

赵文虎, 王文, 梁晏宾. 基于改进乌鸦算法优化 SCN 的 TE 过程故障诊断[J]. 智能计算机与应用, 2024, 14(6): 153-157.
DOI: 10.20169/j.issn.2095-2163.240621

基于改进乌鸦算法优化 SCN 的 TE 过程故障诊断

赵文虎, 王文, 梁晏宾

(新疆工业职业技术学院 机电工程系, 乌鲁木齐 830022)

摘要: 现代化工过程变得越来越复杂, 及时准确发现故障显得尤为重要。针对随机配置网络(SCN)泛化能力差、分类准确率低以及乌鸦搜索算法(CSA)寻优能力差等问题, 提出了改进乌鸦算法(ICSA)优化随机配置网络的田纳西-伊斯曼(TE)过程故障诊断方法, 结合 ICSA 算法和其它优化算法的寻优对比试验以及优化 SCN 网络对 TE 过程中不同故障的分类结果, 可以得出, 将所提方法应用于 TE 过程中, 可以明显提高不同故障的分类准确率, 整个测试集的分类准确率也高达 97.6%, 具有较好的分类效果, 也更符合现代化工生产的需求。

关键词: 化工过程; 随机配置网络; 改进乌鸦算法; 田纳西-伊斯曼过程; 分类准确率

中图分类号: TP183

文献标志码: A

文章编号: 2095-2163(2024)06-0153-05

TE process fault diagnosis based on improved crow algorithm optimizing SCN

ZHAO Wenhui, WANG Wen, LIANG Yanbin

(Department of Mechanical and Electrical Engineering, Xinjiang Industrial Vocational and Technical College, Urumqi 830022, China)

Abstract: Modern chemical processes are becoming more and more complex, so it is particularly important to find faults timely and accurately. Aiming at the problems of poor generalization ability, low classification accuracy of Stochastic Configuration Network and poor optimization ability of Crow Search Algorithm, the Tennessee-Eastman process fault diagnosis method of Improving Crow Search Algorithm to optimize Stochastic Configuration Network is proposed. Combined with the optimization comparison test of Improved Crow Search Algorithm and other optimization algorithms and the classification results of different faults in TE process by optimizing Stochastic Configuration Network, it can be concluded that, applying the proposed method to the TE process can significantly improve the classification accuracy of different faults, and the classification accuracy of the whole test set is also up to 97.6%, which has a good classification effect and is more in line with the requirements of modern chemical production.

Key words: chemical process; stochastic configuration network; improved crow algorithm; Tennessee-Eastman (TE) process; classification accuracy

0 引言

在如今化工过程规模越来越大、工艺越来越复杂、自动化程度越来越高的情况下, 保证化工过程的安全稳定运行就是一个至关重要的研究课题。目前, 应用于化工过程的故障诊断方法有很多。文献[1]针对故障区分难的问题, 提出通过遗传算法(Genetic Algorithm, GA)改进拉斯维加斯包裹式(Las Vegas Wrapper, LVW)方法, 能够及时进行监控, 速度更快, 稳定性更好, 也提高了故障诊断的准确率。文献[2]针对数据维度高、故障难区分、网络易陷入局部最优等问题, 利用改进核 Fisher 判别分析(Kernel Fisher Discriminant Analysis, KFDA)与差

分进化算法(Differential Evolution, DE)对自组织映射(Self-Organizing Map, SOM)网络进行优化, 用在田纳西-伊斯曼(Tennessee Eastman, TE)过程中, 具有较好的分类效果。

本文在现有研究的基础上, 首先针对优化算法全局搜索能力差、寻优精度不高等问题, 利用一些好的优化策略提高算法的寻优能力, 对于乌鸦搜索算法(Crow Search Algorithm, CSA)^[3], 采用混沌思想和 Levy 飞行策略加以改进, 然后针对随机配置网络受参数的影响、分类准确率低的问题, 利用改进乌鸦算法(Improved Crow Search Algorithm, ICSA)优化随机配置网络(Stochastic Configuration Network, SCN), 最后通过 ICSA-SCN 模型对 TE 过程中的故

作者简介: 赵文虎(1995-), 男, 硕士, 助教, 主要研究方向: 参数优化, 故障诊断。Email: 1770462421@qq.com

收稿日期: 2023-05-04

哈尔滨工业大学主办 ◆ 系统开发与应用

障进行分类。实验结果说明,改进乌鸦算法比其它原始算法寻优效果更好,ICSA-SCN模型也能够更准确地各类故障进行分类,有效提高了整体分类准确率,对于化工过程的运行,具有重要的意义,可以有效预防化工过程的故障发生,保障生命安全,减少财产损失。

