

基于非均匀变异算子的状态空间进化算法

凌 哲,李茂军

(长沙理工大学 电气与信息工程学院,湖南 长沙 410114)

摘 要:基于非均匀变异算子的状态空间进化算法(NUMSEA)是一种具有新颖性的实数编码进化算法。针对传统的状态空间进化算法转移矩阵的不足,设计一种基于非均匀变异等算子改进的状态空间转移矩阵。该矩阵突破了传统的状态空间转移矩阵,并在此基础上增加了非均匀变异算子以及非均匀算术交叉算子。通过提取分析每一代的最适值,再左乘新的转移矩阵,能够在原有的最优个体附件进行微小的搜索。进一步实现了转移矩阵随群体中个体适应度值的自适应变化,上一代群体中适值越大的个体在生成新个体时所作的贡献越大,算法的收敛速度也将增加。实验结果表明,改进算法不仅能提升对主效基因挖掘的精确性与平稳性,还能缩短对特征数据的提取时间。

关键词:状态空间算法;转移矩阵;适应度值;大数据

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2018)09-0068-04

doi:10.3969/j.issn.1673-629X.2018.09.015

State Space Evolutionary Algorithm Based on Non-uniform Mutation Operator

LING Zhe, LI Mao-jun

(School of Electrical and Information Engineering, Changsha University of
Science & Technology, Changsha 410114, China)

Abstract:State space evolutionary algorithm based on non-uniform mutation (NUMSEA) is a novel evolutionary algorithm with real number coding. Aiming at the disadvantage of the transfer matrix of traditional state space model, we design an improved state space transfer matrix based on non-uniform mutation, which breaks through the traditional state space transfer matrix and on the basis adds non-uniform mutation operator and non-uniform arithmetic crossover operator. By extracting and analyzing the optimum value of each generation, left multiplying by newly transfer matrix, we can conduct a small search in the original optimal individual attachment. The adaptive change of transfer matrix is further implemented with the individual fitness value in groups. In the previous generation, the more adaptable individuals in the group, the more contributions they make in generating new individuals, and the convergence speed of the algorithm will also increase. The experiment shows that the improved algorithm can not only improve the accuracy and stability of the major gene mining, and shorten the time of extracting characteristic data.

Key words:state space algorithm; transfer matrix; value of fitness; big data

0 引言

传统遗传算法是模拟自然界中生物进化,通过自然选择机制求解最适度值问题的一类自组织、自适应、自识别的人工智能技术,是人工智能和仿生计算技术相结合的产物,包括模拟了自然界中生物的进化过程的进化算法,模拟自然状态下某些生物群体群居行为的群体智能算法,为解决复杂程度高的多目标问题开辟了新的道路。但是传统遗传算法在实际应用中容易陷入一种早熟收敛的现象,算法控制参数比如迭代次

数选择主要依靠经验。上述缺点很大程度上限制了遗传算法的实际应用,如何增加算法的收敛精度以及如何提升算法的收敛速度,是国内外学者积极探讨的问题。例如,文献[1]提出了不同于已有遗传算法的交叉和变异策略,即通过利用诱导和随机相结合的交叉和变异算子进行局部微调,使得种群的一部分个体采用诱导变异,其余个体采用随机变异;对搜索的初始时段采用随机线性组合交叉,结束时段采用部分确定性诱导交叉。文献[2]着重从交叉和变异算子的变换出

收稿日期:2017-09-11

修回日期:2018-01-09

网络出版时间:2018-04-28

基金项目:国家自然科学基金(61074018)

作者简介:凌 哲(1993-),男,硕士研究生,研究方向为智能计算及应用;李茂军,博士,教授,硕导,研究方向为智能计算及其应用。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20180427.1640.042.html>

发,通过增加种群的多样性和选择性,并且以提高算法全局与局部搜索能力加快收敛速度为目的,设计出了一种新的算法。文献[3]提出一种采用实数编码的基于状态空间模型的进化算法。状态空间模型的构建不仅能把种群信息以较少的信息量表述出来,而且能够清楚地表现出在迭代过程中个体的状态变化。文中算法通过构造一个状态进化转移矩阵来替代遗传算法中的交叉与变异算子功能进而产生一组新的群体,通过选种池的选择方式产生较优的群体。相较于遗传算法计算量大、易陷入早熟收敛、全局搜索能力差等缺点^[4],该算法有着计算量小、计算精确度高等优点。并可通过评估转移矩阵的范数来考察算法的全局收敛性和收敛速度^[5-8],突破了传统遗传算法的固有模式。

