

文章编号:1672-3961(2023)02-0000-00 DOI:10.6040/j.issn.1672-3961.0.2022.162

改进果蝇算法优化 BP 神经网络预测汽油辛烷值

韦修喜¹,陶道²,黄华娟^{1*}

(1.广西民族大学人工智能学院,广西 南宁 530006;2.广西民族大学电子信息学院,广西 南宁 530006)

摘要:针对 BP 神经网络存在预测精度不佳、预测结果不稳定的问题,提出改进果蝇算法优化 BP 神经网络(back propagation neural network)的预测模型。将混沌映射、判别因子与变步长机制引入果蝇优化算法(fruit fly optimization algorithm, FOA)中,得到改进后的自适应混沌果蝇优化算法(fruit fly optimization algorithm with chaos and discriminant factors, CDFOA),并利用测试函数对算法进行性能验证。利用 CDFOA 优化 BP 神经网络的初始权值与阈值,构建基于 CDFOA 优化 BP 神经网络对于汽油辛烷值的预测模型 CDFOA-BP。将采集到的 60 组汽油数据输入到预测模型中进行测试分析。预测结果表明,与 FOA-BP 模型、PSO-BP 模型、SSA-BP 模型和 BP 神经网络模型相比,CDFOA-BP 模型在预测精度与预测稳定性上均优于其他 4 种模型,验证该模型的有效性 with 可行性。

关键词:果蝇优化算法;混沌映射;判别因子;函数测试;BP 神经网络;辛烷值

中图分类号:TP183 **文献标志码:**A

引用格式:韦修喜,陶道,黄华娟.改进果蝇算法优化 BP 神经网络预测汽油辛烷值[J].山东大学学报(工学版),2023,53(2)

WEI Xiuxi, TAO Dao, HUANG Huajuan. Optimizing BP neural network to predict gasoline octane number by improved fruit fly algorithm [J]. Journal of Shandong University (Engineering Science), 2023, 53(2)

Optimizing BP neural network to predict gasoline octane number by improved fruit fly algorithm

WEI Xiuxi¹, TAO Dao², HUANG Huajuan^{1*}

(1. School of Artificial Intelligence, Guangxi Minzu University, Nanning 530006, Guangxi, China; 2. College of Electronic Information, Guangxi Minzu University, Nanning 530006, Guangxi, China)

Abstract: Aiming at the problems of poor prediction accuracy and unstable prediction results of back propagation (BP) neural network, a prediction model based on improved fruit fly algorithm and optimized BP neural network was proposed. Chaos mapping, discriminant factor and variable step size mechanism were introduced into the fruit fly optimization algorithm (FOA) to obtain an improved fruit fly optimization algorithm with chaos and discriminant factors (CDFOA), and the performance of the algorithm was verified by function test. The initial weights and thresholds of BP neural network were optimized by CDFOA, and the prediction model CDFOA-BP for gasoline octane number based on CDFOA optimized BP neural network was constructed. 60 groups of gasoline data collected were input into the prediction model for test and analysis. The prediction results showed that compared with FOA-BP model, PSO-BP model, SSA-BP model and BP neural network model, CDFOA-BP model was superior to the other four models in prediction accuracy and prediction stability, which verified the effectiveness and feasibility of the model.

Keywords: fruit fly optimization algorithm; chaotic mapping; discriminant factor; function test; BP neural network; octane number

收稿日期:2022-04-22

基金项目:国家自然科学基金项目(62266007,61662005);广西自然科学基金项目(2021GXNSFAA220068,2018GXNSFAA294068)

第一作者简介:韦修喜(1980—),男,广西百色人,副教授,硕士生导师,主要研究方向为机器学习。E-mail: weixiuxi@163.com

***通信作者简介:**黄华娟(1984—),女,广西崇左人,副教授,博士,计算机学会(CCF)会员(94058M),主要研究方向为机器学习和数据挖掘。

E-mail: hhj-025@163.com

0 引言

随着我国经济的快速发展,人民生活水平的提高,家用汽车逐渐走进了千家万户。汽车保有量的增加,使得我国对汽油的需求量也越来越大。在这样的大环境下,对汽油品质的把控也显得越来越重要,而辛烷值就是衡量汽油品质好坏的重要指标^[1]。传统的汽油辛烷值测定方法主要有两种,实验室法测定辛烷值和行车法测定辛烷值。传统的实验室检测方法存在样品用量大、测试周期长和费用高等问题^[2],不适用于生产控制,特别是在线测试。近红外光谱分析方法(near infra-red, NIR),作为一种快速分析方法,已经广泛应用于农业、制药、生物化工、石油产品等领域^[3],其优越性是无损检测、低成本、无污染、能在线分析,更适合于生产和控制的需要。目前,已经有许多学者提出了新式的汽油辛烷值预测方法。文献[4]采用增压技术在低压缺氧条件下测量汽油的辛烷值,但这种方法的局限性明显且成本较高,不适合推广应用。文献[5]通过变量加权模块捕获汽油精制过程中变量数据的相关性,针对不同变量对汽油辛烷值的重要程度获取其对应的权重,提出了一种基于自适应变量加权的神经网络预测方法。该方法在一定程度上简化了预测变量的个数,提高了预测效率。但是,该算法在预测过程中需要构建复杂的加权模型,耗费时间且不利于广泛应用。文献[6]提出了一种基于随机森林回归算法的研究法辛烷值(research octane number, RON)预测方法。虽然随机森林算法具有非常优秀的预测性能,但该算法的两个随机^[7]也会导致预测结果的不确定性。

