

混合遗传蚁群算法优化 BP 神经网络预测空气质量^①



杜沅昊, 刘媛华

(上海理工大学 管理学院, 上海 200093)

通信作者: 刘媛华, E-mail: liuyuanhua@usst.edu.cn

摘要: 为了进一步提高空气质量指数预测精度, 提出一种混合遗传蚁群算法优化 BP 神经网络的方式对空气质量指数进行预测. 首先初始化蚁群算法的信息素分布, 对不满足适应度条件的进行遗传算法的交叉、变异操作, 进而计算蚁群的状态转移概率和信息素浓度, 当适应度值满足条件要求时, 将寻优结果作为 BP 神经网络的最优权值和阈值, 来改善单一 BP 神经网络的不足. 最后通过运用西安市的空气质量指数日历史数据进行验证, 实验表明, 本文所提模型的各个评价指标相对其他对比模型误差更小, 在预测精度方面具有更高的说服力, 因此能够有效地预测空气质量指数.

关键词: 遗传算法; 蚁群算法; BP 神经网络; 空气质量指数 AQI; 优化

引用格式: 杜沅昊, 刘媛华. 混合遗传蚁群算法优化 BP 神经网络预测空气质量. 计算机系统应用, 2023, 32(4): 223-230. <http://www.c-s-a.org.cn/1003-3254/9042.html>

BP Neural Network Optimized by Hybrid Genetic-ant Colony Algorithm for Air Quality Prediction

DU Yuan-Hao, LIU Yuan-Hua

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: In order to further improve the prediction accuracy of the air quality index, a hybrid genetic ant colony algorithm is proposed to optimize the back propagation (BP) neural network, so as to predict the air quality index. First, the pheromone distribution of the ant colony algorithm is initialized, and crossover and mutation operations of the genetic algorithm are performed if fitness conditions are not met. Then the state transition probability and pheromone concentration of the ant colony are calculated. When the fitness meets the conditions, the optimal results are used as the optimal weights and thresholds of the BP neural network to improve the shortcomings of a single BP neural network. Finally, historical daily data of the air quality index in Xi'an are utilized for verification, and the experiment shows that all evaluation indexes of the model proposed in this study have smaller errors than those of other comparative models and are more convincing in terms of prediction accuracy. Therefore, the proposed model can effectively predict the air quality index.

Key words: genetic algorithm; ant colony algorithm; back propagation (BP) neural network; air quality index AQI; optimization

① 基金项目: 国家自然科学基金 (72071130)

收稿时间: 2022-09-07; 修改时间: 2022-10-10; 采用时间: 2022-10-27; csa 在线出版时间: 2022-12-09

CNKI 网络首发时间: 2022-12-15

近年来由于城市工业生产和交通运输事业的高速发展,以及城市人口的急剧增长,从人类生产与生活中排出的大量废气已对城市环境产生了巨大污染.而目前,城市的大气污染情况亦日趋加重,人类最基本的生存质量正在被损害,甚至会渐渐地危及人类身体健康^[1].因此,国家和社会也越来越重视空气质量的状况.空气质量预测模型的构建对于每个城市环境治理工作来说,同样十分必要^[2].

空气质量的持续改善离不开有效的科学支持和系统的管理决策,有效管控的基础是首先要对空气质量影响因素和变化趋势进行准确预测^[3].空气质量预测模型主要涉及两种类型,分别是数值模型和统计模型,其中数值模型虽然不需要有大量的历史数据就可以完成,但是对研究对象的全面理解是必不可少的;对于统计模型,尽管它对于部分研究对象来说计算简便,但对空气质量指数(AQI)这类非线性和不确定特征较为明显的对象却很难得到良好的预测效果^[4].鉴于此,近年来众多学者开始采用混合模型进行预测,而由于神经网络具有较强的非线性拟合能力,同时将神经网络与其他优化算法进行结合变得越来越普遍.如史学良等^[5]建立了一种EEMD-LSTM预测模型,首先采用EEMD方法将获得的空气质量指数AQI值分成若干有效部分,可以得到不同尺度的IMF分量和残余分量,再利用经PSO改进的LSTM模型进行预测,该模型具有较高的预测效率和精度. Yang等^[6]首先使用主成分分析法对获得的空气质量指数数据进行降维,然后利用遗传算法全局搜索优势,克服神经网络难以得到最优解的缺点,使得研究对象的误差波动越来越小,同时可以看出所提模型在迭代率和预测精度方面的性能较好. Chen等^[7]采用一种经过鲸鱼优化算法改进的Elman神经网络模型对空气质量指数进行预测,即引入自适应惯性权值来提高WOA的全局搜索能力,然后将其应用于对Elman进行优化,提高了预测精度较低的问题. 许毅蓉等^[8]提出一种基于智能型算法来预测空气综合质量指数的方法.利用模拟退火(SA)结合支持向量机的(SVM)智能型算法自动调整参数,最终得到最佳分类准确率. Gilik等^[9]创建了混合CNN+LSTM深度神经网络,利用时空关系预测城市多个位置的空气污染物浓度.通过对模型的超参数进行调整,确定了最合适的模型结构,证明该模型具有更准确地预测性能. Nurcahyanto等^[10]利用递归神经网络(RNN)对PM₁₀