1 状态空间模型进化算法

1.1 算法概述

遗传算法逐渐成为寻找优化方式的路径,通过交叉、选择、变异三个算子来筛选出最优值。文中基于离散系统的状态空间模型,引入遗传算法的理念,构建基于状态空间模型的方程^[9-11],即

$$\mathbf{X}(k+1) = \mathbf{G}\mathbf{X}(k) \quad (1)$$

其中, $\mathbf{X}(k)$ 为第 k 个时刻的采样向量; \mathbf{G} 为状态转移矩阵。

在该算法中, $\mathbf{X}(k)$ 表示进化算法中的群体, $\mathbf{X}(k)$ 为第 k 代群体,含有 N 个分量,每个分量均代表了 1 个个体,每个个体中包括了 M 个变量(个体是通过实数编码的方式产生的)。实际上,状态向量 $\mathbf{X}(k)$ 表示为一个 $N \times M$ 矩阵,矩阵的每一行可视为一个个体,每一个元素表示为变量的实数值。通过基于进化算法中的进化方式来构建状态转移矩阵,在此是通过运用遗传算法中的算子来构建状态转移矩阵 \mathbf{G} 。先是通过随机产生的方式得到初始群体 $\mathbf{X}(0)$,在左乘矩阵 \mathbf{G} 得到群体 $\mathbf{X}(1)$,以此类推可到一系列的 $\mathbf{X}(1)$, $\mathbf{X}(2)$, \dots 。再让群体 $\mathbf{X}(k+1)$ 和 $\mathbf{X}(k)$ 同时进入选种池内,选种池是按照物竞天择、适者生存的思想构建的,通过计算两个群体中 $2N$ 个个体的适应度函数值,并选择其中较大的 N 个个体组成新的 $\mathbf{X}(k+1)$ 群体,然后置为式 1 中的 $\mathbf{X}(k)$,如此循环,直到满足最后的终止条件。通常建立为如图 1 所示的有选择性的闭环模式。

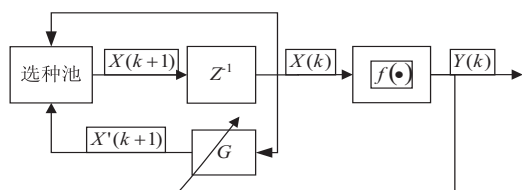


图 1 基于状态空间进化算法的有选择闭环模型

1.2 状态转移矩阵的构造

状态方程 \mathbf{G} 的构造是影响状态空间算法收敛速度和收敛性的决定因素,也是评价算法好坏的标准。状态转移矩阵 \mathbf{G} 的构造可以拟用仿生算法中群体进化的基本方法进行,例如可以模仿遗传算法的遗传算子来构造状态转移矩阵,也可以采用其他算法或者数学方法^[12-14]。对基于传统的遗传算法中的遗传算子构建状态转移矩阵做详细阐述。选择、变异、交叉是遗传算法中的基本操作,由于已经有了适应度评判函数,所以只需要在状态空间转移方程下体现交叉和变异的特点。对于实数编码下的情况,交叉操作一般选取个体间的算术均匀交叉方式,其状态空间转移矩阵方程为:

$$\mathbf{G}_1 = \begin{bmatrix} a_1 & 0 & \cdots & 0 & 1 - a_1 \\ 0 & a_2 & \cdots & 1 - a_2 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & a_{n-1} & \cdots & 1 - a_{n-1} & 0 \\ a_n & 0 & \cdots & 0 & 1 - a_n \end{bmatrix} \quad (2)$$

其中, $0 < a_i < 1$, n 由种群规模确认。

父代种群经过左乘形如 \mathbf{G}_1 的状态转移矩阵后,便实现了遗传算法中个体的交叉操作。从中可得子代群体的 1 号个体是由父代群体中的 1 号个体和 n 号个体交叉产生,由此可推断出只要交叉参数不同,其产生的子代也将有所不同^[15-17]。但是这里完全没有实现遗传算法中变异算子的功能,种群的多样性得不到较好的维持,算法容易出现早期收敛的现象。