作为一种成熟的回归预测模型,BP神经网络具有优秀的非线性映射能力。其神经网络能够自适应建模预测数据,并且具有较高的预测精度,能够有效节约时间,因此有着广泛的应用。此外,作为一种新兴的优化方法,群智能算法已经受到了越来越多学者的关注,其具有智能性和并行式搜索的主要特点,同时具有渐近式寻优、有指导的随机搜索、全局最优解、易于与其他算法结合等基本特性。作为群智能算法之一的果蝇算法,也具备上述优点。本研究针对现有检测方法存在的不足,提出了一种改进果蝇算法优化BP神经网络在汽油辛烷值中的预测模型。

1 基本果蝇优化算法

果蝇优化算法(fruit fly optimization algorithm, FOA)是文献[8]受果蝇觅食行为启发于2012年提出的一种群体智能全局优化算法。果蝇群体在觅食的过程中,主要依靠其敏锐的嗅觉发现食物。当群体中的某一果蝇个体发现食物源时,其它个体便会利用视觉搜索,确定最优个体的位置。同时,其它果蝇个体朝着最优个体的位置靠近。如此反复,当果蝇优化算法达到最大迭代次数时便结束搜寻。基本果蝇优化算法的主要步骤如下。

步骤 1 果蝇种群位置的随机初始化,其中 X 、 Y 均为向量。

$$\text{Init } X_i, \text{Init } Y_i$$

步骤 2 果蝇个体利用嗅觉随机搜索食物,其中 R 为随机数。

$$\begin{cases} X_{i+1} = X_i + R \\ Y_{i+1} = Y_i + R \end{cases} \quad (1)$$

步骤 3 计算果蝇个体与原点之间的距离并求其倒数作为味道浓度判定值

$$\begin{aligned} D_i &= \sqrt{X_i^2 + Y_i^2}, \\ S_i &= 1/D_i. \end{aligned}$$

步骤 4 计算果蝇的味道浓度值

$$W_i = \text{Fitness}(S_i).$$

步骤 5 找出果蝇群体中味道浓度最佳的个体,其中 B 为最佳浓度值, I 为最佳浓度值果蝇下标。

$$[B \ I] = \min W.$$

步骤 6 保留最佳味道浓度值与坐标,果蝇群体利用目视搜索飞往该位置

$$Z = B,$$

$$\begin{cases} X_i = X_I \\ Y_i = Y_I \end{cases}.$$

步骤 7 进入迭代寻优,循环执行步骤 2~6;达到最大迭代次数,返回最优果蝇个体。

2 改进的果蝇优化算法

2.1 自适应混沌果蝇算法

虽然传统的果蝇优化算法具有参数设置少、运行速度快等优点,但由于 FOA 算法采用随机步长机制来进行迭代寻优,导致算法具有一定的盲目性与不稳定性等问题。近些年来,已有许多学者提出了相关的改进算法,文献[9]等提出了双重驱动的果

蝇优化算法,文献[10]提出了自适应果蝇优化算法,文献[11]提出了动态变步长果蝇算法,文献[12]提出了双策略协同进化果蝇优化算法,文献[13]提出了混合果蝇优化算法。

上述改进措施主要侧重于对算法寻优精度的提升,而很少涉略到对其时间复杂度的改善。因此,本研究从以下3个方面进行改进优化,使得改进的自适应混沌果蝇算法(fruit fly optimization algorithm with chaos and discriminant factors, CDFOA)在具有较高寻优精度的同时也具有较快的收敛速度。

(1)原算法采用随机步长进行迭代寻优,由于随机步长机制的局限性,导致算法产生解的多样性不高。为此,引入Cubic混沌映射来调节果蝇群体在迭代过程中的搜索步长,利用Cubic映射的混沌性来增加果蝇种群的多样性,以此来提高算法的寻优能力与稳定性。

(2)为了使改进算法能够具更佳的时间性能,减少算法不必要的时间损失,加快算法的收敛速度。为此,加入判别因子来实现算法收敛的自适应判断,进一步提高算法的寻优性能。

(3)由于原算法在迭代寻优过程中采用随机步长机制,导致算法的寻优过程具有盲目性与无序性,致使算法的寻优效率低下,寻优精度不高。为此,引入变步长机制替换原算法中的随机步长机制,利用log对数的单调递增性使得算法的寻优过程更加高效有序,从而进一步强化改进算法的寻优性能。

2.2 Cubic 映射

针对算法在迭代过程中,采用随机步长机制容易造成算法的性能不稳定与寻优过程的盲目性问题,本研究采用的解决方案是引入最常见的Cubic混沌映射,动态调节果蝇群体的搜索步长,从而使得优化后的果蝇位置具有混沌现象随机性、遍历性和规律性的特点^[14-15]。

标准的Cubic混沌映射函数的可以表示为:

$$f_{n+1} = bf_n^3 - cf_n,$$

式中: b 、 c 均为混沌调节参数,影响着Cubic映射分布的均匀程度与范围。一般 $c \in (2.3, 3)$ 时,Cubic映射产生的序列为混沌状态。此外, $b = 1$, $f_n \in (-2, 2)$; $b = 4$, $f_n \in (-1, 1)$ 。

将标准的Cubic混沌映射进行推广,得到一般的Cubic混沌映射表达式,其系统方程为:

$$f_{n+1} = \rho f_n (1 - f_n^2),$$

式中: ρ 为映射的控制参数,Cubic映射的混沌性与参数 ρ 的取值有着很大的关系^[16]。本研究选取 $\rho = 0.5$ 作为此映射的控制参数。然后,将Cubic混沌映

射作为一个调节参数 α ,加入到式(1)为:

$$\begin{cases} X_{i+1} = X_i + \alpha R \\ Y_{i+1} = Y_i + \alpha R^\circ \end{cases} \quad (2)$$

2.3 判别因子

FOA算法在果蝇寻优阶段主要依据算法所设置的迭代次数来判断是否结束寻优过程。当算法在迭代次数之前就已经寻到最优值时,接下来的寻优过程即可停止。因此,在FOA的迭代寻优过程中加入判别因子 σ ,可减少算法的时间复杂度,提高算法的寻优性能。算法改进部分的伪码如下。

