

◆人工智能算法与应用◆

基于改进麻雀搜索算法的空气质量指数预测*

胡青¹, 龚世才^{1**}, 胡珍²

(1. 浙江科技学院理学院, 浙江杭州 310000; 2. 湖北工业大学理学院, 湖北武汉 430068)

摘要:为更准确地预测空气质量指数(Air Quality Index, AQI), 提出一种基于改进麻雀搜索算法(Improved Sparrow Search Algorithm, ISSA)的AQI预测模型(ISSA-BP)。利用麻雀搜索算法(Sparrow Search Algorithm, SSA)的全局搜索性能对BP神经网络的权值和阈值进行优化, 解决传统BP神经网络在预测AQI过程中出现的收敛速度慢、易陷入局部最优等问题。同时, 针对SSA在优化过程中的缺陷, 引入立方映射和优化策略增强算法的全局搜索及收敛能力, 进一步提高预测性能。应用ISSA-BP模型预测杭州市AQI, 实验结果表明, 与其他模型相比, 该模型的预测精度有显著提升。本研究为大气污染防治提供了新的预测方法。

关键词:空气质量指数预测 麻雀搜索算法 混沌映射 优化策略 BP神经网络

中图分类号: TP301.6 文献标识码: A 文章编号: 1005-9164(2022)04-0642-10

DOI: 10.13656/j.cnki.gkx.20220919.004

近年来, 随着我国工业化进程不断加快, 经济迅猛发展, 人口不断增长, 空气污染问题已经引起全社会的关注^[1]。建立空气质量预测模型对城市环境治理工作有着重要作用, 不仅有助于减轻大气污染对人体的伤害, 还能为环保部门准确预测空气质量提供科学的方法^[2]。

空气质量指数(Air Quality Index, AQI)的传统预测方法有时间序列模型、数理统计模型、线性回归模型、灰色预测方法等^[3,4]。天气、大气污染物等因素对空气质量具有较大影响, 导致空气质量具有明显的非线性、不确定性等特征。因此, 传统的预测方法

很难得到精确的预测结果^[5]。由于神经网络具有较强的非线性拟合能力, 可以较好地完成对AQI的预测。王国胜等^[6]采用BP神经网络模型对AQI进行预测, 结果相较于传统统计方法, 虽然得到了更高的预测精度, 但是单一BP神经网络收敛速度慢且易陷入局部最优解。为了进一步提高预测精度, 有学者利用智能优化算法对其进行优化, 如张旭等^[7]利用粒子群算法(Particle Swarm Optimization, PSO)优化BP神经网络, 提升了模型的泛化能力, 在空气质量指数预测中取得较好效果。杨云等^[8]利用遗传算法(Genetic Algorithm, GA)对BP神经网络进行改进, 建立

收稿日期: 2022-04-04

* 国家自然科学基金项目(11571315, 12101557, 11901525)和浙江省自然科学基金项目(LY20A010005)资助。

【作者简介】

胡青(1998-), 男, 在读硕士研究生, 主要从事机器学习与数据挖掘、统计计算与仿真研究。

【**通信作者】

龚世才(1971-), 男, 教授, 主要从事复杂系统与复杂网络研究, E-mail: scgong@zafu.edu.cn。

【引用本文】

胡青, 龚世才, 胡珍. 基于改进麻雀搜索算法的空气质量指数预测[J]. 广西科学, 2022, 29(4): 642-651.

HU Q, GONG S C, HU Z. Air Quality Index Prediction Based on Improved Sparrow Search Algorithm [J]. Guangxi Sciences, 2022, 29(4): 642-651.

空气质量指数预测模型,提升了BP神经网络的性能,预测结果有更高的精度。综上所述,传统的单一BP神经网络模型的AQI预测效果较差,尽管有些学者利用传统优化算法优化模型,在一定程度上解决了预测精度的问题,但在算法迭代后期仍存在易陷入局部最优、收敛速度慢的缺点。

为解决上述问题,本研究利用改进麻雀搜索算法(Improved Sparrow Search Algorithm,ISSA),对BP神经网络进行训练,以期得到预测精度高、全局搜索能力强的ISSA-BP模型。该模型的主要思路是利用麻雀搜索算法(Sparrow Search Algorithm,SSA)相较于传统优化算法具有收敛速度快、寻优精度高的优点^[9],对BP神经网络的权值及阈值进行优化,同时考虑到SSA在迭代后期,存在种群多样性减少、收敛速度慢以及全局搜索能力弱的缺陷^[10],通过引入立方映射提高初始解的质量,增加初始化麻雀种群的多样性,从而提升算法的收敛能力,同时结合优化策略增强算法的全局搜索能力。使用ISSA-BP模型对杭州市空气质量指数值进行预测,检验该模型的实际应用效果。

1 麻雀搜索算法(SSA)

麻雀搜索算法是模拟麻雀搜索食物过程的一种新型智能算法^[11],通过比较适应度值,不断更新种群中麻雀个体的位置。

在SSA中,发现者搜索范围广,负责引导种群搜索和觅食,发现者的位置更新公式如下:

$$X_{i,d}^{t+1} = \begin{cases} X_{i,d}^t \exp(\frac{-i}{\alpha t_{\max}}), R < ST \\ X_{i,d}^t + QL, R \geq ST \end{cases}, \quad (1)$$

式(1)中, d 表示搜索空间维度, t 表示当前的迭代次数, t_{\max} 表示最大迭代次数, α 是(0,1]间的均匀随机数。 $X_{i,d}^t$ 表示第 t 次迭代下,第 i 只麻雀处于 d 维的位置。 R 为预警值,取值范围为[0,1]。 ST 为警戒值,取值范围为[0.5,1]。 Q 为服从标准正态分布的随机数。 L 为一个单行 d 维的全1矩阵。