浓度、湿度和温度进行年度和季节室内空气质量预测,并结合了实时数据预测的室内空气质量管理系统IAQP方法.通过观测设置的不同传感器的结果可以看出所提出的模型具有良好的结果和可行性.陈峰^[11]提出一种基于蚁群算法,同时加入了支持向量机算法的预测模型,利用蚁群算法的自动搜索能力来优化支持向量机算法里的各种参数,用动态预测模型代替静态模型,从而使得AQI预测结果更加准确. Septiawan等^[12]采用一种可以应用于递归神经网络(RNN)网络架构的BPTT(back propagation through time)算法,通过比较Elman RNN、Jordan RNN和混合网络架构来预测空气污染物浓度的时间序列数据,进而确定空气质量.研究表明文章所提的3种体系结构适用于预测空气质量.

尽管神经网络算法预测效果良好,但是在预测精度方面仍有较大的提升空间.为了进一步提高预测的精确度,本文提出一种混合遗传算法和蚁群算法进而优化BP神经网络的模型.即将遗传算法(GA)加入到蚁群算法(ACO)进而优化BP神经网络.首先利用遗传算法优化蚁群中的信息素分布,然后用蚁群算法搜索神经网络的最优权值和阈值,并采用西安市的空气质量日历史数据进行模型验证.

1 基本算法

1.1 BP神经网络算法

BP神经网络是一种采用误差反向传播(back propagation)算法的多层前馈网络,由输入层、隐含层、输出层组成,自提出以来被应用到众多领域,是目前使用得最为成功的神经网络模型之一^[13].其基本结构图如图1,图中各参数的含义见表1.

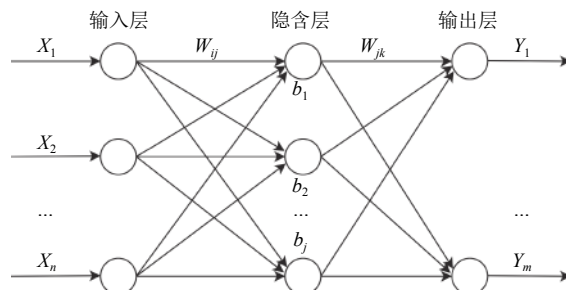


图1 BP神经网络结构图

通过对BP神经网络进行学习的过程可发现,其学习效果和收敛速率主要由初始权值的选择决定.比较合理的初始权值能够使得网络模型很快收敛并且获得

最优解,反之,将会导致一系列典型问题的发生,例如网络收敛速度变慢,较易陷入局部极小值等^[13,14].故本文采用混合遗传蚁群算法 GA-ACO 优化 BP 神经网络建立预测模型,即通过将遗传算法加入到蚁群算法的求解过程,进而求取出最佳信息素,然后调整和优化 BP 神经网络初始权值、阈值,提高了 BP 神经网络的收敛速度,克服了易陷入局部极值的问题,可以得到较好的预测效果.

表1 BP神经网络参数含义

参数	含义
X_1, X_2, \dots, X_n	BP神经网络的输入向量
W_{ij}	从输入层到隐含层的权重
W_{jk}	从隐含层到输出层的权重
b_1, b_2, \dots, b_j	神经网络的阈值
Y_1, \dots, Y_m	神经网络的输出值

1.2 遗传算法与蚁群算法

遗传算法(GA)是一种模拟自然界遗传机制,以生物进化论为理论基础而形成的过程搜索最优解算法^[15].主要包括选择操作、交叉操作、变异操作.在进行选择操作时,个体被选中的概率与适应度值相关,个体适应度值愈突出,入选的可能性愈大,最常用的方法是轮盘赌法^[16].每个个体的适应度值和选择概率如下:

$$F_i = k \left(\sum_{i=1}^n |y_i - o_i| \right) \quad (1)$$

$$p_i = \frac{f_i}{\sum_{i=1}^N f_i} \quad (2)$$

$$f_i = k/F_i \quad (3)$$

其中, F_i 为个体 i 的适应度值; k 为系数; y_i 为实际值; o_i 为预测值; n 为输出层结点个数; N 为种群个体数量.

