

中心引力算法收敛分析及在神经网络中的应用*

孟超¹, 刘三民^{1,2}, 孙知信^{1,3}

¹(南京航空航天大学 计算机科学与技术学院, 江苏 南京 210016)

²(安徽工程大学 计算机与信息学院, 安徽 芜湖 241000)

³(宽带无线通信与传感网技术教育部重点实验室(南京邮电大学), 江苏 南京 210003)

通讯作者: 孙知信, E-mail: sunzx@njupt.edu.cn

摘要: 中心引力优化算法(central force optimization, 简称 CFO)是一种新型的基于天体动力学的多维搜索优化算法. 该算法是一种确定性的优化算法, 利用一组质子在万有引力作用下的运动, 搜索目标函数在决策空间上的最优值. 利用天体力学理论对该算法中质子运动方程进行了深入的研究, 并利用天体力学中万有引力定理对质子运动方程进行了推导, 建立起天体力学与 CFO 算法之间的联系, 通过天体力学中数学分析的方法对该算法中质子收敛性能进行了分析, 最后, 通过严格的数学推导证明出: 无论初始时质子是何种分布, CFO 算法中所有的质子始终都会收敛于 CFO 空间的确定最优解. 作为测试效果, 将 CFO 算法与常见的 BP 训练算法相结合, 提出了 CFO-BP 训练算法, 优化前馈型人工神经网络的权值和结构. 实验结果表明, 采用 CFO-BP 算法优化神经网络比其他常见优化算法有更好的收敛精度和收敛速度.

关键词: 收敛性; 中心引力优化算法; 确定性

中图法分类号: TP183 **文献标识码:** A

中文引用格式: 孟超, 刘三民, 孙知信. 中心引力算法收敛分析及在神经网络中的应用. 软件学报, 2013, 24(10): 2354-2365. <http://www.jos.org.cn/1000-9825/4391.htm>

英文引用格式: Meng C, Liu SM, Sun ZX. Convergence proof for central force optimization algorithm and application in neural networks. Ruan Jian Xue Bao/Journal of Software, 2013, 24(10): 2354-2365 (in Chinese). <http://www.jos.org.cn/1000-9825/4391.htm>

Convergence Proof for Central Force Optimization Algorithm and Application in Neural Networks

MENG Chao¹, LIU San-Min^{1,2}, SUN Zhi-Xin^{1,3}

¹(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

²(College of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China)

³(Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education (Nanjing University of Posts and Telecommunications), Nanjing 210003, China)

Corresponding author: SUN Zhi-Xin, E-mail: sunzx@njupt.edu.cn

Abstract: Central force optimization (CFO) is a new deterministic multi-dimensional search metaheuristic based on the metaphor of gravitational kinematics. CFO is a deterministic algorithm that explores a decision space by “flying” a group of probes whose trajectories are governed by Newton’s laws. Based on in-depth studies on the probes movement governed by the equations of gravitational motion, this paper utilizes Celestial Mechanics theory to deduce moving formulas, establishes the relationship between CFO algorithm and Celestial Mechanics, and analyzes CFO convergence through mathematics analysis of Celestial Mechanics. It concludes that no matter how initial probe distribute, all probes converge deterministically in CFO space with optimal solution. To test CFO’s effectiveness, a hybrid CFO-BP

* 基金项目: 国家自然科学基金(61373135, 60973140, 61170276); 江苏省产学研项目(BY2013011); 江苏省科技型企业创新基金(BC2013027); 江苏省高校自然科学研究重大项目(12KJA520003)

收稿时间: 2012-07-04; 修改时间: 2012-10-19; 定稿时间: 2013-01-25

algorithm is proposed for joint optimization of three-layer feed forward artificial neural network (ANN) structure and parameters (weights and bias). The experimental results show that the proposed hybrid CFO-BP algorithm is better than other algorithms in convergent speed and convergent accuracy.

Key words: convergence; CFO (central force optimization); deterministic

中心引力优化算法(central force optimization,简称 CFO)**是一种确定性的启发式优化算法,该算法可以实行多维全局的优化搜索^[1].这一算法来源于天体力学中的万有引力动力学,它认为所有的物体都是在重力作用下运动.

目前,几乎所有启发式的优化算法都是基于某种自然界的隐喻,比如说最近几年刚出现的 GSO 算法^[2]、比较传统的粒子群优化算法^[3]和蚁群优化算法^[4]等.这些算法都会有一些随机的特征,每一次的运行都会得到不同的结果.CFO 算法是一种完全确定性的优化算法,因为牛顿的万有引力规则和物体的运动都是确定性的.确定性的算法有着许多优点,比如该算法可以采用确定性的数据.算法的每一次执行,使用同样的开始数据可以得到同样的结果数据,该算法在运行期间可以使用反馈的方法实时地调整参数.CFO 算法通过一组质子***在多维空间的搜索,查找最优值.这组质子按照两个来源于天体力学万有引力的迭代公式在空间移动.在 CFO 算法中的质子类似于粒子群算法的粒子和蚁群算法的蚂蚁,但是他们在搜索方式上有很大的差异^[5].

CFO 算法是由 Formato 在 2007 年最先提出来的^[1],目前的研究处于初级阶段.在文献[1]中,Formato 最先提出了基本 CFO 算法,指出在天体物理学中,三维空间中的小天体,往往都是在大质量天体的轨道周围聚集(trap),这与定位目标函数的最大值或最优值很相似.在此基础上,作者提出了 CFO 算法,给出了迭代公式,指出了算法的确定性.因为算法的每一次迭代都是确定的,并且在迭代方程中没有任何随机因子,最后将 CFO 算法应用于一些电磁学中的优化装载问题.在文献[6]中,Green 首次将基本 CFO 算法应用到神经网络训练中对数据进行分类,并且与粒子群算法的神经网络训练进行比较,文中采用了 3 种不同的神经网络的模型,并采用了 Iris 数据集,最后得出结论:CFO 算法是一种新的启发式搜索算法,性能优越.在文献[7]中,Mahmoud 利用 CFO 算法的全局收敛性,将该算法与 Nelder-Mead 算法相结合,提出了一种混合型的算法,并将该算法用于共振频率检测,取得了较好的效果.在文献[8]中,Haghighi 将 CFO 算法应用于受压管道的泄漏检测和摩擦因子校准的问题上,取得了一些研究成果.在其他一些工程领域,该算法也有了一定的应用,参见文献[9,10].但是,这些文章在应用 CFO 算法之前并未证明出算法的收敛性能,算法的应用并没有可靠的理论基础.而且文献[1]的作者指出,CFO 算法的全局搜索能力很强,但是局部搜索能力不够,在接近最优值附近,收敛缓慢.所以,该算法在实际应用之前,必须进行一系列改进,以提升算法的局部收敛性能.