1 随机配置网络

SCN是Wang等学者于2017年提出的一种新的随机神经网络^[4]。与经典神经网络相比,其不同点就在于随机配置了隐含层节点的输入权重和偏差,并且通过在SCN中引入监督机制以建设性的方式评估输出权重。SCN的结构也包含了输入层、隐含层和输出层,和前馈神经网络的结构是比较相似的,如图1所示。

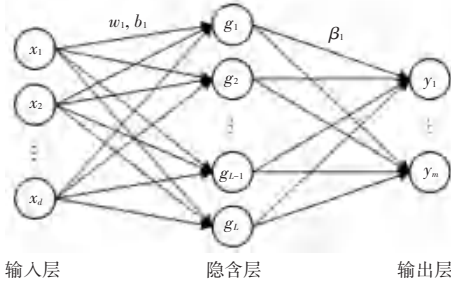


图1 SCN网络结构

Fig. 1 SCN network structure

和随机向量函数连接网络(Random Vector Functional Link network, RVFL)^[5]相比,SCN在随机参数选择、学习效率和泛化能力等方面都有优势。

给定一个训练数据集, $X = \{x_1, x_2, \dots, x_n\}$ 表示输入,其中 $x_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,d}\} \in R^d$, 对应的标签通过 $Y = \{y_1, y_2, \dots, y_n\}$ 表示,其中 $y_i = \{y_{i,1}, y_{i,2}, \dots, y_{i,d}\} \in R^m$ ($i = 1, 2, \dots, n$), 假设构建的SCN模型有 $L-1$ 个隐含层节点,则输出可以通过式(1)求得:

$$f_{L-1} = \sum_{l=1}^{L-1} \beta_l \varphi_l(w_l^T X + b_l) \quad (1)$$

其中, $f_0 = 0$; L 表示隐含节点数; w_l 和 b_l 分别表示第 l 个隐含层节点的输入权重和偏差; $\beta_l = [\beta_{l,1}, \beta_{l,2}, \dots, \beta_{l,m}]^T$ 表示隐含节点的输出权重; φ 表示SCN的激活函数。

根据式(2)计算增加第 L 个隐含节点前SCN的当前剩余误差向量:

$$E_{L-1} = f - f_{L-1} = [E_{L-1,1}, E_{L-1,2}, \dots, E_{L-1,m}]^T \in R^{n \times m} \quad (2)$$

第 L 个隐含节点的输出可通过式(3)计算:

$$h_L = [\varphi_L(w_L^T x_1 + b_L), \varphi_L(w_L^T x_2 + b_L), \dots, \varphi_L(w_L^T x_n + b_L)]^T \quad (3)$$

节点 L 的输入权重和偏差通过式(4)进行随机分配:

$$\xi_{L,j} = \frac{(E_{L-1,j}^T, h_L)^2}{h_L^T h_{L,j}} - (1 - r - \mu_L) \|E_{L-1,j}\|_F^2 \quad (4)$$

其中, $j = 1, 2, \dots, m, 0 < r < 1$ 表示正则化参数; μ_L 被定义为一个非负实数序列,其中 $\mu_L \leq (1 - r)$, $\lim_{L \rightarrow \infty} \mu_L = 0$; $\xi_L = \sum_{j=1}^m \xi_{L,j} \geq 0$ 的最大值确定节点 L 的参数。

输出权重通过式(5)进行评估:

$$\beta^* = [\beta_1^*, \beta_2^*, \dots, \beta_L^*]^T = H_L^+ Y \quad (5)$$

其中, H_L^+ 是 $H_L = [h_1, h_2, \dots, h_L]$ 的 Moore-Penrose 逆。

2 改进乌鸦算法

2.1 乌鸦搜索算法

在乌鸦的生活习性中,有储藏食物的特点,这些食物会藏在特定的地点,不会被发现,因此乌鸦彼此之间又会产生跟踪偷取对方食物的行为,针对乌鸦的这些习性和特点,Alireza等学者^[3]提出了乌鸦搜索算法(CSA)。

在整个算法中,每只乌鸦的藏食地点都是不一样的,在跟踪过程中,存在没有被发现和被发现这2种情况,跟踪偷取食物的乌鸦用 i 表示,被跟踪的乌鸦用 j 表示,因此乌鸦的位置更新也存在2种方式:

(1)如果乌鸦 j 没有发现乌鸦 i ,乌鸦 i 的位置更新公式就可以表示为:

$$x^{i,iter+1} = x^{i,iter} + r_i \times fl^{i,iter} \times (m^{j,iter} - x^{i,iter}) \quad (6)$$

其中, $x^{i,iter}$ 表示在第 $iter$ 次迭代时乌鸦 i 所处的位置; r_i 表示在 $[0, 1]$ 内随机产生的数; $fl^{i,iter}$ 为乌鸦 i 在第 $iter$ 次迭代时移动的距离,较小的移动距离使乌鸦进行局部搜索,较大的移动距离使乌鸦进行全局搜索; $m^{j,iter}$ 为乌鸦 j 在第 $iter$ 次迭代时食物所藏的位置。

(2)如果乌鸦 j 发现了乌鸦 i ,那么乌鸦 j 就会随机地改变自己的位置,避免乌鸦 i 偷取食物,此时乌鸦 i 的位置通过式(7)来表示:

$$x^{i,iter+1} = \begin{cases} x^{i,iter} + r_i \times fl^{i,iter} \times (m^{j,iter} - x^{i,iter}), & r_j \geq AP^{j,iter} \\ \text{a random position,} & \text{otherwise} \end{cases} \quad (7)$$

其中, AP 是感知概率, $AP^{j,iter}$ 为乌鸦 j 在第 $iter$

次迭代时的 AP 值。较小的 AP 值增加激烈化, 较大的 AP 值增加多样化。

2.2 乌鸦搜索算法的改进

2.2.1 混沌思想

混沌^[6]表示的是非线性现象, 在维持多样性上具有比较广泛的应用。因此经过对比, 本文在对种群进行初始化的时候, 选择立方混沌映射产生分布更均匀的种群, 从而增加种群多样性, 表示为:

$$y(n+1) = 4y(n)^2 - 3y(n) \quad (8)$$

立方映射相比于 Logistic 映射结构更简单, 最关键的是可以得到分布更均匀的种群, 乌鸦在整个迭代寻优过程中的活动范围也会变得更大。

2.2.2 Levy 飞行策略

很多优化算法在迭代到后期的时候都容易陷入局部最优, 最好的解决办法就是通过引入 Levy 飞行策略^[7], 使算法的全局搜索能力得到增强。Levy 飞行包含了 2 部分。一部分是距离比较短的随机搜索, 另外一部分是距离比较长的偶尔的跳跃式搜索, 最后更新后的公式可以用式(9)进行表示:

$$x^{i, iter+1} = x^{i, iter} + x^{i, iter} \otimes Levy(\lambda) \quad (9)$$

Levy 飞行的步长公式通过式(10)来表示:

$$Levy(x) = 0.1 \times \frac{r_1 \times \sigma}{|r_2|^{1/\alpha}} \quad (10)$$

其中, r_1, r_2 都是从 $[0, 1]$ 内随机产生的数, α 值取为 1.5, σ 的计算公式为:

$$\sigma = \left\{ \frac{\Gamma(1+\alpha)\sin(\pi\alpha/2)}{\Gamma[(1+\alpha)/2]\alpha^{2(\alpha-1)/2}} \right\}^{1/\alpha} \quad (11)$$

其中, $\Gamma(x) = (x-1)!$ 是 gamma 函数, 这里 $\Gamma(x) = (x-1)!$ 。

2.2.3 ICSA 算法步骤

通过 2 种策略对 CSA 算法进行改进, 研究给出 ICSA 详细流程如下。

步骤 1 对算法中的所有参数初始化;

步骤 2 先随机选择一个乌鸦个体 X_i , 然后利用立方混沌思想对乌鸦的位置和记忆进行初始化, 使整个种群的质量和多样性更好。对于初始迭代过程, 乌鸦藏匿食物的位置就被定义为初始位置;

步骤 3 通过对比所有乌鸦的适应度值, 找出其中的最优解;

步骤 4 按式(5)对乌鸦的位置进行更新;

步骤 5 更新位置后, 检测可行性, 如果位置可行, 则按式(8)对乌鸦的位置进行更新, 如果不可行, 乌鸦位置保持不变;

步骤 6 通过对比更新位置后乌鸦的适应度

值, 再找出其中的最优解;

步骤 7 将新位置和记忆位置的适应度值进行对比, 用更好的代替较差的, 更新乌鸦的记忆;

