

DOI:10.13364/j.issn.1672-6510.20190008

基于自适应布谷鸟与反向传播协同搜索的病害识别系统

王秀清, 陈 琪, 杨世凤
(天津科技大学电子信息与自动化学院, 天津 300222)

摘 要: 为实现番茄病害快速诊断识别, 提出了一种基于自适应布谷鸟搜索算法(cuckoo search, CS)与反向传播(back propagation, BP)协同搜索的病害识别算法(ASCS-BPCA)。首先, 该算法将全局搜索能力强的 CS 与 BP 中反向传播算法结合, 协同搜索最优识别参数, 并在此基础上引入自适应调节步长机制加快布谷鸟搜索算法收敛速度。然后, 以 3 种番茄病害(灰霉病、白粉病和晚疫病)叶片及正常叶片为研究对象, 提取病斑特征集构建 ASCS-BPCA 病害识别模型, 与标准 CS-BP 网络进行结果对比分析。仿真结果表明: ASCS-BPCA 网络平均正确识别率达 90% 以上, 优于同等条件下 CS-BP 算法, 且更加稳定高效。

关键词: 协同搜索; 自适应步长; 特征提取

中图分类号: TP391

文献标志码: A

文章编号: 1672-6510(2020)02-0069-05

Disease Recognition System Based on Cooperative Search of Cuckoo and BP Algorithm

WANG Xiuqing, CHEN Qi, YANG Shifeng

(College of Electronic Information and Automation, Tianjin University of Science & Technology,
Tianjin 300222, China)

Abstract: In order to diagnose and classify tomato diseases rapidly, this research developed an algorithm (ASCS-BPCA) to collaborate self-adaptive step cuckoo search (CS) and back propagation (BP). Firstly, the algorithm combined the global search capability of cuckoo with back-propagation algorithm of BP to optimize weights and thresholds, based on which, it introduced self-adaptive step into the cooperative algorithm to accelerate the convergence speed of cuckoo algorithm. Three tomato diseased leaves and normal leaves were taken as research subjects and classification features of disease spots were extracted as a sample set to construct an ASCS-BPCA classifier. Finally, the classification accuracy of ASCS-BPCA network was compared with that of CS-BP network used in other researches. The results show that the average correct recognition rate of ASCS-BPCA is more than 90%, better than CS-BP algorithm's performance under the same conditions, and ASCS-BPCA is more stable and efficient.

Key words: collaborative search; self-adaptive step; feature extraction

番茄病害会导致产量与作物品质严重下滑, 因而在病害发生时需要快速反应并合理施药。由于农产品生产者依靠经验判断病害时存在一定主观性和局限性, 因此病害识别模型的研究必不可少。

近些年, 对农作物病害识别的研究引起了越来越多学者的关注, 目前研究方向有: 植物叶片特征描述和更优分类算法^[1]等。王雪等^[2]提取植物叶片形状和

颜色特征, 利用 k 近邻算法作为分类识别器, 将黄瓜霜霉病从其他病害中识别出来, 且识别率达到 95%。王临铭等^[3]以 3 种甘肃啤酒大麦病斑图像为研究对象, 提取其 RGB 颜色特征及基于灰度共生矩阵的纹理特征等特征, 利用反向传播 (back propagation, BP) 神经网络作为分类器训练, 最终识别正确率达 86.7%。祁钊等^[4]提取 3 种玉米叶片病害病斑的颜

收稿日期: 2019-01-10; 修回日期: 2019-05-10

基金项目: 天津市应用基础与前沿技术研究计划资助项目(14JCZDJC39000)

作者简介: 王秀清(1967—), 女, 山西人, 教授, lwxqly@163.com

色、纹理及不变矩特征,并采用主成分分析和 SVM 相结合的方法进行分类识别,最终识别精度为 90.74%。夏永泉等^[5]利用随机森林方法对小麦叶片病害进行了分类识别,整体识别准确率达 95%。Soni 等^[6]提出用 PNN 分类器实现植物病害识别,得到了良好的识别效果。王秀清等^[7]利用声发射技术检测植物是否发生病害,也取得了良好的效果。