交叉操作: 一般采用实数交叉法, 第 k 个染色体 a_k 和第 l 个染色体 a_l 同时在 j 位的交叉方式如下:

$$a_{kj} = a_{kj}(1-b) + a_{lj}b \quad (4)$$

$$a_{lj} = a_{lj}(1-b) + a_{kj}b \quad (5)$$

其中, b 是 $[0, 1]$ 的随机数.

变异操作: 通过选择第 i 个个体的第 j 个基因 a_{ij} 进行变异, 具体操作方法如下:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max}) \cdot f(g), & r > 0.5 \\ a_{ij} + (a_{\min} - a_{ij}) \cdot f(g), & r \leq 0.5 \end{cases} \quad (6)$$

$$f(g) = r_2(1 - g/G_{\max})^2 \quad (7)$$

其中, a_{\max} 与 a_{\min} 分别是基因 a_{ij} 的上界和下界, r_2 是一个随机数, g 是当前迭代次数, G_{\max} 是最大进化次数, r_2 是 $[0, 1]$ 之间的随机数.

蚁群算法(ACO)是一种启发式随机搜索算法,它模拟了自然界中蚂蚁寻径过程^[17].基本原理: 蚂蚁在觅食时, 最初是随机择路而行, 同时会分泌一种化学物质, 称为信息素, 残留于所行路径中, 之后, 随着迭代次数增加, 某条路径中信息素越来越多. 最终, 这些蚂蚁会选择某一条信息素最丰富的路走, 如此一来便找到了一条最优的路径^[18,19].

步骤1. 构建解空间. 首先每只蚂蚁的起始位置都是任意的, 然后统计每只蚂蚁 k ($k = 1, 2, \dots, m$) 下一个要访问的城市, 直到全部蚂蚁访问完设置的全部城市. 在 t 时刻, 蚂蚁 k 从城市 i 转移到城市 j 的概率 $p_{ij}^k(t)$ 为:

$$p_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}(t)]^\beta}{\sum_{s \in J_k(i)} [\tau_{is}(t)]^\alpha \cdot [\eta_{is}(t)]^\beta}, & j \in J_k(i) \\ 0, & \text{其他} \end{cases} \quad (8)$$

其中, α 为信息素因子, β 为启发函数因子; i, j 分别为起始点和终点; $\eta_{ij}(t)$ 为两点之间长度的倒数; t 为最大迭代次数; $\tau_{ij}(t)$ 为时间 t 由 i 到 j 的信息素含量; $J_k(i)$ 为所有没被访问过的节点的集合, 完全访问完成后表示一次循环结束.

步骤2. 更新信息素. 记录在当前迭代次数上的最佳解(最短路径), 并对不同城市的路线中的信息素含量做出了调整.

$$\tau_{ij}(t+n) = (1-\rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij} \quad (9)$$

$$\Delta\tau_{ij} = \sum_{k=1}^m \Delta\tau_{ij}^k \quad (10)$$

其中, ρ 为信息素挥发因子, 取值范围一般在 $[0.2, 0.5]$ 之间; $(1-\rho)$ 为信息素残留系数; 式(9)表示在访问过程结束后 m 只蚂蚁在城市 i 到 j 路径上留下的信息素总数. $\Delta\tau_{ij}^k$ 为在本次循环中蚂蚁 k 在路径上留下的信息量, 由下式表示:

$$\Delta\tau_{ij}^k = \begin{cases} \frac{Q}{L_k}, & \text{蚂蚁 } k \text{ 在本次循环中的路径 } (i, j) \\ 0, & \text{其他} \end{cases} \quad (11)$$

其中, Q 为正常数, L_k 为蚂蚁 k 在当前周期内蚂蚁通过路

径的长度。

步骤3. 判断是否结束. 如果迭代次数少于最大迭代次数, 迭代次数加1, 同时清空所有蚂蚁经过路线的记录表, 并返回步骤1; 否则结束计算, 将得到的最优解进行输出处理。

2 混合预测模型的构建方法

2.1 混合 GA-ACO 优化 BP 神经网络算法模型

首先遗传算法和蚁群算法都可以对 BP 神经网络的参数性能进行改善, 蚁群算法与其他启发式算法相比, 具有很强的鲁棒性和搜索较好解的能力, 易与多种启发式算法结合, 以改善算法性能. 遗传算法是一种全局优化搜索算法, 能够避开局部极小点, 而且在进化过程中无需提供所要解决问题的梯度信息。

本文所提出的混合遗传蚁群算法优化 BP 神经网络模型主要是将利用蚁群算法处理的 BP 神经网络的

初始权值和阈值进行适应度计算, 其中对于满足适应度条件信息素的直接进行蚁群算法的整个流程, 并重复迭代直至选择出最优的权值和阈值. 对于不满足适应度条件的就进行遗传算法的选择、交叉、变异操作, 将获得的最优个体按照蚁群算法的状态转移概率公式进行选择路径并记录具体的参数值, 并且进行信息素浓度计算, 再次计算适应度是否满足所设条件, 对上述过程进行循环迭代, 直至获得 BP 神经网络的最优权值和阈值结束。

