文章编号：2095-2163(2022)05-0136-05 中图分类号：TP241 文献标志码：A

受强化学习思想启发的一种结构优化算法

李 徐，张 帆
（上海工程技术大学 机械与汽车工程学院，上海 201620）

摘 要：针对工程以及机械结构中的优化问题，本研究提出一种基于强化学习思想的结构优化算法，该算法受到强化学习中的状态转移模型的启发，将设计变量定义为动作，相应的界限函数作为状态，神经网络的损失值由待优化的目标函数代替，采用神经网络去模拟策略函数，通过反向传播、梯度下降的原理去迭代更新神经网络使其收敛于参数的最优值。基于 Python 语言的案例仿真结果用布谷鸟搜索算法进行了验证。在文章的最后说明了该算法的局限性。

关键词：结构优化；强化学习；神经网络

A structure optimization algorithm inspired by reinforcement learning

LI Xu, ZHANG Fan
(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

【Abstract】 Aiming at the optimization problems in engineering and mechanical structures, this research proposes a structure optimization algorithm based on reinforcement learning, which is inspired by the state transition model in reinforcement learning. The design variables are defined as actions, and the corresponding boundary functions are used as states. The loss value of the neural network is replaced by the objective function to be optimized, the neural network is used to simulate the strategy function, and the neural network is iteratively updated through the principle of back propagation and gradient descent to converge to the optimal value of the parameters. The case simulation results based on Python language are verified by the Cuckoo search algorithm. The limitations of the algorithm are explained at the end of the article.

【Key words】 structural optimization; reinforcement learning; neural network

0 引 言

结构工程中的大多数设计优化问题都是高度非线性的，在复杂的约束条件下涉及许多不同的设计变量，这些约束可以写成简单的界限，如材料特性的范围，也可以写成非线性关系，包括最大应力、最大挠度、最小承载能力和几何形状，这种非线性经常导致多模态响应景观[1]，所以在结构优化问题中，寻求最优参数就变得更加困难。

强化学习的思想来自于条件反射理论和动物学习理论，是一种受到动物学习过程启发而得到的仿生算法，是重要的机器学习方法[2]，因为其具有良好的无监督学习能力，该算法主要被运用于机器人领域以及人工智能领域，在结构优化问题中几乎没有被提及，这是因为如果套用整个的强化学习算法模型，则无法运用在结构优化问题中，但是基于结构优化问题中的设计参数和界限函数与强化学习中的动作、状态很相似，选择某一组设计参数就会对应某一个确切的界限函数值，在强化学习中选择某一动作后转移到某一个状态是一个概率事件，而结构优化问题中是确定性事件，也可认为转移的概率为 1，在强化学习中目标是要找到奖励值最大的策略，而在结构优化问题中，只要找到可以使目标函数最优的设计参数就行，而设计参数的测试与选取也是靠某种策略得到，因此可以将强化学习的动作状态转移模型转换到结构优化寻优中来，更详细的说明在算法原理中有介绍。

布谷鸟搜索(Cuckoo Search ，CS)是一种新的元启发式搜索算法，这个算法是基于一些杜鹃物种的专性繁殖寄生行为，结合了一些鸟和果蝇，是由 Yang 等人[3]开发的，初步研究表明，该算法的应用前景很广，优于现有的遗传算法和粒子群算法[3]，因此在本案例中，用布谷鸟搜索(CS)算法去验证所提出的优化算法，证明了该算法的可行性。本文的

基金项目：上海市科学技术委员会科技支撑医疗器械项目(17441901200)。

作者简介：李 徐(1997-)，男，硕士研究生，主要研究方向：智能控制；张 帆(1980-)，女，博士，副教授，主要研究方向：并联机器人、医疗机器人、先进制造技术。

通讯作者：张 帆 Email：pdssophia@qq.com

收稿日期：2021-09-10

工作主要分为以下 3 个部分：

（1）算法原理中，介绍了该算法的原理，以及算法流程。

（2）案例验证中，用纯数学问题和工程案例去验证该算法的可行性。

（3）结果与讨论中，分析了实验仿真的结果，同时说明了该算法的局限性。

1　算法原理

在算法中，采用神经网络作为策略函数，因其具有极强的泛化能力、非线性映射能力，以及高度的非线性并行性，研究时常被用来作为一种学习器，并广泛运用于图像识别、分类应用中。在强化学习中也充分利用了其特性来作为策略函数，以解决具有连续动作和连续状态的强化学习，例如 Policy Gradient（PG）、Proximal Policy Optimization（PPO）、Deep Deterministic Policy Gradient（DDPG）等。基于此，在本次研究中，采用了神经网络作为策略函数，单层的神经网络模型如图 1 所示。