跟随者会跟随发现者进行觅食以获得更好的适应度,其位置更新公式如下:

$$X_{i,d}^{t+1} = \begin{cases} Q \exp(\frac{X_{\text{worst}}^t - X_i^t}{i^2}), i > \frac{n}{2} \\ X_p^{t+1} + |X_{i,d}^t - X_p^{t+1}| A^+ L, \text{其他} \end{cases}, \quad (2)$$

式(2)中, X_p^{t+1} 为第 $t+1$ 次迭代发现者最优位置,

X_{worst}^t 为第 t 次迭代全局最差位置。 A 为一个单行 d 维的矩阵,其中每个元素随机赋值为1或-1,且 $A^+ = A^T(AA^T)^{-1}$, n 为麻雀总数。

在麻雀觅食的同时会在种群中随机选取一定比例的个体作为警戒者进行侦查预警,其位置更新公式如下:

$$X_{i,d}^{t+1} = \begin{cases} X_{\text{best}}^t + \beta |X_{i,d}^t - X_{\text{best}}^t|, f_i > f_g \\ X_{i,d}^t + K \left(\frac{|X_{i,d}^t - X_{\text{worst}}^t|}{(f_i - f_w) + \xi} \right), f_i = f_g \end{cases}, \quad (3)$$

式(3)中, X_{best}^t 为第 t 次迭代全局最优位置, β 为步长控制参数, K 为[-1,1]的随机数。 ξ 是一个极小常数,用以避免当 $f_i = f_w$ 时,分母为0的情况。 f_i 、 f_g 、 f_w 分别表示当前、最佳和最差适应度。

2 改进的麻雀搜索算法(ISSA)

2.1 立方映射初始化种群

麻雀搜索算法对种群进行初始化时,采用的是随机生成的方式,这种方式会使得麻雀种群分布不均匀,影响后期的迭代寻优。为提高算法全局搜索能力,避免迭代后期种群多样性降低,同时考虑到混沌映射具有随机性、遍历性和规律性等特点^[12,13],利用混沌序列对种群进行初始化。汤安迪等^[14]证明了立方映射的均匀性及遍历性优于经典的Logistic映射,故本研究采用立方映射完成种群的初始化。立方映射公式如下:

$$y(n+1) = 4y(n)^3 - 3y(n), \quad (4)$$

式(4)中, n 为映射次数, $y(n) \in (-1,0) \cup (0,1)$, $y(n)$ 为第 n 次映射值。

为防止通过立方映射产生的值超过BP神经网络的权值及阈值优化范围,利用式(5)将映射值调整到统一区间:

$$X_i = X_{lb} + 0.5(X_{lb} - X_{ub})(y_i + 1), \quad (5)$$

式(5)中, X_{lb} 、 X_{ub} 分别为麻雀种群中个体所处维度的上、下边界; X_i 为麻雀个体实际位置值。

立方映射初始化麻雀种群的步骤如下:

步骤1 设置种群规模 N 、维数 d 、初始值 y_0 。

步骤2 根据式(4)进行迭代计算,产生序列 y ,序列 y 中每个 d 维向量作为一个麻雀个体。

步骤3 根据式(5)对越界值进行处理,使得每个 d 维向量的值变化区间是[-1,1]。

假设种群规模为30,在二维搜索空间中产生的初始化种群分布如图1和图2所示,从图中可以看

出,相比随机初始化的种群序列,通过立方映射产生的初始种群分布更加均匀,遍历性更好。

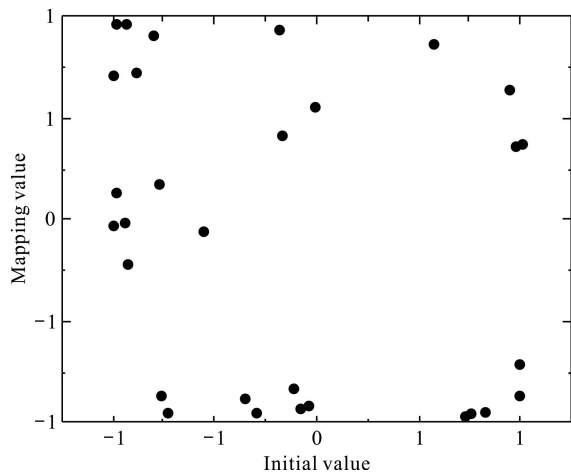


图1 随机初始化种群

Fig. 1 Random initialization population

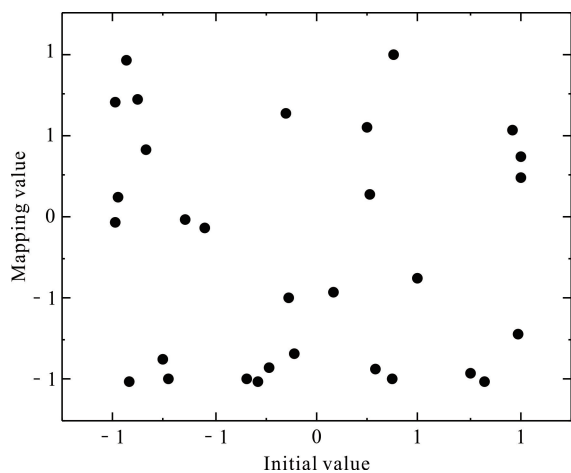


图2 立方映射初始化种群

Fig. 2 Cube mapping initialization population

2.2 蝴蝶优化策略

蝴蝶优化算法 (Butterfly Optimization Algorithm, BOA) 是受到蝴蝶觅食和求偶过程所启发的群智能算法^[15]。在迭代时,蝴蝶将朝着气味最浓的方向移动,该阶段被称为全局搜索阶段。其位置更新方式如下:

$$X_i^{t+1} = X_i^t + (r^2 X_{g_{best}} - X_i^t) f_i, \quad (6)$$

式(6)中, X_i^t 是第 t 次迭代过程中第 i 只蝴蝶的位置, $X_{g_{best}}$ 是当前全局最优位置, f_i 表示第 i 只蝴蝶发出的气味,其取值大小取决于适应度的大小, r 为 $0-1$ 的随机数。