所提混合预测模型流程图如图2所示, 主要目的在于进一步使所取的神经网络初始权值和阈值更加合理, 提高神经网络的收敛速度, 进而找到最优解. 其中 BP 神经网络在此模型中主要起到预测的作用; 在 GA-ACO 优化部分, 主要采用遗传算法来优化蚁群算法, 然后将得到的最优权值与阈值应用于 BP 神经网络. 具体实施步骤如下。

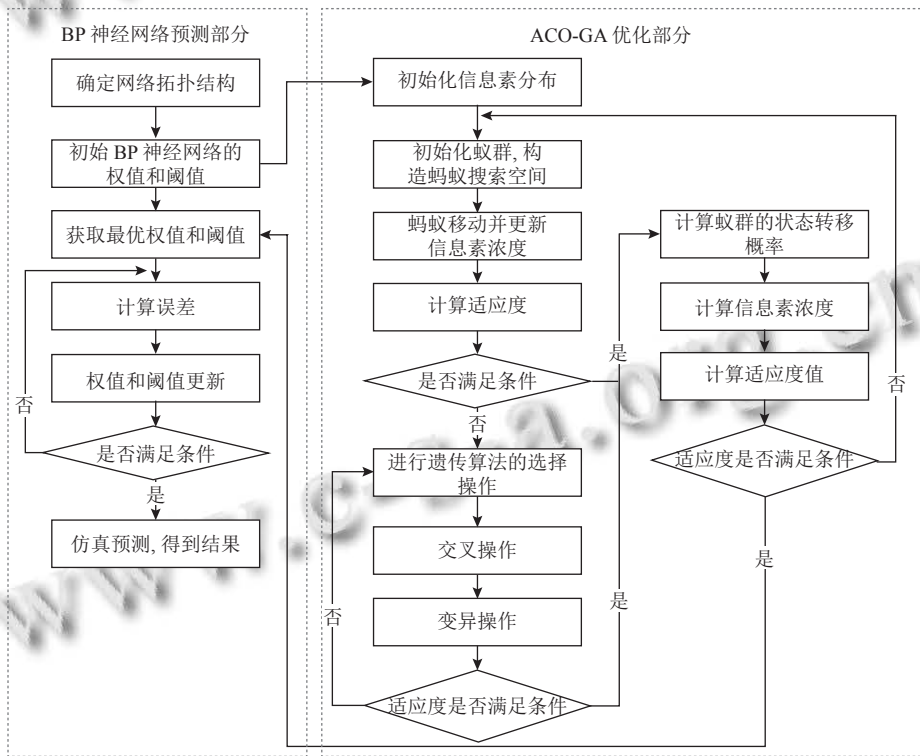


图2 混合 GA-ACO-BP 算法流程图

步骤1. 参数初始化. 对 BP 神经网络的权值 W_{ij} 和 W_{jk} 、阈值 b_1, b_2, \dots, b_j 等进行设定, 并将由上述参数所组成的所有元素集合记为 I_{wi} ; 设定遗传算法的交叉概率 p_c 、变异概率 p_m ; 对蚁群算法的蚂蚁规模 m 、信息素

因子 α 等一系列参数进行设置. 目标误差用 E 表示, 公式如式 (12), 其中 Q_{ij} 为预测值, O_{ij} 为实际值.

$$E_i = \frac{1}{2} \sum_{j=1}^m (Q_{ij} - O_{ij})^2 \quad (12)$$

步骤 2. 对蚁群执行选择交叉变异的操作. 执行蚁群算法时, 首先构造蚂蚁搜索空间, 蚂蚁移动并更新信息素浓度且计算出适应度, 在不满足设置的适应度条件时, 依次执行遗传算法的 3 步操作. 满足设置的适应度条件时, 进行步骤 3.

步骤 3. 对步骤 2 中的蚂蚁计算状态转移概率、更新信息素.

$$\tau_j(I_{wi}) = (1 - \rho) \cdot \tau_j(I_{wi}) + \Delta\tau_j(I_{wi}) \quad (13)$$

$$\Delta\tau_j(I_{wi}) = \sum_{k=1}^m \Delta\tau_j^k(I_{wi}) \quad (14)$$

其中, ρ 为信息素挥发系数; $\Delta\tau_{ij}^k$ 为在本次过程中蚂蚁 k 在路径上产生的信息量.

步骤 4. 将步骤 2, 步骤 3 获得的寻优结果作为 BP 神经网络的参数进行训练; 计算网络的输出误差.