算法

```
for G=1 to g
  for Γ=1 to k
    α=chaos G
    XΓG=X-axis+α Randvalue
    YΓG=Y-axis+α Randvalue
    DistΓ=sqrt(XΓ2+YΓ2)
    SΓ=1/DistΓ
    Smell(Γ)=Fitness(SΓ)
    If Γ ≥ 2 && abs((Smell(Γ)-Smell(Γ-1))) ≤ σ
      break;
    end
  end
end
```

注: G 表示种群的最大迭代次数, Γ 表示果蝇的种群数量, σ 为 10^{-10} 。

2.4 变步长机制

由于FOA算法在寻优过程中采用随机步长机制,容易造成果蝇种群的多样性不高,使得算法的寻优能力不强。因此,在结合前面两种改进的基础上,将lg对数变步长机制引入式(2)中来提高种群的多样性,增强算法的全局搜索能力与稳定性。改进表达式为:

$$\begin{cases} X_{i+1} = X_i + \alpha \lg(10R) \\ Y_{i+1} = Y_i + \alpha \lg(10R)^\circ \end{cases}$$

式中 α 为混沌调节参数。

3 函数测试

3.1 参数设置与试验评价标准

为了验证改进算法的有效性,检验CDFOA算法在寻优精度、收敛速度与稳定性方面的性能表现,选取6个高维多峰函数分别命名为 F_1 、 F_2 、 F_3 、 F_4 、 F_5 、 F_6 ,进行最小值试验。同时,采用FOA算法、

花授粉算法 (flower pollination algorithm, FPA)、飞 鸡群优化算法 (chicken swarm optimization, CSO)
蛾扑火优化算法 (moth-flame optimization, MFO) 与 进行对比测试,测试函数如表 1 所示。

表 1 测试函数及参数的设定
Table 1 Test functions and parameter settings

函数名称	表达式	搜索空间	维度	峰值	最小值
Griewank	$F_1(x) = \sum_{i=1}^d \frac{x_i^2}{4000} - \prod_{i=1}^d \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	$[-600, 600]$	30	多峰	0
Nocontinous-Rastrigin	$F_2(x) = F_2(y) \ (y_i = \{x_i, x_i < 0.5; \text{round}(2x_i/2, x_i \geq 0.5)\})$	$[-5.12, 5.12]$	30	多峰	0
Drop-Wave	$F_3(x) = -\frac{1 + \cos(12\sqrt{x_1^2 + x_2^2})}{0.5(x_1^2 + x_2^2) + 2}$	$[-5.12, 5.12]$	30	多峰	-1
Elliptic UN	$F_4(x) = \sum_{i=1}^d (10^{\frac{6(i-1)}{d-1}}) x_i^2$	$[-100, 100]$	30	多峰	0
Rastring	$F_5(x) = 10d + \sum_{i=1}^d [x_i^2 - 10\cos(2\pi x_i)]$	$[-5.12, 5.12]$	30	多峰	0
Schaffer N.2	$F_6(x) = 0.5 + \frac{\sin^2(x_1^2 - x_2^2) - 0.5}{[1 + 0.001(x_1^2 + x_2^2)]^2}$	$[-100, 100]$	30	多峰	0

试验中各算法种群数量均为 50，独立运行 30 次,设置终止条件为迭代次数达到 500 次。为评价算法优化效果,给出如下判定标准:优化均值 m 为算法运行 30 次后得到的最优值的期望,用来衡量算法寻优平均质量;标准差 s 为算法运行 30 次后得到的最优值与平均最优值之间的标准差,评价算法寻优稳定性;全局最优解 z 为算法运行 30 次得到的全局最优解;全局最差解 w 为算法运行 30 次得到的全局最差解;优化平均耗时 t 为算法 30 次运行时间的平均值,用于评价算法的收敛时间,单位为 s。

3.2 试验结果与分析

对表 1 中的 6 个标准测试函数,各算法分别独立进行 30 次测试。对比算法依次为 CDFOA、FOA、FPA、MFO 与 CSO,测试对比结果见表 2。

表 2 函数测试结果
Table 2 Test functions results

测试函数	评价指标	CDFOA	FOA	FPA	MFO	CSO
F_1	z	0	7.25×10^{-12}	1.85×10^1	5.26×10^{-1}	0
	w	0	7.33×10^{-12}	3.76×10^1	1.81×10^2	3.70×10^{-7}
	m	0	7.28×10^{-12}	2.75×10^1	2.49×10^1	7.40×10^{-9}
	s	0	1.69×10^{-14}	5.47×10^0	5.25×10^1	5.23×10^{-8}
	t/s	4.61×10^{-4}	5.31×10^{-3}	1.36×10^{-2}	5.57×10^{-3}	3.48×10^{-1}
F_2	z	0	4.79×10^{-4}	1.17×10^2	4.43×10^1	0
	w	0	4.85×10^{-4}	2.05×10^2	2.22×10^2	3.66×10^1
	m	0	4.83×10^{-4}	1.56×10^2	1.17×10^2	7.32×10^{-1}
	s	0	1.39×10^{-6}	1.87×10^1	4.80×10^1	5.18×10^0
	t/s	7.24×10^{-4}	4.96×10^{-3}	1.29×10^{-2}	4.72×10^{-3}	3.38×10^{-1}
F_3	z	-1	-9.36×10^{-1}	-9.99×10^{-1}	-1	-1
	w	-1	-9.36×10^{-1}	-9.36×10^{-1}	-9.36×10^{-1}	0
	m	-1	-9.36×10^{-1}	-9.77×10^{-1}	-9.85×10^{-1}	-2.00×10^{-2}
	s	0	2.01×10^{-6}	2.49×10^{-2}	2.74×10^{-2}	1.41×10^{-1}
	t/s	6.66×10^{-4}	5.65×10^{-3}	1.36×10^{-2}	5.48×10^{-3}	3.55×10^{-1}