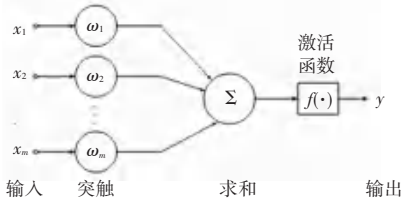


图 1　单层神经网络模型

Fig. 1　Single-layer neural network model

由图 1 可知，图 1 的输出值为：

$$y = f\left(\sum_{i=1}^{m} w_i\, x_i\right) \tag{1}$$

神经网络的更新方式是通过反向传播、梯度下降的原理去更新神经网络的参数，其更新方式的数学原理如下。

设输入样本为：

$$x(k) = (x_1(k), x_2(k), x_3(k), \cdots, x_n(k)) \tag{2}$$

期望输出为：

$$d_o(k) = (d_1(k), d_2(k), d_3(k), \cdots, d_n(k)) \tag{3}$$

隐藏层各神经元的输入输出如下：

$$h_{i_h}(k) = \sum_{i=1}^{n} w_{ih}\, x_i(k) - b_h \quad h = 1,2,3,\cdots,p \tag{4}$$

$$ho_h(k) = f(h_{i_h}(k)) \quad h = 1,2,3,\cdots,p \tag{5}$$

$$y_{i_o}(k) = \sum_{h=1}^{p} w_{ho}\, ho_h(k) - b_o \quad o = 1,2,3,\cdots,q \tag{6}$$

$$yo_o(k) = f(y_{i_o}(k)) \quad h = 1,2,3,\cdots,q \tag{7}$$

$$e = d_o(k) - yo_o(k) \tag{8}$$

求出期望值 $d_o(k)$ 与实际输出值 $yo_o(k)$ 的误差函数的偏导数 $\delta_o(k)$，其公式如下。

输出层可写为：

$$\frac{\partial e}{\partial w_{ho}} = \frac{\partial e}{\partial y_{i_o}} \frac{\partial y_{i_o}}{\partial w_{ho}} \tag{9}$$

$$\frac{\partial y_{i_o}(k)}{\partial w_{ho}} = \frac{\partial\left(\sum_{h}^{p} w_{ho}\, ho_h(k) - b_o\right)}{\partial w_{ho}} = ho_h(k) \tag{10}$$

$$\frac{\partial e}{\partial y\, i_o(k)} = \frac{\partial \left(\frac{1}{2}\sum_{o=1}^{q} ((d_o(k) - yo_o(k))\right)^2}{\partial y_{i_o}} =$$
$$-(d_o(k) - yo_o(k))\, yo_o'(k) =$$
$$-((d_o(k) - yo_o(k))\, f'(y_{i_o}(k)) = \delta_o(k) \tag{11}$$

隐藏层可写为：

$$\frac{\partial h_{i_h}(k)}{\partial w_{ih}} = \frac{\partial\left(\sum_{h}^{p} w_{ih}\, x_i(k) - b_o\right)}{\partial w_{ih}} = x_i(k) \tag{12}$$

$$\frac{\partial e}{\partial h_{i_h}(k)} = \frac{\partial \left(\frac{1}{2}\sum_{o=1}^{q} (d_o(k) - yo_o(k))\right)^2}{\partial ho_h(k)} \frac{\partial ho_h(k)}{\partial h_{i_h}(k)} =$$
$$\frac{\partial \left(\frac{1}{2}\sum_{o=1}^{q} (d_o(k) - f(y_{i_o}(k)))\right)^2}{\partial ho_h(k)} \frac{\partial ho_h(k)}{\partial h_{i_h}(k)} =$$
$$\frac{\partial\left(\frac{1}{2}\sum_{o=1}^{q}\left(d_o(k) - f\left(\sum_{h=1}^{p} w_{ho}\, ho_h(k) - b_o\right)\right)^2\right)}{\partial ho_h(k)} \frac{\partial ho_h(k)}{\partial h_{i_h}(k)} =$$
$$-\sum_{o=1}^{q} (d_o - yo_o(k))\, f'(y_{i_o}(k))\, w_{ho} \frac{\partial ho_h(k)}{\partial h_{i_h}(k)} =$$
$$-\left(\sum_{o=1}^{q} \delta_o(k)\, w_{ho}\right) f'(h_{i_h}(k)) = \delta_h(k) \tag{13}$$