布谷鸟搜索算法(cuckoo search, CS)具有优化参数少,全局搜索能力强等特点,已被用于人体行为识别^[8]、图像分割^[9]、股价预测等领域,但至今仍未用于病害识别领域。孙晨等^[10]提出布谷鸟优化 BP 神经网络(CS-BP)的方法预测股价,该算法将 CS 搜索的最优鸟巢直接作为 BP 网络的权值阈值构建预测模型,但 CS 算法具有后期收敛速度较慢,易陷入局部最优等缺点。因此,本文对 CS-BP 算法加以改进,利用 BP 算法的反向传播特性优化 CS 算法搜索步骤,构造布谷鸟与 BP 网络协同算法(CS and BP collaborative algorithm, CS-BPCA)搜索最佳权值,并引入自适应调节步长机制加快布谷鸟搜索算法收敛速度,构造出自适应步长布谷鸟与 BP 协同搜索算法(self-adaptive step cuckoo search and BP collaborative algorithm, ASCS-BPCA),并将其用于番茄病害识别模型。本文首先选用番茄病害叶片作为研究对象,基于图像处理的高斯混合模型(Gaussian mixture model, GMM)聚类算法分离复杂背景以及 RGB-HSI 坐标变换实现完整病斑分割,再选择病斑部位 3 种颜色空间以及各颜色空间的灰度梯度共生矩阵作为本文特征描述,利用 ASCS-BPCA 作为本文分类识别器,并将识别正确率与 CS-BP 网络进行比较分析。

1 病斑分割与特征提取

1.1 病害图像采集

本文病害图像来源于网络,共收集了 3 种病害及正常叶片各 100 张共 400 张图像,并经专家验证所属病害类别无误。选取图像大都为自然背景,分割病害部位时复杂自然背景的剪除更容易造成误差,但更加贴合实际应用。

1.2 病害叶片背景分离与病斑分割

选用 GMM 聚类算法去除 400 张病害图像的杂草、阴影等复杂背景。利用背景部分与番茄病害叶片部分存在的差异,利用 GMM 计算数据集的概率分布,实现聚类分离带病斑的番茄植株叶片与模糊背景。

(1)将病害图像转为像素点作为原始数据,利用

k-means 聚类方法对样本聚类,得到标记像素点矩阵作为 GMM 隐含类别标签 $z^{(i)}$ 。

(2)已知 $z^{(i)}$ 后,高斯混合模型简化为式(1),式中 $p(x, z)$ 为最大似然估计。

$$L(\varphi, \mu, \Sigma) = \sum_{i=1}^m \log p(x^{(i)} | z^{(i)}; \mu, \Sigma) + \log p(z^{(i)}; p) \quad (1)$$

对 μ 和 Σ 进行求导,得到类别为 j 的样本特征均值 μ_j 和协方差矩阵 Σ_j ,即式(2)和式(3)。

$$\mu_j = \frac{\sum_{i=1}^m I\{z^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m I\{z^{(i)} = j\}} \quad (2)$$

$$\Sigma_j = \frac{\sum_{i=1}^m I\{z^{(i)} = j\} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^m I\{z^{(i)} = j\}} \quad (3)$$

(3)通过不断迭代更新均值 μ_j 和协方差矩阵 Σ_j 计算最大似然估计,直至收敛得到最佳聚类结果,从而分离病害图像的背景部位。

分离复杂背景后,利用 HSI 模型中色度信息提取病斑。HSI 模型反映人的视觉系统观察彩色的方式,模型可以将彩色图像色度(H)、饱和度(S)、强度(I)分离,在图像分割领域应用优势很大。病害叶片 RGB 坐标到 HSI 坐标色度(H)转换公式为式(4)。

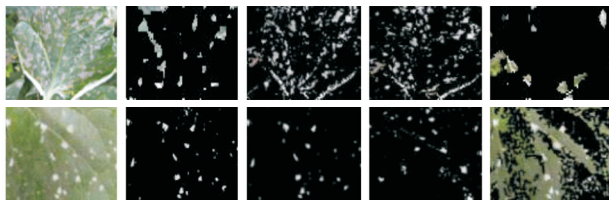
$$H = \begin{cases} \theta & G \geq B \\ 2\pi - \theta & G < B \end{cases} \quad (4)$$

$$\text{式中: } \theta = \cos^{-1} \left(\frac{(R-G) + (R-B)}{2\sqrt{(R-G)^2 + (R-B)(G-B)}} \right)$$