为了实现种群的多样性,引入变异操作,将上述矩阵进行变换,如下所示:

$$\mathbf{G}_2 = \begin{bmatrix} 1 - \beta & 0 & \cdots & 0 & 0 \\ 0 & a_2 & \cdots & 1 - a_2 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & a_{n-1} & \cdots & a_{n-1} & 0 \\ a_n & 0 & \cdots & 0 & 1 - a_n \end{bmatrix} \quad (3)$$

其中, $\beta = 0.028 + 0.01\text{rand}()$ (经由实验确认), $\text{rand}()$ 为在 $[0, 1]$ 符合均匀分布的随机数,父代群体左乘以形如 \mathbf{G}_2 的状态转移矩阵后,交叉和变异的遗传算子得以实现^[18-19]。

2 基于改进的非均匀变异的状态转移矩阵

2.1 矩阵的设计

上述的状态空间转移矩阵的构造仅仅是基于两个个体间的交叉和单个个体多点变异的状态转移矩阵,

并且种群的一号个体不能经过交叉产生,不能体现自然选择的特点。

为此将进行如下改进。

$$G_3 = \begin{bmatrix} \varepsilon_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \varepsilon_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \varepsilon_3 & 0 & \cdots & 0 \\ a_4 & 0 & 0 & 1 - a_4 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 - a_n \end{bmatrix} \quad (4)$$

文中在此基础上改进之前的变异算子,使得种群尽可能地搜索到含有优良基因型的个体,同时为了避免陷入局部搜索的情况,由一个变异点变成三个点,最大程度上能够达到全局最优解。由于在选种池的操作中,选取 $2N$ 个体里适度值大的 N 个个体并按适度值从高到低排列组成新的 $X(k+1)$ 种群,这样下一代排列靠前的个体往往是最优个体。

为了便于研究,设定上一代最优个体的基因值为

$$U_{\text{best}}^k, \tau_i = \frac{U_{\text{best}}^k}{X_i^k}, X_i^k \text{ 表示第 } k \text{ 代的第 } i \text{ 个个体的基因值。}$$

对于前三个个体,再产生下一代,新的基因型 X_i^{k+1} 由下列公式决定。

$$X_i^{k+1} = \begin{cases} X_i^k + (U_{\text{best}}^k - X_i^k)(1 - r^{(1-\frac{1}{T})^b}), r \leq \frac{1}{10} \\ a_i X_i^k + (1 - a_i) U_{\text{best}}^k, r > \frac{1}{10} \end{cases} \quad (5)$$

其中, $0 < a_i < 1$; r 为 $[0, 1]$ 范围内符合均匀分布的一个随机数; T 为最大进化代数; b 为系统参数,决定了随机数扰动对进化代数 t 的依赖程度。

对照上述公式, ε_i 的设置如下:

$$\varepsilon_i = \begin{cases} 1 + (\tau_i - 1)(1 - r^{(1-\frac{1}{T})^b}), r \leq \frac{1}{10} \\ a_i + (1 - a_i)\tau_i, r > \frac{1}{10} \end{cases} \quad (6)$$

采用非均匀变异操作后,在原有最优个体附近区域进行微小搜索,能够精准快速到达全局最优解,并根据优胜劣汰的自然选择法则,以一定概率并入了算术交叉算子,使得适应度高的优秀个体能够以较大概率参与到下一次的迭代过程中。这样情况下的子代对父代的延伸程度越高,种群多样性也越高。

2.2 约束处理

在算法的迭代过程中,随着算法的搜索区域不断增大,可能导致子代的个体元素不在可行域的范围之内,故对其约束如下:

当 $X_i^k > U_{\text{max}}$ 时,使得

$$X_i^k = \begin{cases} U_{\text{max}} - < (U_{\text{max}} - U_{\text{min}}) \frac{(X_i^k - U_{\text{max}})}{(X_i^k - U_{\text{min}})} \xi >, \\ < \frac{X_i^k - U_{\text{max}}}{U_{\text{max}} - U_{\text{min}}} > = 2k, k \in z \\ U_{\text{min}} + < (U_{\text{max}} - U_{\text{min}}) \frac{(X_i^k - U_{\text{max}})}{(X_i^k - U_{\text{min}})} \xi >, \\ < \frac{X_i^k - U_{\text{max}}}{U_{\text{max}} - U_{\text{min}}} > = 2k + 1, k \in z \end{cases} \quad (7)$$