利用输出层各神经元的 $\delta_o(k)$ 和隐藏层各神经元的输出来修正链接权值 $w_{ho}(k)$。研究中推得的数学公式可表示为：

$$\Delta w_{ho}(k) = -\mu \frac{\partial e}{\partial w_{ho}} = \mu\, \delta_o(k)\, ho_h(k) \tag{14}$$

$$w_{ho}^{N+1} = w_{h0}^{N} + n\delta_o(k)\, ho_h(k) \tag{15}$$

利用隐藏层各神经元的 $\delta_h(k)$ 和输入层各神经元的输入修正权连接。研究中推得的数学公式可表示为：

$$\Delta w_{ih}(k) = -\mu \frac{\partial e}{\partial w_{ih}} = -\mu \frac{\partial e}{\partial h_{i_h}(k)} \frac{\partial h_{i_h}(k)}{\partial w_{ih}} = \partial_h(k)\, x_i(k) \tag{16}$$

$$w_{ih}^{N+1} = w_{ih}^{N} + n\delta_h(k)\, x_i(k) \tag{17}$$

其中，n 是学习率。

通过式(17)就可以实现梯度下降，从而找到最优的目标函数值。

对于一个工程或机械结构优化问题，会涉及到设计参数 a，需要优化的目标函数 $f(a)$，以及会因为参数改变而引起其他性能变化的函数，在这里称为界限函数 $g(a)$，一般会要求 $g(a)$ 在某一范围 k 内。

虽然该算法启发与强化学习，但是在寻优的过程中没有正向意义上的奖励值，而是用优化机会作为奖励值，只要满足某一状态就可得到一次优化的机会，在强化学习中使用了基于神经网络的强化学习算法，如 PPO、DDPG 算法。在训练神经网络时，是对一批数据输入进行训练，这是由强化学习的算法决定的，现实中均值是无法确定的，需要通过求平均值来近似均值，而在结构优化中，只是要找到满足界限函数这一状态下的最优目标参数这一动作值即可，因此用本文提出的算法只能一次输入一个状态值(界限函数)到神经网络中，得到什么样的权重参数不重要，重要的是得到的设计参数是否是最优。

在提出的算法中，运用了强化学习的状态转移和策略函数的思想，以及神经网络的特性去优化目标函数，在有界限函数 $g(a)$ 的情况下找到最优解的前提是其设计参数必须先满足界限函数，因此首先要找到满足界限函数的参数空间位置，然后在这个参数空间中寻找最优的目标参数。在该算法中，先随机选择一组满足界限函数的初始状态，用该状态作为神经网络的输入，其输出是该状态下选择的动作(设计参数)，将该动作作为界限函数的自变量，计算出下一个状态 $g(a)$，与强化学习中的状态定义不同，这里将满足界限函数值这一类作为一个状态，不满足作为另一个状态，如果下一个状态在界限函数范围 k 中，将该状态作为新的神经网络的输入，通过反向传播，梯度下降优化目标函数，如果超出了界限函数的 k 范围，就以上一次的状态作为新的输入，通过反向传播、梯度下降的方式去更新动作输出，直到输出的动作值在界限函数值的 k 范围之内，重复以上过程，直到训练结束，算法流程如图 2 所示。