步骤 5. 根据获得的结果, 判断是否满足目标要求. 记录误差最小的一组权值、阈值, 判断是否满足要求. 直到满足时退出条件.

2.2 混合预测模型的实施步骤

首先, 剔除无效数据并将采集到的有效数据进行归一化处理.

其次, 确定最佳隐藏层个数, 并确定神经网络结构. 将数据应用在本文所构建的混合模型中.

最后, 分别采用 BP 网络模型、GA-BP 网络模型、ACO-BP 网络模型、GA-ACO-BP 网络模型对空气质量指数进行预测, 并对不同模型的预测结果进行对比分析.

3 基于组合模型的空气质量预测

3.1 数据预处理

本篇文章采用的是西安市 2021 年 1 月–2022 年 5 月的空气质量指数日历史数据, 数据来源于环境专业知识服务系统和中国气象局官网. 根据《环境空气质量标准》可知, 用来测量空气质量的污染物主要有 6 种, 分别为 SO_2 、 NO_2 、 PM_{10} 、 $\text{PM}_{2.5}$ 、 O_3 、 CO . 本文数据包含上述 6 种污染物数据以及 AQI 数据信息. 剔除问题数据 (2022 年 3 月 3 日–13 日数据丢失) 之外共 502 条数据, 如表 2 所示.

数据首先需要采用 `mapminmax()` 函数进行归一化处理, 目的在于消除因量纲不同对预测结果造成的影响, 如式 (15) 所示:

$$x^* = \frac{(x - x_{\min})(y_{\max} - y_{\min})}{x_{\max} - x_{\min}} + y_{\min} \quad (15)$$

其中, y_{\max} , y_{\min} 为设定的上界与下界, 分别为 1 和 0, x 为待归一化样本数据, x_{\min} 和 x_{\max} 分别为样本数据的最小值和最大值, x^* 为样本数据归一化处理后的结果.

表 2 西安市空气质量指数数据 (部分)

Time	O ₃	CO	NO ₂	SO ₂	PM _{2.5}	PM ₁₀	AQI
2021-01-01	26	1.1	66	11	71	108	95
2021-01-02	23	1.7	77	20	129	183	170
2021-01-03	24	1.6	86	19	125	181	165
2021-01-04	8	1.9	85	19	159	231	209
2021-01-05	35	1.4	68	16	119	170	156
2021-01-06	52	0.8	44	11	54	115	83
2021-01-07	50	0.5	30	10	20	83	67

3.2 仿真条件及结果分析

3.2.1 实验条件及评价指标

本文的仿真实验是在 Matlab 2016a 平台上进行的, 首先选择了单隐藏层的 BP 神经网络, 并根据实际需要选取网络模型的所有参数, 其中输入层结点为 6, 输出层结点为 1, 隐藏层结点根据经验公式来确定, 如式 (16):

$$l = \sqrt{m+n} + a \quad (16)$$

其中, l 表示隐藏层结点个数; m 为输入层结点个数; n 为输出层结点个数; a 为 [1, 10] 之间的任意整数. 由此可知隐藏层结点个数的有效范围为 [4, 12], 在此范围内进行试验并挑选出最优值. 本文模型所用 4 种评价指标如式 (17)–式 (19) 所示, 其中 $RMSE$ 、 $MAPE$ 、 MAE 的值越小, 说明预测模型的性能越好; 式 (20) 中 EC 为均等系数, 其值介于 0–1 之间, 若 EC 能够处于 0.9–1 的范围, 则表示预测效果较为理想^[20,21]. 具体公式为:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^*| \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2} \quad (18)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y_i^*|}{y_i} \quad (19)$$

$$EC = 1 - \frac{\sqrt{\sum_{i=1}^n (y_i - y_i^*)^2}}{\sqrt{\sum_{i=1}^n (y_i)^2} + \sqrt{\sum_{i=1}^n (y_i^*)^2}} \quad (20)$$

其中, n 为预测样本总数, y_i 是实际值, y_i^* 是预测值。

3.2.2 空气质量指数预测及分析

采用本文模型预测空气质量指数时, 将 502 条数据按 7:3 的比例分为训练集和测试集。随后将数据分别输入到 BP、GA-BP、ACO-BP、GA-ACO-BP 模型中进行实验。其中, 根据上述隐藏层结点数的经验公式并结合准确率公式, 可得表 3, 准确率公式为:

$$\text{准确率} = \left(1 - \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y_i^*|}{y_i} \right) \times 100\% \quad (21)$$