众所周知,一种新的算法在广泛应用之前,必须进行收敛性能的分析证明工作.只有对算法进行收敛性能的理论分析证明,才能使算法有了可靠的应用基础.目前还没有针对这方面的工作.本文从天体力学的角度,针对算法中的质子运动方程进行了分析研究,利用天体力学中数学分析的方法对质子的运动方程进行了推导,建立起天体力学与 CFO 算法之间的联系,最后,通过天体力学的数学分析的方法推导出算法收敛的结论,完全从理论上证明了 CFO 算法中质子的收敛性.CFO 算法具有很强的全局搜索能力,但是局部搜索能力欠缺,在接近最优解附近搜索缓慢.将 CFO 算法与局部收敛性能很强的算法——BP 算法相结合,提出了混合型 CFO-BP 算法.将该算法应用于训练三层前馈型人工神经网络,最后的实验结果表明,CFO 算法与其他几种算法相比有更强的全局搜索能力,而 BP 算法的局部收敛能力很强,将两者结合后的 CFO-BP 算法,训练神经网络能够取得较好的泛化能力.

本文第 1 节介绍基本 CFO 算法.第 2 节从天体力学的角度对算法中质子运动方程作一推导,建立起天体力学与 CFO 算法的联系.第 3 节是采用天体力学数学分析的方法对算法收敛性加以证明.第 4 节是 CFO-BP 算法训练神经网络的问题.第 5 节是对 CFO-BP 算法的性能测试.第 6 节是结论.

** 该算法最近几年才出现,还没有相关的中文论文,我们根据算法来源于天体的万有引力,所以暂译为中心引力算法.

*** 文献[1]中,作者都是采用了 probe,即探测器,我们根据天体力学的一些理论暂译为质子,即有质量的粒子.

1 基本 CFO 优化算法

CFO 算法定位目标函数 $f(x_1, x_2, \dots, x_{N_d})$ 最优值, 在区域 $\Omega: \min(x_k) \leq x_k \leq \max(x_k), 1 \leq k \leq N_d$, x_k 是 N_d 维空间的变量, N_d 是 CFO 决策空间的维数, $\min(x_k)$ 和 $\max(x_k)$ 分别表示第 k 维最小值与最大值. 算法包含两个质子运动方程式(1)和公式(2), 这两个方程决定了质子在一系列迭代中的运动轨迹.

$$\mathbf{a}_{j-1}^p = G \sum_{\substack{k=1 \\ k \neq p}}^{N_p} U(M_{j-1}^k - M_{j-1}^p) \cdot (M_{j-1}^k - M_{j-1}^p)^\alpha \times \frac{(\mathbf{R}_{j-1}^k - \mathbf{R}_{j-1}^p)}{\|\mathbf{R}_{j-1}^k - \mathbf{R}_{j-1}^p\|^\beta} \quad (1)$$

$$\mathbf{R}_j^p = \mathbf{R}_{j-1}^p + \mathbf{a}_{j-1}^p, j \geq 1 \quad (2)$$

其中, \mathbf{a}_{j-1}^p 是质子 p 在第 $j-1$ 迭代步的加速度; $\mathbf{R}_j^p = \sum x_k^{p,j} \mathbf{e}_k$ 是质子 p 在第 j 步的位置向量, 其中, $x_k^{p,j}$ 是它在第 k 维的坐标, \mathbf{e}_k 是沿 x_k 轴的单位向量, 参数 $1 \leq p \leq N_p$ 和 $0 \leq j \leq N_t$ 分别表示了质子和迭代步, N_p 和 N_t 代表总共的质子数目和总共的迭代步数; $M_{j-1}^p = f(x_1^{p,j-1}, x_2^{p,j-1}, \dots, x_{N_d}^{p,j-1})$ 是质子 p 在当前位置在迭代步 $j-1$ 时的目标函数适应值, 而其他质子在每一步的迭代都有适应值 $M_{j-1}^k, k=1, \dots, p-1, p+1, \dots, N_p$; G 是 CFO 算法的重力常数; U 是一个单步函数, $U(z) = \begin{cases} 1, & z \geq 0 \\ 0, & \text{otherwise} \end{cases}$; α 和 β 是常数, 分别是 1 和 3, 有时, 为了简化计算, β 取值 1.

在 CFO 算法搜索空间中, “质量”的定义与力学中质量的定义不同, CFO 算法中的质量是一个根据质子位置变化的、用户自定义的目标函数值, 在具体实施算法时质量定义为

$$MASS_{\text{CFO}} = U(M_{j-1}^k - M_{j-1}^p) \cdot (M_{j-1}^k - M_{j-1}^p)^\alpha.$$

亦即将目标函数适应值之间的差值定义为质量. 具体算法参见文献[1].

2 天体力学对 CFO 算法迭代公式的推导

CFO 算法的核心是对于加速度 \mathbf{a} 的求解, 本节从天体力学的角度对 CFO 算法的加速度公式作一推导, 目的是建立起天体力学与 CFO 算法的联系.