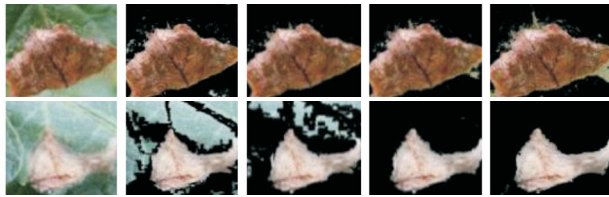
归一化色度分量,将其作为色度调节因子 h 去除绿色部分,分离出病斑。分割结果如图 1 所示。图中第 1 列为分类样本原图,第 2—5 列为调节因子 $h \in [1/7, 0.45]$ 、 $h \in [1/6, 0.5]$ 、 $h \in [1/5, 0.6]$ 、 $h \in [1/4, 0.65]$ 时病斑分割结果。

由图 1 可知: $h \in [1/6, 0.5]$ 时图 1(b) 中第 2 行分割效果不太理想,但其他都能分离出清晰病斑部位; $h \in [1/5, 0.6]$ 时图 1(b) 中第 2 行分割效果好,但其他图像都有少量绿色区域残留, h 选取其他范围时分割效果会变差。因此,使用算法 ASCS-BPCA 进行病斑分割时,选择 $h \in [1/6, 0.5]$ 为基本值,效果较差时在 $[1/6, 0.6]$ 间微调,可快速优良地分割本文中 3 种病害病斑。经实验,本算法分割 100 张图像病斑在 Matlab 中运行耗时 2.036 s,执行速度较快,且适应度广,适

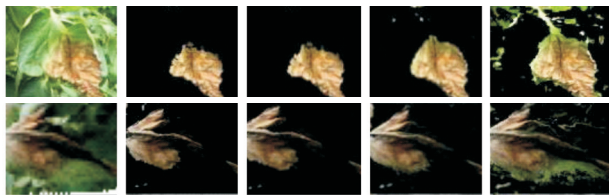
合快速提取大量病害叶片病斑.



(a) 白粉病



(b) 灰霉病



(c) 晚疫病

图 1 叶片病斑分割

Fig. 1 Segmentation of disease spots

1.3 提取病斑识别特征参数

本文以分割出的 400 张病斑图像为样本, 提取病斑特征集. 分别提取颜色直方图 RGB 的 R 、 G 、 B 颜色成分, 颜色空间 HSV 的色调、饱和度、明度, 颜色模型 HSI 中的色度、饱和度、强度共 9 种颜色特征; 利用灰度-梯度共生矩阵方法^[11]分别提取 RGB 直方图, HSV 颜色空间通道 H 、 S 、 V , HSI 颜色模型通道 H 、 S 、 I 的灰度平均、逆差距、惯性、混合熵、梯度熵、灰度熵、相关性、灰度均方差、梯度平均等共 45 个纹理特征. 利用 Tamura 方法^[12]提取病斑 RGB 坐标下的粗糙度、对比度共 2 个统计特征, 综上本文共提取 400 张病斑图像的 56 个特征向量, 组成维度空间 [400, 56] 的特征集, 选取每种病害图像各 65 张共 260 张作为验证集构成番茄叶片病斑识别网络模型, 并将剩余图像作为测试集.

2 ASCS-BPCA 的具体设计

2.1 布谷鸟搜索算法

标准布谷鸟搜索算法是通过自然界布谷鸟寄生育雏 (brood parasitism) 行为的启发提出的一种群智能搜索算法, 每个鸟巢代表一个候选解, 通过基于莱

维 (Levy) 飞行随机游走方式搜索新解代替旧解, 最终搜索种群中的最优解, 即式 (5).

$$\text{Levy} = \alpha \cdot s \cdot (x_n - x_b) \quad (5)$$

式中: x_n 为旧巢 n ; x_b 为当前最优解; s 为 Levy 飞行步长, 即式 (6).

$$s = \frac{u}{|v|^{1/\beta}} \quad (6)$$

其中给出变量 $\beta \in [0.3, 1.99]$, u 和 v 都服从正态分布 $u \sim N(0, \sigma_u^2)$, $v \sim N(0, \sigma_v^2)$

2.2 BP 神经网络

神经网络是由多个神经元组成的多层感知机 (multi layer perceptron, MLP), 用反向传播算法训练, 即 BP 神经网络. 其结构包含输入层、隐藏层、输出层, 本文构造的神经网络模型如图 2 所示.