由式(1)可知,当 $R < ST$ 时,虽然随着迭代次数的增大,取值分布变得较为均匀,在一定程度上提高了其收敛速度,但是发现者的每一维都在变小并逐渐收敛于 0,因此其搜索能力在逐渐降低,这会导致算

法在迭代后期易陷入局部极值的情况。故本研究利用 BOA 的全局搜索阶段位置更新策略改进式(1),改进后的公式如下:

$$X_{i,d}^{t+1} = \begin{cases} X_{i,d}^t + (r^2 X_{best}^t - X_{i,d}^t) f_i, & R < ST \\ X_{i,d}^t + QL, & R \geq ST \end{cases} \quad (7)$$

对改进前后的公式分别进行迭代,得到麻雀种群中个体的分布如图 3 和图 4 所示。从图中可以看出,改进后的策略在迭代后期保证了麻雀种群个体分布的多样性,在提高收敛速度的同时,在一定程度上扩大了搜索空间,提升了算法的全局寻优能力。

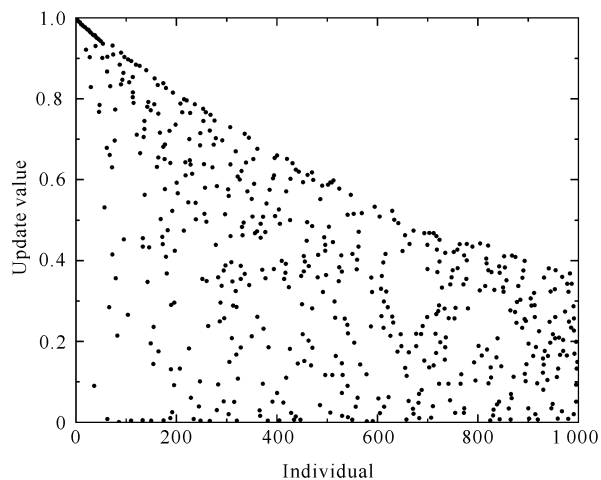


图3 改进前发现者搜索策略

Fig. 3 Discoverer search strategy before improvement

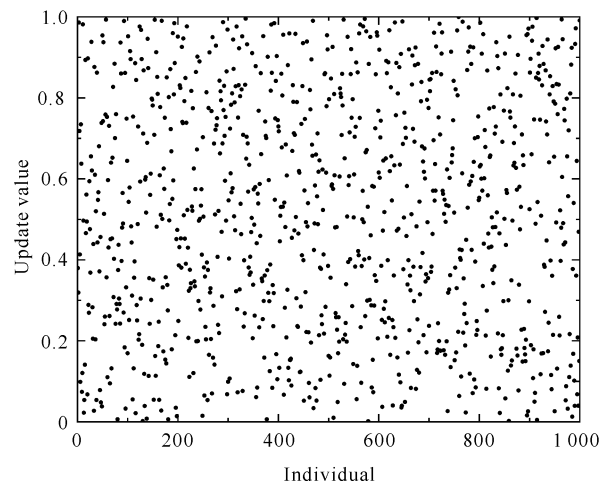


图4 改进后发现者搜索策略

Fig. 4 Discoverer search strategy after improvement

2.3 正余弦搜索策略

根据式(2)可知,跟随者在向发现者最优位置靠近时,会导致种群个体短时间聚集,虽然提高了算法的收敛速度,但是易出现算法陷入局部最优。正余弦函数不仅易于实现,而且有着良好的收敛性。利用其

振荡特性进行寻优可以降低算法易陷入局部最优解的风险。改进后跟随者的位置更新公式如下:

$$X_{i,d}^{t+1} = \begin{cases} X_{i,d}^t + S_1 \cdot \sin(S_2) \cdot |S_3 \cdot X_{\text{best}}^t - X_{i,d}^t|, S_4 < 0.5 \\ X_{i,d}^t + S_1 \cdot \cos(S_2) \cdot |S_3 \cdot X_{\text{best}}^t - X_{i,d}^t|, S_4 \geq 0.5 \end{cases}, \quad (8)$$

$$S_1 = a \left(1 - \frac{t}{t_{\max}}\right), \quad (9)$$

式(8)(9)中: a 为大于1的常数,本研究中的 a 取2; S_1, S_2, S_3, S_4 均为服从均匀分布的随机数; $S_2 \in (0, 2\pi), S_3 \in (0, 2), S_4 \in (0, 1); t_{\max}$ 为最大迭代次数。

2.4 高斯扰动策略

高斯分布在局部空间搜索性能好,对全局最优个体使用高斯扰动有利于算法跳出局部极值点,同时也增强了算法全局搜索性能。使用高斯扰动策略生成新个体的公式如下:

$$X_{\text{Gbest}}^t = X_{\text{best}}^t (1 + \text{Gaussian}(u, \sigma^2)), \quad (10)$$

式(10)中, X_{best}^t 为当前全局最优位置, X_{Gbest}^t 为高斯变异后得到的个体位置。 u 和 σ^2 为高斯分布的均值和方差,本研究中的 u 取0, σ^2 取1。