当 $X_i^k < U_{\text{min}}$ 时,使得

$$X_i^k = \begin{cases} U_{\text{max}} - < (U_{\text{max}} - U_{\text{min}}) \frac{(U_{\text{min}} - X_i^k)}{(U_{\text{max}} - X_i^k)} \xi >, \\ < \frac{U_{\text{min}} - X_i^k}{U_{\text{max}} - U_{\text{min}}} > = 2k + 1, k \in z \\ U_{\text{min}} + < (U_{\text{max}} - U_{\text{min}}) \frac{(U_{\text{min}} - X_i^k)}{(U_{\text{max}} - X_i^k)} \xi >, \\ < \frac{U_{\text{min}} - X_i^k}{U_{\text{max}} - U_{\text{min}}} > = 2k, k \in z \end{cases} \quad (8)$$

其中, $[U_{\text{min}}, U_{\text{max}}]$ 为变量的取值范围; ξ 表示寻值的精度限定; $< >$ 表示向下取整。

3 仿真实验

实验的仿真平台为 MATLAB 2013a。实验使用了 2 个经典函数进行测试,如下所示:

(1) Camel 函数。

$$\begin{cases} f_1 = (4 - 2.1x^2 + \frac{x^4}{3})x^2 + xy + (-4 + 4y)y^2 \\ -5.12 \leq x, y \leq 5.12 \end{cases} \quad (9)$$

(0.089 8, -0.712 6) 为最小值点,最小值为 -1.031 628。

(2) Rastrigrin 函数。

$$\begin{cases} f_2 = 20 + x^2 - 10\cos(2\pi x) + y^2 - 10\cos(2\pi y) \\ -5.12 \leq x, y \leq 5.12 \end{cases} \quad (10)$$

有多个极小值点,但只有一个全局最小值 0,在 (0,0) 处。

在本次测试中,设定算法参数中的种群规模大小 100,迭代的次数为 60,每个测试函数运行 50 次。为了对比结果,实验采用两种算法进行比较,分别为一般状态转移矩阵下的状态空间进化算法(evolutionary algorithm based on state-space model, SEA)和基于改进的非均匀变异转移矩阵的状态空间进化算法(state-

space evolutionary algorithm based on non-uniform mutation, NUMSEA)。对以上两个测试函数分别从平均最优解、平均收敛代数、标准差(平均最优解与全局最优解之差)等方面进行比较。

结果如表 1 所示。

表 1 测试结果

函数	算法	平均最优解	平均收敛代数	标准差
f_1	SEA	1.030 8	29	0.000 8
	NUMSEA	1.031 6	22	0
f_2	SEA	0.001 1	23	0.001 1
	NUMSEA	0.000 0	12	0

图 2 和图 3 分别显示了运用文中算法 (NUMSEA) 对求解上述函数优化问题时的最优解对应的适应度函数值随迭代次数变化的曲线。

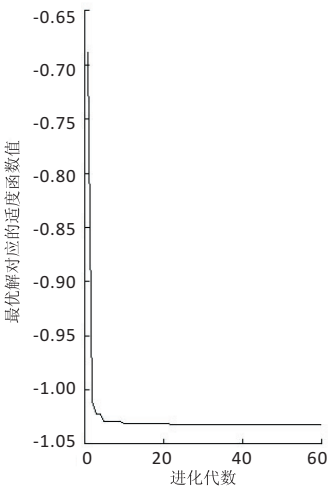


图 2 求解 f_1 优化问题时适应度函数值变化曲线

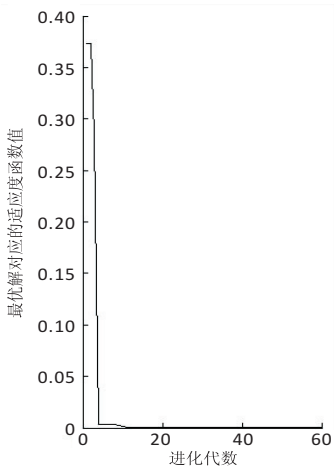


图 3 求解 f_2 优化问题时适应度函数值变化曲线

实验结果表明,NUMSEA 算法在 2 个函数的优化中明显优于之前所述的 SEA 算法,算法的收敛速度相对较快,且最优值更为接近理论值,并且能在短时间内精确定位,充分体现了其高效性。通过 50 次的测试可以看出,在稳定性方面较之前的算法有较大的提升,计