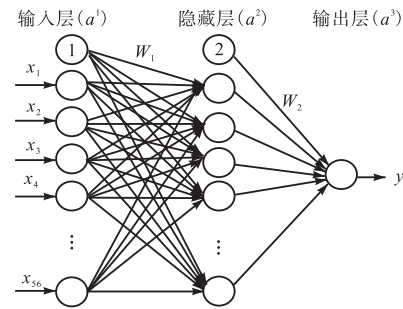


图 2 BP 模型

Fig. 2 Model of BP

其中第 1 层为输入层, 输入值 $a_i^1 = x_i$, 利用前向传播算法计算 l 层第 j 个神经元输出有 $a_j^l = g(z_j^{l-1})$, 其中 $z_j^{l-1} = \theta_j^{l-1} a_i^{l-1}$, $g(z)$ 为 sigmoid 函数, $\theta_l \in R^{s_{l+1} \times s_l}$.

在第 L 层即输出层中, 对于每一个训练集样本分类数 y^l , 有误差 $\delta^l = a^l - y^l$, 由反向传播算法从右到左计算得到第 l 层误差, 见式 (7).

$$\delta^l = \left((\theta^l)^T \delta^{l+1} \right) \cdot a^l \cdot (1 - a^l) \quad (7)$$

此时对于每个神经元权值更新公式见式 (8).

$$W_{ij}^l = W_{ij}^l + \lambda a_j^l \delta_i^{l+1} + \mu C_{ij} \quad (8)$$

式中: W_{ij}^l 表示 l 层神经元 i 、 j 间权值.

为加快学习效率, 引入矫正矩阵机制, 更新矫正矩阵公式, 见式 (9).

$$C_{ij} = a_j^l \delta_i^{l+1} \quad (9)$$

2.3 CS-BPCA 的设计

由于 CS 算法寻优后期收敛较慢、易陷入局部最小化, 本文结合 BP 神经网络反向传播能力, 自动调整权值阈值特点, 提出了布谷鸟与 BP 协同算法 (CS-BPCA), 该算法有利于让 CS 寻优后期跳出局部最

小,改善易陷入局部最小缺点.

该协同网络中,本文构建的三层 BP 网络(图 2) 权值阈值 $\{w^1\}\{b^2\}\{w^2\}\{b^3\}$, 其中 $\{b^2\}\{b^3\}$ 分别为隐藏层输出层阈值,由编码公式(10)组成布谷鸟搜索算法中巢向量 x_{bp} .

$$x_{bp} = [w^1, b^2, w^2, b^3] \quad (10)$$

与 CS 算法搜索的最优鸟巢位置 x_{cs} 根据式(11)比较判别

$$x^b = \begin{cases} x^{bp} & \text{若 } f(x^{bp}) < f(x_b^{cs}) \\ x_b^{cs} & \text{若 } f(x_b^{cs}) < f(x^{bp}) \end{cases} \quad (11)$$

得到更优解向量 x^b ,将 x^b 由式(5)和式(8)迭代更新并不断比较择优,直至结束条件为止,最终搜索到 CS-BPCA 网络中最优解向量.该算法中利用 BP 算法反向传播特性自动更新 x_{bp} ,产生更优的 x_{bp} 替代 x_{cs} 巢向量,将 x^b 利用更新式(8)自动修正权值得到新巢 x^{bp} 和 $f(x^{bp})$,利用 CS 算法式(5)更新 25 个巢向量并择出最优新巢 x_b^{cs} 和 $f(x_b^{cs})$,再由式(10)得到更优巢向量 x^b ,不断循环迭代,直到 $f(x_b) < \varepsilon$,训练达到设定误差精度或迭代次数达到最大迭代次数为止.协同网络最终输出向量 x^b 解码得到最优权值阈值.

2.4 ASCS-BPCA 的设计

CS 算法中,由于 Levy 飞行产生步长越大,越容易搜索到全局最优,但同时降低搜索精度,甚至会出现震荡现象,为增强其自适应性,对本文构造的 CS-BPCA 引入自适应度步长调节机制,见式(12)^[13].

$$d_i = \frac{\|n_i - n_{best}\|}{d_{max}} \quad (12)$$

式中: n_i 表示第 i 个鸟窝位置; n_{best} 表示最优鸟窝位置; d_{max} 表示最优鸟窝与其余鸟窝间最大距离值.