表 2(续)

测试函数	评价指标	CDFOA	FOA	FPA	MFO	CSO
F_4	z	2.32×10^{-19}	2.74×10^{-4}	1.77×10^7	3.89×10^5	0
	w	2.54×10^{-19}	2.80×10^{-4}	4.49×10^7	9.58×10^7	1.97×10^{-18}
	m	2.45×10^{-19}	2.77×10^{-4}	2.86×10^7	2.98×10^7	3.95×10^{-20}
	s	5.01×10^{-21}	1.14×10^{-6}	6.94×10^6	2.81×10^7	2.79×10^{-19}
	t/s	1.05×10^{-3}	8.56×10^{-3}	1.63×10^{-2}	7.91×10^{-3}	4.46×10^{-1}
F_5	z	0	4.79×10^{-4}	1.18×10^2	7.56×10^1	0
	w	0	4.85×10^{-4}	1.98×10^2	2.08×10^2	0
	m	0	4.82×10^{-4}	1.60×10^2	1.51×10^2	0
	s	0	1.55×10^{-6}	1.92×10^1	3.44×10^1	0
	t/s	7.21×10^{-4}	4.91×10^{-3}	1.34×10^{-2}	4.77×10^{-3}	3.27×10^{-1}
F_6	z	0	1.56×10^{-13}	4.07×10^{-7}	0	0
	w	0	1.77×10^{-13}	6.63×10^{-3}	0	0
	m	0	1.59×10^{-13}	1.51×10^{-3}	0	0
	s	0	3.45×10^{-15}	1.78×10^{-3}	0	0
	t/s	5.32×10^{-4}	5.70×10^{-3}	1.35×10^{-2}	4.82×10^{-3}	3.42×10^{-1}

由表 2 可知,CDFOA 各项指标均优于 FOA。FOA 的方差值具有较大的起伏,表明原算法的稳定性较差。相较而言,CDFOA 的方差值则波动较小,证明了改进算法具有较好的稳定性。在寻优精度方面,CDFOA 算法均优于 FOA。除函数 F_4 外,其余测试函数均直接寻到最优值。在 6 组测试指标中,CDFOA 的性能基本排在第一位。虽然 MFO 与 CSO 在 F_5 、 F_6 的测试中也表现出了极佳的精确度与稳定性,但算法的收敛时间均低于 CDFOA 算法 1 个数量级以上。由此可见,改进策略对原算法在寻优精度、收敛速度与稳定性方面都得到了提升,表明了算法改进的有效性与可行性。

4 CDFOA-BP 模型的构建

4.1 BP 神经网络

由于 BP 神经网络(back propagation neural network)优秀的非线性映射能力,在进行数据预测时无需建立精准的数学模型,且具有较高的预测精度,因此在数据预测方面有着较为广泛的应用^[17-20]。基本的 BP 神经网络主要包括信号的前向传播和误差的反向传播两个过程。即计算误差输出时,按从输入层、隐藏层到输出的方向进行,将得到的预测结果误差作为阈值反向回馈给神经网络,神经网络再根据误差来动态调整神经网络的各项权值。具体的工作原理如图 1 所示。

由图 1 可知,BP 神经网络的初始权值与阈值都是随机产生的,利用梯度下降的方法进行不断修正。因此,BP 神经网络性能的好坏非常依赖于初始

权值,这也导致了 BP 神经网络预测性能的不稳定与收敛速度慢的问题。针对 BP 神经网络存在的这些不足,引入 CDFOA 算法优化神经网络的初始权值与阈值,改善网络的上述情况,提高 BP 神经网络的预测准确度。

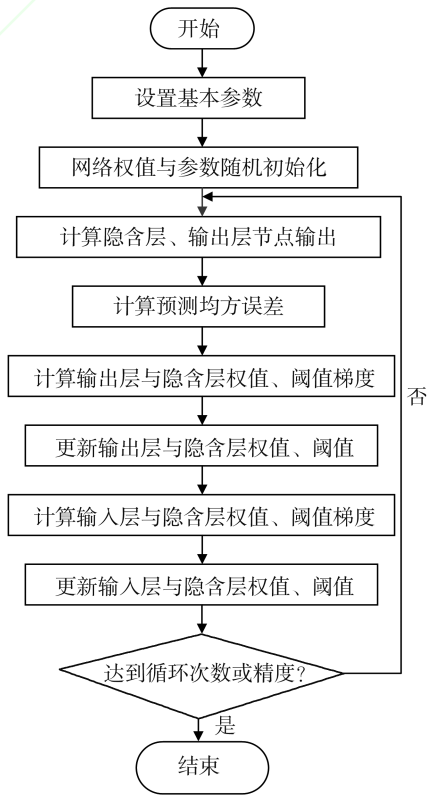


图 1 BP 神经网络流程图
Fig.1 Flow chart of BP neural network

4.2 CDFOA-BP 算法的构建

CDFOA 优化 BP 神经网络的主要思路是利用

改进果蝇算法的全局寻优能力优化 BP 神经网络的学习能力,将原本随机初始化的权值与阈值,使用果蝇个体在寻优空间中的位置来替代,提高权值与阈值产生的多样性。寻找最优的权值与阈值初始化 BP 神经网络,提高网络预测的准确性与稳定性。