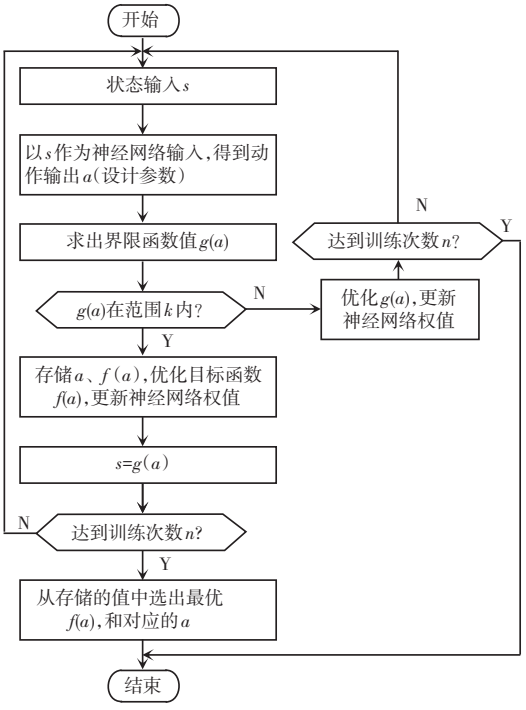


图 2 算法流程图

Fig. 2 Algorithm flow chart

2 案例验证

2.1 纯数学问题验证

为了验证该算法可以找到最优解，考虑如下一个数学问题，有 4 个参数 (a,b,c,d)，$a \in [0,4]$，$b \in [0,5]$，$c \in [0,4]$，$d \in [0,6]$，有函数：

$$g_1(a,b,c,d) = a + b + c + d \tag{18}$$

$$g_2(a,b,c,d) = a * b * c * d \tag{19}$$

求在 $g_1(a,b,c,d) \leqslant 9$，$g_2(a,b,c,d) \leqslant 20$ 的情况下函数 $f(a,b,c,d)$ 的最小值：

$$f(a,b,c,d) = (d-6)^2 * (a-2)^2 + (b-2)^2 + 1/(c-1)^2 \tag{20}$$

可以知道，在不考虑 g_1，g_2 的情况下 d 可以取值 6，b 取值 2，c 取值 4，而 a 可以任意取值就可得到最优解；当考虑 g_1，g_2 的情况下，d 明显不能取值 6，a 取值 2，b 取值 2，c 取值 4 同样可以得到最优解。通过所提出的算法求得其参数见表 1，和预想的结果几乎一样，同时在相同运算时间内比使用 CS 算法得到的值更精确。

表 1 实验数据

Tab. 1 Experimental data

f	g_1	g_2	a	b	c	d
0.111 113 22	8.955 725	15.296 792	2.000 060 8	1.999 685 2	3.999 974	0.956 353 8

该数学问题的目标函数不是很复杂，满足 g_1，g_2 的参数空间占比较大，且连续，可以看见收敛过程非常快（见图 3），f 值与 g_1，g_2 相对应的散点三维图见图 4，很明显在目标函数值下降的过程中，最小值对应的点越来越集中。

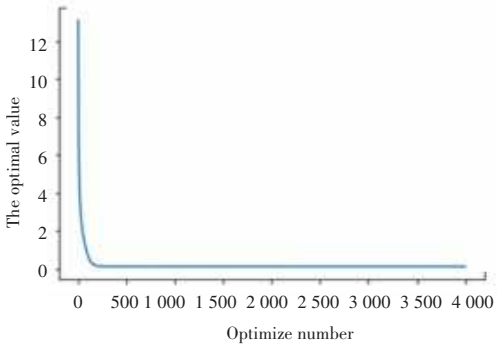


图 3　收敛过程

Fig. 3　Convergence process

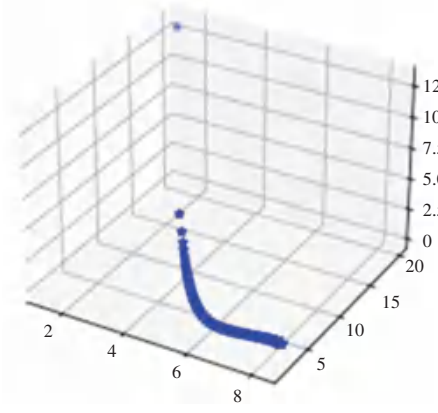


图 4　散点图

Fig. 4　Scatter plot

2.2　工程实例

最小化工字梁的垂直偏转：使用一个包含 4 个变量的设计问题来测试该算法，此案例来自报告[4]中的原始问题（已修改），目标是最小化工字梁的垂直偏转（见图 5）。梁的相关参数：P = 5 600 kN，Q = 550 kN。并且同时满足给定载荷下的横截面积和应力约束。当梁的长度（L）和弹性模量（E）分别为 5 200 cm 和 523 104 kN/cm^2 时，最小化垂直挠度 $f(x) = PL^3/48EI$。因此，该问题的目标函数可写作如下形式：

