而基于 BPNN 的预测数据与实际数据偏离较大,有些数据甚至差距非常明显。

表 1 在 PM2.5 下 MSE、MAPE、PJB 的比较

| 算法 | MSE | MAPE | PJB/% |
|---------|-------|------|-------|
| BPNN | 41.02 | 0.13 | 87 |
| ABC-BP | 23.76 | 0.10 | 90 |
| KABC-BP | 12.88 | 0.07 | 93 |

表 2 在 SO₂ 下 MSE、MAPE、PJB 的比较

| 算法 | MSE | MAPE | PJB/% |
|---------|------|------|-------|
| BPNN | 4.85 | 0.18 | 82 |
| ABC-BP | 2.85 | 0.14 | 86 |
| KABC-BP | 1.91 | 0.11 | 89 |

表 1 中 PM2.5 在均方误差的指标上最小,只有 12.88,平均绝对百分比误差只有 0.07,也就是平均预测准确度为 93%;表 2 中 SO₂ 在均方误差指标上只有 1.91,平均绝对百分比误差只有 0.11,也就是平均预测准确度为 89%。表 1 和表 2 的整体指标数据还表明,基于人工蜂群反向传播神经网络算法的天气质量预测准确度比反向传播神经网络算法的准确度高,比基于改进人工蜂群反向传播神经网络算法预测准确度低。

4 结束语

空气质量预测受多种因素的影响,文中通过构建一个非线性递减函数实现了基本人工蜂群算法中雇佣蜂和跟随蜂搜索范围的动态调整,从而提出了一种改进人工蜂群算法,并将改进算法和反向传播神经网络相结合,提出了一种人工蜂群和反向传播神经网络相结合的混合算法,并应用该算法实现了天气质量数据预测。实验结果表明,该算法取得了较为理想的效果。

参考文献:

[1] 孙家仁,许振成,刘煜,等.气候变化对环境空气质量影响的研究进展[J].气候与环境研究,2011,16(6):805-814.

[2] 陈世权,崔晓丹,黄严.空气污染预报方法简介[J].黑龙江环境通报,2005,29(1):47-48.

[3] 吴有训,彭慕平,刘勇.基于径向基函数网络的宣城市空气质量预测[J].安徽师范大学学报:自然科学报,2011,34

(4):374-379.

[4] 雷孝恩.大气污染数值预报基础和模式[M].北京:气象出版社,1998:301-321.

[5] 郭庆春,何振芳,李力,等.人工神经网络在 API 预报中的应用[J].陕西农业科学,2011,57(2):124-126.

[6] 李学桥,马莉.神经网络工程应用[M].重庆:重庆大学出版社,1996.

[7] RAGA G B,MOYNE L L E. On the nature of air pollution dynamics in Mexice City-I. nonlinear analysis[J]. Atmospheric Environment,1996,30(23):3987-3993.

[8] FENG Qi,WU Shengjun,DU Yun,et al. Improving neural network prediction accuracy for PM10 individual air quality index pollution levels [J]. Environmental Engineering Science, 2013,30(12):725-732.

[9] JIANG Zhifang, MAO Bing, MENG Xiangxu,et al. An air quality forecast model based on the BP neural network of the samples self organization clustering [C]//6th international conference on natural computation. Yantai:IEEE,2010:1523-1527.

[10] 周国亮,刘希玉,武鲁英. BP 神经网络模型在空气质量级别评价中的应用[J]. 计算机工程与设计,2009,30(2):392-394.

[11] WANG Zhu,WANG Lingfeng. Intelligent control of ventilation system for energy - efficient buildings with CO2 predictive model[J]. IEEE Transactions on Smart Grid,2013,4(2):686-693.

[12] 王芳,程水源,李明君,等. 遗传算法优化神经网络用于大气污染预报[J]. 北京工业大学学报,2009,35(9):1230-1234.

[13] THOMAS M A,DEVASTHALE A. Sensitivity of free tropospheric carbon monoxide to atmospheric weather states and their persistency:an observational assessment over the nordic countries[J]. Atmospheric Chemistry and Physics, 2014, 14(21):11545-11555.

[14] RAJAB J M,MATJAFRI M Z,LIM H S. Air surface temperature correlation with greenhouse gases by using airs data over peninsular malaysia[J]. Pure and Applied Geophysics,2014, 171(8):1993-2011.

[15] MAGDON-ISMAIL M,ATIYA A F. The early restart algorithm [J]. Neural Computation,2000,12(6):1303-1312.

[16] KARABOGA D. An idea based on honey bee swarm for numerical optimization[D]. Kayseri:Erciyes University,2005.

[17] LIU Dujun,WANG Huajun,WANG Sen, et al. Quaternion-based improved artificial bee colony algorithm for color remote sensing image edge detection [J]. Mathematical Problems in Engineering,2015,2015:138930.