表1 基准测试函数

Table 1 Benchmark function

函数名称 Function name	测试函数 Test function	维度 Dimension	搜索范围 Search scope	最优值 Optimal value
Sphere	$F_1(x) = \sum_{i=1}^n x_i^2$	30	$[-100, 100]$	0
Schwefel 1.2	$F_2(x) = \sum_{i=1}^n (\sum_{j=1}^n x_j)^2$	30	$[-100, 100]$	0
Schwefel 2.22	$F_3(x) = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	30	$[-10, 10]$	0
Rosenbrock	$F_4(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	30	$[-30, 30]$	0
Step	$F_5(x) = \sum_{i=1}^n (x_i + 0.5)^2$	30	$[-100, 100]$	0
Griewank	$F_6(x) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos(\frac{x_i}{\sqrt{i}}) + 1$	30	$[-600, 600]$	0
Rastrigin	$F_7(x) = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10]$	30	$[-5.12, 5.12]$	0
Ackley	$F_8(x) = -20 \exp\left(0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i)^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n (\cos(2\pi x_i))\right) + 20 + e$	30	$[-32, 32]$	0

2.5.3 算法寻优能力分析

为了去除偶然性带来的误差,分别利用上述4种算法对8个测试函数进行30次独立试验,以0.01为取值间隔,通过参数值在对应取值范围内的不断遍

2.5 算法性能测试

2.5.1 算法初始化参数设置

为了验证ISSA的寻优效果与稳定性,本研究选取遗传算法(Genetic Algorithm, GA)^[16]、粒子群算法(Particle Swarm Optimization, PSO)^[17]、SSA以及ISSA进行寻优结果的对比,同时为了保持所有初始种群分布的均匀性与多样性,随机产生的每维取值范围均是 $[-1, 1]$ 的30维向量 y_0 ,作为种群的第一个个体,后续种群个体利用上述立方映射公式进行种群初始化。设置种群中的边界范围为 $[-3, 3]$,通过上述立方映射的边界值处理公式将种群中所有个体的取值范围统一为 $[-1, 1]$,其中所有算法的种群数量设置为30,最大迭代次数为1000。

2.5.2 测试函数选取

为了进一步验证ISSA的寻优性能,采用8个具有不同特征的基准测试函数进行测试。其中, $F_1 - F_5$ 是单峰函数, $F_6 - F_8$ 是多峰函数,测试函数维度为30,寻优的最优值为0,8个基准测试函数及其具体信息如表1所示。

历,选取30次最优值的平均值作为算法的寻优结果,选取历史最优值对应的算法参数值作为算法参数的最佳取值。算法的最佳参数值及测试函数的寻优结果分别如表2和表3所示。

表2 算法最佳参数值

Table 2 Optimal parameter value of algorithm

算法 Algorithm	参数 Parameter	取值范围 Value range	最佳取值 Optimal value
GA	Crossover probability P_c	[0,1]	0.8
	Mutation probability P_m	[0,1]	0.2
PSO	Individual acceleration constant C_1	[0,4]	2.0
	Social acceleration constant C_2	[0,4]	2.0
	Inertia weight ω	[0,1]	0.9
SSA	Warning value ST	[0.5,1]	0.8
	Alarm value R	[0,1]	0.2
	Proportion of vigilantes SD	[0,1]	0.2
ISSA	Warning value ST	[0.5,1]	0.8
	Alarm value R	[0,1]	0.2
	Proportion of vigilantes SD	[0,1]	0.2

由表3可知,ISSA对于单峰函数 $F_1 - F_3$ 的优化效果非常明显,其寻优的指标值远超其他算法,而对于单峰测试函数 $F_4 - F_5$,虽然ISSA的改进效果略优于SSA,但其改进效果要明显优于GA与PSO,

表3 测试函数寻优结果

Table 3 Optimization results of test function

函数 Function	算法 Algorithm	最优值 Optimal value	平均值 Average value	标准差 Standard deviation	函数 Function	算法 Algorithm	最优值 Optimal value	平均值 Average value	标准差 Standard deviation
F_1	GA	2.11	4.43	1.98	F_5	GA	18.89	70.35	45.69
	PSO	0.55	0.75	0.71		PSO	0.87	1.15	0.24
	SSA	$6.5e-178$	$1.2e-91$	$1.74e-91$		SSA	$1.7e-09$	$9.9e-06$	$1.6e-05$
	ISSA	$1.3e-200$	$4.9e-112$	$1.1e-111$		ISSA	$5.0e-10$	$5.4e-06$	$6.1e-06$
F_2	GA	184.08	228.28	228.28	F_6	GA	58.9	66.8	7.63
	PSO	11.24	14.31	14.31		PSO	154	198	29.9
	SSA	$2.4e-95$	$3.1e-51$	$3.5e-51$		SSA	0	0	0
	ISSA	$9.4e-179$	$9.2e-61$	$2.1e-60$		ISSA	0	0	0
F_3	GA	4.19	5.98	1.97	F_7	GA	0.02	0.02	0.007
	PSO	160.05	321.10	321.10		PSO	0.04	0.06	0.01
	SSA	$3.2e-18$	$1.1e-15$	$2.4e-15$		SSA	0	0	0
	ISSA	$5.3e-23$	$4.8e-17$	$1.1e-18$		ISSA	0	0	0
F_4	GA	3 392.3	6 516.9	2 806.5	F_8	GA	0.01	0.02	0.009
	PSO	172.16	387.35	316.58		PSO	0.02	0.5	0.714
	SSA	$1.4e-09$	0.000 9	0.001 9		SSA	$8.9e-16$	$1.6e-15$	$1.6e-15$
	ISSA	$1.8e-14$	0.000 2	0.000 4		ISSA	$8.9e-16$	$8.9e-16$	0

对于多峰函数 F_8 ,ISSA的寻优效果与SSA差别不大,但寻优效果优于GA与PSO十几个数量级。不论是单峰函数还是多峰函数,ISSA在多次寻优的过程中,平均值和标准差均比其他算法更小,表明ISSA拥有更好的全局搜索能力。