表 3 不同隐层节点数的 GA-ACO-BP 神经网络对应的准确率

隐层节点数	4	5	6	7	8	9	10	11	12
准确率 (%)	87.6	91.8	91.5	92.2	92.5	93.1	97.6	88.9	96.1

通过表 3 可知, 当隐藏层节点个数为 10 时, 准确率最高。下面依次对 GA-BP、ACO-BP 对比模型进行实验, 首先根据实验经验, 通过使用控制变量的方式, 分别在遗传算法和蚁群算法中对所需参数在一定范围内进行调试, 将调试 10 次之后所获得的 MAE、MASE、MAPE、EC 值进行取平均值, 通过比较选择出最优的参数。

通过表 4、表 5 中 GA-BP、ACO-BP 的参数调试性能对比, 筛选出以上两个模型中评价指标较为优秀的参数, 并应用到本文所提的混合模型中, 得到的参数性能对比结果如表 6 所示。

表 4 GA-BP 参数调试性能对比

P_c	P_m	平均MAE	平均MSE	平均MAPE	平均EC
0.8	0.1	6.5070	43.2625	0.0418	0.9668
	0.2	3.4447	26.0694	0.0310	0.9789
0.7	0.1	13.2252	481.1514	0.1388	0.9286
	0.2	7.4407	134.6240	0.0636	0.9540
0.6	0.1	5.0881	62.8027	0.0454	0.9681
	0.2	8.0335	161.7195	0.0683	0.9488
0.5	0.1	7.3057	137.0145	0.0701	0.9548
	0.2	6.2937	118.9470	0.0533	0.9566

根据表 6 中筛选出的不同参数的运行结果对比, 最终可以确定混合模型的最优参数值见表 7。

将 BP、GA-BP、ACO-BP 与本文提出的 GA-ACO-BP 模型进行对比, 如图 3。对比图 3(a) 和图 3(b) 可以看出在 GA-BP、ACO-BP 模型使用的测试集数据中, 预测值和实际值在图像上所反映的仍有很多天的 AQI 数据值差距比较明显。但是通过图 3(c) 可以看出, GA-ACO 优化的 BP 神经网络模型大大减小了这种差距,

使得预测值和实际值的拟合度变得更高, 效果也是这几种模型中最好的。

表 5 ACO-BP 参数调试性能对比

α	β	ρ	平均MAE	平均MSE	平均MAPE	平均EC
1	3	0.3	4.5251	38.0387	0.0387	0.9748
		0.5	8.0255	106.4405	0.0743	0.9568
	4	0.3	5.7096	61.7588	0.0460	0.9676
		0.5	4.9364	51.6286	0.0432	0.9705
2	3	0.3	4.2031	34.3862	0.0396	0.9760
		0.5	3.9407	30.0457	0.0306	0.9777
	4	0.3	5.7053	67.8913	0.0506	0.9659
		0.5	5.1920	61.5420	0.0413	0.9677
3	3	0.3	7.2111	118.3074	0.0655	0.9546
		0.5	5.6354	63.5222	0.0492	0.9678
	4	0.3	12.9460	364.9036	0.0130	0.9175
		0.5	5.6123	105.4635	0.0633	0.9575
4	3	0.3	6.0522	112.4594	0.0575	0.9558
		0.5	4.9408	68.4154	0.0517	0.9660
	4	0.3	4.8996	62.6765	0.0428	0.9672
		0.5	8.6570	138.2478	0.0724	0.9508

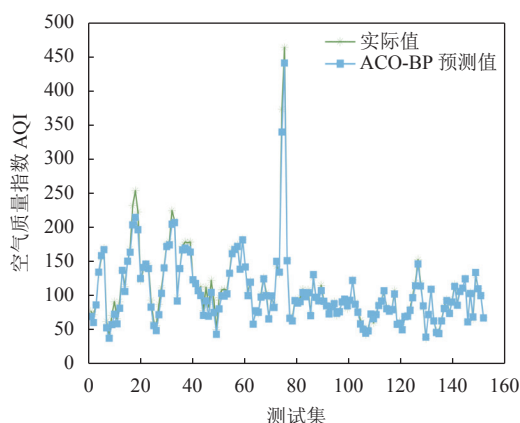
表 6 GA-ACO-BP 参数调试性能对比

P_c	P_m	α	β	ρ	平均MAE	平均MSE	平均MAPE	平均EC
0.8	0.2	1	3	0.3	2.7108	13.5490	0.0256	0.9802
0.8	0.2	1	4	0.5	2.5843	14.1638	0.0249	0.9845
0.8	0.2	2	3	0.5	2.3759	12.0764	0.0238	0.9858
0.8	0.2	4	3	0.5	3.8926	27.0460	0.0363	0.9788
0.8	0.2	4	4	0.3	5.5239	70.1854	0.0489	0.9656
0.6	0.1	1	3	0.3	3.1988	18.9931	0.0317	0.9821
0.6	0.1	1	4	0.5	2.7193	15.9012	0.0261	0.9838
0.6	0.1	2	3	0.5	2.8076	14.6287	0.0281	0.9844
0.6	0.1	4	3	0.5	3.1703	18.3447	0.0321	0.9825
0.6	0.1	4	4	0.3	3.3717	23.3496	0.0320	0.9803