步骤 8 如果达到最大迭代次数, 记录最优解, 输出对应的乌鸦位置, 否则转到步骤 4。

2.2.4 性能测试仿真实验

对于提出的 ICSA 算法, 通过 2 个 Benchmark 基准测试函数^[8]来验证算法的优越性能, 2 个函数的表达式分别如下。

(1) Schaffer 函数。函数表达式为:

$$f(x) = 0.5 + \frac{\sin^2(x_1^2 - x_2^2) - 0.5}{[1 + 0.001(x_1^2 + x_2^2)]^2} \quad (12)$$

(2) Rosenbrock 函数。函数的表达式为:

$$f(x) = \sum_{i=1}^D [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \quad (13)$$

为了增加 ICSA 算法的说服力, 选择粒子群算法^[9](Particle Swarm Optimization, PSO)、萤火虫算法^[10](Firefly Algorithm, FA)、灰狼算法^[11](Grey Wolf Optimizer, GWO)、蝙蝠算法^[12](Bat Algorithm, BA)和 CSA 算法^[13]与其进行对比, 最后从迭代寻优曲线的对比结果中说明 ICSA 的优越性。

对 ICSA 中的参数进行设置: 乌鸦种群数 30, 最大迭代次数 500, AP 值取 0.05, β 设为 1.5。另外 4 种算法的种群数和最大迭代次数分别也都设为 30 和 500。6 种算法分别在 2 个函数下的迭代寻优过程如图 2 和图 3 所示。

从图 2 和图 3 中的曲线可以发现, 6 种算法具有很明显的差异, 在整体对比中, ICSA 算法与其它 5 种算法相比较, ICSA 算法在 2 个函数上的适应度值是最小的, 说明其寻优的效果是最好的。并且 ICSA 算法在 2 个函数上的迭代曲线都在 100 代以内就完成了收敛, 速度更快, 并且全程波动较小。综上所述, ICSA 算法的性能是更优的, 也充分说明了 2 种策略对 CSA 算法的改进达到了预期。

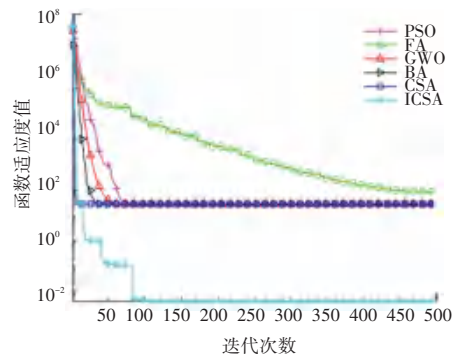


图 2 Schaffer 函数下的适应度曲线

Fig. 2 Fitness curve under Schaffer function

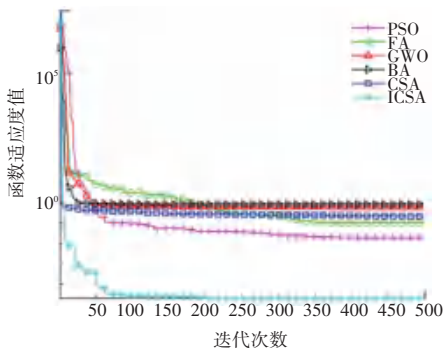


图3 Rosenbrock 函数下的适应度曲线

Fig. 3 Fitness curve under Rosenbrock function

图2和图3中的曲线能够反映大致的走向,虽然比较直观,但是没有更多数据的支撑,所以将更多的数据通过列表的方式表现出来,既直观、又具体,在反映性能的指标中选择平均值和方差数据进行对比,分别反映了收敛速度和稳定性。研究仿真中的平均值和方差数据都取20次实验的平均值,使误差最小,最后得到的数据见表1。

表1 6种算法的测试结果

Table 1 Test results of 6 algorithms

函数名	算法	平均值	方差
Schaffer 函数	PSO	8.234 8e+03	6.215 6e+05
	FA	1.582 1e+07	3.935 7e+14
	GWO	5.484 4e+04	0.492 8e+09
	BA	0.496 3e+06	2.165 0e+12
	CSA	2.485 2e+05	7.354 2e+08
	ICSA	2.368 8e+00	9.364 2e+01
Rosenbrock 函数	PSO	4.435 7e-01	2.483 7e-01
	FA	8.023 0e+04	2.721 1e+07
	GWO	4.023 4e+02	1.632 7e+04
	BA	1.413 0e+02	9.782 4e+04
	CSA	7.812 1e+03	6.610 2e+05
	ICSA	2.325 7e-03	3.985 1e-06