假设 CFO 空间是由 $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N_d})$ 这组两两正交的单位向量所张成的一个 N_d 维空间, N_p 个质子 $P_i (i=1, 2, \dots, N_p)$, m_i 为质子 P_i 的质量, $(x_1^i, x_2^i, \dots, x_{N_d}^i)$ 为质子 P_i 的 CFO 空间中的坐标, 则任意两个质子 P_i 和 P_j 间的距离为公式(3), 公式(4)给出了 N_p 个质子中任意两个质子的万有引力方程.

$$r_{ij} = \left(\sum_{k=1}^{N_d} (x_k^i - x_k^j)^2 \right)^{\frac{1}{2}} \quad (3)$$

$$F = G \frac{m_i m_j}{r_{ij}^2} \quad (4)$$

其中, x_k^i 表示质子 i 的第 k 维坐标. 在 N_d 维空间中, \mathbf{F} 是一个向量 $\mathbf{F} = \{F_{x_1}, F_{x_2}, \dots, F_{x_{N_d}}\}$, 进一步可以得到公式(5):

$$\mathbf{F} = \left\{ G \frac{m_i m_j}{r_{ij}^2} \frac{x_1^i - x_1^j}{r_{ij}}, G \frac{m_i m_j}{r_{ij}^2} \frac{x_2^i - x_2^j}{r_{ij}}, \dots, G \frac{m_i m_j}{r_{ij}^2} \frac{x_{N_d}^i - x_{N_d}^j}{r_{ij}} \right\} \quad (5)$$

牛顿第二运动定律 $\mathbf{F} = m\mathbf{a}$, 加速度 \mathbf{a} 是一个矢量. 设 $\mathbf{a} = \{a_1, a_2, \dots, a_{N_d}\}$, 则任意一个质子 P_i 的加速度第 k 维分量如公式(6)所示:

$$a_k^i = \frac{G \frac{m_i m_j}{r_{ij}^2} \frac{x_k^i - x_k^j}{r_{ij}}}{m_i} = G \frac{m_j}{r_{ij}^2} \frac{x_k^i - x_k^j}{r_{ij}} \quad (6)$$

由于万有引力是普遍存在的, N_p 个质子中任意两个质子都存在万有引力, 所以对于任意一个质子, 受到了除了本身以外其他质子的万有引力, 于是, 加速度分量写成公式(7).

$$a_k^i = \sum_{\substack{j=1 \\ j \neq i}}^{N_p} G \frac{m_j}{r_{ij}^2} \frac{x_k^i - x_k^j}{r_{ij}} \quad (7)$$

公式(7)与公式(1)相比只有对于质量的处理上有所不同,在公式(1)中,质子的质量是一个自定义的值:

$$MASS_{CFO} = U(M_{j-1}^k - M_{j-1}^p) \cdot (M_{j-1}^k - M_{j-1}^p)^\alpha, \alpha = 1,$$

其中, U 为单步函数.

根据 U 函数的定义,当质子 k 所在位置的目标函数值大于质子 p 所在位置的目标函数值时,才有质子 k 对 p 的引力存在,否则,该引力为 0. CFO 算法中质量定义的另一个不同之处在于目标函数值的差: $(M_{j-1}^k - M_{j-1}^p)$, 而没有使用目标函数值本身 M_{j-1}^k . 这样是为了避免距离非常接近质子的过度的引力,具体可参见文献[1].

设 N_p 个质子目标函数在任意一个迭代步满足不等式:

$$f_1 \leq f_2 \leq f_3 \leq \dots \leq f_{N_p-1} \leq f_{N_p},$$

其中, f_i 是第 i 个质子的目标函数值.

在算法中, U 函数的目的是使得目标函数值大的质子对目标函数值小的质子的引力作用,所以可将质子运动方程写成如下方程组形式:

$$\left. \begin{aligned} \ddot{x}_k^1 &= G \cdot m_2 \frac{x_k^2 - x_k^1}{r_{21}^3} + \dots + G \cdot m_{N_p-1} \frac{x_k^{N_p-1} - x_k^1}{r_{N_p-1,1}^3} + G \cdot m_{N_p} \frac{x_k^{N_p} - x_k^1}{r_{N_p,1}^3} \\ &\dots \\ \ddot{x}_k^{N_p-2} &= G \cdot m_{N_p-1} \frac{x_k^{N_p-1} - x_k^{N_p-2}}{r_{N_p-1,N_p-2}^3} + G \cdot m_{N_p} \frac{x_k^{N_p} - x_k^{N_p-2}}{r_{N_p,N_p-2}^3} \\ \ddot{x}_k^{N_p-1} &= G \cdot m_{N_p} \frac{x_k^{N_p} - x_k^{N_p-1}}{r_{N_p,N_p-1}^3} \\ \ddot{x}_k^{N_p} &= 0 \end{aligned} \right\} \quad (8)$$

其中, \ddot{x}_k^i 表示第 i 个质子,第 k 维坐标对时间 t 的二阶导数,即加速度; \dot{x}_k^i 表示对时间 t 的一阶导数; m_i 表示第 i 个质子的质量. 在下面的收敛分析过程中,由于一次迭代过程中质子的目标函数值不发生变化,所以总是假设质量为单位质量.

3 CFO 算法收敛性分析

本节首先介绍柯西-比卡定理,然后根据质子运动方程公式(8)将此定理作相应的修改,得到一个推论,根据推论结合质子运动方程组,证得 CFO 算法收敛的结论.