2.5.4 算法收敛能力分析

通过绘制基准测试函数迭代进化曲线,可进一步对比分析各个算法的收敛能力,从而更加精准地评判其预测性能,图5-12为所有算法迭代1 000次后的进化曲线图。

由图5-12可知,相比于其他几种算法,一方面,ISSA迭代更少次数进入相同精度,表明了引入的立方映射提高了初始解的质量,使得种群分布更加均匀;另一方面,当算法的寻优值存在不同程度的停滞时,ISSA的寻优值呈现出波动下降,表明优化策略对算法的改进使得算法有能力跳出局部极值,在全局空间内搜索最优解。以上改进均有效提高了ISSA的收敛速度,使得ISSA相较于其他的算法拥有更强的收敛能力。

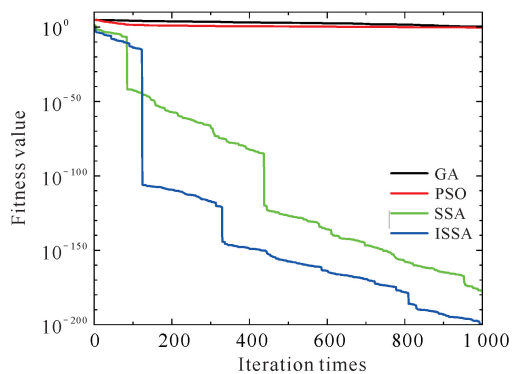


图5 F_1 进化曲线
Fig. 5 F_1 evolution curve

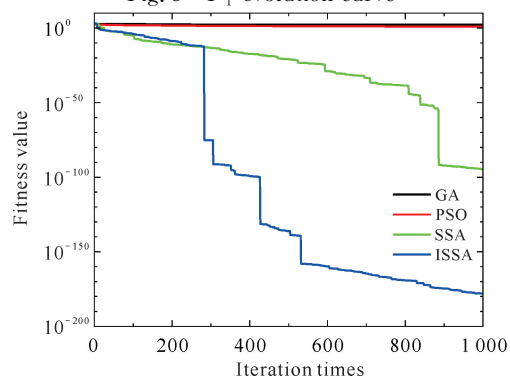


图6 F_2 进化曲线
Fig. 6 F_2 evolution curve

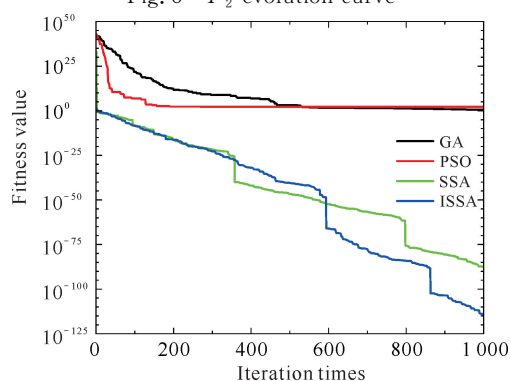


图7 F_3 进化曲线
Fig. 7 F_3 evolution curve

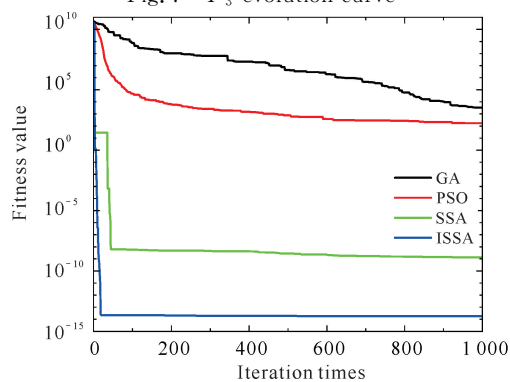


图8 F_4 进化曲线
Fig. 8 F_4 evolution curve

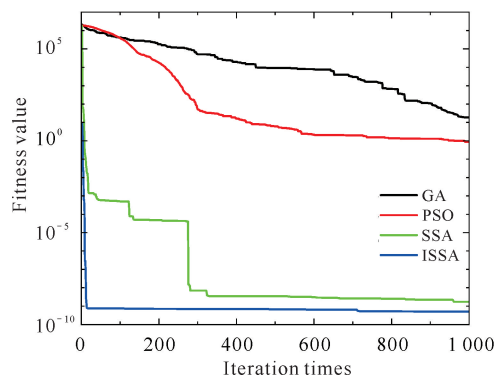


图9 F_5 进化曲线
Fig. 9 F_5 evolution curve

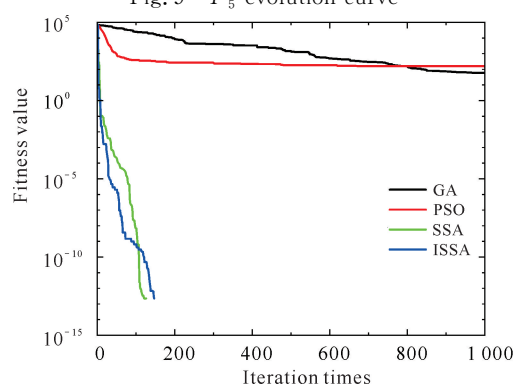


图10 F_6 进化曲线
Fig. 10 F_6 evolution curve

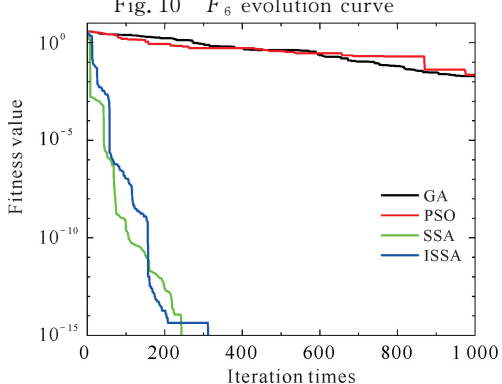


图11 F_7 进化曲线