研究中,对比了表1中6种算法在2个函数上的平均值,可以看出ICSA算法的平均值数据要比其它5种算法小于几个数量级,充分说明了收敛速度最快的就是ICSA算法,与图1和图2的实验结果相契合。同样在方差数据的对比结果中,ICSA算法也要比其它5种算法小于几个数量级,又说明了稳定性最好的也是ICSA算法,ICSA算法的优越性再一次得到了验证。

3 ICSA-SCN 模型故障分类仿真实验

3.1 基于ICSA优化SCN的故障分类模型

传统的SCN网络对TE过程数据集中的故障不

能进行准确有效分类,为了解决这一问题,采用的方法就是通过ICSA算法去优化SCN中的正则化参数 r 和输入权重与偏差的比例因子 λ 。优化参数,构建ICSA-SCN模型的具体步骤如下。

步骤1 对TE过程数据集进行归一化处理;

步骤2 利用立方混沌思想初始化参数,将式(14)给出的均方根误差作为适应度函数:

$$RMSE = \sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / N} \quad (14)$$

步骤3 通过训练集数据,在ICSA优化SCN的过程中,以均方根误差的值最小作为目标,通过式(15)、式(16)得到最优的 r, λ :

$$f_{\text{best}} = \min f(x_i^t) \quad (15)$$

$$x_{\text{best}}^t = (x_{\text{best},r}^t, x_{\text{best},\lambda}^t) \quad (16)$$

步骤4 利用ICSA-SCN模型对测试集中的故障进行分类并输出结果。

综上所述,得到的ICSA-SCN流程如图4所示。

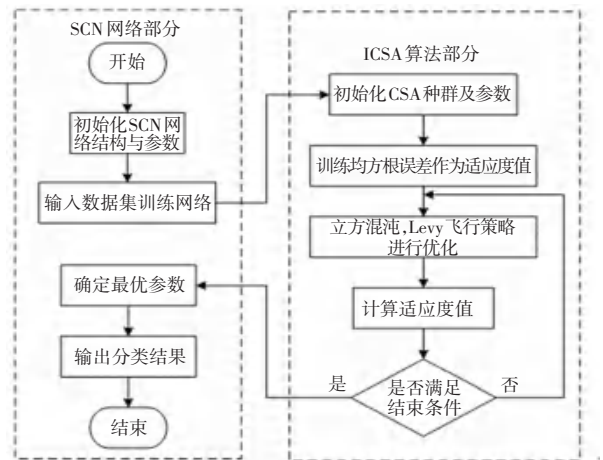


图4 ICSA-SCN 流程

Fig. 4 ICSA-SCN flow chart

3.2 TE过程数据

本文选择的数据集是TE过程数据集,是由Downs等学者^[14]提出来的数据集中的故障共有21种,数据集又分为训练集和测试集。对于训练集,每种故障的样本数都是480个,对于测试集,每种故障的样本数都是960个。由于存在不可识别的故障3、9和15以及常量故障21,所以研究中选择用于实验的是18种工况数据,并且选取的测试集中的每种故障样本数都是200个。

3.3 结果分析

通过ICSA-SCN模型对选取的测试集中的18种工况数据进行分类,为了使结果更加直观具体,利

用混淆矩阵进行表示,如图5所示。

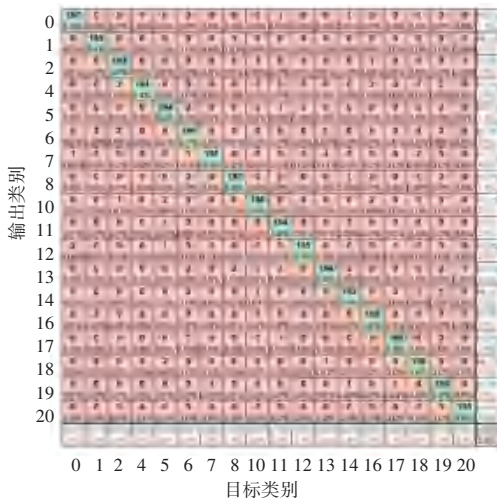


图5 混淆矩阵分类图

Fig. 5 Confusion matrix classification diagram

观察图5中数据,可以发现,每种故障分类正确和分类错误都有一个明确的表示,上面是样本数,下面是样本数占总数的比例,最下面一行是每种故障的分类准确率,故障2和故障14的分类准确率最低,是96.5%,最后总的分类准确率也高达97.6%,因此本文提出的ICSA-SCN模型能够更加有效地应用在TE过程的故障诊断中,解决了SCN网络分类难的问题。