柯西-比卡定理. 设 n 阶常微分方程组

$$\frac{dz_i}{dt} = Z_i(z) (i=1, 2, 3, \dots, n) \quad (9)$$

右端函数 Z_i 为不显含自变量 t 的已知函数,而且在域

$$|z_i - \bar{z}_i| < a_i (i=1, 2, 3, \dots, n) \quad (10)$$

内解析,其中, $\bar{z} = (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_n)$ 为某个确定点, a_i 为一组正实数. 若存在一组正实数 $A_i (i=1, 2, 3, \dots, n)$, 使得在公式(10)条件确定的域内有:

$$|Z_i(z)| < A_i (i=1, 2, 3, \dots, n) \quad (11)$$

则公式(9)存在唯一的解,在域

$$|t - \bar{t}| < T \quad (12)$$

内解析,其中, \bar{t} 为给定的数, T 为 a_i/A_i 中的最小值,而且对所有满足公式(12)的 t 值,相应的 $z_i(t)$ 也满足公式(11),并有:

$$\lim_{t \rightarrow \bar{t}} z_i(t) = \bar{z}_i (i = 1, 2, 3, \dots, n) \tag{13}$$

此定理中,自变量 t 和 $z=(z_1, z_1, \dots, z_n)$ 都是复变量, $Z_i(z)$ 也认为是复变函数.若应用到本文 CFO 搜索空间中,必须要加上如下限制条件:

- 1) 因 $z=(z_1, z_1, \dots, z_n)$ 表示质子的坐标分量,故当 $t=0$ 时, N_p 个质子中第 i 个、第 k 维的坐标初值 x_k^i (z_i 的一部分)是实的而且是有限的,并使得质子间相应的距离不等于 0,而且是正的有限的实数;
- 2) 对于质子的速度分量来说, $t=0$ 时的初值 \dot{x}_k^i (z_i 的一部分)应为有限的实数,而且与坐标的初值一起,使得组成的总动量矩分量 $c=(c_1, c_2, \dots, c_{N_d})^{(11-13)}$ 以及总能量常数 K 都是实的有限数.

另外, x_k^i 和 \dot{x}_k^i 在 $t=0$ 的某个邻域内都是自变量 t 的解析单值函数.也就是说,在此领域内,质子间的距离都是大于 0 的实数.

质子坐标和速度的分量在 $t \rightarrow \bar{t}$ 时,坐标和速度的极限值分别为 \bar{x}_k^i 和 $\dot{\bar{x}}_k^i$, 都大于 0.

本文引入一个正数 ν ,把任意两个质子间的距离 \bar{r}_i 大于 0 的条件写为

$$\bar{r}_i \geq 14\nu \tag{14}$$

其中,因子 14 的引入完全是为了后面推导方便而取.在这样的条件下,根据柯西-比卡定理,质子的坐标和速度(即 z_i)可以展开为 $t - \bar{t}$ 的幂级数,并在区域 $|t - \bar{t}| < T$ 内绝对收敛.

为了找出柯西-比卡定理中的 a_i 和 A_i (统一用 ν 来表示),根据 CFO 搜索空间中的质子运动方程公式(8)来具体加以讨论.公式(8)可改写为

$$\frac{dx_k^i}{dt} = \dot{x}_k^i, \quad \frac{d\dot{x}_k^i}{dt} = \frac{1}{m_i} \frac{\partial U}{\partial x_k^i} \tag{15}$$

其中, $U = G \sum_{\substack{j=1 \\ r_{ji} > r_i}}^{N_p} \frac{m_j}{r_{ji}}$.

为了求出这些右端函数在 \bar{t} 邻域内的上界 A_i ,先设:

$$|x_k^s - \bar{x}_k^s| < \nu_0, \quad |\dot{x}_k^s - \dot{\bar{x}}_k^s| < \nu'_0 \tag{16}$$

其中, $s=1, 2, \dots, N_p$; ν_0 和 ν'_0 也是两个正实数,相当于柯西-比卡定理中的 a_i ,但是大小可以统一用 ν 和质子组运动常数来表示.

公式(15)右端函数中,主要是质子间相互距离的倒数.若以其中 r_{ji} 为例,则有:

$$\begin{aligned} r_{ij}^2 &= (x^i - x^j) \cdot (x^i - x^j) \\ &= [(\bar{x}^i - \bar{x}^j) + (x^i - \bar{x}^i) - (x^j - \bar{x}^j)] \cdot [(\bar{x}^i - \bar{x}^j) + (x^i - \bar{x}^i) - (x^j - \bar{x}^j)] \\ &= (\bar{x}^i - \bar{x}^j) \cdot (\bar{x}^i - \bar{x}^j) + \dots \\ &= \bar{r}_{ij}^2 + P(x^i - \bar{x}^i, x^j - \bar{x}^j) \end{aligned} \tag{17}$$

这里, P 表示所含引数的多项式,系数为整数以及 $(\bar{x}^i - \bar{x}^j)$. $|P|$ 显然应小于其中每项都取绝对值之后的和.但若用条件公式(16)以及 $|\bar{x}^i - \bar{x}^j| \leq r_{ij}$, 应有 $|P| < 12\bar{r}_{ij}\nu_0 + 12\nu_0^2$.

因此,若展开 $r_{ij}^{-1}, r_{ij}^{-3}, \dots$ 为 $|x^s - \bar{x}^s|$ 的幂级数(x^s 表示任意一个质子坐标 $s=1, 2, \dots, N_p$),只要满足条件:

$$12\bar{r}_{ij}\nu_0 + 12\nu_0^2 < \bar{r}_{ij}^2 \tag{18}$$

即

$$\nu_0 < \frac{\bar{r}_{ij}}{6 + 4\sqrt{3}} = \frac{\bar{r}_{ij}}{12.9282\dots} \tag{19}$$

故,取

$$\nu_0 = \frac{\bar{r}_{ij}}{14} \tag{20}$$

是符合要求的.在此条件下,由公式(17)可得:

$$r_{ij}^2 > \bar{r}_{ij}^2 - |P| > \bar{r}_{ij}^2 - 12\bar{r}_{ij}v_0 - 12v_0^2 = \frac{4}{49}\bar{r}_{ij}^2,$$

即

$$r_{ij} > \frac{2}{7}\bar{r}_{ij} \quad (21)$$

并有:

$$|x_k^i - x_k^j| \leq |\bar{x}_k^i - \bar{x}_k^j| + |x_k^i - \bar{x}_k^i| + |x_k^j - \bar{x}_k^j| < \frac{8}{7}\bar{r}_{ij} \quad (22)$$

则可得到:

$$\left| \frac{x_k^i - x_k^j}{r_{ij}^3} \right| < \frac{49}{\bar{r}_{ij}^2} = \frac{1}{4v_0^2} \quad (23)$$

比较公式(20)和公式(14)可知,应有:

$$v_0 > v \quad (24)$$

所以,公式(23)可以写为

$$\left| \frac{x_k^i - x_k^j}{r_{ij}^3} \right| < \frac{1}{4v^2} \quad (25)$$

相应的条件公式(16)可以改为

$$|x_k^i - \bar{x}_k^i| < v \quad (26)$$

对于质子组中其余的质子有相同的结果,因此有:

$$\left| \frac{1}{m_i} \frac{\partial U}{\partial x_k^i} \right| < \frac{M}{4v^2} \quad (27)$$

其中, $M = \sum_{s=1}^{N_p} m_s$.