Fig. 11 F_7 evolution curve

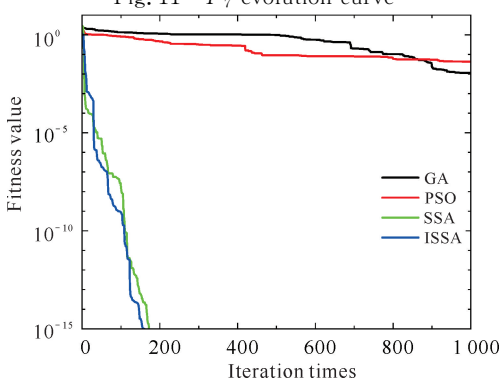


图12 F_8 进化曲线
Fig. 12 F_8 evolution curve

3 ISSA-BP 模型

3.1 BP 神经网络原理

BP 神经网络是一种采用误差反向传播算法进行学习的多层前馈网络^[18,19],其结构包括输入层、隐含层、输出层。BP 神经网络的拓扑结构如图 13 所示。

3.2 ISSA-BP 模型原理

ISSA-BP 模型的主要思想是利用改进后的麻雀搜索算法取代梯度下降法在全局范围内对 BP 神经网络中的权值和阈值进行大范围搜索,通过适应度值的更新不断调整训练误差。本模型将测试集中的均方误差作为适应度函数值。迭代完成后通过解码形成最优权值及阈值,同时结合改进后算法收敛速度快、泛化能力强的特点对 BP 神经网络进行训练,最

后得到预测性能更好的神经网络模型。ISSA-BP 模型的流程图如图 14 所示。

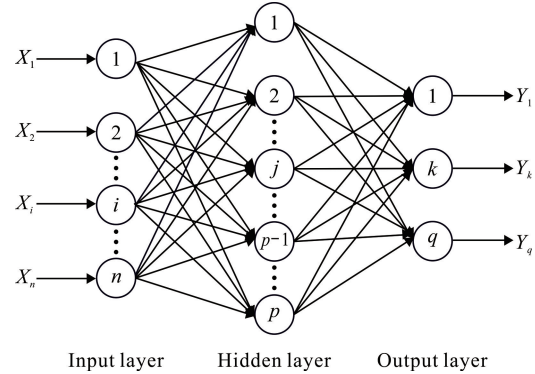


图 13 BP 神经网络拓扑结构

Fig. 13 BP neural network topological structure

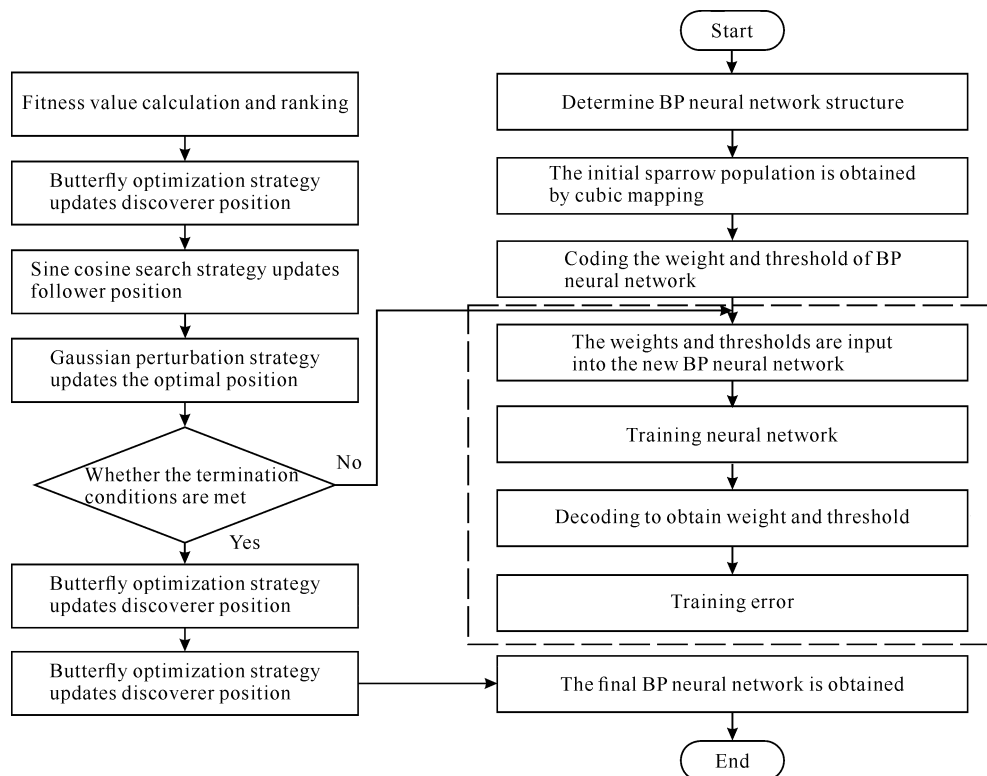


图 14 ISSA-BP 模型流程图

Fig. 14 Flow chart of ISSA-BP model

4 基于 ISSA-BP 模型的 AQI 预测

4.1 数据采集与预处理

本研究数据是从中华人民共和国生态环境部网站(<https://www.mee.gov.cn>)实时监测的杭州市空气质量数据中获取的,采样时间为 2020 年 9 月 1 日至 2022 年 1 月 13 日,共计 500 组样本数据。训练集为随机选取的 450 组数据,测试集为剩余的 50 组