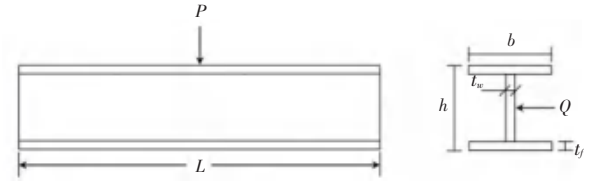


图 5　梁结构

Fig. 5　Beam structure

$$f(b,h,t_w,t_f)=\frac{5\ 000}{\frac{t_w\ (h-2\ t_f)^{\ 3}}{12}+\frac{b\ t_f^3}{6}+2b\ t_f\left(\frac{h-t_f}{2}\right)^2} \tag{21}$$

以横截面积小于 300 cm^2 为基准，有：

$$g_1 = 2bt_w + t_w(h - 2t_f) \leqslant 300 \tag{22}$$

如果梁的容许弯曲应力为 56 kN/cm^2，应力约束如下：

$$g_2 = \frac{18h \times 10^4}{t_w(h - 2t_f)^{\ 3} + 2b\ t_w(4t_f^2 + 3h(h - 2\ t_f)\ )} + \frac{15b \times 10^3}{(h - 2t_f)\ t_w^3 + 2t_w\ b^3} \leqslant 6 \tag{23}$$

其中，设计参数空间满足 $10 \leqslant h \leqslant 80$，$10 \leqslant b \leqslant 50$，$0.9 \leqslant t_w \leqslant 5$，$0.9 \leqslant t_f \leqslant 5$。

目标函数最终是收敛的（见图 6），在近乎相同时间内该算法得到的最优值和相关参数与布谷鸟搜索算法得到的值见表 2。为了排除该算法得到的最优值具有偶然性，运行得到了其他 3 组数据，仿真结果显示目标函数收敛在同一值附近，如图 7 所示。

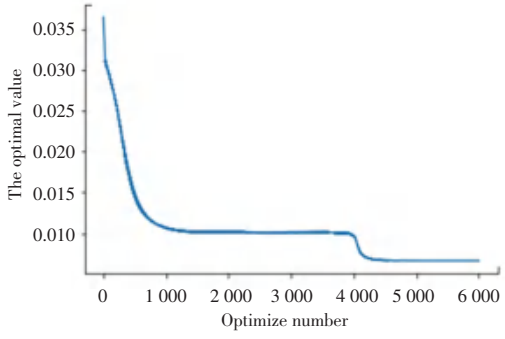


图 6　收敛过程

Fig. 6　Convergence process

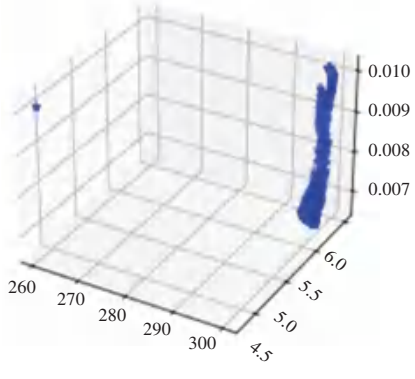


图 7　散点图

Fig. 7　Scatter plot

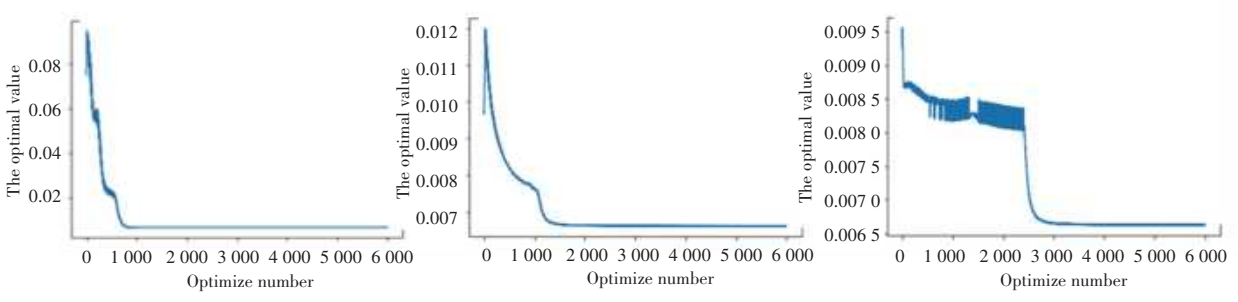
表 2 本文提出的算法与布谷鸟搜索算法得到的数据对比

Tab. 2 Comparison of the results between the proposed algorithm in this paper and the Cuckoo Search algorithm