算量上有一定程度的缩减,如果能适当调整其控制参数,或将具有更好的全局搜索能力。

4 结束语

为了精准地从大数据环境下样本化、随机化的数据中提取有效信息,设计了一种具备伸缩性与并行处理功能的算法。针对一般转移矩阵下状态空间进化算法的不足,提出了一种基于非均匀变异的状态空间进化算法。通过引入改进的非均匀变异算子实现种群多样化,在一定程度上提高了算法的收敛速度与收敛精度。经过对 2 个多峰值的经典函数的优化实验,证明了该算法的可行性与稳定性,表现出了比较突出的寻优性能,缩减了处理数据的时间。虽然仿真实验结果令人满意,但是仿真环境是设为一个理想的条件下,并且实验的次数是有限的。因此,进一步构造更为良好的状态转移矩阵,以及验证算法的全局收敛性和空间遍历性,仍需要进行不断研究。

参考文献:

[1] 刘习春,喻寿益.局部快速微调遗传算法[J].计算机学报,2006,29(1):100-105.

[2] 谢燕丽,许青林,姜文超.一种基于交叉和变异算子改进的遗传算法研究[J].计算机技术与发展,2014,24(4):80-83.

[3] 李茂军,贾玲.一种基于状态空间模型的进化算法[J].计算技术与自动化,2014,33(2):85-88.

[4] 孙田雨,史峥.采用改进遗传算法优化 FIR 数字滤波器设计[J].计算机工程与应用,2017,53(17):108-111.

[5] 王鼎湘,李茂军,李雪,等.基于状态空间模型进化算法的全局收敛性分析[J].计算机应用,2014,34(10):2816-2819.

[6] 曹道友,程家兴.基于改进的选择算子和交叉算子的遗传算法[J].计算机技术与发展,2010,20(2):44-47.

[7] 卢厚清,陈亮,宋以胜,等.一种遗传算法交叉算子的改进算法[J].解放军理工大学学报:自然科学版,2007,8(3):250-253.

[8] 李雪,李茂军,王鼎湘,等.基于状态空间模型的智能优化算法及其应用[J].计算技术与自动化,2015,34(1):34-38.

[9] DU Jinling, WANG Chunxiao, ZHANG Feng. Multi-objective optimization of bus dispatching based on improved genetic algorithm[C]//2011 seventh international conference on computational intelligence and security. Hainan, China: IEEE, 2011.

[10] CHEN Jianli, ZHU Wenxing, ALI M M. A hybrid simulated annealing algorithm for nonslicing VLSI floorplanning[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 2011, 41(4):544-553.

3 结束语

针对危险源识别算法中无法实现在线学习的问题,提出了一种基于在线序列极限学习机的危险源识别算法 HI-OSELM。在建立训练网络模型时,由于算法随机生成网络输入连接权值和隐层偏置,因而可能导致网络模型结构设计不佳,进而导致识别结果准确度不够高。为此,引入自适应差分进化的智能优化方法,将输入权值和隐层偏置作为种群成员,经过种群变异、交叉、选择,不断进化,获得最优参数。网络完成初始学习后,进入在线学习阶段,新数据可以一个接一个地到达,也可以几十个、几百个同时到达。仿真实验表明,该算法对危险源具有很好的识别准确率,并且具有快速、自学习的特点,而且对以往数据不重新学习,只学习新数据,不浪费计算资源,缓解了内存压力。在用差分进化优化参数时算法实现了变异策略的自适应更新,但是,一些参数如变异因子和交叉因子的选取是在参考许多文献及同类型实验后选取了最为合适的值直接设置在算法中的,今后将继续考虑让这些也能够完成自适应进化,以期获得更高的准确率。

参考文献:

[1] ICAO. DOC 9859; safety management manual (SMM) [S]. Montreal: International Civil Aviation Organization, 2013.