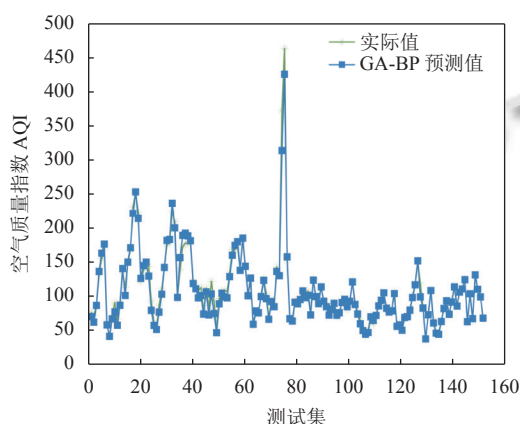
表 7 GA-ACO-BP 模型参数设置

算法	参数名称	值
遗传算法	交叉概率 p_c	0.8
	变异概率 p_m	0.2
蚁群算法	蚂蚁数量 m	20
	信息素因子 α	2
	启发函数因子 β	3
	信息素挥发因子 ρ	0.5
	信息素常数 Q	1
	最大迭代次数 t	30

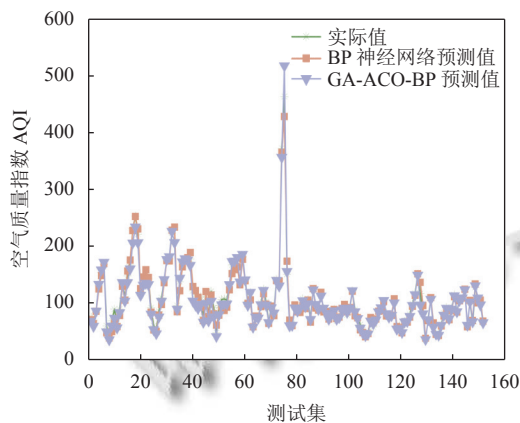
通过引入的评价指标, 对各个模型的预测结果进一步分析。根据设定的参数值, 从表 8 中可以看出本文所提出的 GA-ACO-BP 混合模型的 ave-MAE、ave-MSE、ave-MAPE 在 4 种模型中都是最小的, 分别为 2.3759、12.0764、0.0238; ave-EC 值较对比模型来说是最大的, 为 0.9858; 说明效果最佳。因此用本文提出的模型对空气质量指数进行预测是更加合适的。



(a) ACO 优化 BP 神经网络预测效果对比图



(b) GA 优化 BP 神经网络预测效果对比图



(c) BP 神经网络预测和 GA-ACO-BP 预测效果对比图

图3 不同模型的预测结果对比图

表8 模型在不同评价指标下的误差对比

模型名称	平均MAE	平均MSE	平均MAPE	平均EC
GA-ACO-BP	2.3759	12.0764	0.0238	0.9858
ACO-BP	3.9407	30.0457	0.0306	0.9777
GA-BP	3.4447	26.0694	0.0310	0.9789
BP	6.0348	68.2592	0.0571	0.9658

4 结论

本文主要利用混合遗传蚁群算法优化 BP 神经网络来预测空气质量指数, 通过使用西安市的空气质量指数日历史数据对所提模型进行验证, 结合预测对比图以及误差评价指标可知, 混合遗传蚁群算法优化的 BP 神经网络预测效果更好, 拟合能力更加突出. 可见该模型在分析空气质量指数方面具有一定优势, 对相关部门判断空气质量具有重要意义.

虽然本文提出的模型在一定程度上提高了预测精度, 但是空气质量指数受很多因素的影响, 其中很大一部分因素是与人类活动有关的, 比如汽车的尾气排放、工厂的有害气体排放、城市的发展密度、各个地区的地形地貌等. 所以在后续的研究工作中, 应更加全面的考虑其他影响因素, 进一步提高预测精度.