这就是公式(15)运动方程中后半方程的右端函数在 \bar{r} 邻域内的上界.前半方程右端函数就是速度分量,可用能量积分来求出它们的上界^[13-16].质子运动的能量积分可由运动方程得到.该积分为

$$\sum_{i=1}^{N_p} m_i \dot{x}^i \cdot \dot{x}^i - 2U = h,$$

其中, h 为积分常数.由能量积分可得速度分量的上界为

$$|\dot{x}_k^i| < \sqrt{\frac{M^2}{21mv} + \frac{M}{4} |K|} \quad (28)$$

其中, K 为能量常数, m 为质子的质量.由此,若取:

$$v_0' = \sqrt{\frac{M^2}{21mv} + \frac{M}{4} |K|} \quad (29)$$

则公式(16)中的速度分量应为

$$|\dot{x}_k^i| < 2v_0' \quad (30)$$

这就是公式(15)前半段方程右端函数的上界.

按照柯西-比卡定理的结果,时间收敛半径 T 是 a_i/A_i 中的最小者.对于方程组公式(15), a_i/A_i 只有两个,即 $\frac{4v^2v_0'}{M}, \frac{v}{2v_0'}$, 则 $\frac{4v^2v_0'}{M} - \frac{v}{2v_0'} = \frac{v(8v_0'^2 - M)}{2Mv_0'} = \frac{v}{2v_0'} \left(\frac{8}{21} \frac{M}{m} + 2v|K| - 1 \right)$.

其中利用了公式(29).

上式右端括号内各项之和应为正值,而质子组中质子个数 $N_p > 2$,表明后者较小,因此对于公式(15)这样的运动方程,可以得到下列推论.

推论 1. 当时间 t 趋于某有限值 \bar{t} 时,质子坐标 x_k^i 趋于有限的极限 \bar{x}_k^i ,若在此极限值时质子间的相互距离 \bar{r}_{ij} 满足公式(14),则质子坐标分量可以展开为 $t-\bar{t}$ 的幂级数,并在

$$|t-\bar{t}| < T = \frac{v}{2v_0'} = \frac{v}{\sqrt{\frac{4M^2}{21mv} + M|K|}} \quad (31)$$

时绝对收敛.

公式(14)保证了运动方程在公式(31)所确定的范围内为 t 的解析函数;而公式(26)也保证了相互距离 r_{ij} 可以展开为 $t-\bar{t}$ 的幂级数,仍在公式(31)的范围内绝对收敛.根据公式(8)所给出的质子运动方程,得到下面的推论 2.

推论 2. 根据推论 1,结合质子运动方程,质子间的距离必在有限的时间内趋于 0,质子组必定收敛于一个确定的最优解.

反证法:假设不趋于 0,则根据质子运动方程必在质子组中存在着这样的两个质子 i 和 j ,且 $f_i \neq f_j$.于是,存在质子 i 对质子 j 的引力或是质子 j 对质子 i 的引力,使得质子 j 向着质子 i 的方向移动或是相反.由于质子的距离不趋于 0,所以这样的过程永远进行下去,与推论 1 矛盾,在有限的时间必定收敛到一个确定值.根据质子的运动方程,算法必收敛到最优值. \square

4 混合型的 CFO-BP 算法及其前馈型人工神经网络的训练

CFO 算法最终会收敛到最优解.本文将 CFO 算法与 BP 算法相结合,提出了 CFO-BP 算法应用于训练三层前馈型人工神经网络.本节首先介绍混合型的 CFO-BP 训练算法,然后介绍三层前馈型人工神经网络的基本结构、编码方式、目标函数.

4.1 混合型的 CFO-BP 算法

CFO 算法是一种全局优化算法,它有很强的全局寻优能力^[1],但也存在着一些缺点,在全局最优解附近,搜索速度缓慢,即局部寻优能力不够.而 BP 算法正好相反,BP 算法是基于梯度下降法,最早由 Werbos 提出^[17],之后, Rumelhart^[18]进一步对其加以完善.该算法有着很强的局部寻优能力,但是该算法全局寻优能力很差.通过将两种算法相结合,本文提出一种新的混合型的 CFO-BP 算法.该算法的基本思想是:在搜索的开始阶段,采用 CFO 算法加快全局搜索速度,当连续几次迭代后,目标函数值不再变化或是变化很小,且小于一个预先定义好的数值,此时可以认为搜索已经接近全局最优解.所以,搜索过程改为梯度下降搜索法,即采用 BP 算法在局部范围内搜索.

在 CFO-BP 算法开始时,初始化一组随机分布的质子,这组质子按照公式(1)和公式(2)的方程在全局空间搜索.然后再采用 BP 算法在全局最优解附近进行局部搜索,采用这种方式搜索,该算法可以更快地找到最优解.具体的算法如下:

1. 初始化质子的位置向量和加速度向量为 0 的向量;
2. 设置 CFO 算法结束的条件,即目标函数变化的最小数值,当小于这个值时 CFO 算法结束.本文设定为 0.01;
3. 计算初始时质子所在位置的目标函数适应值;
4. 开始迭代:
 - 4.1. 用公式(1)计算加速度;
 - 4.2. 用公式(2)计算质子的位置, R_j^p 目标函数值;
 - 4.3. 重复循环,直至满足 CFO 算法的结束条件;
5. 使用 BP 算法在目标函数值附近搜索,如果搜索结果好于当前结果,则输出;否则,输出 CFO 算法的搜索结果.