[2] SMITH B E, DE JONG H H, EVERDIJ M H C. A prognostic method to identify hazards for future aviation concepts [C]//26th congress of international council of the aeronautical sciences. [s. l.]: [s. n.], 2008: 3950-3965.

[3] LIANG Nanying, HUANG Guangbin, SARATCHANDRAN P, et al. A fast and accurate online sequential learning algorithm for feedforward networks [J]. IEEE Transactions on Neural Networks, 2006, 17(6): 1411-1423.

[4] HUANG Guangbin, WANG Dianhui, LAN Yuan. Extreme learning machines: a survey [J]. International Journal of Ma-

chine Learning and Cybernetics, 2011, 2(2): 107-122.

[5] HUANG Guangbin, ZHOU Hongming, DING Xiaojian, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions on Systems, Man, and Cybernetics, 2012, 42(2): 513-529.

[6] ZHANG Nan, DING Shifei, SHI Zhongzhi. Denoising Laplacian multi-layer extreme learning machine [J]. Neurocomputing, 2016, 171(C): 1066-1074.

[7] ZHOU Hongming, HUANG Guangbin, LIN Zhiping, et al. Stacked extreme learning machines [J]. IEEE Transactions on Cybernetics, 2015, 45(9): 2013-2025.

[8] LI K, ZHANG J, XU H, et al. A semi-supervised extreme learning machine method based on co-training [J]. Journal of Computational Information Systems, 2013, 9(1): 207-214.

[9] HUANG Gao, SONG Shiji, WU Cheng. Orthogonal least squares algorithm for training cascade neural networks [J]. IEEE Transactions on Circuits and Systems, 2012, 59(11): 2629-2637.

[10] 方力智, 张翠芳, 易芳. 基于改进差分进化算法的 RBF 神经网络优化方法 [J]. 成都大学学报: 自然科学版, 2009, 28(3): 231-233.

[11] SUN Chengfo, ZHOU Haiyan, CHEN Liqing. Improved differential evolution algorithms [C]//IEEE international conference on computer science and automation engineering. Zhangjiajie, China: IEEE, 2012: 142-145.

[12] 王永皎. 改进自适应差分进化算法求解大规模整数任务分配 [J]. 计算机应用, 2012, 32(8): 2165-2167.

[13] 徐涛, 郭威, 吕宗磊. 基于快速极限学习机和差分进化的机场噪声预测模型 [J]. 电子与信息学报, 2016, 38(6): 1512-1518.

[14] 冉东梅, 蒋泰, 刘东杰. 基于自适应进化极端学习机的室内定位算法 [J]. 计算机工程与设计, 2016, 37(3): 788-792.

(上接第 71 页)

[11] 毛盈旎. 基于状态空间模型仿生算法的多目标无功优化 [D]. 长沙: 长沙理工大学, 2009.

[12] TANG Minan, REN Enen, ZHAO Chunyan. Route optimization for bus dispatching based on genetic algorithm-ant colony algorithm [C]//2009 international conference on information management, innovation management and industrial engineering. Xi'an, China: IEEE, 2009.

[13] 江中央, 蔡自兴, 王勇. 求解全局优化问题的混合自适应正交遗传算法 [J]. 软件学报, 2010, 21(6): 1296-1307.

[14] 阚超豪. 多向学习自适应的粒子群算法 [J]. 计算机工程与应用, 2013, 49(6): 23-28.

[15] GRONIEWSKY A. Exergoeconomic optimization of a thermal power plant using particle swarm optimization [J]. Thermal Science, 2013, 17(2): 509-524.

[16] 刘靖洁, 陈桂明, 刘小方. 基于矩阵编码的遗传算法研究 [J]. 计算机工程, 2011, 37(13): 160-162.

[17] 姜明佐, 张新立, 吴涛, 等. 基于云控制的混沌种群自适应遗传算法 [J]. 计算机工程, 2014, 40(1): 186-190.

[18] 董丽丽, 龚光红, 李妮, 等. 基于云模型的自适应并行模拟退火遗传算法 [J]. 北京航空航天大学学报, 2011, 37(9): 1132-1136.

[19] 刘正龙, 杨艳梅. 基于遗传算法的数值优化约束问题的研究 [J]. 计算机系统应用, 2013, 22(5): 139-142.