由此式引入自适应调整步长策略,见式(13).

$$s_i = s_{min} + (s_{max} - s_{min})d_i \quad (13)$$

式中: s_{max} 为 Levy 飞行步长最大值, s_{min} 为最小值.

由式(12)和式(13)引入本文协同算法,构造自适应步长布谷鸟与 BP 协同搜索算法(ASCS-BPCA).

3 病害识别的实验结果与分析

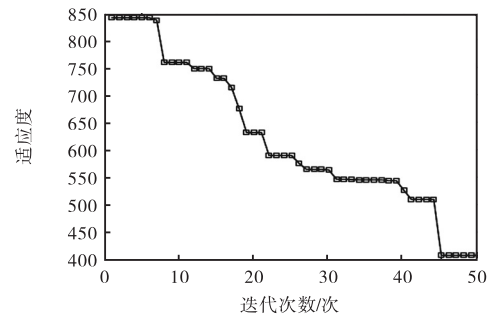
3.1 仿真实验

为了分析 ASCS-BPCA 的病害识别模型有效性与优越性,用 Matlab 2014 仿真软件进行病害识别仿真实验.将样本特征集[400, 56]均匀抽取[260, 56]作为训练集,其余作为测试集构建识别模型.将协同搜索得到的最优鸟巢 x^b 作为网络权值阈值,构建本文

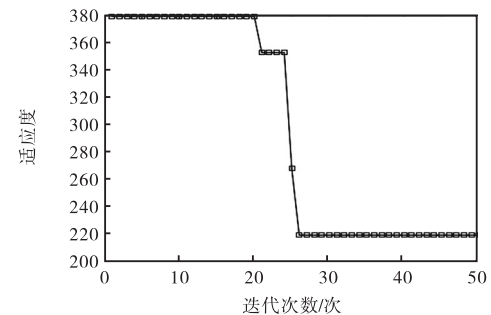
的 ASCS-BPCA 识别模型,并在相同条件下构造 CS-BP 识别模型作为对比.其中鸟巢规模为 30,迭代次数为 50.

3.2 结果与分析

将两种算法构造模型分别连续进行 5 次独立重复实验,主要从网络性能和番茄病害识别正确率分析,模型适应度随迭代次数的变化如图 3 所示.由图 3 可知: ASCS-BPCA 仅需 26 次迭代后算法收敛,用时 28.3 s,此时适应度值降为最低;而 CS-BP 算法至少需 45 次迭代后才收敛,至少用时 48.3 s,且最低适应度值仍比 ASCS-BPCA 高很多.由此可知 ASCS-BPCA 收敛速度有了较大提升,且训练效果更好.



(a) CS-BP



(b) ASCS-BPCA

图 3 识别网络性能图

Fig. 3 Performance of recognition network

对于两种算法,在训练后保存训练网络,并将测试集[140, 56]作为输入,对改进算法和传统算法进行了多次迭代计算,分析改进算法和传统算法的每次执行的平均运行时间、算法收敛所需的平均时间和识别准确率.分别进行连续不断的测试并保存连续 5 次的测试结果.选择迭代次数为 35 次、50 次和 65 次的识别结果进行对比(表 1),用于对比两种算法的识别准确率,并验证 ASCS-BPCA 算法收敛速度快于传统算法.

由表 1 可见:两种算法迭代次数相同时, ASCS-BPCA 的识别准确率均高于 CS-BP 的,说明 ASCS-

BPCA 识别准确率确实优于传统算法;ASCS-BPCA 迭代 35 次、50 次和迭代 65 次时平均识别准确率都在 90% 以上,而 CS-BP 算法平均识别准确率都在

75% 左右,且最低只有 47.1%,应是寻优时陷入局部最小所致.由此说明传统的 CS-BP 算法不太稳定.