BP 算法有一个参数称为学习速率,该参数可以控制算法局部最优的收敛速度.在实际应用中,往往通过经

验或是理论的方法来设定学习速率,在本文中,设定学习速率 η 为 0.5.

4.2 CFO-BP算法训练人工神经网络

4.2.1 三层前馈型人工神经网络

典型的三层前馈型人工神经网络如图 1 所示. $X(x_1, x_2, \dots, x_l)$ 和 $Y(y_1, y_2, \dots, y_m)$ 分别是输入 l 个节点和输出 m 个节点,隐藏层有 h 个节点.

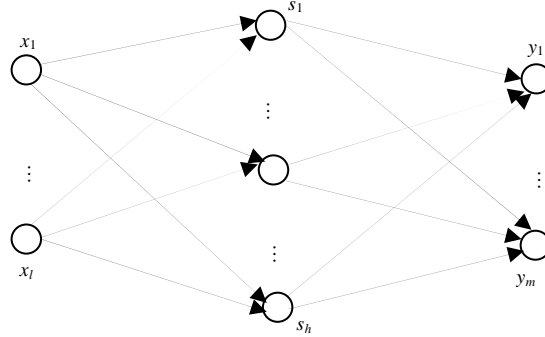


Fig.1 Typical three layer feedforward neural network

图 1 典型的三层前馈型人工神经网络

隐藏层输入节点值和输出节点值分别是公式(32)和公式(33).

$$s_j = f\left(\sum_{i=1}^l w_{ji}x_i + b_j\right), 1 \leq j \leq h \tag{32}$$

$$y_k = f\left(\sum_{j=1}^h w_{kj}s_j + b_k\right), 1 \leq k \leq m \tag{33}$$

其中 f 是 sigmoid 函数, w_{ji} 代表输入层节点与隐藏层节点之间的连接权值,而 w_{kj} 代表隐藏层节点与输出层节点之间的连接权值, b_j 和 b_k 分别表示隐藏层节点和输出层节点的偏差.

定义目标函数 fitness 为均方差函数公式(34).

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \tag{34}$$

其中, X_i 是期望输出, \bar{X} 是实际输出, n 是样本数目.

三层前馈型神经网络的训练算法主要是对神经网络的权值和偏差进行不断的修改,以降低目标函数值,即不断地减少均方差函数值.

4.2.2 编码方案

在 CFO-BP 训练算法编码方案中,每一个质子都由一个矩阵来表示.假设网络的结构是 2-3-4,则相应的每个粒子的矩阵是公式(35).

$$probe(i) = [w_{lh}, b_h, w_{hm}, b_m]$$

$$w_{lh} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix}, b_h = \begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \end{pmatrix}, w_{hm} = \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \end{pmatrix}, b_m = \begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \\ b_{41} \end{pmatrix} \tag{35}$$

在 CFO-BP 算法执行过程中,不断地进化参数.每个质子由上面公式(35)表示,其中, w_{lh} 表示输入层与隐藏层间的链接权值, w_{hm} 表示隐藏层与输出层间的链接权值, b_h 表示隐藏层的偏差, b_m 表示输出层的偏差.采用这种编码方案,相应地,其解码方案也非常简单,本文不再详述.

5 实验结果与性能分析

实验共分两个部分:第 1 部分是对 CFO 算法的全局收敛性能的测试,而 BP 算法的局部收敛性能可以参见文献[17,18],这里不再详述;第 2 部分是对 CFO-BP 算法性能的分析.

5.1 CFO 算法的全局收敛性能分析

下面以一个多峰函数为例说明 CFO 算法具有很强的全局收敛性能.

函数定义为 $f = \sum_{i=1}^{N_d} [x_i \sin(\sqrt{|x_i|})]$, $-500 \leq x_i \leq 500$, $N_d = 30$, $f_{\max}(x) = 12569.5$.

该函数的最优值分布在决策空间的主对角线上,图 2 显示了各种算法在维数为 30、CFO 的质子数为 240、PSO 的粒子数为 30 时的收敛轨迹.

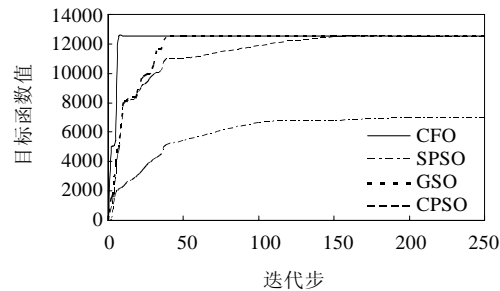


Fig.2 30 dimension convergence curves of Schwefel function

图 2 30 维的 Schwefel 函数的收敛曲线

从图 2 中可以看出:

- CFO 算法在迭代将近 10 步时,接近全局最优值 12 569.5,然而在 10 步以后变化趋缓,看不出明显的变化,在 200 多步才真正收敛到全局最优值;
- 而标准 PSO 算法会陷入局部最小,错过全局最优值;
- 基于混沌的粒子群算法 CPSO 可以收敛到全局最优,但是收敛速度缓慢;
- 种群优化算法 GSO 收敛速度比 CPSO 要快,但是比起 CFO 算法收敛速度要慢.

图 2 说明了 CFO 算法比起其他几种常见的最优化算法能够更快地收敛到全局最优解的附近,也即 CFO 算法比其他几种算法有更强的全局搜索能力.但是,当 CFO 算法的搜索质子在接近最优值时,需要在局部范围内进行搜索,此时,CFO 算法的局部搜索能力不够,所以导致在接近最优值附近时搜索缓慢,几乎看不到最优值的明显变化.