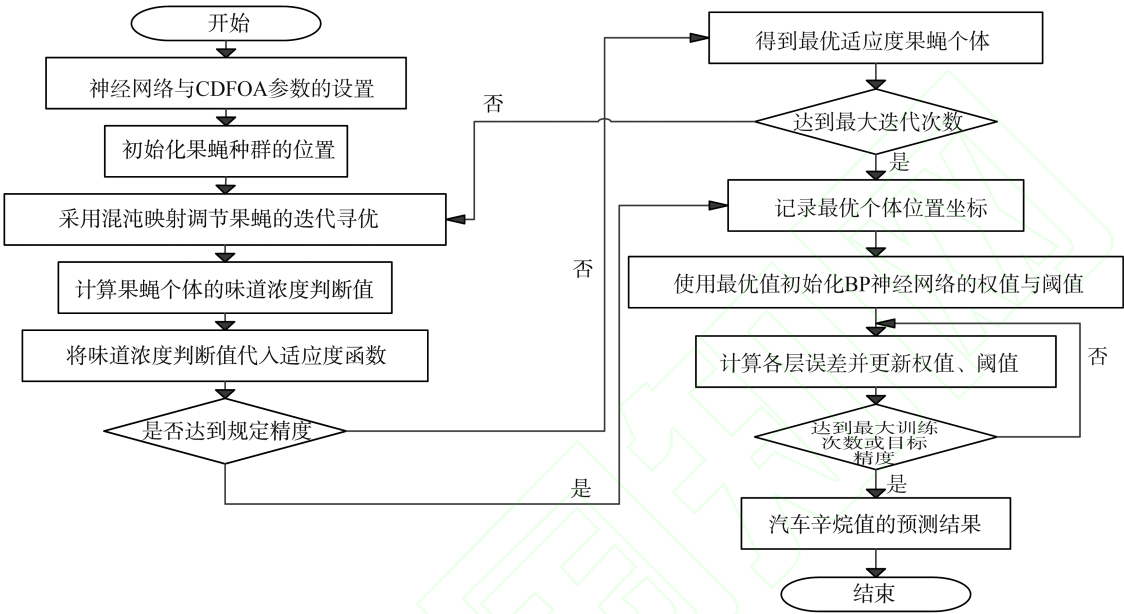


图 2 CDFOA-BP 算法流程图
Fig.2 CDFOA-BP algorithm flow chart

4.3 CDFOA-BP 算法复杂度分析

对于 BP 神经网络,假设输入层、隐含层与输出层的节点数量分别为 k,h,j 。根据 BP 神经网络的映射规则,则它的时间复杂度为 $o(k \times h \times j)$ 。

由 2.3 节的算法伪码可以求出 CDFOA 算法的时间复杂度,其算法的时间复杂度主要由迭代次数、种群数量与适应度函数 3 个要素共同决定。基本 FOA 算法的时间复杂度为 $o(\Gamma \times G \times f)$,其中 Γ 为种群数量、 G 为迭代次数、 f 为适应度函数。相较于 FOA 算法,CDFOA 算法的时间复杂度改变主要集中在算法的改进部分,即迭代过程。通过引进判别因子来判断算法在寻优过程中的收敛状态,避免不必要的算法性能损失。在 CDFOA 算法中,由于判别因子的介入,使得优化后的种群数量 Γ' 必定小于或等于 Γ 。因此,CDFOA 算法的时间复杂度 $o(\Gamma' \times G \times f)$ 也必定小于或等于 $o(\Gamma \times G \times f)$ 。

根据上述推论,将各算法的适应度函数时间复杂度替换为 BP 神经网络的时间复杂度,则改进后的 CDFOA-BP 算法的时间复杂度 $o(k \times h \times j \times \Gamma' \times G)$ 必定小于等于 FOA-BP 算法的时间复杂度 $o(k \times h \times j \times \Gamma \times G)$ 。因此,CDFOA-BP 算法的时间复杂度相较于其余原算法,有一定的减少,有助于算法性

能的提升。首先将果蝇个体视作神经网络中的权值与阈值,神经网络的预测误差绝对值之和作为 CDFOA 算法的适应度函数,其次利用 CDFOA 迭代寻优取代 BP 神经网络的梯度修正下降法,最后利用适应度最佳的果蝇个体初始化神经网络的权值和阈值,基本流程如图 2 所示。

能的提升。

5 试验分析

5.1 数据预处理

试验样品由 60 个汽油样品组成,试验数据主要包括 401 个波长的特征值,由傅里叶近红外光谱分析(near infra-red, NIR)得到。数据的采集过程中,在 900~1700 nm 波长范围内,间隔 2 nm 取一个波长点作为特征值,共 401 个特征值,通过特征值预测辛烷值。

采用模型预测汽油辛烷值时,将 60 组数据按照 5:1 的比例划分为训练集与测试集,即前 50 组为训练集,后 10 组为测试集。由于输入数据的单位不一样,导致数据范围差别很大,神经网络收敛慢,训练时间长。此外,数据范围大的输入作用显得更大,数据范围小的输入作用就会显得不明显。基于以上两点原因,对所有输入数据 $[-1,1]$ 归一化:

$$y=2\frac{x-x_{\min}}{x_{\max}-x_{\min}}-1,$$

(3)

式中: x 为样本数据, x_{\max} 为样本数据的最大值, x_{\min} 为样本数据的最小值, y 为归一化后的数据。

5.2 模型参数设置与试验结果

CDFOA-BP 模型中的 BP 神经网络采用 3 层结构,其中输入层的节点数为 401 个,分别代表对应输入数据的 401 个波长的特征值;输出层的节点数为 1,对应汽油辛烷值的预测值。隐含层的节点个数,一般由经验公式推得。需要特别说明的是,经验公式推出的隐含层节点数只能作为参考值,并不一定是最佳值。根据常见的几种经验公式,确定隐含层的取值为

$$q=\log_2^u,$$
 (4)

式中: u 为输入层节点的个数, q 为隐含层节点的个数。

本研究在经验值为 9 的情况下进行进一步的调试,最终确定隐含层节点数为 5,BP 神经网络的预测性能相对更优。

采用 CDFOA-BP 模型对汽油辛烷值进行预测,同时使用 FOA-BP、PSO-BP、SSA-BP 与 BP 预测模型进行对比试验,以此检验改进模型的预测性能。此外,本试验所用到的对比算法,粒子群算法 (particle swarm optimization, PSO)与麻雀搜索算法 (sparrow search algorithm, SSA)的时间复杂度也均为 $o(I\times G\times f)$ 。5 种模型的参数设置为:迭代次数为 30 代,种群数量为 10,学习速率为 0.1,目标精度为 0.000 01。