算法	f	g_1	g_2	h	b	t_w	t_f
本文算法	0.006 63	297.787 05	5.755 534 6	79.992 24	49.991 600 0	1.764 555 7	4.998 628 6
布谷鸟搜索算法	0.006 63	299.264 89	5.726 330 0	80.000 00	50.000 000 0	1.756 060 0	5.000 000 0

3 结束语

实验仿真结果表明,在利用该算法寻优过程中,其目标函数值会因要考虑界限函数的值,可能出现一个振荡下降的过程(见图 8),起初会有一个搜索阶段,这个阶段是为了寻找满足界限函数值的区域,当收敛到满足要求的界限值位置时,目标函数值会以一个相对光滑的趋势下降,直到收敛到最小值,呈现一条相对粗细一致的线条。从图 8 中可以看到,其收敛过程存在一定的差异,但最后都收敛到了最优值,经过分析发现,其原因就在于神经网络的初始化参数和初始化的状态 s 值不同,较好的初始化位置,会使其在寻优过程中收敛曲线更加地平稳光滑,甚至会影响其收敛速度,因此可以认为,不同的初始化状态值 s 和神经元参数对收敛过程具有一定的影响。



(a) 仿真结果 1 (b) 仿真结果 2 (c) 仿真结果 3

图 8 实验仿真结果

Fig. 8 Experimental simulation results

另外,该算法具有一定的局限性,当满足界限函数的参数空间在整个参数空间占比较小,且位置较分散的时候,很难收敛到最优值,有时甚至无法收敛到满足界限函数的参数,这时需要不断改变初始值去求解最优值,同时如果面临参数空间中,有多个局部最优解的情况,此时初始化权值参数和初始化状态值 s 没有选择到合适值,就可能无法找到全局最优解。但是对于全局最优解单一,满足界限函数的参数空间占比足够大的结构优化问题,该算法则可以更高效地找到最优解。

参考文献

[1] GANDOMI A H, YANG X S, ALAVI A H. Cuckoo search algorithm: A metaheuristic approach to solve structural optimization problems [J]. Engineering with Computers, 2013, 29(1):17-35.

[2] 张汝波,顾国昌,刘照德. 强化学习理论、算法及应用 [J]. 控制理论与控制应用,2000,17 (05):637-642.

[3] YANG X S, DEB S. Cuckoo search via Lévy flights [C]// Proceedings of World Congress on Nature & Biologically Inspired Computing. Piscataway:IEEE, 2009:210-214.

[4] GOLD S, KRISHNAMURTY S. Trade-offs in robust engineering design[C]// Proceedings of the 1997 ASME Design Engineering Technical Conferences. California, USA: Design Engineering Division,1997:1-8.

(上接第 135 页)

5 结束语

蓝牙遥控六足机器人是基于 STM32F103C8T6 单片机作为控制中心,通过 HC-05 蓝牙模块进行无线遥控,PCA9685 对舵机直接控制。开机后,蓝牙自动配对成功,按下遥控器按键,机器人便执行相应动作。为适应可能发生的不确定地形情况,本次设计着重研究了 JY901 姿态传感器的开发与实现,能够使机器人在一些地形上具备更强的适应性,提高了六足机器人的稳定性。

参考文献

[1] 王倩. 六足仿生机器人步态规划与控制系统研制[J]. 哈尔滨:哈尔滨工业大学, 2007.

[2] 矫军. 计算机控制系统在机器人技术中的应用[J]. 才智, 2014 (19):323.

[3] 蔡自兴. 机器人学的发展趋势和发展战略[J]. 机器人技术与应用, 2001 (04): 11-16.

[4] 高峻峣. 国外军用地面机器人系统综述[J]. 机器人, 2003,25 (z1): 746-750,755.