将 CFO 算法与局部搜索能力很强的 BP 算法^[17,18]相结合,可以实现优势互补,使得 CFO-BP 算法既具有 CFO 算法很强的全局搜索能力,又具有 BP 算法良好的局部搜索能力.

5.2 CFO-BP 算法的性能分析

本文采用 CFO-BP 算法训练如图 1 所示的人工神经网络,来拟合几个复杂的多峰函数:

$$f_1 = \sum_{i=1}^{N_d-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2], -30 \leq x_i \leq 30, N_d = 4,$$

$$f_2 = \sum_{i=1}^{N_d} x_i^2, -100 \leq x_i \leq 100, N_d = 8,$$

$$f_3 = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4, -5 \leq x_i \leq 5, i = 1, 2,$$

$$f_4 = 0.5\sin(\pi x_1^2)\sin(2\pi x_2), -1 \leq x_i \leq 1, i = 1, 2,$$

并且与目前几种流行的 PSO 算法、GSO 算法和 BP 算法进行比较,对应的神经网络是一个 $i-S-1$ 的结构,其中, S 是隐藏层的节点数, i 是搜索空间的维数.随机生成 100 个训练样本和 50 个测试样本.

算法具体的参数设置如下:

- 对于 CFO 算法,质子数是 240,初始时质子随机分布;
- PSO 算法种群规模设置为 40,其余设置参见文献[3].

PSO-BP 算法中,PSO 算法全局搜索迭代 200 次,BP 算法局部搜索迭代 1 000 次,而基本 BP 算法对于这几个拟合函数迭代 1 000 次.

表 1 中列出了拟合函数在几种算法下的训练误差 $E1$ 和测试误差 $E2$ 的比较,选择隐藏层节点数是 7.从表 1 中可以看出,混合型的 CFO-BP 算法比其他算法能够取得更高的精确度.

Table 1 The comparison analysis of several common optimization algorithm

表 1 几种常见的优化算法对于 4 个函数的误差分析比较

函数	项目	BP	CFO	PSO-BP	CFO-BP
f_1	$E1$	0.04581	$3.12e-04$	$4.25e-04$	$4.85e-06$
	$E2$	0.3124	$8.87e-02$	$7.91e-02$	$1.21e-02$
f_2	$E1$	$1.2e-02$	$1.12e-02$	$1.23e-02$	$6.94e-03$
	$E2$	$1.3e-02$	$1.18e-02$	$1.84e-02$	$7.31e-03$
f_3	$E1$	$4.01e-04$	$3.84e-03$	$2.68e-03$	$4.01e-05$
	$E2$	0.2114	$6.82e-02$	$5.12e-02$	$8.12e-04$
f_4	$E1$	$6.12e-02$	$3.24e-02$	$3.14e-02$	$2.14e-03$
	$E2$	0.1245	$4.12e-01$	$5.68e-01$	$1.24e-02$

下面以 f_1 函数为例,说明各种常见优化算法的性能优势.

图 3 显示了隐藏层节点数是 7 的情况下,各种算法训练误差的比较曲线.图中列出了前 400 次迭代的情况,后面几百次迭代误差变化不够显著,所以未画出.从图中可以看出,BP 算法陷入了局部收敛,主要是因为 BP 算法具有很强的局部收敛性能,但是全局搜索性能较差,所以很容易陷入局部收敛的陷阱中.其他几种算法都可以全局收敛,其中,CFO 算法全局搜索能力很强,在迭代初始收敛速度很快,但是由于局部搜索能力不够,所以中后期收敛缓慢;PSO-BP 算法利用了 PSO 算法的全局搜索能力,同时也利用了 BP 算法的局部搜索能力,所以取得了较好的性能;CFO-BP 算法中,CFO 算法比 PSO 算法有更好的全局搜索能力,所以收敛速度更快,精度更高.在图 3 所示的曲线中,CFO-BP 算法能够取得最好的搜索性能.

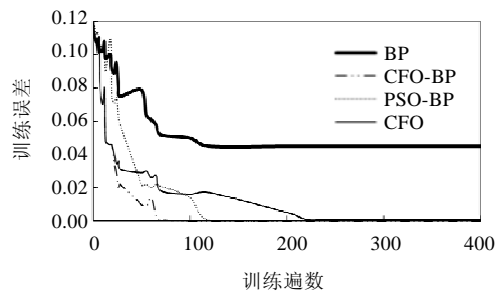


Fig.3 The comparison curves of several optimization algorithm

图 3 各种算法的训练误差比较曲线

选择终止条件是: $MSE \leq 0.02$,其余设置如前所述.

图 4 显示了拟合 f_1 时,隐藏层节点数与训练时间的关系曲线.

从隐藏层节点数与训练时间的关系曲线可以看出,CFO 算法与 PSO-BP 算法的训练时间很接近,这主要是因为 CFO 算法全局收敛速度比 PSO 算法要快,但是局部收敛速度比 BP 算法要慢,而算法执行中开始时全局搜索,然后转为局部搜索,所以总体上看,CFO 算法与 PSO-BP 算法训练时间相差不大.而 CFO-BP 算法无论在隐藏

层节点数是多少,都可以取得较好的时间性能.

图5显示了隐藏层节点数与测试误差的关系,对于CFO-BP算法,当隐藏层节点数为11时,网络的泛化能力最好.从总体上看,CFO-BP算法的泛化能力强于其余几种常见的算法.