参考文献

- 雷明丽, 张艳飞. 浅析空气质量的影响因素和防治措施. 城市建设理论研究 (电子版), 2012, (14).
- 刘闯, 王帅, 林宏, 等. 沈阳市冬季环境空气质量统计预报模型建立及应用. 中国环境监测, 2014, 30(4): 10-15. [doi: 10.3969/j.issn.1002-6002.2014.04.002]
- 齐甜方, 蒋洪迅, 石晓文. 面向多源数据沈阳市 PM_{2.5} 浓度预测研究及实证分析. 系统工程, 2018, 36(5): 104-115.
- 甘露情, 刘媛华. 基于 BP-SVR 组合模型的空气质量指数预测. 软件导刊, 2020, 19(10): 80-83.
- 史学良, 李梁, 赵清华. 基于改进 LSTM 网络的空气质量指数预测. 统计与决策, 2021, 37(16): 57-60. [doi: 10.13546/j.cnki.tjyjc.2021.16.012]
- Yang RJ, Hu XQ, He LJ. Prediction of Shanghai air quality index based on BP neural network optimized by genetic algorithm. Proceedings of the 2020 13th International Symposium on Computational Intelligence and Design (ISCID). Hangzhou: IEEE, 2020. 205-208.
- Chen Z, Cai DW. Based on a prediction method for improving WOA-Elman air quality prediction. Proceedings of the 2nd International Conference on Computing and Data Science. Stanford: ACM, 2021. 43.
- 许毅蓉, 连金海, 张小蓉, 等. 运用智能型算法预测空气综合质量指数的研究. 福建电脑, 2022, 38(6): 17-21. [doi: 10.16707/j.cnki.fjpc.2022.06.005]
- Gilik A, Ogrenci AS, Ozmen A. Air quality prediction using CNN+LSTM based hybrid deep learning architecture. Environmental Science and Pollution Research, 2022, 29(8): 11920-11938. [doi: 10.1007/s11356-021-16227-w]

- 10 Nurcahyanto H, Prihatno AT, Alam MM, *et al.* Multilevel RNN-based PM₁₀ air quality prediction for industrial Internet of things applications in cleanroom environment. *Wireless Communications and Mobile Computing*, 2022, 2022: 1874237.
- 11 陈峰. 基于蚁群算法和支持向量机的空气质量指数预测模型研究 [硕士学位论文]. 金华: 浙江师范大学, 2017.
- 12 Septiawan WM, Endah SN. Suitable recurrent neural network for air quality prediction with backpropagation through time. *Proceedings of the 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS)*. Semarang: IEEE, 2018. 1–6.
- 13 李长安, 卢雪琴, 吴忠强, 等. 基于蚁群算法优化反向传播神经网络的港口吞吐量预测. *计量学报*, 2020, 41(11): 1398–1403. [doi: [10.3969/j.issn.1000-1158.2020.11.14](https://doi.org/10.3969/j.issn.1000-1158.2020.11.14)]
- 14 杨客. 遗传算法优化的BP神经网络在连云港港口吞吐量预测中的应用研究 [硕士学位论文]. 深圳: 深圳大学, 2017.
- 15 吴慧静, 赫晓慧. 基于GA-BP神经网络的空气质量指数预测研究. *安徽师范大学学报(自然科学版)*, 2019, 42(4): 360–365. [doi: [10.14182/J.cnki.1001-2443.2019.04.011](https://doi.org/10.14182/J.cnki.1001-2443.2019.04.011)]
- 16 杨云, 杨毅. 基于GA-BP神经网络的供暖期空气质量指数预测分析. *陕西科技大学学报*, 2016, 34(4): 171–176, 186. [doi: [10.19481/j.cnki.issn1000-5811.2016.04.032](https://doi.org/10.19481/j.cnki.issn1000-5811.2016.04.032)]
- 17 Ding MJ, Zhang SZ, Zhong HD, *et al.* A prediction model of the sum of container based on combined BP neural network and SVM. *Journal of Information Processing Systems*, 2019, 15(2): 305–319.
- 18 Irani R, Nasimi R. An evolving neural network using an ant colony algorithm for a permeability estimation of the reservoir. *Petroleum Science and Technology*, 2012, 30(4): 375–384. [doi: [10.1080/10916466.2010.483442](https://doi.org/10.1080/10916466.2010.483442)]
- 19 杨锐锐, 王颖. 蚁群算法的基本原理及参数设置研究. *南方农机*, 2018, 49(13): 38–39. [doi: [10.3969/j.issn.1672-3872.2018.13.023](https://doi.org/10.3969/j.issn.1672-3872.2018.13.023)]
- 20 吴铁锋. 改进BP神经网络的城市交通流预测研究. *交通科技*, 2010, (5): 92–94. [doi: [10.3963/j.issn.1671-7570.2010.05.031](https://doi.org/10.3963/j.issn.1671-7570.2010.05.031)]
- 21 孟飞, 兰巨龙, 胡宇翔. 基于改进的量子粒子群优化小波神经网络的网络流量预测. *计算机应用研究*, 2015, 32(5): 1450–1453. [doi: [10.3969/j.issn.1001-3695.2015.05.042](https://doi.org/10.3969/j.issn.1001-3695.2015.05.042)]

(校对责编: 牛欣悦)