为了验证改进 FOA 算法的有效性,本研究对 CDFOA 算法与 FOA 算法的适应度曲线进行对比,适应度曲线的迭代过程如图 3 所示。

由图 3 可知,FOA 算法的最终适应度值为 0.93,而 CDFOA 算法的最终适应度值为 0.68。

相比于原算法,改进算法具有更佳的寻优性能。FOA 算法迭代到 15 代左右时就陷入了局部最优值,CDFOA 则迭代到 25 代以后才收敛,表明改进算法的全局寻优能力更强。因此,CDFOA 算法无论是在寻优精度,还是在全局搜索能力上都显著优于原算法,验证了改进算法的有效性。

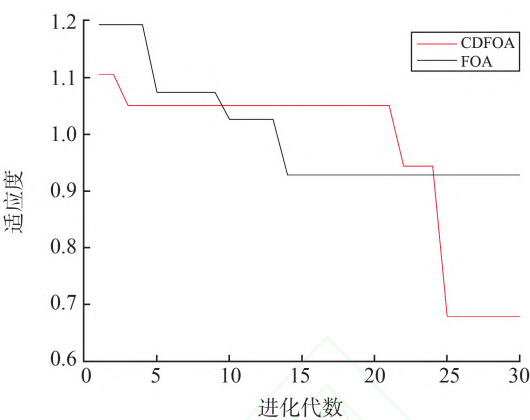


图3 CDFOA 与 FOA 的适应度曲线
Fig.3 The fitness curve of CDFOA and FOA

由于机器学习模型普遍存在过拟合的现象,因此本研究将针对改进模型的拟合度进行分析,CDFOA-BP 模型的学习曲线如图 4 所示。

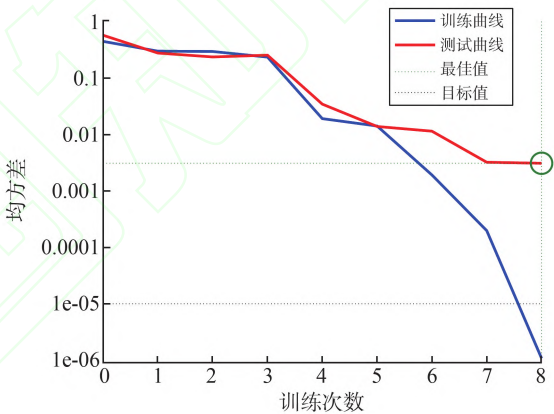


图4 CDFOA-BP 的训练曲线
Fig.4 The training curve of CDFOA-BP

由图 4 可知,CDFOA-BP 模型的训练误差曲线随着训练次数而下降,同时测试 (泛化) 曲线也随之下降,这表明训练后的模型对测试数据集具有较好的预测准确度,模型未出现过拟合现象。

使用数据集对改进模型与其它 4 种模型分别进行测试,预测结果如表 3 所示,预测误差曲线如图 5 所示。

表 3 5 种模型预测结果对比
Table 3 Comparison of prediction results of 5 models

模型预测值	样本 1	样本 2	样本 3	样本 4	样本 5	样本 6	样本 7	样本 8	样本 9	样本 10
实际辛烷值	88.100 0	87.600 0	88.350 0	85.100 0	85.100 0	84.700 0	87.200 0	86.600 0	89.600 0	87.100 0
CDFOA-BP 预测值	87.792 0	87.074 3	88.479 8	85.519 6	85.824 7	84.460 8	87.424 8	86.545 3	89.244 6	87.488 5
FOA-BP 预测值	87.570 4	87.460 3	90.338 9	84.804 5	84.963 8	84.341 0	86.833 7	85.833 3	88.399 6	86.600 4
PSO-BP 预测值	87.464 8	87.464 8	87.464 8	86.349 7	86.349 7	84.976 6	87.464 8	87.464 8	88.532 4	87.464 8
SSA-BP 预测值	88.833 3	88.578 2	88.833 3	87.372 6	87.502 5	87.464 2	88.244 6	88.592 9	88.832 1	88.833 3
BP 预测值	89.656 3	88.111 8	89.579 6	86.836 7	86.680 7	85.420 9	89.095 3	87.787 6	90.521 9	88.961 7

由表 3 可知,相较于 FOA-BP 模型、PSO-BP 模型、SSA-BP 神经网络与 BP 神经网络,使用 CDFOA

优化后的 BP 神经网络预测误差的绝对值之和最小,这表明 CDFOA 算法提高 BP 神经网络的预测

精度。

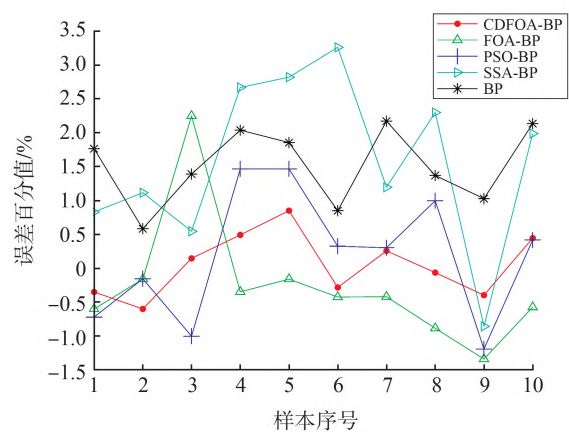


图 5 预测误差曲线
Fig.5 Prediction error curves

由图 5 的预测误差曲线可知,使用 BP 神经网络进行预测时,其预测曲线存在较大波动,这表明原 BP 神经网络的预测性能不稳定。相比之下,CDFOA-BP 模型预测曲线则较为平缓,这表明 CDFOA 算法提高了 BP 神经网络的预测稳定性。