数据。将 $PM_{2.5}$ 、 PM_{10} 、 SO_2 、 CO 、 NO_2 、 O_3 纳入预测体系中,作为输入变量对 AQI 进行预测。

由于数据的单位或数量级的不同,可能导致收敛速度变慢、训练时间变长等问题。因此,要对样本数据进行归一化处理,归一化公式如下:

$$X'_k = \frac{X_k - X_{\min}}{X_{\max} - X_{\min}}, \quad (11)$$

式(11)中, X_{\min} 为训练集中最小值, X_{\max} 为训练集中

最大值, X_k 为训练集中任一数值, X'_k 为归一化之后的数值。

4.2 BP 神经网络结构

BP 神经网络结构的确定是数据输入以及网络训练的关键, BP 神经网络结构包括输入层、输出层以及隐含层 3 个部分。其中, 输入层为 6 个节点, 输出层为 1 个节点, 隐含层节点数是决定 BP 神经网络预测精度的重要因素^[20], 神经网络中确定隐含层的神经元数量通常使用经验公式(12)进行计算。

$$q = \sqrt{m + n} + a, \quad (12)$$

式(12)中, q 为隐含层的节点个数, m 为输入层的节点个数, n 为输出层的节点个数, a 的取值为 1-10 的常数。

通过不断改变隐含层神经元数目, 并代入神经网络训练, 得到的均方误差见表 4。由表 4 可知, 当隐含层节点数为 9 时, 均方误差最小, 因此, 本研究选择的隐含层节点数为 9。由此可得训练的网络结构为 6-9-1。

表 4 隐含层节点个数对应的均方误差

Table 4 Mean square error corresponding to the number of hidden layer nodes

节点数 Number of nodes	均方误差 Mean square error	节点数 Number of nodes	均方误差 Mean square error
3	0.025 4	8	0.001 3
4	0.003 4	9	0.000 6
5	0.000 8	10	0.001 2
6	0.001 5	11	0.001 1
7	0.000 7	12	0.001 7

4.3 仿真实验及结果分析

为检验 ISSA-BP 模型的性能, 分别用 BP 神经网络、遗传算法优化的 BP 神经网络(GA-BP)、粒子群算法优化的 BP 神经网络(PSO-BP)、麻雀搜索算法优化的 BP 神经网络(SSA-BP)作为对照模型进行仿真实验。本研究利用 MATLAB 进行预测仿真, 5 种模型的仿真预测结果如图 15 所示。

由仿真结果可知, ISSA-BP 模型的预测曲线的拟合程度高于其他 4 种模型, 预测值更加接近真实值, 因此, ISSA-BP 模型对 AQI 的预测更加精准。为了对所建立的空气预测模型的性能进行全面而系统的分析, 本研究选取均方根误差(Root Mean Square Error, RMSE)、平均绝对误差(Mean Absolute Error, MAE)、平均绝对百分比误差(Mean Absolute

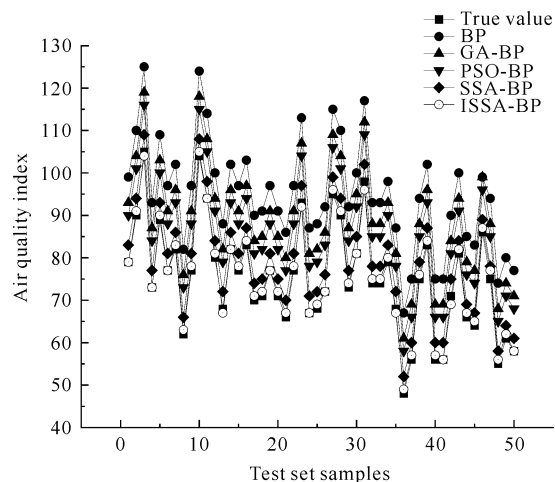


图 15 模型仿真预测结果

Fig. 15 Model simulation and prediction results

Percentage Error, MAPE)、拟合优度 (R^2) 作为模型优越性的衡量指标。计算公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (13)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (14)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|, \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (16)$$

式(13)-(16)中, \hat{y}_i 为测试集样本的预测值, y_i 为测试集样本的真实值, \bar{y} 为测试集样本平均值, n 为测试集样本个数。

利用训练好的模型测试 AQI, 得到 5 种预测模型性能的评价结果如表 5 所示。由表 5 可知, 经过相关算法优化后的组合模型的预测效果明显优于单一的 BP 神经网络模型。其中, 在组合模型中, ISSA-BP 模型的 RMSE、MAE、MAPE 数值最小。相比于 BP、GA-BP、PSO-BP、SSA-BP 模型, ISSA-BP 模型的 RMSE 分别降低了 70.8%、50.2%、45.2%、34.6%, MAE 分别降低了 74.6%、57.9%、54.2%、43.3%, MAPE 分别降低了 68.2%、53.1%、49.1%、25.2%。决定系数 R^2 为 0.99, 最接近 1, 拟合程度最优。因此, 本研究提出的 ISSA-BP 模型的预测精度最好, 预测误差最小, 能够有效地应用于空气质量指数的预测中。

表5 预测模型性能评价结果

Table 5 Performance evaluation results of prediction model

模型 Model	均方根 误差 (RMSE)	均值绝 对误差 (MAE)	平均绝对百 分比误差 (MAPE)	拟合 优度 (R^2)
BP	4.15	3.66	3.55	0.89
GA-BP	2.43	2.21	2.41	0.93
PSO-BP	2.21	2.03	2.22	0.94
SSA-BP	1.85	1.64	1.51	0.96
ISSA-BP	1.21	0.93	1.13	0.99