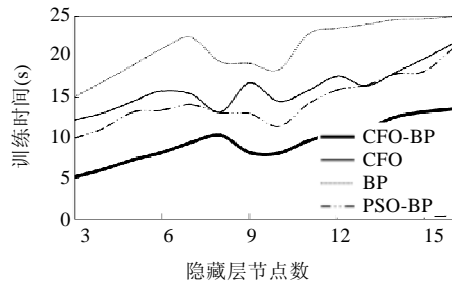


Fig.4 The relation curves of the number of hidden layer nodes and training time

图4 隐藏层节点数与训练时间的关系曲线

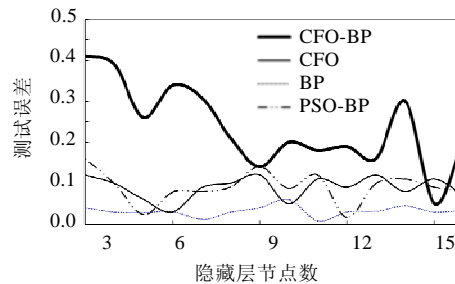


Fig.5 The relation curves of the number of hidden layer nodes and testing error

图5 隐藏层节点数与测试误差的关系曲线

6 结论

文献[1]最早提出了CFO迭代算法,但到目前为止,针对CFO算法收敛性能的理论分析证明据我们所知还未出现.本文正是利用天体力学数学分析方法,对CFO算法中质子运动规律进行了分析,并从理论上严格证明了CFO算法无论初始的质子如何分布,CFO算法都收敛到最优解.最后,为了测试CFO算法的性能,本文将该算法与BP算法相结合,提出了CFO-BP神经网络训练算法,去训练三层前馈型人工神经网络.实验结果说明,CFO算法具有较好的全局搜索能力,BP算法具有很好的局部搜索能力,将两者结合后的混合CFO-BP算法具有良好的泛化能力.

References:

- [1] Formato RA. Central force optimization: A new metaheuristic with applications in applied electromagnetics. *Prog Electromagnetics Research PIER*, 2007,77:425-449. [doi: 10.2528/PIER07082403]
- [2] He S, Wu QH, Saunders JR. Group search optimizer: An optimization algorithm inspired by animal searching behavior. *IEEE Trans. on Evolutionary Computation*, 2009,13(5):973-990. [doi: 10.1109/TEVC.2009.2011992]
- [3] Shen YX, Wang GY, Zeng CH. Correlative particle swarm optimization model. *Ruan Jian Xue Bao/Journal of Software*, 2011,22(4): 695-708 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3728.htm> [doi: 10.3724/SP.J.1001.2011.03728]
- [4] Neto RFT, Filho MG. An ant colony optimization approach to a permutational flowshop scheduling problem with outsourcing allowed. *Computers & Operations Research*, 2011,38(9):1286-1293. [doi: 10.1016/j.cor.2010.11.010]
- [5] Formato RA. Central force optimization with variable initial probes and adaptive decision space. *Applied Mathematics and Computation*, 2011,2(17):8866-8872.
- [6] Green II RC, Wang LF, Alam M. Training neural networks using central force optimization and particle swarm optimization: Insights and comparisons. In: *Proc. of the Expert Systems with Applications*. 2011. [doi: 10.1016/j.eswa.2011.07.046]
- [7] Mahmoud KR. Central force optimization: Nelder-Mead hybrid algorithm for rectangular microstrip antenna design. *Electromagnetics*, 2011,31(8):578-592. [doi: 10.1080/02726343.2011.621110]
- [8] Haghghi A, Ramos HM. Detection of leakage freshwater and friction factor calibration in drinking networks using central force optimization. *Water Resour Manage*, 2012,26:2347-2363. [doi: 10.1007/s11269-012-0020-6]
- [9] Ragot S, Kovesi B, Guyader AL. Noise feedback coding revisited: refurbished legacy codecs and new coding models. *ZTE Communications*, 2012,10(2):34-44.
- [10] Duan LY, Chen J, Wang CY, Ji RR, Huang TJ, Gao W. Key technologies in mobile visual search and MPEG standardization activities. *ZTE Communications*, 2012,10(2):57-66.

- [11] Szebehely VG. Theory of Orbits—The Restricted Problem of Three Bodies. New York: Academic Press, 1969.
- [12] Stiefel EL, Scheifele G. Linear and Regular Celestial Mechanics. Berlin: Springer-Verlag, 1971.
- [13] Siegel CJ, Moser JK. Lectures on Celestial Mechanics. Berlin: Springer-Verlag, 1972.
- [14] Li LS. Secular effect of evolution of the orbits of binaries induced by the variation of gravitational constant with time (the case for the elliptical orbit). Acta Astronomica Sinica, 2011,52(3):242–250 (in Chinese with English abstract).
- [15] Tang JS, Liu L. The application of quasi-mean-element-method to LEO under additional perturbation due to change of coordinate system. Acta Astronomica Sinica, 2010,51(3):75–84 (in Chinese with English abstract).
- [16] Valsecchi GB, Milani A, Gronchi GF, Chesley SR. Resonant returns to close approaches: Analytical theory. Astronomy & Astrophysics, 2003,408(3):1179–1196. [doi: 10.1051/0004-6361:20031039]
- [17] Werbos PJ. Beyond regression: New tools for predictions and analysis in the behavioral science [Ph.D. Thesis]. Harvard University, 1974.
- [18] Rumelhart DE, Hinton GE, Williams RJ. Parallel Distributed Processing: Exploration in the Microstructure of Cognition. Cambridge: MIT Press, 1986. 318–362.

附中文参考文献:

- [3] 申元霞,王国胤,曾传华.相关性粒子群优化模型.软件学报,2011,22(4):695–708. <http://www.jos.org.cn/1000-9825/3728.htm> [doi: 10.3724/SP.J.1001.2011.03728]
- [14] 李林森.引力常数随时间变化对双星轨道演变的长期效应.天文学报,2011,52(3):242–250.
- [15] 汤靖师,刘林.近地卫星运动的坐标系附加摄动在拟平均根数法中的处理.天文学报,2010,51(1):75–84.



孟超(1977—),男,江苏镇江人,博士生,主要研究领域为智能计算,云计算,云安全.
E-mail: mz848@163.com



孙知信(1964—),男,博士,教授,博士生导师,主要研究领域为网络安全.
E-mail: sunzx@njupt.edu.cn



刘三民(1978—),男,博士生,主要研究领域为网络安全.
E-mail: aqlsm@163.com