5.3 预测结果分析

为进一步评价辛烷值预测模型的性能,选取均方误差 E 、平均绝对误差 A 、平均绝对百分比误差 P 作为评价指标,其计算公式为:

$$E=\frac{1}{n}\sum_{\gamma=1}^n(y_{\gamma}-y_{\gamma}^*)^2, \tag{5}$$

$$A=\frac{\sum_{\gamma=1}^n|y_{\gamma}-y_{\gamma}^*|}{n}, \tag{6}$$

$$P=\frac{100\%}{n}\sum_{\gamma=1}^n\left|\frac{y_{\gamma}^*-y_{\gamma}}{y_{\gamma}}\right|, \tag{7}$$

式中: n 表示预测样本的个数, y_{γ} 表示第 γ 个样本的真实值, y_{γ}^* 表示第 γ 个样本的预测值。

利用式(5)~(7)对表 3 中的 5 种预测模型的预测结果进行计算分析,得到 e_E 、 e_A 、 e_P 3 组指标数据,其中 e_E 用于验证模型预测误差平方的期望值, e_A 用于验证模型的预测精准度, e_P 用于验证模型的平均偏离程度,其结果如表 4 所示。

表 4 5 种预测模型性能对比

预测模型	e_E	$e_P/\%$	e_A
CDFOA-BP	0.147 70	0.39	0.337 0
FOA-BP	0.690 30	0.72	0.628 2
PSO-BP	0.649 60	0.81	0.699 4
SSA-BP	3.941 75	2.04	1.760 0
BP	1.954 90	1.52	1.320 3

由表 4 可知,CDFOA-BP 模型预测精度优于 FOA-BP 模型, e_E 值由 0.690 3 下降到了 0.147 7, e_P 值由 0.72% 降到 0.39%, e_E 值与 e_P 值分别提高 0.542 6 和 0.33%。而且,这两种优化模型均优于 PSO-BP、SSA-BP 和 BP 模型,体现了 FOA 算法对于 BP 神经网络优化的可行性以及 CDFOA-BP 模型对于汽油辛烷值预测的高准确性。

从模型的稳定性角度分析,BP 神经网络的预测机制决定了其预测性能具有不稳定性,在使用 FOA 算法进行优化后,BP 神经网络的预测稳定性有了一定幅度的提升。 e_A 由原来的 1.320 3 下降到了 0.628 2,而且综合性能明显优于 SSA-BP 模型与 BP 模型。

由于 FOA 算法采用随机步长机制进行迭代寻优,导致算法的寻优过程具有一定的盲目性与不稳定性。尤其是面对高维度问题时,其寻优效果往往不理想,寻优精度不佳、易陷入局部最优。因此,FOA-BP 模型相较于 PSO-BP 模型并没有多大的优势。针对 FOA 算法存在的不足之处,分别采用 Cubic 混沌映射、判别因子与变步长机制对 FOA 算法进行改进,并将其应用于 BP 神经网络之中。改进后的 CDFOA-BP 模型相较于其他 4 种模型,无论是在预测精度还是稳定性方面均为最佳。综上所述,CDFOA-BP 模型能够实现对汽油辛烷值的准确预测,且预测模型具有较高的稳定性。

6 结语

为了解决传统 BP 模型预测精度不高、预测性能不稳定的问题,本研究使用 CDFOA 算法优化 BP 神经网络的初始权值与阈值,提出了基于 CDFOA-BP 模型的汽油辛烷值预测模型,具体结论如下。

(1)针对 FOA 算法采用随机步长进行迭代寻优,导致其寻优效率不高的问题,引入混沌映射与变步长机制,以此来改善果蝇在迭代寻优过程中的盲目性与无序性。为了避免算法不必要的时间性能损失,引入判别因子实现算法的自适应收敛。随后,采用 6 个高维多峰函数检测算法改进的有效性。测试结果表明,CDFOA 算法具有更高的寻优精度与更快的收敛速度,证明了改进策略的有效性。

(2)将 CDFOA 与 BP 神经网络进行结合,构建起基于 CDFOA-BP 的汽油辛烷值预测模型。随后,将相关数据输入模型并与其它模型进行对比测试。试验结果表明,CDFOA-BP 模型在汽油辛烷值的预测中具有相对更佳的预测精度和模型稳定性,可以

满足工程的实际应用。

本研究提出的 CDFOA-BP 模型在一定程度上改善了 BP 神经网络所存在的不足之处,为相关人员提供了一种新的汽油辛烷值检测方法。

参考文献:

- [1] TIAGO D, RODOLFO O, PEDRO S, et al. Predictive analytics in the petrochemical industry: research octane number (RON) forecasting and analysis in an industrial catalytic reforming unit[J]. Computers and Chemical Engineering, 2020, 139: 106912.
- [2] 丁怡曼,薛晓康,范宾,等. 基于 PLS-红外光谱的汽油辛烷值测定方法研究[J]. 化学研究与应用, 2021, 33(5): 863-867.
- DING Yiman, XUE Xiaokang, FAN Bin, et al. Determination of gasoline octane number based on PLS-Infrared spectroscopy[J]. Chemical Research and Application, 2021, 33(5): 863-867.
- [3] 褚小立,陈瀑,李敬岩,等. 近红外光谱分析技术的最新进展与展望[J]. 分析测试学报, 2020, 39(10): 1181-1188.
- CHU Xiaoli, CHEN Pu, LI Jingyan, et al. Progresses and perspectives of near infrared spectroscopy analytical technology[J]. Journal of Instrumental Analysis, 2020, 39(10): 1181-1188.
- [4] 夏良. 低压环境汽油辛烷值测定技术研究[J]. 仪器仪表用户, 2021, 28(9): 6-9.
- XIA Liang. Study on determination of octane number of gasoline in low pressure environment[J]. Instrumentation, 2021, 28(9): 6-9.
- [5] 赵林,李希,谢永芳,等. 基于自适应变量加权的汽油辛烷值预测方法[J]. 控制与决策, 2022, 37(10): 2738-2744.
- ZHAO Lin, LI Xi, XIE Yongfang, et al. Prediction method of gasoline octane number based on adaptive variable weighting[J]. Control and Decision, 2022, 37(10): 2738-2744.
- [6] 郑斌,孙洪霞,王维民. 基于随机森林回归的汽油研究法辛烷值预测[J]. 石油炼制与化工, 2020, 51(12): 69-75.
- ZHENG bin, SUN Hongxia, WANG Weimin. Prediction of gasoline research octane number based on random forest regression[J]. Petroleum Processing and Petrochemical, 2020, 51(12): 69-75.
- [7] 杨丰瑞,罗思烦,李前洋. 基于特征选择的极限随机森林算法研究[J]. 计算机应用研究, 2020, 37(9): 2625-2628.
- YANG Fengrui, LUO Sifan, LI Qianyang. Research on extreme random forest algorithm based on feature selection [J]. Application Research of Computers, 2020, 37(9): 2625-2628.
- [8] PAN W T. A new fruit fly optimization algorithm: taking the financial distress model as an example [J]. Knowledge-Based Systems, 2012, 26(2): 69-74.
- [9] 凤丽洲,王友卫,韩琳琳,等. 双重驱动的果蝇优化算法及其在 PID 控制器中的应用[J]. 控制与决策, 2021, 36(9): 2225-2233.
- FENG Lizhou, WANG Youwei, HAN Linlin, et al. Double drive fruit fly optimization algorithm and its application in PID controller[J]. Control and Decision, 2021, 36(9): 2225-2233.
- [10] 黄小莉,陈静娴,胡思宇. 基于自适应果蝇优化算法的 K-means 聚类[J]. 国外电子测量技术, 2021, 40(6): 14-20.
- HUANG Xiaoli, CHEN Jingxian, HU Siyu. K-means clustering based on adaptive fruit fly optimization algorithm[J]. Foreign Electronic Measurement Technology, 2021, 40(6): 14-20.
- [11] 谭晶晶. 基于动态变步长果蝇算法优化支持向量机的轴承故障诊断方法[J]. 机械设计与研究, 2021, 37(1): 102-105.
- TAN Jingjing. A fault diagnosis method based on svm optimized by improved fruit fly optimization algorithm [J]. Machine Design and Research, 2021, 37(1): 102-105.
- [12] 石建平,刘国平,李培生,等. 双策略协同进化果蝇优化算法及其应用[J]. 计算机集成制造系统, 2022, 28(5): 1482-1495.
- SHI Jianping, LIU Guoping, LI Peisheng, et al. Double strategies co-evolutionary fruit fly optimization algorithm and its application [J]. Computer Integrated Manufacturing Systems, 2022, 28(5): 1482-1495.
- [13] 曾宇容,王林,王思睿. 基于混合果蝇优化算法的选址-库存-配送集成优化研究[J]. 运筹与管理, 2022, 31(3): 24-30.
- ZENG Yurong, WANG Lin, WANG Sirui. Integrated optimization of location-inventory-delivery problem using hybrid fruit fly optimization algorithm [J]. Operations Research and Management Science, 2022, 31(3): 24-30.
- [14] FAN Yibo, JIN Yize, MENG Zihao, et al. Pixels and panoramas: an enhanced cubic mapping scheme for video/image-based virtual-reality scenes[J]. IEEE Consumer Electronics Magazine, 2019, 8(2): 44-49.
- [15] 吴远超,舒乃秋,关向雨. 基于 LSSVM 和混沌理论改进人工蜂群优化算法的 GIS 母线接头温度预测[J]. 电测与仪表, 2018, 55(20): 8-14.
- WU Yuanchao, SHU Naiqiu, GUAN Xiangyu. Temperature prediction method of GIS bus connector based on

improved artificial bee colony optimization LSSVM algorithm[J]. Electrical Measurement & Instrumentation, 2018, 55(20): 8-14.

[16] 张孟健,张浩,陈曦,等. 基于 Cubic 映射的灰狼优化算法及应用[J]. 计算机工程与科学, 2021, 43(11): 2035-2042.

ZHANG Mengjian, ZHANG Hao, CHEN Xi, et al. Grey wolf optimization algorithm based on Cubic mapping and its application[J]. Computer Engineering & Science, 2021, 43(11): 2035-2042.

[17] DONGMEI H, SHIQING H, XUHUI H, et al. Prediction of wind loads on high-rise building using a BP neural network combined with POD[J]. Journal of Wind Engineering and Industrial Aerodynamics, 2017, 170: 1-17.

[18] ALJARAH I, FARIS H, MIRJALILI S. Optimizing connection weights in neural networks using the whale optimization algorithm[J]. Soft Computing, 2018, 22(1): 1-15.

[19] 刘畅,刘天琪,陈振寰,等. 基于 KPCA 和 BP 神经网络的短期负荷预测[J]. 电测与仪表, 2016, 53(10): 57-61.

LIU Chang, LIU Tianqi, CHEN Zhenhuan, et al. Short-term power load forecasting based on kernel principal component analysis and BP neural network[J]. Electrical Measurement & Instrumentation, 2016, 53(10): 57-61.

[20] 王克杰,张瑞. 基于改进 BP 神经网络的短期电力负荷预测方法研究[J]. 电测与仪表, 2019, 56(24): 115-121.

WANG Kejie, ZHANG Rui. Research on short-term power load forecasting method based on improved BP neural network[J]. Electrical Measurement & Instrumentation, 2019, 56(24): 115-121.

(编辑:李骏)