5 结束语

本研究提出了一种改进的麻雀搜索算法 (ISSA), 并基于 BP 神经网络构建了 AQI 预测模型 (ISSA-BP)。通过改进后的麻雀搜索算法来优化 BP 神经网络权值及阈值, 与未优化的 BP 神经网络相比, 显著提升了其预测精度, 可有效预测 AQI。所建立的 ISSA-BP 模型不仅解决了传统 BP 神经网络在预测 AQI 方面的不足, 更弥补了利用麻雀搜索算法在优化模型时的缺陷。下一步研究重点是将 ISSA-BP 模型应用到实际工程领域中, 以进一步验证其性能。

参考文献

- [1] 黄伟建, 李丹阳, 黄远. 面向空气质量的时空混合预测模型[J]. 计算机应用, 2020, 40(11): 3385-3392.
- [2] 刘炳春, 郑红梅, 张斌. 基于 IG-LASSO 模型的城市空气质量指数混合预测研究[J]. 环境科学与技术, 2017, 40(11): 144-148.
- [3] 杨云, 付彦丽. 关于空气中 PM_{2.5} 质量浓度预测研究[J]. 计算机仿真, 2016, 33(3): 413-418.
- [4] 杨丰玉, 王宝英, 陈英, 等. 一种考虑风力作用的 KNN 城市 AQI 预测算法[J]. 计算机应用研究, 2019, 36(6): 1679-1682, 1722.
- [5] 石晓文, 蒋洪迅. 面向高精度与强鲁棒的空气质量预测 LSTM 模型研究[J]. 统计与决策, 2019, 35(16): 49-53.
- [6] 王国胜, 郭联金, 董晓清, 等. 深圳市区空气污染的人工神经网络预测[J]. 环境工程学报, 2015, 9(7): 3393-3399.
- [7] 张旭, 杜景林. 改进 PSO-GA-BP 的 PM_{2.5} 浓度预测[J]. 计算机工程与设计, 2019, 40(6): 1718-1723.
- [8] 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[J]. 陕西科技大学学报(自然科学版), 2016, 34(4): 171-176, 186.
- [9] 李黄曼, 张勇, 张瑶. 基于 ISSA 优化 SVM 的变压器故障诊断研究[J]. 电子测量与仪器学报, 2021, 35(3): 123-129.
- [10] 毛清华, 张强. 融合柯西变异和反向学习的改进麻雀算法[J]. 计算机科学与探索, 2021, 15(6): 1155-1164.
- [11] 王海瑞, 鲜于建川. 改进麻雀搜索算法在分布式电源配置中的应用[J]. 计算机工程与应用, 2021, 57(20): 245-252.
- [12] LIU L F, SUN S Z, YU H Y, et al. A modified Fuzzy C-Means (FCM) clustering algorithm and its application on carbonate fluid identification [J]. Journal of Applied Geophysics, 2016, 129: 28-35.
- [13] 张伟康, 刘升, 任春慧. 混合策略改进的麻雀搜索算法[J]. 计算机工程与应用, 2021, 57(24): 74-82.
- [14] 汤安迪, 韩统, 徐登武, 等. 基于混沌麻雀搜索算法的无人机航迹规划方法[J]. 计算机应用, 2021, 41(7): 2128-2136.
- [15] 高文欣, 刘升, 肖子雅, 等. 柯西变异和自适应权重优化的蝴蝶算法[J]. 计算机工程与应用, 2020, 56(15): 43-50.
- [16] 蔡恒雨, 郑启龙. 深度学习节点分配的遗传算法设计[J]. 小型微型计算机系统, 2021, 42(6): 1158-1164.
- [17] 钱晓宇, 方伟. 基于局部搜索的反向学习竞争粒子群优化算法[J]. 控制与决策, 2021, 36(4): 779-789.
- [18] 王伟, 代红, 赵斯祺. 基于特征优化和 BP 神经网络的入侵检测方法[J]. 计算机工程与设计, 2021, 42(10): 2755-2761.
- [19] LIU Y Q, QIU D C, LI H L, et al. The intrusion detection module utilizing LE and modified PSO-BP [C]// IEEE BEIJING SECTION. Proceedings of 2017 IEEE 8th international conference on software engineering and service science. Beijing, China: Institute of Electrical and Electronics Engineers, 2017: 318-321.
- [20] 刘文杰, 邢彦锋, 陆瑶. 基于改进的 BP 神经网络胶铆接头的力学性能预测研究[J]. 农业装备与车辆工程, 2021, 59(3): 12-16.

Air Quality Index Prediction Based on Improved Sparrow Search Algorithm

HU Qing¹, GONG Shicai¹, HU Zhen²

(1. School of Science, Zhejiang University of Science and Technology, Hangzhou, Zhejiang, 310000, China; 2. School of Science, Hubei University of Technology, Wuhan, Hubei, 430068, China)

Abstract: In order to predict Air Quality Index (AQI) more accurately, an AQI prediction model (ISSA-BP) based on Improved Sparrow Search Algorithm (ISSA) is proposed. The global search performance of Sparrow Search Algorithm (SSA) is used to optimize the weight and threshold of BP neural network, which solves the problems of slow convergence speed and easy to fall into local optimum in the process of predicting AQI by traditional BP neural network. At the same time, aiming at the defects of SSA in the optimization process, the cubic mapping and optimization strategy are introduced to enhance the global search and convergence ability of the algorithm and further improve the prediction performance. ISSA-BP model was used to predict AQI in Hangzhou. The experimental results show that the prediction accuracy of the model is significantly improved compared with other models. This study provides a new prediction method for air pollution prevention and control.

Key words: air quality index prediction; sparrow search algorithm; chaotic mapping; optimization strategy; BP neural network

责任编辑: 梁 晓



微信公众号投稿更便捷

联系电话: 0771-2503923

邮箱: gxkx@gxas.cn

投稿系统网址: <http://gxkx.ijournal.cn/gxkx/ch>