表 1 网络识别结果对比
Tab. 1 Comparison of network recognition results

迭代次数	算法	测试样本数	每次识别准确率/%					平均准确率/%	算法平均运行时间/s	平均收敛时间/s
			1	2	3	4	5			
35	ASCS-BPCS	140	89.3	90.0	92.9	88.6	90.7	90.3	40.6	25.0
	CS-BP	140	70.0	80.0	63.6	74.3	80.7	73.7	39.9	未收敛
50	ASCS-BPCS	140	92.1	90.7	90.0	91.4	89.3	90.7	57.2	26.0
	CS-BP	140	79.3	84.3	81.4	83.5	47.1	75.1	58.0	49.4
65	ASCS-BPCS	140	90.0	90.0	89.3	89.3	92.1	90.1	75.3	28.3
	CS-BP	140	79.3	63.3	82.9	78.6	77.9	76.5	73.2	55.5

ASCS-BPCA 迭代 35 次内算法就能收敛,平均收敛时间 25 s;而 CS-BP 算法迭代次数至 35 时未能收敛,依据图 3 也说明了 CS-BP 算法至少需 45 次迭代后才能收敛.增加迭代次数至 65 后,ASCS-BPCA 迭代平均收敛时间 28.3 s,CS-BP 算法迭代平均收敛时间 55.5 s,表明改进算法收敛速度有了较大提升.

通过仿真实验分析表明:本文构造的 ASCS-BPCA 确能使 CS 算法寻优跳出局限,实现更好的全局搜索能力,且能提高整体性能,使平均识别准确率 达到 90% 以上.

4 结 语

本文将患有白粉病、灰霉病、晚疫病的番茄病害叶片及健康叶片共 400 张图像作为识别样本,其中 260 张用于训练,140 张用于测试验证.利用图像处理方法进行背景分离,并选择 RGB-HSI 转换算法实现病斑分割,分析其最优色度调节因子范围,提取出优质完整的病斑.对病斑部位提取颜色特征、纹理特征等 56 种特征参数,组成特征集用于构建病害识别模型.对 CS-BP 算法进行优化,提出 ASCS-BPCA 算法,该算法借助反向传播算法跳出 CS 寻优局限性,实现更优权值搜索,并引入自适应步长调节机制,加快收敛速度.仿真实验表明该改进算法收敛速度确有提升,且识别率高,网络性能优于 CS-BP 网络.

参考文献:

- [1] 恩德,忽胜强.基于集成神经网络的植物叶片识别方法[J].浙江农业学报,2015,27(12):2225-2233.
[2] 王雪,马卓,王欣,等.基于颜色和形状特征的黄瓜霜霉病自动识别研究[J].安徽农业大学学报,2013,

40(6):1071-1075.

- [3] 王临铭,高晓阳,李红岭,等.基于神经网络的大麦病害识别研究[J].甘肃农业大学学报,2015,50(2):173-176.
[4] 祁钊,江朝晖,杨春合,等.基于图像技术的玉米叶部病害识别研究[J].安徽农业大学学报,2016,43(2):325-330.
[5] 夏永泉,王兵,支俊,等.基于随机森林方法的小麦叶片病害识别研究[J].图学学报,2018,39(1):57-62.
[6] Soni P,Chahar R. A segmentation improved robust PNN model for disease identification in different leaf images[C]// IEEE International Conference on Power Electronics. Delhi: IEEE, 2016: 4-6.
[7] 王秀清,赵娜,陈浩,等.无线传感网络温室环境监测病害诊断系统的构建[J].天津科技大学学报,2013,28(2):50-54.
[8] 马伟.布谷鸟搜索算法优化特征和分类器参数的人体行为识别[J].微电子学与计算机,2016,33(5):102-105.
[9] 张晓琳,张冲,杨涛.基于改进布谷鸟算法的火焰图像阈值分割算法[J].微电子学与计算机,2017(1):102-105.
[10] 孙晨,李阳,李晓戈,等.基于布谷鸟算法优化 BP 神经网络模型的股价预测[J].计算机应用与软件,2016(2):276-279.
[11] 肖达,王润民,邹孝,等.基于 Gabor 变换和灰度梯度共生矩阵的超声无损测温研究[J].传感技术学报,2017,30(11):1684-1688.
[12] 吕晓琪,郭金鸽,赵宇红,等.基于图像分割的 Tamura 纹理特征算法的研究与实现[J].中国组织工程研究,2012,16(17):3160-3163.
[13] 郑洪清,周永权.一种自适应步长布谷鸟搜索算法[J].计算机工程与应用,2013,49(10):68-71.

责任编辑:郎婧