为了进一步验证ICSA-SCN模型具有更好的分类性能,将SCN、CSA-SCN、ICSA-SCN进行对比,评价指标选择训练精度和测试精度,每项指标数据都是独立运行20次再取平均值,最后的结果见表2。

表2 3种模型的对比

Table 2 Comparison of three models %

模型	训练精度	测试精度
SCN	92.21	90.14
CSA-SCN	96.19	94.86
ICSA-SCN	98.96	97.64

表2是针对3种模型的训练精度和测试精度进行对比,3种模型是一个递进关系,可以看到不经过任何优化的SCN模型分类精度是最低的,CSA-SCN模型的分类精度比SCN的高,ICSA-SCN模型的分类精度是最高的。由此可以得出在ICSA算法的作用下,SCN网络的分类性能得到了很大的提高,可以更准确地将各类故障区分开,具有明显优势。

4 结束语

针对SCN网络分类难的问题,利用ICSA算法

对SCN中的2个重要参数进行优化,ICSA算法是在CSA算法的基础上利用混沌思想和Levy飞行策略改进的一种算法,具有更好的优化性能。仿真结果包含了2部分。一部分是ICSA与PSO、FA、GWO、BA、CSA在2个测试函数上的迭代寻优对比,另一部分是ICSA-SCN模型对TE过程数据集中的故障进行分类,结果显示ICSA算法收敛速度更快,稳定性更好,比其它算法具有更好的寻优性能,ICSA-SCN模型也可以更准确地将各类故障分开,比SCN、CSA-SCN模型具有更高的分类准确率,效果更显著,充分说明了ICSA-SCN模型的优越性。

参考文献

- [1] 耿志强,李俊,曹原,等. GA-LVW算法在化工过程故障诊断中的应用[J]. 北京化工大学学报, 2022, 49(6): 101-109.
- [2] 李国友,张新魁,才士文,等. 基于改进KFDA与DE优化SOM的故障诊断模型及其化工过程诊断[J]. 化工进展, 2022, 41(4): 1793-1801.
- [3] ALIREZA A. A novel metaheuristic method for solving constrained engineering optimization problems; Crow search algorithm[J]. Computers & Structures, 2016, 169(21): 1-12.
- [4] WANG Dianhui, LI Ming. Stochastic configuration networks: Fundamentals and algorithms[J]. IEEE Transactions on Cybernetics, 2017, 47(10): 3466-3479.
- [5] ZAYED M E, ZHAO Jun, LI Wenjia, et al. Predicting the performance of solar dish stirling power plant using a hybrid random vector functional link/chimp optimization model[J]. Solar Energy, 2021, 222(1): 1-17.
- [6] 赵耿,马艳艳,马英杰. 基于混沌序列的SCMA码本设计及系统性能分析[J]. 通信学报, 2019, 40(2): 137-144.
- [7] 刘晓明,沈明玉,侯整风. 基于Levy飞行的萤火虫模糊聚类算法[J]. 计算机应用, 2019, 39(11): 3257-3262.
- [8] ARC F N, KAYA E. Comparison of meta-heuristic algorithms on benchmark functions[J]. Academic Perspective Procedia, 2019, 2(3): 508-517.
- [9] MAIHEMUTI S, WANG W, WANG H, et al. Voltage security operation region calculation based on improved particle swarm optimization and recursive least square hybrid algorithm[J]. Journal of Modern Power Systems and Clean Energy, 2021, 9(1): 138-147.
- [10] 任静敏,潘大志. 混沌小生境萤火虫算法求解有界背包问题[J]. 西南师范大学学报, 2020, 45(11): 59-65.
- [11] 龙文,伍铁斌,唐明珠,等. 基于透镜成像学习策略的灰狼优化算法[J]. 自动化学报, 2020, 46(10): 148-164.
- [12] 赵青杰,王龙葛,李捷,等. 基于蝙蝠优化算法的网络入侵检测模型[J]. 计算机科学与应用, 2018, 8(11): 1650-1656.
- [13] 辛梓芸,张达敏,陈忠云,等. 多段扰动的共享型乌鸦算法[J]. 计算机工程与应用, 2020, 56(2): 55-61.
- [14] DOWNS J J, VOGEL E F. A plant-wide industrial process control problem[J]. Computers and Chemical Engineering, 1993, 17(3): 245-255.